Light Verb Constructions in ELEXIS-WSD – Annotation, Comparisons and Issues

Cvetana Krstev
Association for Language
Resources and Technologies
Belgrade, Serbia
cvetana@jerteh.rs

Ranka Stanković
University of Belgrade
F. of Mining and Geology
Belgrade, Serbia
ranka@rgf.rs

Aleksandra Marković
Institute for
the Serbian Language SASA
Belgrade, Serbia
malexa39@gmail.com

This paper deals with light verb constructions and their annotation in ELEXIS- sr, the Serbian extension of the ELEXIS-WSD corpus. In Section 1, general introductory remarks are given about these constructions, the notion of light verbs, and their treatment and further classification in the PARSEME annotation guidelines (subtypes LVC.full and LVC.cause). Section 2 offers an insight into ELEXIS-WSD corpus, annotated with VMWEs for several languages, with a remark that these VMWEs were not further subcategorised into finer classes. For this paper, we classified them ourselves to facilitate comparisons of the LVCs annotated in ELEXIS-sr. Tools and resources used for the automatic annotation of ELEXIS-sr are presented in Section 3, as well as the results of manual checking. In Section 4, we offer a comparison of LVCs in four ELEXIS-WSD sub-collections: Serbian, Bulgarian, Slovene, and English. We use Serbian as a starting point for this comparison, as it has been thoroughly annotated with MWEs (and NEs). We present the results of the comparison of all the occurrences of LVCs in the Serbian extension with their occurrences and annotation both in ELEXIS-WSD and Parseme sub-corpora for other languages.

An important conclusion is that the most equivalents among LVCs are between Serbian and Bulgarian, closely related Slavic languages (a total of 34 equivalents), while between Serbian and Slovene, also Slavic, there are 11 equivalents, as between Serbian and English. It seems that this could be explained by the number of VMWES and LVCs annotated, or by the strategy used by different annotators.

Keywords: *Light verb constructions, annotation, ELEXIS-WSD, Serbian, Bulgarian, Slovene, English*

1. Light verb constructions – LVCs

1.1 General

This study deals with the identification and annotation of light verb constructions (LVCs) in the Serbian extension of ELEXIS-WSD corpus. One of our aims is to take a comparative look at these constructions in Serbian and their different translations in the ELEXIS corpora for some other languages. Our study could also serve as a contribution to the study of LVCs in the Serbian language.

Light verbs, first mentioned in (Jespersen 1965),[1] was a frequent research subject (a rather detailed overview of related work can be found in (Stoyanova, Leseva, and Todorova 2016). While the term 'light verb construction' is used in English, in other languages, the terminology for this notion is not uniform (for example, as (Wittenberg 2016) notes in German, or as we will demonstrate below in Serbian).

The first mention of the phenomenon we are dealing with in this paper in Serbian was in (Radovanović 1990). The author names it 'predicate decomposition' and treats it as a language universal and a part of global nominalization processes in language. The author also mentions that syntactic models with predicate decomposition show a considerable extent of phraseologisation and that their lexical components exhibit a kind of collocation. Relevant to our research is that, as the author observes, nominalization processes and predicate decomposition are frequently represented in certain functional styles, namely those that favour abstract and intellectual, general language use, like those of official documents, scientific prose, and publicistic literature. This fact is relevant because the nature of the ELEXIS corpus we are working on (see Section 2) supports this observation: it lacks the language of belles-lettres, as well as that of everyday use. Similar observations are found in (Samardžić 2007), where the author uses the term 'light verb constructions' for the first time for the Serbian language and says it is a widespread linguistic phenomenon, which, although very productive, exhibits some collocational properties, and because of this requires special attention in translation, second language learning and teaching, as well as in lexicography. Based on the analysis of a small sample of parallel sentences, the conclusion showed that LVCs in English cannot be translated into Serbian word-for-word and are always translated with a single perfective verb. Besides these, there was a bundle of research dedicated to this and related problems (periphrastic predicates in Slavic languages (Topolińska 1982), and in Serbian (Lazić-Konjik 2006); on predicate decomposition (Ivić 1988), to mention some of them).

In (Wittenberg 2016), the author mentions what she considers the essential characteristics of light verb constructions. Namely, they are complex predicates composed of the light or semantically bleached verb and its nominal part, the event nominal. As this author says, most of the predicative meaning of LVCs comes from the event nominal, which assigns semantic roles to the subject, besides the light verb itself. In these constructions,

---

1 Otto Jespersen was the first author who used the term *light verb*. However, it seems that he used it conditionally, since the word *light* is in quotation marks.

a phenomenon known as argument sharing occurs: a subject is not an agent only of the light verb but also of the event nominal. There is a suspension of the canonical one-to-one correspondence between syntactic positions and semantic roles. On the syntactic level, LVCs behave similarly to their non-light counterparts (for example, in the sentences *The woman is giving the man a kiss* and *The woman is giving the man a book* both predicates are ditransitive). But LVCs exhibit a mismatch between semantic and syntactic representation levels: "The linguistic structure of light verb constructions looks syntactically like in non-light constructions, but semantically like in base verbs." (Wittenberg 2016). This author also thinks that the specific semantics and structure of LVCs, as well as an undefined repertoire of LVCs in languages, make their recognition and annotation an important task.

1.2 LVC in Parseme corpora

PARSEME corpus version 1.3 represents a multilingual corpus comprising 26 languages (including Serbian, English, Bulgarian and Slovene) that is annotated with verbal multiword expressions (VMWE) (Savary et al. 2023). PARSEME annotation guidelines[2] define multi-word expressions (MWE) as continuous or discontinuous sequences of words that show some degree of orthographic, morphological, syntactic, or semantic idiosyncrasy with respect to what is considered general grammar rules of a language. The component words of such a sequence have to include a headword and at least one other syntactically related word and at least two of its components have to be lexicalized. The most salient property of MWEs is semantic non-compositionality, that is, it is often impossible to deduce the meaning of the whole unit from the meanings of its parts and from its syntactic structure.

A verbal MWE is a multiword expression whose neutral form[3] is such that it has a distribution of a verb, a verbal phrase or a verbal clause, and its syntactic head is a verb.

Parseme distinguishes the following categories of VMWEs:

- universal categories:
  - *light verb constructions* (LVC);
  - *verbal idioms* (VID);

- quasi-universal categories:
  - *inherently reflexive verbs* (IRV);
  - *idiomatic verb-particle constructions* (IVPC);
  - *multiverb constructions* (MVC);

- language-specific categories, defined for some particular languages;

- an optional experimental category, *inherently adpositional verbs* (IAV).

Light verb constructions have the following two characteristics:

- They are formed by a verb v and a single or compound noun n, which either directly depends on v or is introduced by a preposition.

---

2 Parseme annotation Guidelines
3 In Parseme, a neutral form of a MWE is its least syntactically marked form which preserves its meaning.

