

Morphosemantic analysis of Croatian nouns formed with the prefix *pred-*

Marta Petrak

Faculty of Humanities and Social Sciences
University of Zagreb (Croatia)

mpetrak@ffzg.unizg.hr

Abstract

In this paper, we present a morphosemantic study of Croatian nouns formed with the prefix *pred-* ‘in front of’. The nouns were retrieved from the new CLASSLA-web corpus, the largest extant one for Croatian, and subsequently analysed both formally (word-formation wise) and semantically from a cognitive linguistic point of view. Cognitive linguistics considers prefixes to be polysemic units which exhibit behaviour similar to that of the traditional “lexical” categories such as nouns or verbs. Our analysis has demonstrated that the derived nouns demonstrate interesting regularities: prefixation frequently yields a temporal reading of the prefix *pred-*, prefix-suffix combination frequently results in nouns denoting places situated in front of something, and suffixation creates nouns with a number of extended meanings based on metaphor and/or metonymy. Detailed analysis of the morphosemantic potential of a specific word-formation element are rare in the literature, but contribute to a better understanding of the mechanisms of the construction of the lexicon.

Keywords: word formation, semantic motivation, Croatian, *pred-*

1 Introduction

The goal of this paper is to present an exhaustive and comprehensive morphosemantic analysis of nouns formed with the Croatian prefix *pred-* (or its allomorph *pret-*).¹ *Pred-* is a native Croatian prefix which can form nouns, verbs and adjectives (Babić, 2002).

Such analyses are quite rare in the extant literature, but are highly important because they explore the morphosemantic potential of one specific derivational element, in this case a prefix, and analyse how it is used in the construal of the lexicon. By analysing both the formal level (i.e. the various word formation mechanisms it enters into) and the semantic level (i.e. the meanings that the prefix *pred-* realizes in its derivatives, and the mechanisms on which these meanings are based), we gain a detailed insight into its morphosemantic potential.

This study is rooted in the cognitive linguistic framework, which has been of the leading currents in linguistics in the past several decades. When it comes to word-formation, however, cognitive linguistic tenets have only recently started to be applied to it (Onysko and Michel, 2010). The question was raised by some authors whether cognitive linguistics is well-equipped to deal with word-formation issues. Ungerer (2010) believes that it “has the potential to stimulate word-formation research” because it “can provide both the theoretical background and the empirical tools” necessary to do so. The underlying goal of all cognitive linguistic efforts related to word formation would be, as Ungerer (2010) puts it, “the semanticization of word-formation analysis”. This process, it needs to be noted, is not a novelty introduced by cognitive linguistics, but it was largely emphasized within this theoretical framework. In a similar vein, it needs to be said that cognitive linguistics is compatible with some basic tenets of the structuralist approach to word formation (Onysko and Michel, 2010; Štekauer and Lieber, 2005; cf. Raffaelli, 2004).

A cognitive linguistic approach to word formation, which is also adopted in this study, is in line with the basic cognitive linguistic tenets, among which the most important ones are prototype theory (e.g. Rosch, 1975), radial categories (Lakoff, 1987) and the theory of conceptual metaphor and metonymy (e.g. Lakoff and Johnson, 1980). All these facts could be subsumed under the idea that the formation of

¹ In our study, we have disregarded the standard spelling rules related to the assimilation according to voicing (cf. Babić et al. 2007), due to which some words are spelled with a *t* rather than with a *d*. We have eliminated from our lists all double entries and have retained only one spelling version of the noun, the more frequent one.

words “is in essence a cognitive phenomenon” (Štekauer and Lieber, 2005, Onysko and Michel, 2010). In other words, word-formation is not merely a linguistic issue, but as all language faculties, it is inextricably linked to human cognitive capacities and the principles of conceptual organization.

As we have previously said, the central tenet of cognitive linguistics is that all aspects of word-formation are treated as meaningful (Ungerer, 2010), which is in line with the axiom of the centrality of meaning (Langacker, 1987). As such, this framework considers prepositions and prefixes to be polysemic radial categories (e.g. Lakoff, 1987; Tyler and Evans, 2003; Šarić, 2008; Matovac, 2013; Petrak, 2021), which is quite different from the homonymous approach to these linguistic elements that had prevailed in traditional approaches to language (Peytard, 1975; Belaj, 2008).

The remainder of the paper is structured as follows. In Section 2 we showcase studies related to the prefix *pred-*. Section 3 presents our methodology, which is corpus-based. In Section 4 we analyse and discuss the results we have obtained, and Section 5 brings some concluding remarks.

