EURALEX XIX

Congress of the European Association for Lexicography

Lexicography for inclusion

7-11 September 2021

Ramada Plaza Thraki

Alexandroupolis, Greece

www.euralex2020.gr

Proceedings Book
Volume 1

Edited by Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

EURALEX Proceedings

ISSN 2521-7100

ISBN 978-618-85138-1-5

Edited by: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris

2020 Edition

CROATPAS: A Lexicographic Resource for Croatian Verbs and its Potential for Croatian Language Teaching

Marini C., Ježek E.

University of Pavia, Italy

Abstract

This paper revolves around CROATPAS (Marini & Ježek 2019), a digital lexicographic resource for Croatian verbs able to frame verbal polysemy and metonymic shifts, which is currently being developed at the University of Pavia. Just like its Italian sister resource T-PAS (Ježek et al. 2014), CROATPAS is a corpus-derived collection of verb argument structures whose argument slots have been manually annotated using a specific set of semantic labels called Semantic Types. At the moment, the resource contains 101 verb entries linked to 457 different verb senses (called *patterns*) and over 22,000 annotated corpus lines (Marini & Ježek 2020). The possible applications of CROATPAS are endless. However, given the status of Croatian as an *under-resourced* and Less Commonly Taught Language, this paper focuses on its potential as a language teaching tool, putting forward some hypothetical vocabulary and grammar teaching suggestions. Even though CROATPAS is still in the early stages, its user-friendly interface, bilingual nature and focus on verb semantics bode well for its future as a tool for the teaching of Croatian as a Foreign Language.

Keywords: Croatian; semantic resource; verb; language teaching; Less Commonly Taught Language

1 Introduction

In this paper, we present a new corpus-derived lexicographic resource for Croatian called CROATPAS (Marini & Ježek 2019). The resource focuses on verb semantics and polysemy; it currently contains 101 verb entries, 457 verb senses and is linked to over 22,000 annotated corpus lines (Marini & Ježek 2020). CROATPAS relies on a sound theoretical background and a well-established lexicographic methodology, which are thoroughly dealt with in section 2. Even though its possible applications are endless, here we explore its potential as a language teaching tool, putting forward some of the vocabulary and grammar teaching activities already offered by its pattern inventory (see § 3). In section 4, an overview of the currently available resources for Croatian verbs is provided, stressing their differences with respect to CROATPAS. In light of its user-friendly interface, bilingual nature and focus on verb semantics, we believe our resource has the potential to become a truly useful tool for the teaching of Croatian as a Foreign Language.

2 CROATPAS

The CROATian Typed Predicate Argument Structures resource (CROATPAS, Marini & Ježek 2019) is a digital lexicographic resource containing a corpus-based collection of manually annotated Croatian verb valency structures with the addition of Semantic Type labels on their argument slots (henceforth, SemTypes). Since each semantically typed verb argument structure – informally called *pattern* – is linked to a different verb sense, the resource is primarily tailored for investigating verbal polysemy. For instance, *[Human]*$_{nominative}$ *pije [Beverage]*$_{accusative}$ and *[Human]*$_{nominative}$ *pije [Drug]*$_{accusative}$ are two of the identified patterns of the Croatian verb *piti* (English, *to drink*), which correspond to the following senses: that of *drinking a beverage* and that of *swallowing a medicine*, respectively. The resource is also able to frame recurrent metonymic shifts in verb arguments, which are encoded in *sub patterns* nested within the patterns they stem from (see § 2.2).

At the moment, CROATPAS contains 101 verb entries, which are linked to 457 patterns, 106 metonymic sub patterns and over 22,000 annotated corpus lines (Marini & Ježek 2020). Like its Italian sister project T-PAS (Ježek et al. 2014, containing 1160 analysed verbs, 8000 senses, and approximately 190,000 annotated concordances), the Croatian resource is being developed at the University of Pavia with the technical support of *Lexical Computing Ltd.* and will be available online in late 2020. The four components the resource relies on are:

- a representative corpus of contemporary Croatian
- a set of semantic labels to tag argument slots with
- a sound corpus-based lexicographic methodology to identify different verb senses
- the adequate corpus tools