- The noun n is predicative and refers to an event (e.g. decision, visit) or a state (e.g. fear, courage). Predicative nouns are nouns that have semantic arguments, that is, they express predicates whose meaning is only fully specified by their semantic arguments.

Two sub-categories of VMWEs are recognized that define two different categories of LVCs:

- Type LVC.full: The verb v is "light" in that it contributes to the meaning of the whole only by bearing morphological features: person, number, tense, mood and morphological aspect. This implies that v's syntactic subject is n's semantic argument. Examples:[4]
  - EN *to make a presentation* – a semantic argument of a noun (*presentation*) is a syntactic subject, a presenter;
  - BG *давам изявление* lit. to give a statement, 'to make a statement' – a semantic argument of a noun (*изявление* – statement) is a syntactic subject;
  - SL *imeti predavanje* lit. to have a lecture, 'to give a lecture' – a semantic argument of a noun (*predavanje* – lecture) is a syntactic subject, a lecturer;
  - SR *донети одлуку* lit. to bring a decision, 'to make a decision' – a semantic argument of a noun (*одлука* – decision) is a syntactic subject.

- Type LVC.cause: The verb v is "causative" in that it indicates that the subject of v is the cause or source of the event/state expressed by n. The noun n has semantic arguments expressed as non-subject elements in the sentence, and the subject of the verb brings additional information, indicating the cause or source of the event/state. Examples:
  - EN *to grant rights* – *X* has the right to *Y*, the granter is not a semantic argument of *rights*, but it causes *X* to have the right to do *Y*;
  - BG *давам възможност* 'to give an opportunity' – *X* has an opportunity to do *Y*, the giver is not a semantic argument of *възможност* 'opportunity', but it causes *X* to have the opportunity for *Y*;
  - SL *narediti konec <nečemu>* lit. to make an end <to something> 'to end <something>', *X* has an end, the syntactic subject is not a semantic argument of *konec* 'end', but it causes *X* to ends;
  - SR *задати главобољу* lit. to cause headache 'to give a headache', *X* has a headache, the syntactic subject is not a semantic argument of *главобоља* 'headache', but it causes *X* to have it.

The Guidelines themselves contain tests that allow for the distinction of VMWEs from other MWEs, and then the distinction between various types of VMWEs.[5] We will not present these tests in detail here, but will mention briefly the tests used for LVCs. These tests

---

4 All examples in this subsection are taken from the Parseme Guidelines 1.3 website.
5 Specific tests for categorizing verbal MWEs

will be applied if the generic decision tree for verbal MWE candidates has determined that the candidate contains a unique verb v as functional syntactic head of the whole, that this verb has a unique dependent, has no lexicalized subject and the morphosyntactic category of the dependent is an extended nominal phrase n.

- LVC.0 – *Noun is abstract*: Is the noun n (single or compound) abstract? A "no" answer rejects the candidate as an LVC.

- LVC.1 – *Noun is predicative*: Does the noun n have at least one semantic argument, implying that it is a predicative noun? A "no" answer rejects the candidate as an LVC.

- LVC.2 – *Verb's subject is noun's semantic argument*: Is the subject of the verb a semantic argument of the noun n? The answer "no" leads to test LVC.5.

- LVC.3 – *Verb with light semantics*: Is v semantically light, that is, is the semantics that v adds to n restricted to: (i) what stems from its morphological features (e.g. future, plural, perfective aspect, etc.), (ii) pointing at the semantic role of n played by v's subject? A "no" answer rejects the candidate as an LVC. A "yes" answer or "unsure" leads to the next test.

- LVC.4 – *Verb reduction*: Is it possible to build an NP without the verb, in which v's subject s becomes n's dependent. A "yes" answer means that it is an LVC.full, a "no" answer rejects the candidate.

- LVC.5 – *Verb's subject is noun's cause*: Is the subject of the verb expressing the cause of the predicate expressed by the noun? A "yes" answer means that it is an LVC.cause, a "no" answer rejects the candidate.

1.3 LVCs in the Parseme: SR, EN, SL, BG

All four languages that we are dealing with in this paper are represented in the PARSEME corpus version 1.3. The size of corresponding sub-corpora measured in tokens, as well as the types of annotated VMWEs and their number differ significantly (see Table 1).

Lng.	Tokens	VID	IRV	LVC		VPC		IAV	MVC
				full	cause	full	semi		
BG	480,413	1,260	3,223	1,909	222	0	0	90	0
EN	124,203	187	0	333	51	368	53	71	51
SL	586,187	724	1626	239	64	0	0	710	0
SR	87,367	269	564	402	69	0	0	0	0

Table 1
Number of occurrences of VMWEs of different type in the Parseme corpus 1.3 for BG, EN, SL and SR.

We will briefly report on some research conducted in connection with the Parseme corpus.

Authors in (Gantar et al. 2019a) report on the structural and semantic classification of VMWEs in Slovene. Quantitative analyses of 3,364 sentences annotated with VMWEs showed that the least frequent category by type was LVC.cause (2%), LVC.full being right after with 7%, and the most frequent was IRV, 48%. The distinction between these two types of LVCs in the NLP context is explained in the next subsection. Interesting is their finding that LVC.full and LVC.cause are the least diverse categories (when talking about different VMWEs, which indicates that they form a closed class, with a limited list of their lexical components). Qualitative analysis showed: that combinations of a verb and a PP are more typical for the LVC.cause category; a relatively limited set of nouns is found in the annotated examples, some of which occur exclusively in LVC.cause (EN: 'effect', 'influence', 'help'), while others are characteristic for LVC.full type (en 'possibility', 'role', 'opinion'). The latter are more diverse, speaking of semantic classes. Among the most frequent verbs in LVCs are 'to have', 'to be', and 'to give'. This observation follows our results (see Section 3).

In (Gantar et al. 2019b), authors mention that LVCs are among MWEs (just like some verb + particle combinations, and some compounds, like *bus driver*) which can be included in dictionaries as lexical units, although semantically transparent.[6] LVCs appear with different degrees of idiomaticity, verbs in these constructions are sometimes void of meaning and can be paraphrased with the verbal form of the noun complement (*take a walk* vs. *walk*). Authors mention that morphology and syntax of LVCs can be unpredictable (e.g., there is only a limited number of nouns light verbs can combine with). The place of these constructions in the dictionary micro- and macro-structure varies; sometimes the LVCs are given as separate entries, sometimes under particular senses (lexical units), and sometimes among other MWEs, in the phrase section.

The semi-automatic compilation of the Dictionary of Bulgarian MWEs, among which nominal and verbal ones were predominant, was described in (Koeva et al. 2016). Since Bulgarian, like Serbian, is a morphologically rich language, many issues need to be addressed in the appropriate description of MWEs, including LVCs. One characteristic of LVCs is that they often take modifiers (BG: *vzemam (trudno/vazhno) reshenie* 'to make a (difficult/important) decision').