2 Previous work

The prefix *pred-* features in several works: Šipka (1989) mentions it as a means of forming antonymic pairs of motivated words, in pairs with the prefixes *poslije-* ‘after’ and *post-* ‘after’. Vulić (2020), in her paper on motivated words in contemporary literary works of Burgenland Croats, mentions that *pred-* is used to form masculine and feminine nouns, without further details.

Apart from these papers, *pred-* is the main subject of Belaj’s (2008) study which deals with verbs formed with that prefix. The author concludes that these verbs form a set unified by the *pre-locativity superschema*, which explains both the prototypical and less prototypical meanings and meaning extensions of the prefix.

In his dissertation, Matovac (2013) analyses the preposition *pred*, from which the prefix *pred-* originated, and claims that together with the prepositions *nad*, *pod* and *za*, *pred* makes the basis of the study of Croatian orientational prepositions. Furthermore, Matovac (2013) remarks that, unlike the prepositions *nad* ‘above’ and *pod* ‘under’, which describe relations on the vertical axis, and which have been abundantly analysed within cognitive linguistics, prepositions referring to the horizontal axis have received much less attention or have, as in the case of *pred*, even been completely neglected. The fourth preposition, *za* ‘behind; for’, has received a lot of attention in traditional grammars due to its specific usage (*ibid.*).

The goal of this paper is to deal with nouns containing the prefix *pred-* in order to shed more light on this previously rather unexplored set of formally unified words. In addition, building upon Ullmann’s (1966) ideas that lexicon is a product, essentially, of morphological and semantic motivation, and especially the hypothesis that these two types of motivation are not and should be not regarded separately (Koch and Marzo, 2007; Raffaelli, 2013, 2015), but that formal (morphological) motivation always accompanies semantic (conceptual) motivation, we set out to explore the mechanisms which gave rise to the part of the Croatian lexicon construed with the prefix *pred-*.

3 Methodology

In line with the usage-based approach, which is predominant in the cognitive linguistic framework, this study is based on real language data as found in large corpora. More precisely, the data for our research was retrieved from the brand new CLASSLA-web.hr corpus (Kuzman and Ljubešić, 2023), which is currently the largest extant Croatian corpus. It contains 2.7 billion tokens crawled primarily from the national top-level internet domain *.hr*.

In order to retrieve all *pred-* prefixed nouns, we did a graphic search for nouns starting with the graphemes *pred* and *pret*. The reason for this lies in the fact that the CLASSLA corpora are not derivationally segmented.² The minimal frequency threshold was set at two occurrences in order to retrieve all the extant nouns (except for hapaxes), which enabled us to reach both high- and low-frequency types. Low-frequency words are known to sometimes possess properties different from

² Derivational segmentation is seldom conducted in NLP because it has not been widely used so far, and due to the fact that it is rather difficult to perform (Nikola Ljubešić, private communication).

higher-frequency words, which may be linked to a different degree of lexicalisation, new meanings, newly developed properties, etc.

The corpus search resulted in two initial lists, one containing *pred-* (3 424) and the other containing *pret-* nouns (1 062). The lists contained a lot of noise. For the types below 10 occurrences, we have first examined whether these were indeed words or simply typos (spelling of prepositional phrases together with the noun). Also, several adjectives and verbs, as well as typos, were found in the lists. There were a few words which belong to closely related Southern Slavic languages, such as Serbian and Slovene (e.g. *preduzetništvo*, *predsednik*), as well as a quite large number of double entries differentiated by the use of hyphen³. All such words were excluded from the final list.

In addition, the largest part of the manual work consisted in checking the semantics of the *pred-* and *pret-* elements, as there were words in the lists which did not actually contain the prefix, but its homograph (e.g. *pretvorba* < *pre-* + *tvoriti* > ‘conversion’, *pretjerivanje* < *pre-* + *tjerati* > ‘exaggeration’, *predikacija* < *predikat*, Lat. *praedicatum* > ‘predicate’, *predavanje* ‘lecture’ < *pre-* + *dati* >, etc.). All the words were checked in the dictionary available at the Croatian Language Portal’s (Hrvatski jezični portal) website,⁴ while those that are not present in it were checked both in the corpus and online.

Having in mind all the tedious work that needs to be done in order to clear corpus lists in studies involving derivational morphemes, it would prove extremely useful to have corpora tagged at the morphological level. One of the possible solutions would be to take derivational morphology training data and train a model for morphological segmentation. In such a case, new transformer models such as BERTi⁵ could prove to be quite useful, especially for semantic recognition (Nikola Ljubešić, private communication).