As far as the corpus choice is concerned, the Croatian Web as Corpus (hrWac 2.2, Ljubešić & Erjavec 2011) was chosen as the reference corpus for CROATPAS in order to maximise compatibility with the Italian T-PAS resource (Ježek et al. 2014), since the latter is linked to a reduced version of the Italian Web as Corpus (ItWaC, Baroni & Kilgarriff 2006). As for the semantic labels, CROATPAS takes advantage of a shallow ontology of approximately 180 SemTypes developed for the T-PAS project and called System of Semantic Types (Ježek 2019). In terms of corpus tools, the main software CROATPAS makes use of is a pattern editing environment linked to the *Sketch Engine* (Kilgarriff et al. 2014) called *Skema* (Baisa et al. 2020). Skema was originally developed by *Lexical Computing Ltd.* for the Italian T-PAS resource, but

it was later customised not only to the needs of CROATPAS, but also of several other projects focused on different languages, such as the *Woordcombinaties* projects for Dutch verbs (Colman & Tiberius 2018). Finally, as far as methodology is concerned, the resource relies on a customised version of Corpus Pattern Analysis (CPA, Hanks 2013), a lexicographic methodology resting on the idea that meaning should be mapped onto its prototypical contexts of use, which was first put into practice in the *Pattern Dictionary of English Verbs* (Hanks & Pustejovsky 2005). CPA usually follows four steps: 1) 250 corpus lines are randomly sampled for each verb; 2) the different verb senses are identified through extensive lexicographic analysis; 3) pattern strings are created in Skema labelling argument slots with the right SemTypes and, finally, 4) numbers are assigned to the corpus lines exemplifying each identified pattern, in order for each semantically tagged valency structure to be justified by corpus evidence.

2.1 Understanding Patterns

From a theoretical point of view, both T-PAS and CROATPAS rely on the Generative Lexicon framework and its principles for strong compositionality, namely the principle of co-composition and the principle of Semantic Type Coercion (Pustejovsky 1995 & 1998; Pustejovsky & Ježek 2008; Ježek 2016).
According to the principle of co-composition, lexical items expressing verb arguments are to be considered as semantically active in the contextual generation of verb meaning as the verb itself. In light of this, verbal polysemy can be traced back to the compositional operations taking place between the verb and the SemTypes associated to its surrounding arguments. For instance, as we can see from Figure 1, the Croatian verb *voziti* (English, *to drive*) takes on different meanings depending on what is said to be *driven*.

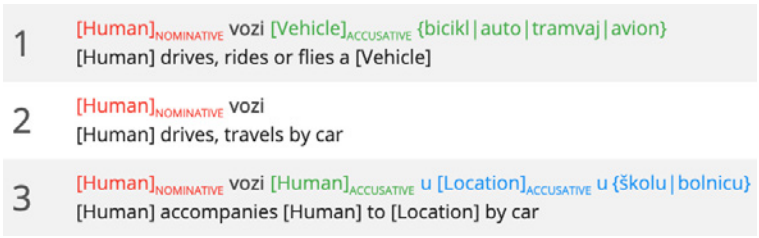

Figure 1: Some of the patterns connected to the verb *voziti* (English, *to drive*) in CROATPAS

If a *[Human] drives a [Vehicle]* as in pattern (1), then he or she is "operating that vehicle in order for it to move". However, if a *[Human] drives another [Human] to a [Location]* as in pattern (3), then he or she is "accompanying that person to a certain destination, usually by car". The meaning of each pattern string is explained in English in the line underneath the pattern, which takes the name of *sense description*. The choice of the right SemType for each argument slot in a pattern is made by the lexicographer on the basis of the lexical items found in that slot in the corpus lines linked to that pattern: for instance, the SemType [Vehicle] in pattern (1) is justified by corpus examples featuring direct objects such as {*bicikl* = bicycle | *auto* = car | *tramvaj* = tram | *avion* = plane}, which make up a the so-called *lexical set* for the SemType [Vehicle] in this context.

2.2 Understanding Subpatterns

As for Semantic Type Coercions, that is how metonymic shifts are called in a Generative Lexicon perspective (Pustejovsky & Ježek 2008). Coercions take place when a verb's selectional requirements in terms of semantic typing are not satisfied by one of its arguments, but no change in verb meaning is observed. For instance, if we look at the corpus line highlighted in Figure 2 – *Ako ne voziš BMW, ti si nitko i ništa* (English, *If you do not drive a BMW, you are nobody and nothing*) – we can see that, even though the example conveys the same meaning encoded in pattern 1 from Figure 1, the direct object of *voziti* (English, *to drive*) is not a [Vehicle] *per se*, but a [Business Enterprise] producing [Vehicle]s.