In (Leseva et al. 2024), the authors focus on developing a uniform approach to the description of MWEs, intending to create an electronic bilingual lexicon (BG-RO) of MWEs. The lexicon is derived from Bulgarian and Romanian wordnets, and verbal MWEs are being covered so far. The work offers the following description levels: lexical, derivational, morphological, syntactic, semantic, contextual, and stylistic. The authors conclude that among VMWEs, LVC and VID cases pose several challenges for proper description and analysis (on the other hand, IRVs have regular structure, word order and syntactic properties). As for the internal syntactic structure of LVCs, for the LVC.full type, most expressions exhibit the V+obj structure (EN: *give check*), while LVC.cause type displays two internal structure types, V+xcomp (EN: *make public*, *make equal*), and V+[case+obl] (EN: *put into circulation*). As for the external syntactic structure of LVCs, LVC.cause type has a valence frame with

---

6 "Many word combinations are very frequent and fixed in structure, but in terms of predictability of meaning, they dwell in the gray area between free combinations and MWEs (e.g. *dark chocolate*), and their classification and inclusion in dictionaries is rather arbitrary" (Gantar et al. 2019b, 7).

a subject and an obligatory object, e.g. "Bank puts money into circulation''. The LVC.full valence frame can contain only the subject, or the subject and a nominal, or a clause, etc.

2. ELEXIS-WSD corpus

ELEXIS-WSD is a parallel sense-annotated corpus in which content words (nouns, adjectives, verbs, and adverbs) have been assigned senses for 10 languages: Bulgarian (BG), Danish (DA), English (EN), Spanish (ES), Estonian (ET), Hungarian (HU), Italian (IT), Dutch (NL), Portuguese (PT), and Slovene (SL).[7] The list of sense inventories is based on WordNets for DA (Pedersen et al. 2023), EN, IT, NL, Wiktionary is used for ES, and national digital dictionaries are used for BG, ET, HU, PT, and SL (Martelli et al. 2021).

All corpora were morpho-syntactically tagged, and to a certain extent, multi-word expressions (MWE) and named entities (NE) were also annotated. The number of different MWEs and NEs annotated per language and the number of different senses associated with them is represented in Table 2. We can observe that the number of annotated MWEs and NEs differs significantly per language; e.g., 7 MWEs for Hungarian compared to 440 for Danish. It should also be noted that the different types of MWEs and NEs were not distinguished. Also, for some languages, for example, Slovene, NEs were annotated as MWEs.

Lang.	MWE		NE	
	lemma	sense	lemma	sense
Bulgarian	299	465	2	2
Danish	440	477	440	459
English	179	309	1	1
Spanish	36	40	4	8
Estonian	177	217	112	145
Hungarian	7	7	6	6
Italian	41	42	0	0
Dutch	33	37	27	27
Portuguese	113	115	14	15
Slovenian	385	451	0	0
Total	1,710	2,160	606	663

Table 2
Number of MWEs and NEs in the repository; the second and fourth columns present the number of unique lemmas in the WSD, while the third and fifth columns present the number of unique senses.

Since this paper deals with comparing verbal multi-word expressions in Serbian and their usage in Bulgarian, Slovene, and English, we analysed the types of VMWEs annotated in the ELEXIS-WSD corpus for these three languages. As we already explained, annotated MWEs in ELEXIS-WSD were not classified into finer categories, so we classified ourselves based on the morphosyntactic tagging, VMWE syntactic structure and information

---

7 Parallel sense-annotated corpus ELEXIS-WSD 1.1

obtained from the Parseme corpus (for VMWEs that occur in it, that is, that are annotated as VMWEs in it). The results are presented in Table 3. We can observe that in Bulgarian and Slovene corpora, most annotated VMWEs are reflexive verbs, while in English, verb-particle constructions prevail.

	IRV		VPC		LVC.full		VID		VERB	
Lang.	**L**	**S**	**L**	**S**	**L**	**S**	**L**	**S**	**L**	**S**
BG	141	306	/	/	5	5	9	10	17	17
EN	/	/	34	156	2	2	3	3	2	3
SL	75	130	/	/	/	/	1	1	2	3

Table 3
Number of VMWEs in the repository for BG, EN, and SL, per lemmas (L) and senses (S); VERB refers to VMWEs that could not be categorized.

The Serbian part of the ELEXIS-WSD corpus, dubbed ELEXIS-SR, is being prepared in the scope of Working Group 2 of the UniDive COST action.[8] The translation from the English corpus of 2,024 sentences has been completed. Care was taken to ensure that the translation was in idiomatic, natural Serbian language while maintaining at the same time the meaning of the sentences in English. Tokenization, lemmatization, and POS-tagging were done automatically (Stanković et al. 2020; Stanković, Škorić, and Šandrih Todorović 2022), and controlled by at least three evaluators. It has been annotated with MWEs (see Section 3), and NEs that have also been linked to Wikidata knowledge database (Nešić et al. 2024). The repository of senses has been prepared on the basis of the Serbian WordNet (Krstev et al. 2025; Stanković et al. 2018), while the mapping of words (including MWEs) to senses is work in progress.

3. Annotation of MWE in the ELEXIS-SR

Automatic annotation of the Serbian set of 2,024 sentences with MWEs was done using different resources and tools. Resources consisted of morphological e-dictionary of Serbian simple and multi-word units, while tools were based on systems of finite-state automata that rely on these e-dictionaries (Krstev 2008).

- The e-dictionary of non-verbal MWEs (nominal, adjectival and functional) was used to annotate this type of MWEs. This dictionary contains all inflected forms of MWEs, if the words are subject to inflection, and associates them with the part-of-speech, lemma, and morphosyntactic category codes (Krstev et al. 2013). Among them were 444 nominal MWE occurrences, 80 preposition occurrences, 44 adverb occurrences, 35 conjunction occurrences, and 2 adjective occurrences.

---

8 COST Action CA21167 UniDive – universality, diversity and idiosyncrasy in language technology

Type	Tot.	TP	FN	FP	Precision	Recall	F1
NID	724	413	311	31	0.93	0.57	0.71
AdjID	5	2	3	0	1.00	0.40	0.57
AdpID	78	73	5	7	0.91	0.94	0.92
AdvID	83	38	45	6	0.86	0.46	0.60
ConjID	52	34	18	1	0.97	0.65	0.78
IRV	290	195	95	10	0.95	0.67	0.79
LVC.full	82	37	45	3	0.93	0.45	0.61
LVC.cause	5	1	4	9	0.10	0.20	0.13
VID	51	11	40	0	1.00	0.22	0.35
PronID	7	-	7	-	-	-	-
PartID	6	-	6	-	-	-	-
NV.LVCfull	2	-	2	-	-	-	
Total	1385	804	581	67	0.92	0.58	0.71

Table 4
Total number of MWEs per type; true positives, false negatives, false positives; precision, recall and F1 measure. True positives are correctly recognised MWEs, their scope and their type. For the explanation of types, see Parseme Annotation Guidelines 2.0.