Once all the manual work has been done, our search yielded a total of 1 006 nouns. When compared to the initial 4 486 types yielded by our corpus search, we can see that only about 22% of this initial bulk quantity of words were actually *pred-* prefixed nouns, and we can observe the large amount of noise and semantic errors which had to be manually checked.

Below is a list of the 50 most frequent types, accompanied by the number of their occurrences in the corpus and frequency per million:

1. *predsjednik* ‘chair, director (m.)’, 1054257, 388.38086
2. *predmet* ‘object’, 468947, 172.75677
3. *predstavnik* ‘representative (m.)’, 414332, 152.63699
4. *prednost* ‘advantage’, 335072, 123.43817
5. *predstava* ‘theater piece’, 333850, 122.98799
6. *predsjednica* ‘chair, president (f.)’, 210374, 77.50030
7. *predstavljanje* ‘presentation’, 164495, 60.59880
8. *preduvjet* ‘precondition’, 64727, 23.84497
9. *predsjedništvo* ‘presidency’, 57161, 21.05771
10. *predrasuda* ‘prejudice’, 41029, 15.11479
11. *predložak* ‘template’, 35058, 12.91512
12. *predstavnica* ‘representative (f.)’, 34965, 12.88086
13. *predak* ‘ancestor’, 30167, 11.11331
14. *predviđanje* ‘prediction’, 28200, 10.38868
15. *predsjedanje* ‘presiding’, 17285, 6.36767
16. *predstojnik* ‘director, head (m.)’, 16638, 6.12932
17. *predgovor* ‘foreword’, 16533, 6.09064
18. *predstavništvo* ‘representative body’, 15737, 5.79740
19. *predlagatelj* ‘proposer’, 15716, 5.78966
20. *predodžba* ‘idea, conception’, 15578, 5.73883

³ The hyphenation problem could easily be solved automatically, as an idea for the development of derivational tools.

⁴ <https://hjp.znanje.hr/>

⁵ <https://huggingface.co/classla/bcms-bertic-ner>

21. *predgrađe* ‘suburbs’, 14587, 5.37375
22. *predvorje* ‘vestibule’, 13894, 5.11845
23. *predvodnik* ‘leader’, 13743, 5.06282
24. *predznak* ‘sign, omen’, 13467, 4.96115
25. *predujam* ‘advance’, 11316, 4.16873
26. *predlaganje* ‘proposition’, 10936, 4.02875
27. *predočenje* ‘presentation’, 8406, 3.09671
28. *predsoblje* ‘anteroom’, 7906, 2.91251
29. *predznanje* ‘prior knowledge’, 7522, 2.77105
30. *predstojnica* ‘director, head (f.)’, 7299, 2.68890
31. *predškolac* ‘preschooler’, 6515, 2.40008
32. *predgrupa* ‘band playing before the main one’, 6354, 2.34077
33. *predjelo* ‘appetizer’, 6016, 2.21625
34. *predstavka* ‘petition’, 5685, 2.09431
35. *predškola* ‘preschool’, 5543, 2.04200
36. *predigra* ‘foreplay’, 4994, 1.83975
37. *predvečerje* ‘eve’, 4918, 1.81176
38. *predbilježba* ‘pre-reservation’, 4849, 1.78634
39. *predračun* ‘invoice’, 4694, 1.72924
40. *predsezona* ‘preseason’, 4659, 1.71634
41. *predostrožnost* ‘precaution’, 4340, 1.59883
42. *predsjedatelj* ‘chairman (m.)’, 3980, 1.46620
43. *predugovor* ‘precontract’, 3661, 1.34869
44. *predlagač* ‘proponent’, 2788, 1.02708
45. *predvodnica* ‘leader (f.)’, 2533, 0.93314
46. *prednarudžba* ‘pre-order’, 2525, 0.93019
47. *predmetnica* ‘subject’, 2354, 0.86720
48. *predvoditelj* ‘leader (m.)’, 2353, 0.86683
49. *predbroj* ‘telephone prefix’, 2221, 0.81820
50. *predpojačalo* ‘preamplifier’, 2119, 0.78062

The number of 1 006 nouns formed with *pred-* (*pret-*) demonstrates that *pred-* has quite a large potential in the construction of the Croatian lexicon. Moreover, the prefix seems to be (much) more used in the formation of nouns than in that of verbs, which have been found to be much less numerous in some previous studies. For instance, Petrak (2021) found only 32 *pred-* verbs upon an analysis of the three Croatian most relevant Croatian corpora at the time: hrWaC, previously the largest Croatian web corpus, HNK, the Croatian national corpus, and the Riznica corpus. However, the frequency limit in her study was set at 10 occurrences, so the number of low-frequency *pred-* verbs remains unknown, and has probably influenced this large disparity in the results obtained.