Figure 2: Some of the metonymic corpus lines linked to the sub pattern 1.1.m for the verb *voziti* (English, *to drive*) in CROATPAS.

When someone refers to a [Vehicle] in terms of the [Business Enterprise] – and specifically the *Automobile Company* – that produces it, then we are witnessing a Semantic Type Coercion from [Business Enterprise] → [Vehicle], which can be listed as an instance of the pervasive metonymy *Producer/Product* (Pustejovsky 1995: 25). Both T-PAS and CROATPAS encode Coercions as sub patterns ending in ".m" (which stands for *metonymic*), as you can see in Figure 3.

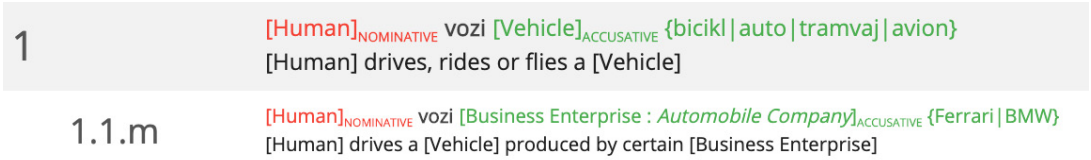

Figure 3: Pattern 1 and its metonymic sub pattern 1.1.m for the verb *voziti* (English, *to drive*) in CROATPAS

3 CROATPAS as a Language Teaching Tool

After presenting the main features of CROATPAS, we now turn to its applications. Its potential uses are countless and they entail, *inter alia*, corpus-based linguistic research on verbal polysemy and metonymies (Marini & Ježek 2020), but also computational applications (e.g. machine translation enhancement) based on multilingual pattern linking with other monolingual CPA-inspired resources, such as T-PAS for Italian (Ježek et al. 2014), PDEV for English (Hanks & Pustejovsky 2005) or *Woordcombinaties* for Dutch (Colman & Tiberius 2018).

Moreover, given the status of Croatian as both an *under-resourced* (Tadić et al. 2014) and a Less Commonly Taught Language (LCTL) in need of attention (Mikelić Preradović et al. 2019), CROATPAS could also become a useful tool for teachers of Croatian as a Foreign Language. In the following, we focus on the latter application.

According to the widely accepted principles of Communicative Language Teaching (Brown & Lee 2015: 31), the best practice in language teaching promotes exposure of language learners to real-life communication and meaningful input embedded in naturally occurring language, which is exactly what corpus-based language teaching has been offering over the past few decades, especially thanks to tools such as corpus concordances and collocation lists (*ivi*, 62). More recently, due to the increasing integration of technology in our daily lives and the development of the Web, we have also witnessed the rise of Computer Assisted Language Learning (CALL), which is characterised, among other things, by an increased "interactive communication and collaboration [among language learners] via the Internet" (*ivi*, 238).

Although existing research has proved the effectiveness of corpus usage on both grammar learning, vocabulary learning and language awareness (Chan & Liou 2005; Liu & Jiang 2009; Lee et al. 2018), research on LCTLs is scarce (Ward 2016). According to the results of a recent study on teaching Croatian as a Foreign Language involving learners with different native languages and proficiency levels (Mikelić Preradović et al. 2019), student response to corpus-based material in experimental classes was mostly positive: beginners showed higher levels of appreciation for the introduction of corpora than advanced learners, while most intermediate learners enjoyed discovering corpus patterns, but were easily overwhelmed.

Given the well-known difficulties connected to learning verb/noun collocations even at advanced level (Nesselhauf 2003) and the attested positive influence of bilingual CALL tools on verb/noun collocation learning (Chan & Liou 2005), we believe that using CROATPAS in the classroom could help learners improve their understanding and usage of Croatian verb patterns.

In the rest of this section, we suggest some vocabulary and grammar teaching activities which could already be carried out with CROATPAS (see § 3.1 and 3.2), and we discuss the possibility of integrating some SkELL-inspired features in its interface (see § 3.3).