- A system for the recognition of verbal MWEs based on e-dictionaries, rules, and the repertoire of VMWEs annotated in the Serbian part of the PARSEME Corpus Release 1.3 (Savary et al. 2023) retrieved 266 occurrences, distributed by type: IRV – 205, LVC.full – 40, VID – 11, and LVC.cause – 10. This system was developed for Unitex[9], a program that uses electronic dictionaries and finite state transducers (FST) for corpus analysis. For this purpose, a collection of FSTs was developed that recognises and annotates various types of VMWEs. In simplified terms, the rule for verbs of type LVC.full, which use the verb *dati* and its imperfective counterpart *davati* 'to give', would be:

  ```
  ((<dati.V>|<davati.V>) <WORD>{0,n} (primer|ocenu|mišljenje|...))|
  ((primer|ocenu|mišljenje|...) <WORD>{0,n} (<dati.V>|<davati.V>))
  ```

  In this expression `<dati.V>` and `<davati.V>` recognize all inflective forms of corresponding verbs, `<WORD>{0,n}` recognizes occurrences of 0 to n arbitrary word forms, and `primer, ocenu, mišljenje,...` are some of the predicative nouns used with these verbs ('example', 'assessment', 'opinion',...) in the expected inflected forms.

- A system for the recognition of adjectival and verbal similes described in (Krstev, Jaćimović, and Vitas 2020; Krstev, Stanković, and Marković 2023) did not retrieve a single simile in this set of sentences, which could be expected given the factual genre of sentences.

---

9 Lexicon-based Corpus Processing Suite
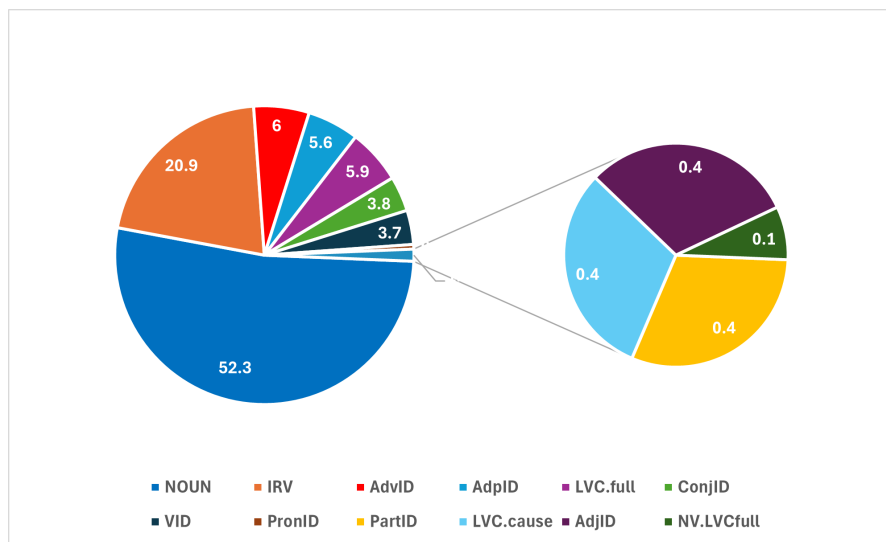
Figure 1
The distribution of MWE types in the manually verified dataset.

All automatically annotated MWEs were manually checked, and missing annotations were identified. Parseme guidelines were used to check the VMWE annotations, that is, decision trees presented in Section 1.2. Guidelines were not used for other MWE annotations (nominal and functional) because they were still in preparation at the time of this experiment, and we relied on our resources for the Serbian language. These resources record more than 20,000 MWEs of these types obtained from traditional sources and corpora research (Krstev et al. 2013). Of 871 automatically recognized MWEs, 817 were confirmed by manual check as MWEs, giving a precision of 0.94. 1385 MWEs were finally identified in ELEXIS-sr, which gives a recall of 0.59.

In some cases, the recognition was incorrect because the noun phrase was not associated with a correct verb. For instance, in sentence 38: *Ovo iskustvo imalo je snažan uticaj na njegov kasniji rad.* 'This experience had a strong influence on his later work.', system recognised as a VMWE *imati iskustvo* 'have experience' instead of *imati uticaj* 'have influence'. Such cases are a consequence of the fact that the system works locally on text that was not syntactically parsed.

Among these 817 correctly identified MWEs, only 13 were incorrectly classified. In 7 cases, the incorrect classification was due to the confusion between adverbs and prepositions, e.g. *u toku*_ADP 'during' and *u toku*_ADV 'ongoing'. In the remaining 6 cases, verbal MWEs were not appropriately classified as LVC.full, LVC.cause or VID. When we take classification into account, 804 MWEs were correctly classified, giving a precision 0.92. The annotation and check results per MWE type are presented in Table 4.

Nominal idioms (NID) are the most numerous in the entire set (52.63%), followed by inherently reflexive verbs (IRV) (20.9%) (see Figure 1). Except for pronominal and particle

idioms,[10] that were not recognised by the automatic procedures, the most omissions, relative to the total number, occurred among verbal idioms (VID, $R = 0.22$), while the fewest occurred among prepositions (AdpID, $R = 0.94$). The most incorrect recognitions occurred among causative light verb constructions (LVC.cause, $P = 0.1$), while none occurred among adjectives (AdjID, $P = 1.00$), and very few among conjunctions (ConjID, $P = 0.97$) (see Figure 2).
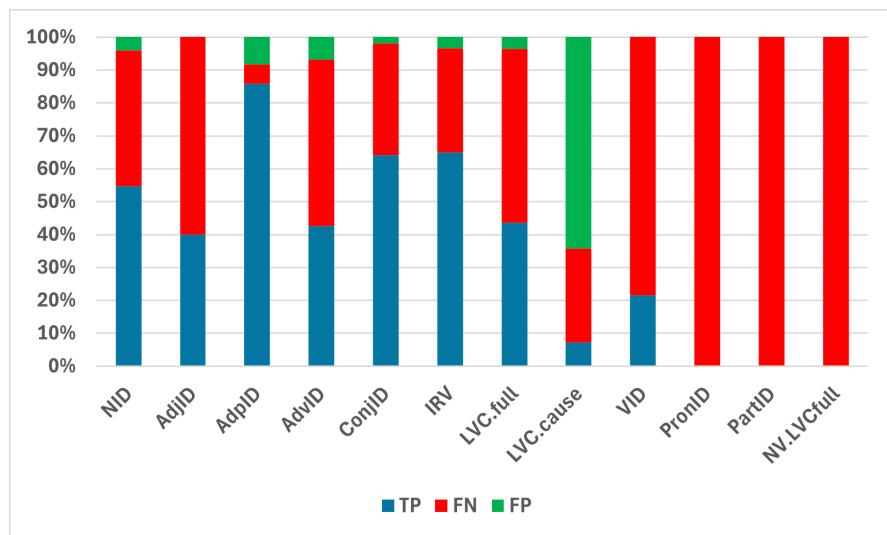

Figure 2
The percentage of true positives, false negatives and false positives per MWE type.