Once we have obtained the list of *pred-* nouns from the CLASSLA corpus, we set out to study the following: 1) which word-formation processes *pred-* is found in, 2) what semantic mechanisms such formations are based on, 3) is there a correlation, i.e. are there regularities in the pairings of word-formation processes and semantic mechanisms, 4) is there any difference between high- and low-frequency nouns. We shall explore all these questions in the following section.

4 Results and analysis

In this section, we shall first take a detailed look into the word-formation types found in the corpus, and shall then turn to their semantic analysis.

4.1. Word-formation types

Below is a table demonstrating the part of the four word-formation types found in our corpus: prefixation, suffixation, prefix-suffix combination and backformation.

	<i>Word-formation mechanism</i>	<i>No.</i>	<i>%</i>
1	prefixation	846	83,5
2	suffixation	129	13
3	prefix-suffix combination	27	3
4	backformation	5	0,5
<i>Total</i>		1 006	100%

Table 1: Word-formation types found in the CLASSLA-web.hr corpus

a) Prefixation

Prefixation is the prevalent word-formation type in our corpus, with more than 80% of all nouns stemming from it (e.g. *predpranje* ‘prewash’, *predromanika* ‘pre-Romanesque period’, *predznanje* ‘prior knowledge’, *predobrada* ‘preprocessing’, *predgovor* ‘preface’). All these nouns were formed by simply attaching the prefix *pred-* to a nominal base: *pranje* ‘wash’, *romanika* ‘Romanesque period or style’, *znanje* ‘knowledge’ and *obrada* ‘processing’ respectively. At the semantic level, a great majority of these nouns realise the temporal meaning, which points to a regularity we could label as following: [*pred*-[N]]_N ‘before N’. We shall explain in detail the mechanisms underlying every meaning in the semantic part below (See 4.2).

b) Suffixation

Suffixation, which is otherwise the most productive and prevalent word-formation process both generally and in nouns in Croatian (cf. Babić, 2002), amounts to 13% of all formations (e.g. *predsjednik* ‘president’, *prethodnik* ‘predecessor’, *predmetak* ‘prefix’, *predsjedatelj* ‘chairman (m.)’, *predvoditeljica* ‘leader (f.)’, *predradnica* ‘foreman (f.)’, etc.).

Semantics-wise, most of the suffixed words are such that the suffixed nouns get a new, metaphorical reading, unlike the prefixation cases, in which the meaning of the prefix frequently and regularly extends the meaning of the nominal base towards the temporal domain. Here is an example: *predsjednik* ‘president’ is formed from *pred-* ‘before’ + *-sjed-* ‘sit’ and the agentive suffix *-nik*, where ‘sitting in front of other people’ stands for ‘presiding’ according to the LEADING IS BEING AHEAD metaphor (cf. Petrak, 2021). Some of the prefixed derivatives are exceptions to this rule, such as *predmetak* ‘prefix’, formed from *pred-*, *-met-* ‘put’ and the suffix *-ak*, which literally reads as ‘that which is put in front of (something)’. This meaning also includes the part for whole metonymy because a single property, i.e. standing in front of something else, is taken to designate the whole entity, i.e. a prefix.

c) Prefix-suffix combination

Prefixation and suffixation representing around 93% of all formations, the rest are, clearly, rarely used formation types. Such a finding is in line with the extant literature, according to which suffixation is the most frequent mechanism in the formation of Croatian nouns, while back-formation and prefix-suffix combinations are rather rare (Babić, 2002). In our corpus, prefix-suffix combinations represent 3% of all derivatives, with examples such as *predslovlje* ‘preface’, *predsoblje* ‘anteroom, antechamber’, *predgrađe* ‘suburbs’, *predgorje* ‘foothills’, *predvečerje* ‘early evening’.