3.1 Teaching Vocabulary with *Lexical Sets* and SemTypes

In addition to being linked to corpus examples providing evidence for their existence, CROATPAS's patterns feature a selection of manually identified verb collocates in each argument slot (see Figure 4). These collocate lists take the name of *lexical sets* (Hanks & Ježek 2008), they are portrayed in braces after their respective SemType and can be exploited in vocabulary teaching activities focusing on the semantic areas of their respective SemTypes.

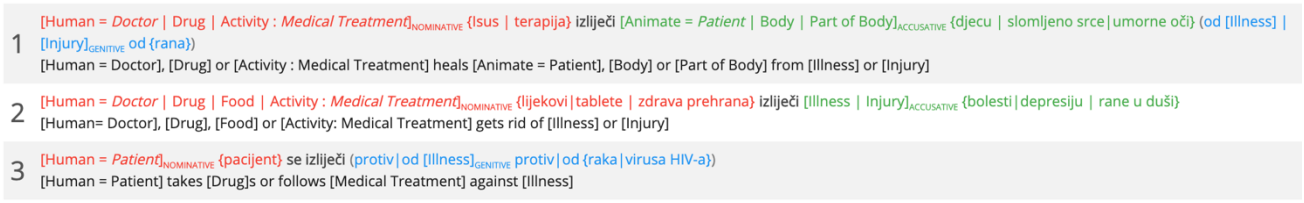

Figure 4: The first three patterns of the verb *izliječiti* (English, *to heal*) in CROATPAS

For instance, since the verb *izliječiti* (English, *to heal*) involves SemTypes such as [Drug], [Activity: Medical Treatment], [Illness] and [Injury] in its pattern strings, the lexical sets of these SemTypes are bound to feature collocates pertaining to the medical semantic area, such as *lijekovi* (English, *meds*), *terapija* (English, *therapy*), *rana* (English, *wound*), *rak* (English, *cancer*), *bolest* (English, *illness*). If users could access a SemType search menu able to query the pattern inventory, they could easily retrieve all the verbs featuring the desired SemType and all its related terminology.

Moreover, teachers could exploit the fact that each lexical item featuring in a lexical set is adapted to its grammatical context both in terms of morphological inflection and required preposition. For instance, if we look at pattern 3 from

Figure 4, we can see that all the items in the lexical set of the SemType [Illness] bear a genitive singular ending, as required by both the preposition *protiv* and *od*: *od raka; protiv virusa HIV-a* (English, *against cancer; against the HIV virus*).

3.2 Teaching Case Inflection with *Patterns*

Being a Slavic language, Croatian is equipped with a case system consisting of seven different morphological cases, namely nominative, genitive, dative, accusative, vocative, locative and instrumental (Barić et al. 1997: 101).
Since Croatian does not have a fixed word order, in patterns such as the one in Figure 5 portraying the verb *dočekati* (English, *to welcome*), where both arguments are semantically typed as [Human], it is *case* that allows us to understand the grammatical relations between sentence components, i.e. which of the two arguments is the *welcoming* subject expressed by the nominative case and which is the *welcomed* object in the accusative case.

Figure 5: Pattern 1 of the verb *dočekati* (English, *to welcome*) in CROATPAS

Even though a resource mainly focusing on verbal polysemy like CROATPAS might be more suitable for intermediate and advanced learners, both the graphical rendering of case markings as bottom-right indexes and the argument colour-coding strategy might be useful devices for teachers to introduce absolute beginners to the concept of case, especially before they have internalized all the different inflectional endings for the different classes of nouns.

3.3 Combining *Patterns*, Concordances and Collocations

In order to become an ever more user-friendly tool for teachers and learners of Croatian as a Foreign Language, CROATPAS could also take example from the Dutch *Woordcombinaties* project[1] (Colman & Tiberius 2018), which – to our knowledge – is the only verb-centred CPA-inspired lexicographic resource equipped with both an inventory of pattern-meaning pairs and some of the features of a SkELL (*Sketch Engine for Language Learning*) interface (Kilgarriff et al. 2015). As you can see from Figure 6, *Woordcombinaties* users have access to:

- corpus concordance lines portraying Good Dictionary Examples for each verb, i.e. prototypical *example sentences* (Dutch, *voorbeeldzinnen*) automatically extracted by the GDEX algorithm developed by Kilgarriff et al. (2008);
- the verbs' *combination possibilities* (Dutch, *combinatiemogelijkheden*), i.e. a list of the lexical items they are usually found together with divided per word class, sketching their grammatical and collocational behaviour;
- the verbs' different senses encoded in *patterns* (Dutch, *patronen*) based on their valency structures, albeit without overt semantic typing.