Five different causative LVCs were annotated in the corpus. The verb *izazvati* was used twice – *izazvati reakciju* 'arouse reaction' and *izazvati protivljenje* 'draw opposition'. The other three verbs are: *doneti* (*slavu*) 'bring (glory)', *praviti* (*problem*) 'cause (trouble)', and *napraviti* (*revoluciju*) 'make (revolution)'.

Among 82 VMWEs annotated as LVC.full, there are 56 different. The most frequent is *dobiti ime* 'lit. to get a name' with 11 occurrences, and *imati uticaj* 'have influence' with 5 occurrences. In these 56 VMWEs appear 33 different verbs. The most frequently used verb in this category is *imati* 'to have' and its negation *nemati*, occurring in 11 different VMWEs with the frequency 22. They are followed by verbs *dobiti* and *dati* 'to give' occurring in 5 different VMWEs (frequency 15) and 6 different VMWEs (frequency 6), respectively.

The noun *podrška* 'support' occurs in three different LVCs: *imati podršku* 'to have support', *dobiti podršku* 'to get support', *pružati podršku* 'to provide support'. Nouns occurring in two different LVCs are: *ime* 'name' (*dati ime* 'to give name', *dobiti ime* 'to get name'), *mišljenje* 'opinion' (*imati (visoko) mišljenje* 'to have (high) opinion', *izneti mišljenje* 'to express

---

10 Particle MWEs are not included in the current version of the Parseme Guidelines v2.0. MWEs like *na primer* 'for example' and *sve u svemu* 'all in all' were annotated as PartID.

opinion'), *nagrada* 'award' (*dobiti nagradu* 'to get award', *dodeliti/dodeljivati nagradu* 'to award prize'), *posao* 'job' (*dobiti posao* 'to get job', *obavljati posao* 'to do job'), *pravo* 'right' (*imati/nemati pravo* 'to have right', *dati pravo* 'to grant right'), *problem* 'problem' (*imati problem* 'to have problem', *rešiti problem* 'to solve problem'), *uslov* 'condition' (*zadovoljavati uslov* 'to satisfy condition', *ispunjavati uslov* 'to fulfill condition').

Some verbs occur in pairs: perfective and imperfective verbs. Those are *dodeliti* and *dodeljivati* 'to award', both used with the same noun *nagrada* 'award', and *izneti* and *iznositi* 'to state' used with different but semantically related nouns: *mišljenje*, *stav* and *viđenje* 'opinion'.

Two MWEs are annotated in this corpus as deverbal nouns derived from VMWEs, specifically LVC.full (NV.LVCfull).[11] These are *postizanje cilja* 'achieving a goal' connected to *postizati cilj* 'to achieve a goal' and *pružanje podrške* 'providing support' connected to *pružati podršku* 'to provide support'. However, the mentioned VMWEs do not occur in our corpus.

Light verb constructions are not systematically identified in Serbian monolingual dictionaries (either under verbal senses, among phrases, usage examples or collocations). The check was done for three of the most frequently used verbs in our corpus.[12] None of the 11 VMWEs that occur in our corpus using the verb *imati* 'to have' is mentioned in its dictionary entry listing 12 senses. Two LVCs evidenced in our corpus for the verb *dobiti* 'to get' are listed as usage examples in the verb's entry: *dobiti ime* 'to get a name', illustrating the sense *steći neku oznaku* 'to acquire a mark, label' and *dobiti posao* 'to get a job' related to the sense *biti postavljen na neku dužnost, položaj, biti proizveden, unapređen u neko zvanje* 'to be appointed to a certain position, to be promoted to a certain rank'. One LVC with the verb *dati*, namely *dati pravo* 'to grant right', is mentioned as a usage example of this verb in the sense *priznati, dodeliti* 'to recognize, to grant', as a part of the phrase *dati pravo građanstva* 'to grant citizenship'. The entry of this verb also states that "a verb with a noun as an object has the meaning of a verb derived from that noun." It holds for four annotated LVCs using this verb: *dati donaciju* 'to make a donation' ↔ *donirati* 'to donate', *dati doprinos* 'to make a contribute' ↔ *doprineti* 'to contribute', *dati ime* 'to give a name' ↔ 'imenovati', 'to name', *dati ocenu* 'to give a rating' ↔ *oceniti* 'to rate'.

4. Comparision of light verb constructions in the ELEXIS-WSD: sr, en, sl, bg

We compared the occurrences of LVCs in four ELEXIS-WSD sub-collections: Serbian, English, Slovene, and Bulgarian. We have taken the Serbian sub-collection as a starting point for the comparison, since it has been carefully annotated with MWEs and NEs, as explained in Section 3, while in the currently available sub-collections for other languages, the annotation of MWEs was not given priority, as discussed in Section 2. The numeric results of the comparison are presented in Table 5. The results are presented for all the occurrences of LVCs in the ELEXIS-sr, since a LVC's equivalent in another language can in one case be

---

11 This category is introduced in the version 2.0 of Parseme guidelines.
12 For this check, we used *Речник српскога језика*. 2. изд. Нови Сад: Матица српска. 2018.

a LVC of the same type, while in the other case it can be a VMWE of another type, or not a VMWE at all. For instance, the equivalents of the LVC *imati pravo* are in ELEXIS-EN a LVC 'to have right' (twice) and an adjective 'eligible' (once), while in ELEXIS-SL they are a LVC 'imeti pravico' (twice) and an adverb 'lahko' (once).

Lang.	LVC.full				LVC.cause			
	LVC candidate	ELEXIS MWE	Parseme same	diff.	LVC candidate	ELEXIS MWE	Parseme same	diff.
SR		82	24	3		5	1	0
EN	54	1	8	3	4	0	1	0
SL	40	2	6	5	3	0	0	0
BG	57	6	25	9	4	0	0	1

Table 5
Comparison of LVCs in SR, EN, SL and BG in ELEXIS and Parseme corpora.