A majority of these derivatives realise a concrete spatial meaning (*predsoblje* ‘antechamber’, *predgorje* ‘foothills’, etc.), but some have a temporal meaning (*predvečerje* ‘eve’ < *pred-* + *večer* ‘evening’ + *-je*>, *predzorje* < *pred-* + *zor(a)* ‘dawn’ + *-je*> ‘dusk’,). Most of these nouns receive the suffix *-je*, while only a few appear with other suffixes as well: *predoltarnik* < *pred-* + *oltar* ‘altar’ + -

nik > ‘altar frontal’, *predkolumbijanac* <*pred-* + *Kolumbo* ‘Columbus’ + *-ijanac* > ‘pre-Columbian’, *pred-bolonjac* <*pred-* + *Bolonja* + *-ac* > ‘student before the Bologna Process’.

In this group, two proper nouns are present: *Predkavkazje* ‘fore-Caucasus’ and *Predalpe* ‘Pre-Alps’, in which the prefix retains its concrete meaning ‘in front of’, and is attached to an oronym (*Kavkaz* ‘the Caucasus’ and *Alpe* ‘the Alps’).

d) Back-formation

The last type of derivatives is formed via back-formation, which is quite rare with only 0,5% of the total number of nouns (e.g. *predstava* ‘theatre piece’, *pretklon* ‘the forward lean’, *predosjećaj* ‘presentiment’, *predrasuda* ‘prejudice’). These nouns can have both concrete (*pretklon* <*pretkloniti* <*pred-* + *kloniti* > ‘to lean forward’ >) and abstract meanings (*predstava* <*pred-* + *-stav-* ‘put’ + *-a* >). The noun *predstava* is a result of a metonymy in which the act of putting something in front of someone, i.e. of “putting” a literary piece on stage, is taken to designate the whole theatre piece. In the noun *predrasuda* <*pred-* + *rasuditi* ‘judge’ + *-a* >, the prefix has a temporal reading ‘before’, and the meaning of the whole noun is based on encyclopaedic knowledge that prejudice are ideas made before facts are taken into consideration, in which the IDEAS ARE OBJECTS metaphor⁶ is also present, according to which prejudice is conceptualized as objects that are placed in front of the act of judgement.

4.2. Semantic analysis

When it comes to the semantic part of our study, we have first concluded that the analysed nouns can be divided into three large groups: 1) those in which the prefix retains its prototypical spatial meaning (e.g. *predbroj* ‘lit. in front-number; telephone prefix’), 2) the group in which only the prefix gets an extended meaning, which is temporal in the largest number of cases (e.g. *predugovor* ‘precontract’), and 3) the group in which the noun as a whole gets an extended meaning (e.g. *predlagač* ‘proponent’ is formed from *pred-* ‘in front’ and *ložiti* ‘arrange’). These three meanings correspond to the concrete-spatial-metaphorical & metonymical labelling as demonstrated in Table 2.

	<i>Meaning type</i>	<i>No.</i>	<i>%</i>
1	concrete (spatial)	110	11
2	temporal	801	80
3	metaphorical & metonymical	95	9,4
<i>Total</i>		1 006	100%

Table 2: Meaning types of the prefix *pred-* in the analysed derivatives

Even though the first and prototypical meaning of the prefix *pred-* is concrete and spatial, more precisely ‘in front of something’,⁷ as the meaning of the corresponding preposition *pred* from which the prefix has originated, in our corpus the prefix is prevalently used with a temporal meaning ‘before’. In other words, in derived nouns, *pred-* shows a clear semantic shift toward more abstract meanings, which is obviously a result of prefixation.

4.2.1. Concrete meaning

Nouns in which *pred-* retains its concrete meaning ‘in front of’ are, for example, the following: *predgrađe* <*pred-* + *grad* ‘town’ + *-je* > ‘suburbs’, *predvorje* <*pred-* + *dvor* ‘court; courtyard’ + *-je* > ‘vestibule’ and *predbroj* <*pred-* + *broj* ‘number’ > ‘telephone prefix’. Most of these nouns are a result of the prefix-suffix combination, but there are also some examples of prefixation (*predbroj*).

⁶ https://metaphor.icsi.berkeley.edu/pub/en/index.php/Metaphor:IDEAS_ARE_OBJECTS

⁷ https://hjp.znanje.hr/index.php?show=search_by_id&id=eVhiWxc%3D&keyword=pred

4.2.2. Temporal meaning

The largest semantic group is formed of nouns in which the prefix has the temporal meaning '(occurring) before (something)'. This meaning is a result of the PAST IS BEHIND metaphor (Lakoff and Johnson, 1980), according to which events happening in the future are conceptualized as being in front of the speaker. For instance, in examples such as *predjelo* 'appetizer', *predradnja* 'activity preceding another one' or *predtestiranje* 'pre-testing', appetizers, previous activities and pre-testings are conceptualized as objects placed before a later meal, activity or testing. This metaphor is a rather frequent one in Indo-European languages, and it is an entailment of the very common TIME IS SPACE metaphor (e.g. Radden, 2003).