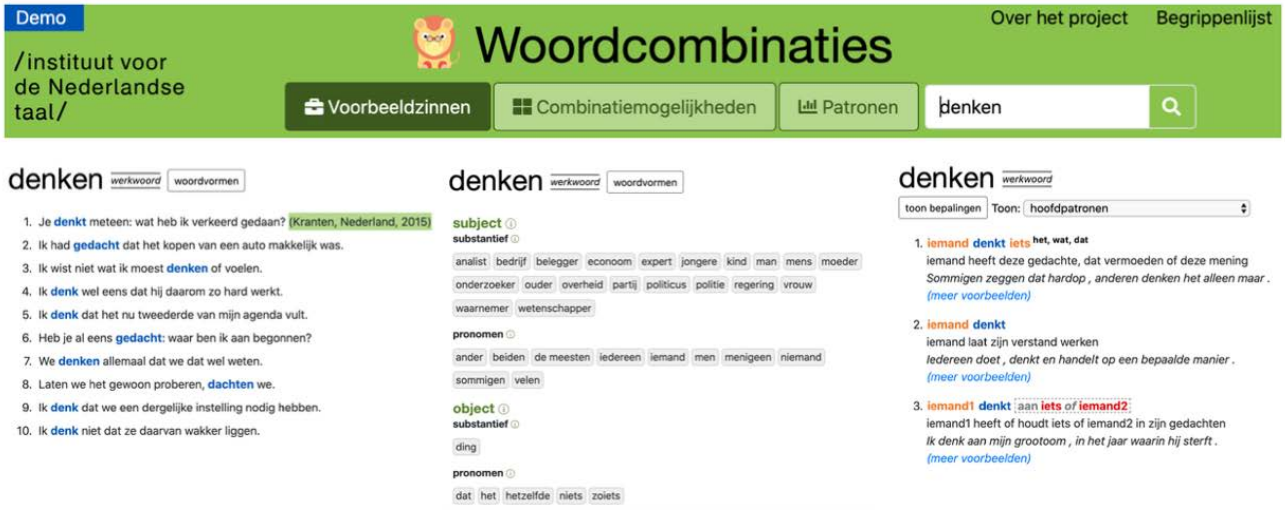

Figure 6: Some example sentences, combination possibilities and verb patterns of the Dutch verb *denken* (English, *to think*) taken from the *Woordcombinaties* project interface

In the case of Croatian, it might be interesting to bring this hybrid approach a step further integrating other SkELL-inspired features in the tool interface, such as the possibility to generate word clouds displaying verb collocates divided not only per verb argument but also per specific verb sense, i.e. *pattern*.

[1] http://woordcombinaties.ivdnt.org/ Website last visited [05/05/2020].

4 Comparing CROATPAS with other Resources for Croatian Verbs

As far as Croatian verbs are concerned, the main currently available resources are: 1) the e-Glava[2] Valency Database of Croatian Verbs (Birtić et al. 2017); 2) the Croatian Valence Lexicon of Verbs[3] (CroVallex, Mikelić Preradović et al. 2009); 3) the Croatian Derivational Lexicon[4] (CroDeriV, Šojat et al. 2014; Filko et al. 2019) and 4) the Nooj Croatian Dictionary of Aspectual Derivatives[5] (Kocijan et al. 2018; Šojat et al. 2018).

The e-Glava project focuses on a sample of 57 psychological verbs, whose senses have been extracted from pre-existing Croatian dictionaries and linked to corpus-based examples providing evidence of their valency patterns. For each verb sense, arguments are described on a morphological, syntactic and semantic level. In terms of morphological and syntactic description, e-Glava and CROATPAS encode similar information. However, the two resources differ most when it comes to semantics. As a matter of fact, e-Glava does not provide English sense descriptions nor Semantic Type labels, but only monolingual sense-based argument periphrases enriched with semantic categories from a small non-hierarchical set. For instance, the nominative subject of the first pattern of the verb *bojati se* (English, *to be afraid*) is described as *onaj koji osjeća strah* (English, *the one who feels fear*) and is enriched by the categories *živo, osoba, skupina ljudi, životinja* (English, *animate, person, human group, animal*).