In Table 5, results are presented separately for the light word constructions of type LVC.full and LVC.cause. The number of phrases in one of the analysed languages that correspond to LVCs in Serbian and can be candidates for LVC are given in column "LVC candidate". We treated as LVC candidates phrases that have a unique verb as the functional syntactic head, one nominal phrase as the syntactic dependent that is not the subject of the phrase, and that besides this satisfy tests LVC.0 and LVC.1 (see Subsection 1.2). This means that the referent of the nominal phrase has abstract and predicative meaning. The column "ELEXIS MWE" gives the number of LVC candidates that were annotated as MWE in the ELEXIS corpus of the corresponding language. The column "Parseme" gives the number of LVCs that were retrieved in the Parseme corpus (Savary et al. 2023): the number of LVCs that were always assigned the same type as in ELEXIS-SR is given in column "same", while the number of VMWEs (both LVC and VID) that were always or in some cases assigned a different type are given in column "diff." When retrieving VMWEs from Parseme corpus, we overlooked differences in conjugation or declension. We also treated as a match those cases in which the used verb has a different aspect, for instance in Serbian *ispuniti obavezu* (perfective) and *ispunjavati obavezu* (imperfective) 'fulfill an obligation' or in Bulgarian *изпълня ангажимент* (perfective) and *изпълнявам ангажимент* (imperfective). For instance:

- A LVC.full *iskoristiti prednost* used in sentence 588 (+1 in column 'ELEXIS MWE' for SR) has as an equivalent in EN a LVC candidate *take advantage* (+1 in column 'LVC candidate' for EN) that was annotated as MWE in ELEXIS-EN (+1 in column 'ELEXIS MWE' for EN). This LVC occurs in Parseme-EN three times: twice it is annotated as LVC.full and once as VID (+1 in column 'Parseme diff.' for EN)

- A LVC.full *dati donaciju* used in sentence 1950 (+1 in column 'ELEXIS MWE' for SR) has as an equivalent in EN a LVC candidate *make donation* (+1 in column 'LVC candidate' for EN), it was not annotated as MWE in ELEXIS-EN

(+0 in column 'ELEXIS MWE' for EN), but it occurs in Parseme corpus where it is annotated as LVC.full (+1 in column 'Parseme same' for EN).[13]

The phrase *imati pravo*, which occurs three times in ELEXIS-SR and is annotated as LVC.full, is the only phrase for which in other languages the same equivalent phrase was used and annotated as LVC.full.[14] One example is (sentence number 1790):

SR   Narod *ima pravo*_{LVC.full} da od ombudsmana traži intervenciju.

EN   The people *have right*_{LVC.full} to ask the ombudsman for intervention.

SL   Ljudje *imajo pravico*_{LVC.full} zaprositi varuha človekovih pravic za posredovanje.

BG   Хората *имат право*_{LVC.full} да помолят омбудсмана да се намеси.

The phrase *imati za cilj* 'to have aim', which also occurs three times in ELEXIS-SR and is annotated as LVC.full never has as the equivalent in other languages the LVC.full phrase nor any VMWE phrase of other type.[15] One example is (sentence number 178):

SR   Evropska komisija *ima za cilj*_{LVC.full} da nagradama podstakne prekograničnu cirkulaciju popularnog muzičkog repertoara...

EN   With the awards, the European Commission *aims* to stimulate the cross-border circulation of popular music repertoire...

SL   Z nagrado *želi* Evropska komisija spodbuditi čezmejno kroženje repertoarja popularne glasbe...

BG   С тези награди ЕК *цели* да се стимулира разпространението на популярна музика в различните държави...

The phrase *imati/nemati uticaj* 'to have (no) effect', which occurs six times in ELEXIS-SR and is annotated as LVC.full, has in one case as the equivalent in other languages the VMWE phrase, but its category varies.[16] The example is (sentence number 1551):

SR   Vaspitni stil roditelja izgleda *nema* veliki *uticaj*_{LVC.full},...

EN   Parenting style seems to *have* no major *effect*_{LVC.full},...

SL   Zdi se, da slog starševstva *nima* veliko *vpliva*_{LVC.cause},...

BG   Изглежда родителският стил *няма* голям *ефект*_{LVC.full/VID},...[17]

---

13 This LVC.full occurs in Parseme-EN once.

14 The similar phrase is used in other languages and annotated as LVC.full in the Parseme corpus, for example: HR *imati pravo*, FR *avoir droit*, PL *mieć prawa*.

15 The similar phrase is used in other languages and annotated as LVC.full in the Parseme corpus, for example: HR*imati cilj*, FR *avoir but*, PL *mieć cel*, SL *imeti cilj* (although it was not used in ELEXIS-SL.).

16 The similar phrase used in other languages is annotated as LVC.full in the Parseme corpus, for example: HR*imati utjecaj*, FR *avoir influence*, PL *mieć wpływ*.

17 In Bulgarian Parseme corpus this VMWE is twice annotated as LVC.full, once as VID, and one occurrence is skipped.

An interesting issue arises with this phrase, which has equivalents of different verbal types. We conclude that this particular LVC causes problems in that its meaning is causative, but its type is LVC.full. This is similar to the cases which contain typical causative verbs, but their type is LVC.full. There is a subtle distinction between the cause subject as a canonical argument (LVC.full) and its "external", non-canonical use (LVC.cause).

As mentioned in Section 3, the verb *imati/nemati* is used in the largest number of LVCs (11). In cases when an LVC is also used in another language, it is always *to have* in English, *imeti* in Slovene, and *имам/нямам* and *окажа/оказвам* in Bulgarian. The verb *dati* used in six LVCs, has equivalents in English with verbs 'to give' and 'to make', and in Bulgarian with verbs 'имам', 'дам/давам' and 'подавам'. The verb *dobiti* 'to get', used in five different LVCs, has no equivalents in English and Slovene, but two equivalents in Bulgarian: *dobiti ime* 'to give name' ↔ 'получавам име' and *dobiti posao* 'to get job' ↔ 'наема на работа'. In the last case, there is no true equivalence, since the Bulgarian phrase is an exact translation of English 'to hire for a job'. Serbian light verb constructions with *imati* and *dati* that have equivalents in other languages are represented in Table 6.

Serbian LVC		Equivalent
imati/nemati pravo	EN	to have (no) right
	SL	imeti pravico
	BG	имам право
imati/nemati uticaj	EN	to have (no) effect
	SL	imeti vpliv
	BG	окаже влияние, имам влияние, окажа/оказвам въздействие
imati/nemati problem	EN	have (no) problem
	SL	imeti težave
imati/nemati dejstvo	EN	to have (no) effect
	BG	оказвам влияние
imati/nemati šansu	SL	imeti možnost
imati pristup	BG	имам достъп
dati pravo	EN	to give right
	BG	давам право
dati donaciju	EN	to make donation
dati ocenu	BG	дам оценка
dati doprinos	BG	имам принос
dati ostavku	BG	подавам оставка

Table 6
Serbian LVCs with verbs *imati* and *dati* and their equivalents in EN, SL, BG.