In addition, we would tentatively say that *pred-* formations in which the prefix has a temporal meaning are probably quite recent formations, which might be the result of calquing from English (cf. *pre-selection*, *pre-finance*, *precontract*, *foreplay*, etc.), but this question should be further studied.

4.2.3. Metaphorical and metonymical

The third group in Table 2 is labelled metaphorical and metonymical because it results from various metaphors and metonymies. We have used this designation for all other meanings which are extensions of the prototypical one, and which we have grouped under a single label due to the fact that they represent less than 10% of all meanings. The leading meaning in this group is a result of the LEADING IS BEING AHEAD metaphor, which is an entailment of the GOOD IS IN FRONT metaphor (Belaj, 2008)⁸ found in English. This metaphor has a metonymical basis because the property of being in front is used to refer to 'leading' (SALIENT PROPERTY FOR CATEGORY, Littlemore, 2010). Examples in which these metaphor and metonymy occur are the following: *predmolitelj* < *pred-* + *moliti* 'pray' > 'leading celebrant', *predradnik* < *pred-* + *radnik* 'worker' > 'foreman', *predšasnik* 'predecessor', *predvođenje* < *pred-* + *voditi* 'lead' + *-nje* > 'leading', *predsjedništvo* < *predsjedati* 'preside' < *pred-* + *sjediti* 'sit' > 'presidency'.

Another relatively important subgroup in this group are nouns such as *predokus* < *pred-* + *okus* 'taste' > 'foretaste', *predosjećanje* < *pred-* + *osjećati* 'feel' + *-nje* > 'presentiment', *predskazivanje* < *pred-* + *kazati* 'tell' + *-nje* > 'foretelling' or *predskazatelj* < *pred-* + *kazati* 'tell' + *-telj* > 'foreteller'. The meaning of these nouns is based on the PART FOR WHOLE metonymy, according to which foretelling or presentiments are conceptualised as saying, tasting or feeling something before something else. The metaphor PAST IS BEHIND is also present in the meaning of these nouns. In the noun *predokus* 'foretaste', the metaphor COGNIZING IS EATING⁹ is also present, according to which mental activities are conceptualized as tasting.

A third subgroup that should be mentioned here is the one consisting of the nouns *predočenje* 'presentation', *predodžba* 'idea, conception', *predočivost* 'imageability' and the like. These lexemes were all formed from the verb *predočiti* 'display, put forward', which comes from the root *oči* 'eyes', and their meaning is the result of the common metaphor KNOWING IS SEEING¹⁰, on the basis of which mental activities are conceptualised as the act of seeing. In these nouns, presenting, displaying and imaging something is conceptualised as putting objects in front of someone's eyes.

In the conclusion to this part, we can say that in nominal derivatives, the prefix *pred-* sometimes retains its prototypical spatial meaning 'before', but in the majority of cases it is extended into the temporal domain via the TIME IS SPACE metaphor. In this second type of derivatives, only the prefix gets a metaphorical reading, while the meaning of the base remains the same. There is a third group of derivatives, in which the whole construction [prefix + noun] gets an extended meaning, motivated by metaphor and/or metonymy. Such insights are in line with the data from the literature on the preposition *pred*, according to which it extends its meaning to the temporal domain, but other than that, it does not develop many non-spatial meanings (Matovac, 2013).

⁸ It is very close to the very similar the LEADING IS A FORCE MOVING AN OBJECT FORWARD metaphor, cf. https://metaphor.icsi.berkeley.edu/pub/en/index.php/Metaphor:LEADING_IS_A_FORCE_MOVING_AN_OBJECT_FORWARD.

⁹ Cf. https://metaphor.icsi.berkeley.edu/pub/en/index.php/Metaphor:COGNIZING_IS_EATING.

¹⁰ Cf. https://metaphor.icsi.berkeley.edu/pub/en/index.php/Metaphor:KNOWING_IS_SEEING.