The data in Table 5 indicate that a significant number of candidates were not marked as LVCs in the Parseme corpus. However, we cannot conclude that these candidates cannot be considered LVCs in the analysed languages, because our analysis did not include checking whether these candidates occurred in that corpus at all. One such example is (sentence 1699):

SR   *Dobila*_{LVC.full} je *podršku* mnogih bogatih i uticajnih sponzora.

EN   It *won*_{LVCcandidate} the *support* of many wealthy and influential backers.

SL   *Uživalo*_{LVCcandidate} je *podporo* številnih bogatih in vplivnih podpornikov.

BG   Беше *спечелена*_{LVCcandidate} *подкрепата* на много заможни и влиятелни поддръжници.

However, the equivalent LVC.full is annotated in the Polish part of the Parseme corpus: *otrzymywać wsparcie*, in the Bulgarian part (using a different verb): *получа/получавам подкрепа*, as well as in the French part: *recevoir soutien*. It should also be noted that the phrases *win / get support* do not occur in the Parseme-EN, and neither do *uživati podporo* in Parseme-SL (however, the near synonym *pridobiti podporo* occurs once in Parseme-SL and is not annotated as VMWE).

It should be stated here that some of the annotated LVCs in the ELEXIS-SR raise doubt. We can take as an example LVC.full *izneti/iznositi mišljenje/stav/viđenje* 'express/present opinion/view'. These phrases pass test LVC.3 (the subject of the verb is a semantic argument of the noun), but can we say that the paraphrase 'subject's opinion/view' expresses the same meaning? Our position here was that 'opinion/view' has to be expressed much as a lecture has to be given. We see in Table 7 that various similar phrases were used across languages, none of which was annotated as LVC.full (or other type of VMWE) either in Parseme or in the ELEXIS corpus. Corpus search using the GrewMatch tool[18] (Guillaume 2021) reveals that *express/present/put forth opinion/view* is not used in Parseme-EN, *predstaviti/izraziti stališče/mnenje* is used four times in Parseme-SL and never annotated as VMWE, while *представя/изразя-(се)/изкажа поглед/гледен точка/мнение* occurs 14 times in Parseme-BG in the form *изразя-(се) мнение* and it is 10 times annotated as LVC.full and 7 times skipped, while the form *изкажа мнение* occurs twice and is both times skipped. Finally, we should add that the annotation of Parseme-SR was not more consistent: *izneti/iznositi mišljenje/stav/viđenje* occurs 6 times: three occurrences of *izneti stav* were annotated as LVC.full, one is skipped, while both single occurrences of *izneti mišljenje* and *izneti viđenje* were skipped.

Sent.	ELEXIS-SR	ELEXIS-EN	ELEXIS-SL	ELEXIS-BG
23	iznositi viđenje	present view	predstaviti sališče	представя поглед
156	izneti stav	express view	izraziti stališče	изразя гледен точка
1731	izneti mišljenje	put forth opinion	biti mnjenja	изкажа мнение

Table 7
LVC.full *izneti/iznositi mišljenje/stav/viđenje* and its equivalents in ELEXIS-EN/SL/BG.

The data from Table 5 further show that the most equivalents among LVCs are between Serbian and Bulgarian (a total of 34), while between Serbian and English, as well as Serbian and Slovene, there are a total of 11 equivalents. This does not seem to depend on the size of

---

18 Grew-match

the Parseme corpus, as the Slovenian part is the largest but has significantly fewer LVCs than the Bulgarian, which is shorter in length (see Table 1). The Serbian corpus is the smallest of all but has a number of LVCs comparable to Slovene and English. It may be that the type of text plays a role in this or even the strategy that the annotators applied when annotating their corpora.

5. Conclusion

This paper deals with light verb constructions and their annotation in ELEXIS-sr, the Serbian extension of the ELEXIS-WSD corpus. We made general introductory remarks about these constructions, the notion of light verbs, and their treatment and further classification in the PARSEME annotation guidelines (subtypes LVC.full and LVC.cause).

Section 2 offers an insight into the ELEXIS-WSD corpus, annotated with VMWEs for several languages, with a remark that these VMWEs weren't further sub-categorized into smaller classes. For this paper, we classified them ourselves to be able to make comparisons of the LVCs annotated in ELEXIS-sr.

In Section 3 we presented tools and resources used in the automatic annotation with MWEs of ELEXIS-sr, as well as the results of manual checking. There are 5 VMWEs annotated as LVC.cause (all different) and 82 annotated as LVC.full (among them 56 different), with 33 different verbs, the most frequent being *imati* 'to have', followed by *dobiti* 'to get' and *dati* 'to give'. As for the nominal part of LVCs, the noun *podrška* 'support' appears in three different LVCs, and nouns *ime* 'name', *mišljenje* 'opinion', *nagrada* 'award', *posao* 'job', *pravo* 'right', *problem* 'problem', *uslov* 'condition' occur in two different LVCs each. We checked whether these and other LVCs are identified and represented in the macro- or microstructure of a monolingual Serbian dictionary and conclude that, as a rule, LVCs are usually not identified (with a few exceptions, which occur in different parts of the dictionary structure).

In Section 4, we offer a comparison of LVCs in four ELEXIS-WSD sub-collections: Serbian, Bulgarian, Slovene, and English. We use Serbian as a starting point for this comparison, as it has been manually checked for annotation with MWEs (and NEs). We present the results for all the occurrences of LVCs in the Serbian extension. Sometimes LVC has an equivalent LVC in another language, sometimes VMWE of another type, and sometimes a single word.

We took LVC candidates from the ELEXIS subcollections of other languages, as well as from the Parseme corpus, separately for LVC.cause and LVC. full types. These comparisons gave us the following insight: Only one phrase, *imati pravo*, annotated in ELEXIS-sr as LVC.full, has as its equivalent the same phrase, annotated as LVC.full in other sub-collections. On the contrary, sr *imati za cilj*, 'to have aim', never has either LVC.full phrase as the equivalent, or any other VMWE type.

An important conclusion is that the most equivalents among LVCs are between Serbian and Bulgarian, closely related Slavic languages (a total of 34 equivalents), while between Serbian and Slovene, also Slavic, there are 11 equivalents, as between Serbian and English. It seems that this could be explained by the number of VMWEs and LVCs annotated, or by the strategy used by different annotators.

This research showed that although LVCs are a universal phenomenon, their repertoire in each language has to be established separately, and their translation from one language to another has an idiomatic character. This is the reason why facts collected through lexicological inventorying facilitate the work on LVCs.

The results obtained from the presented research suggest that its extension to other languages and to other types of VMWEs could yield interesting results. Considering that the annotation of a multilingual corpus with other types of top-level MWEs (nominal, adjectival, adpositional, etc.) is being prepared within the framework of the UniDive COST action, the research could be extended in that direction as well.

Acknowledgments

References

Gantar, Polona, Špela Arhar Holdt, Jaka Čibej, and Taja Kuzman. 2019a. Structural and Semantic Classification of Verbal Multi-Word Expressions in Slovene. *Contributions to Contemporary History*, 59(1):99–119.