Similar conclusions were signalled in a previous study (Petрак, 2021) with respect to the semantic structure of *pred-* verbs. More precisely, while the study did not find any *pred-* verbs in which the prefix retains its concrete spatial meaning, *pred-* verbs could be grouped into two groups identical to the second (temporal) and third (metaphorical and metonymical) groups in our study. In the first group, the prefix has a temporal reading, while the base retains its concrete meaning (e.g. *predugrijati* < *pred-* + *ugrijati* ‘heat’ > ‘preheat’), and in the second group, verbs develop an extended meaning as a whole, in which the two most frequent meanings are ‘leadership’ and ‘foretelling’ (e.g. *predvoditi* < *pred-* + *voditi* ‘lead’ ‘leading’ >; *predvidjeti* < *pred-* + *vidjeti* ‘see’ ‘foretell’ >).

4.3. Word-formation – semantic pairings

In this part of our analysis, we examine whether there are specific word-formation – semantic pairings or regularities in the formation of the Croatian lexicon with the prefix *pred-*.

Prefixed nouns exhibit a clear and straightforward preference for the temporal meaning. In other words, there is a formal-semantic regularity according to which [*pred-* + Noun] frequently results in nouns with a temporal meaning ‘before N’.

The suffixed nouns group is the most heterogenous one, in which various meanings appear: concrete (*predmetak* ‘prefix’), temporal (*predškolka* ‘preschooler (f.)’) and metaphorical/metonymical (*predvodnik* ‘leader’), based on different metaphors and/or metonymies.

In prefix-suffix formation, *pred-* is typically used in its prototypical, spatial meaning to denote spaces that are located in front of another spatial landmark (e.g. *predvorje* ‘antechamber, entrance hall’ < *pred-* ‘before’ + *dvor-* ‘court’ + *-je* > denotes a space located in front of a large(r) front door), and in such formations the suffix *-je* is typically used. In other words, [*pred-* + N + *-je*] frequently results in nouns with a spatial meaning ‘the space in front of N’.

Backformations are very heterogenous and exhibit no clear regularities.

4.4. Comparison of high- and low-frequency types

Lastly, we need to address the question of whether there is any difference in the behaviour of high- and lower-frequency nouns. Below is a table summarizing word-formation data for low-frequency nouns:

	Word-formation type	No.	%
1	prefixation	491	96
2	suffixation	11	2,1
3	prefix-suffix combination	10	1,9
4	backformation	0	0
	Total	512	100

Table 3: Word-formation types in low-frequency nouns

Low-frequency nouns exhibit the highest percentage of prefixation (96%, whereas it was 83,5% for high-frequency nouns), followed by suffixation, 2,1% (as opposed to 13% in high-frequency nouns), prefix-suffix combination 1,9% (as opposed to 3%). Backformations have not been found in the low-frequency range, while about 0,5% of high-frequency types exhibit that word-formation mechanism.

It can be concluded that low-frequency nouns formed with *pred-* are predominantly and almost systematically formed through prefixation, while other word-formation types are rather rare. Suffixation and prefix-suffix combination represent almost equal shares (2,1 and 1,9%, respectively), while there was a larger difference in the high-frequency group (13% and 3%, respectively).

On the semantic level, in the largest majority of the lower-frequency nouns, *pred-* has a temporal meaning (e.g. *predčistište* ‘ante-Purgatory’, *predijagnoza* ‘pre-diagnosis’, *predglaćanje* ‘pre-ironing’) according to the TIME IS SPACE metaphor mentioned before.

Data provided in this sub-section points to some differences between the high- and lower-frequency range, and explain why lower-frequency types (in this case, all nouns except for hapaxes) should be included in research on derivational processes.

5 Concluding remarks

Grammar-lexicon continuum

Our detailed study of the morphosemantic potential of the prefix *pred-* has demonstrated that it is found in numerous nouns whose morphosemantic structures exhibit interesting regularities. On the formal level, the prefix enters into four types of word-formation processes: prefixation, which is the prevalent one, suffixation, prefix-suffix formation and back-formation, the latter two being rather rare. Semantics-wise, the prefix in *pred-* nouns seldom retains its prototypical spatial meaning, and such nouns are usually the result of prefix-suffix combination. In the largest number of nouns, which are the result of prefixation, *pred-* gets an extended, temporal meaning according to the TIME IS SPACE metaphor. Suffixation usually produces nouns which get an extended meaning as a whole, based on metaphor, metonymy or both of these mechanisms. These are interesting morphosemantic regularities that have been found in the part of the Croatian nominal lexicon constructed with the prefix *pred-*. Our study of *pred-* nouns adds up to research proving that formal and semantic motivation are indeed intertwined (cf. Koch and Marzo, 2007), and sheds some more light on the semantic structure of motivated words, which remains a rather unexplored area of morphology (cf. Onysko and Michel, 2010).