Gantar, Polona, Lut Colman, Carla Parra Escartín, and Héctor Martínez Alonso. 2019b. Multiword expressions: between lexicography and NLP. *International Journal of Lexicography*, 32(2):138–162.

Guillaume, Bruno. 2021. Graph Matching and Graph Rewriting: GREW tools for corpus exploration, maintenance and conversion. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 168–175, Association for Computational Linguistics, Online.

Ivić, Milka. 1988. Još o dekomponovanju predikata. *Južnoslovenski filolog*, 44:1–5.

Jespersen, Otto. 1965. A modern English grammar on historical principles.

Koeva, Svetla, Ivelina Stoyanova, Maria Todorova, and Svetlozara Leseva. 2016. Semi-automatic compilation of the dictionary of Bulgarian multiword expressions. In *Proceedings of the Workshop on Lexicographic Resources for Human Language Technology*, pages 86–95.

Krstev, Cvetana. 2008. *Processing of Serbian – Automata, Texts and Electronic Dictionaries*. University of Belgrade, Faculty of Philology, Belgrade.

Krstev, Cvetana, Jelena Jaćimović, and Duško Vitas. 2020. Analysis of Similes in Serbian Literary Texts (1840-1920) Using Computational Methods. In *Proceedings of the Fourth International Conference Computational Linguistics in Bulgaria (CLIB 2020)*, pages 31–41, Institute for Bulgarian Language "Prof. Lyubomir Andreychin", Bulgarian Academy of Sciences, Sofia, Bulgaria.

Krstev, Cvetana, Ivan Obradović, Ranka Stanković, and Duško Vitas. 2013. An approach to efficient processing of multi-word units. *Computational Linguistics: Applications*, pages 109–129.

Krstev, Cvetana, Ranka Stanković, and Aleksandra Marković. 2023. Multiword Expressions – Comparative Analysis Based on Aligned Corpora. In *Book of Abstracts of the UniDive 1st general meeting, 16-17 March 2023, Paris-Saclay University, France*.

Krstev, Cvetana, Ranka Stanković, Aleksandra Marković, and Milica Ikonić Nešić. 2025. Progress in SR-ELEXIS Semantic Annotation: Focusing on Multiword Expressions, Named Entities, and Sense Repository. In *Book of Abstracts of the UniDive 3rd general meeting, 29-30 January 2025, Budapest, Hungary*.

Lazić-Konjik, Ivana. 2006. Struktura, funkcija i leksikografska obrada perifrastičnih predikata (na primerima iz dnevne štampe). *Zbornik Matice srpske za filologiju i lingvistiku*, 49(1):219–304.

Leseva, Svetlozara, Verginica Barbu Mititelu, Ivelina Stoyanova, and Mihaela Cristescu. 2024. A uniform multilingual approach to the description of multiword expressions. *Multiword expressions in lexical resources: Linguistic, lexicographic, and computational perspectives*, pages 73–116.

Martelli, Federico, Roberto Navigli, Simon Krek, Jelena Kallas, Polona Gantar, Svetla Koeva, Sanni Nimb, Sussi Olsen, Margit Langemets, Kristina Koppel, et al. 2021. Designing the ELEXIS Parallel Sense-Annotated Dataset in 10 European Languages. In *Proceedings of the eLex 2021 conference*, pages 377–395, Lexical Computing.

Nešić, Milica Ikonić, Saša Petalinkar, Ranka Stanković, Miloš Utvić, and Olivera Kitanović. 2024. SrpCNNeL: Serbian Model for Named Entity Linking. In *2024 19th Conference on Computer Science and Intelligence Systems (FedCSIS)*, pages 465–473, IEEE.

Pedersen, Bolette, Sanni Nimb, Sussi Olsen, Thomas Troelsgård, Ida Flörke, Jonas Jensen, and Henrik Lorentzen. 2023. The DA-ELEXIS corpus - a sense-annotated corpus for Danish with parallel annotations for nine European languages. In *Proceedings of the Second Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2023)*, pages 11–18, Association for Computational Linguistics, Tórshavn, the Faroe Islands.

Radovanović, Milorad. 1990. *Spisi iz sintakse i semantike.* Izdavačka knjižarnica Zorana Stojanovića.

Samardžić, Tanja. 2007. Light verb constructions in English and Serbian. *English Language and Literature Studies–Structures across Cultures*, pages 59–73.

Savary, Agata, Cherifa Ben Khelil, Carlos Ramisch, Voula Giouli, Verginica Barbu Mititelu, Najet Hadj Mohamed, Cvetana Krstev, Chaya Liebeskind, Hongzhi Xu, Sara Stymne, Tunga Güngör, Thomas Pickard, Bruno Guillaume, Eduard Bejček, Archna Bhatia, Marie Candito, Polona Gantar, Uxoa Iñurrieta, Albert Gatt, Jolanta Kovalevskaite, Timm Lichte, Nikola Ljubešić, Johanna Monti, Carla Parra Escartín, Mehrnoush Shamsfard, Ivelina Stoyanova, Veronika Vincze, and Abigail Walsh. 2023. PARSEME corpus release 1.3. In *Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023)*, pages 24–35, Association for Computational Linguistics, Dubrovnik, Croatia.

Stanković, Ranka, Miljana Mladenović, Ivan Obradović, Marko Vitas, and Cvetana Krstev. 2018. Resource-based WordNet Augmentation and Enrichment. In *Proceedings of the Third International Conference Computational Linguistics in Bulgaria (CLIB 2018)*, pages 104–114, Institute for Bulgarian Language "Prof. Lyubomir Andreychin", Bulgarian Academy of Sciences, Sofia, Bulgaria.

Stanković, Ranka, Branislava Šandrih, Cvetana Krstev, Miloš Utvić, and Mihailo Škorić. 2020. Machine Learning and Deep Neural Network-Based Lemmatization and Morphosyntactic Tagging for Serbian. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3954–3962, European Language Resources Association, Marseille, France.

Stanković, Ranka, Mihailo Škorić, and Branislava Šandrih Todorović. 2022. Parallel Bidirectionally Pretrained Taggers as Feature Generators. *Applied Sciences*, 12(10).

Stoyanova, Ivelina, Svetlozara Leseva, and Maria Todorova. 2016. Towards the automatic identification of light verb constructions in Bulgarian. In *Proceedings of the Second International Conference on Computational Linguistics in Bulgaria (CLIB 2016)*, pages 28–37, Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences, Sofia, Bulgaria.

Topolińska, Zuzanna. 1982. Perifrastični predikatski izrazi na međuslovenskim relacijama. *Južnoslovenski filolog*, 38:35–49.

Wittenberg, Eva. 2016. *With light verb constructions from syntax to concepts*, volume 7. Universitätsverlag Potsdam.