References

- Stjepan Babić, Dalibor Brozović, Ivo Škarić, and Stjepko Težak. 2007. *Glasovi i oblici hrvatskoga književnoga jezika*. Nakladni zavod Globus, Zagreb.
- Stjepan Babić. 2002. *Tvorba riječi u hrvatskome književnome jeziku*. Globus and HAZU, Zagreb.
- Branimir Belaj. 2008. Pre-locativity as the schematic meaning of the Croatian verbal prefix *pred-*. *Jezikoslovlje* 9 (1-2): 123–140.
- Peter Koch and Daniela Marzo. 2007. A two-dimensional approach to the study of motivation in lexical typology and its first application to French high-frequency vocabulary. *Studies in Language* 31 (2): 259–291.
- Tanja Kuzman and Nikola Ljubešić. 2023. CLASSLA-web: Bigger and better web corpora for Croatian, Serbian and Slovenian on Clarin.si concordancers. <https://www.clarin.si/info/k-centre/classla-web-bigger-and-better-web-corpora-for-croatian-serbian-and-slovenian-on-clarin-si-concordancers/>
- George Lakoff and Mark Johnson. 1980. Conceptual metaphor in everyday language. *The Journal of Philosophy* 77 (8): 453–486.
- George Lakoff. 1987. *Women, Fire and Dangerous Things. What Categories Reveal About the Mind*. University of Chicago Press, London, Washington, DC.
- Janet Littlemore. 2010. *Metonymy: Hidden Shortcuts in Language, Thought and Communication*. Cambridge University Press, Cambridge.
- Darko Matovac. 2013. *Semantika hrvatskih prijedloga*. PhD dissertation. Sveučilište J. J. Strossmayera, Osijek, Filozofski fakultet.
- Alexander Onysko and Sascha Michel. 2010. Introduction: unravelling the cognitive in word formation. In *Cognitive Perspectives on Word Formation*. De Gruyter Mouton, pages 1–28.
- Marta Petrak. 2021. Complex interplay of metaphor and metonymy: the case of *pred-* prefixed verbs. In *2nd International Conference for Young Researchers in Cognitive Linguistics (YRCL) Book of Abstracts*. Universidad de Alcalá, page 37.
- Jean Peytard. 1975. *Recherches sur la préfixation en français contemporain*. Librairie Honoré Champion, Paris.
- Günter Radden. 2003. The Metaphor TIME AS SPACE across Languages. In *Übersetzen, Interkulturelle Kommunikation, Spracherwerb und Sprachvermittlung - das Leben mit mehreren Sprachen. Festschrift für Juliane House zum 60. Geburtstag. Zeitschrift für Interkulturellen Fremdsprachenunterricht*. AKS-Verl, pages 226–239.
- Ida Raffaelli. 2004. Odnos strukturalne semantike prema kognitivnoj. *Suvremena lingvistika* 57-58(1-2):67–92.

- Ida Raffaelli. 2013. The model of morphosemantic patterns in the description of lexical architecture. *Lingue e linguaggio* 1:47–72.
- Ida Raffaelli. 2015. *O značenju. Uvod u semantiku*. Matica hrvatska, Zagreb.
- Eleanor Rosch. 1975. Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General* 104: 193–233.
- Ljiljana Šarić. 2008. *Spatial Concepts in Slavic. A Cognitive Linguistic Study of Prepositions and Cases*. Harrasowitz Verlag, Wiesbaden.
- Danko Šipka. 1989. Tvorbena sredstva antonimizacije. *Jezik: časopis za kulturu hrvatskoga književnog jezika* 37(5): 139–145.
- Pavol Štekauer and Rochelle Lieber. 2005. *Handbook of Word-Formation*. Springer, Dordrecht.
- Andrea Tyler and Vyvyan Evans. 2003. *Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition*. Cambridge University Press, Cambridge.
- Stephen Ullmann. 1966. Semantic universals. In *Universals of language*. MIT Press, Cambridge MA, pages 217–262.
- Friedrich Ungerer. 2010. Word-Formation. In *The Oxford Handbook of Cognitive Linguistics*. Oxford University Press, Oxford.
- Sanja Vulić. 2020. Tvorbeno motivirane riječi u suvremenim gradišćansko-hrvatskim književnim djelima. *Čakavska rič: Polugodišnjak za proučavanje čakavske riječi* XLVIII (1-2): 9–31.