The first project aimed at building a valency lexicon of Croatian verbs was CroVallex, a resource combining valency theory and frame semantics and containing approximately 1800 high-frequency verbs. In CroVallex, each verb entry is linked to as many valence frames as its number of senses. In each frame, argument slots are labelled with deep role labels (e.g. AGT for Agent, RESL for Result) and morphosyntactic numeric indexes encoding case markings (an idea we borrowed and adapted to the needs of CROATPAS). Overall, these features make CroVallex a highly specialised resource for competent users, primarily aimed at linguistic research – all the more so since it is entirely monolingual.

As for CroDeriV, this resource is centred around Croatian derivational morphology. At the moment, CroDeriV 1.0 consists of an online morphological database containing data about the morphological structure and derivational relatedness of approximately 14,500 Croatian verbs (Šojat et al. 2014). A redesign of the database and its online query interface is currently on-going to include also non-verbal lemmas (Filko et al. 2019). CroDeriV is the perfect tool to learn more about how Croatian verbs can be decomposed into lexical, derivational and inflectional morphemes, and how those verbs belonging to the same derivational families are morphologically connected with one another.

The same could be said for the Nooj Dictionary of Aspectual Derivatives, an on-going project aimed at creating an online database to investigate verb derivation chains and the affixation mechanisms involved in Croatian verb derivation. The database currently contains approximately 4,000 entries and is able to recognise over 377,603 inflected forms

Overall, we can say that all the above-mentioned resources bear some similarity with the CROATPAS project, but they either tend not to focus on potential accessibility problems by non-native users without a background in linguistics or to concentrate on different linguistic aspects, such as derivational morphology. Therefore, we can say that when compared to these resources CROATPAS stands out for its utmost focus on verb polysemy, its bilingual nature and its compatibility with the other CPA-inspired projects, thus positioning itself as a complementary resource to the existing ones.

5 Concluding Remarks

In this paper, we have introduced CROATPAS (CROATian Typed Predicate Argument Structures), a digital semantic resource currently containing 101 Croatian verb entries linked to 457 corpus-derived semantically labelled valency structures (i.e. *patterns*) and over 22,000 annotated corpus lines (Marini & Ježek 2019 & 2020).

CROATPAS is tailored for investigating verbal polysemy, since each pattern it contains is linked to a different verb sense, but it allows for a variety of different applications. After elaborating on the resource's theoretical underpinnings and providing some pattern examples (see § 2), we focused on its potential as a language teaching tool (see § 3).

As a matter of fact, once equipped with a SemType query option, CROATPAS could already be used in vocabulary teaching activities to access the most important lexical items pertaining to the different semantic areas connected to its SemType inventory (e.g. [Food], [Beverages], [Animals]s, etc.). Moreover, its user-friendly rendering of case markings and argument colour-coding strategy allows for a gentle introduction to the concept of case.

Finally, after comparing it with the other currently available resources for Croatian verbs (see § 4), we can say that CROATPAS is a faceted resource, which stands out for its user-friendly interface, bilingual nature and focus on verb semantics and polysemy. In time, it could become a truly useful tool for the teaching of Croatian as a Foreign Language.

6 References

Baisa, V., Tiberius, C., Ježek, E., Colman, L., Marini, C. & Romani, E. (2020). Skema: A new tool for corpus-driven lexicography. In *Proceedings of the 19th EURALEX International Congress.* Alexandroupolis, Greece.

Barić, E., Lončarić, M., Malić, D., Pavešić, S., Peti, M., Zenčević, V. & Znika, M. (1997). *Hrvatska Gramatika*. Zagreb: Školska knjiga.

Baroni, M., Kilgarriff, A. (2006). Large Linguistically Processed Web Corpora for Multiple Languages. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics.* Trento, Italy.

Birtić, M., Brač, I. & Runjaić, S. (2017). The Main Features of the e-Glava Online Valency Dictionary. In *Proceedings of*

[2] http://valencije.ihjj.hr/ Website last visited [05/05/2020].

[3] http://theta.ffzg.hr/crovallex/ Website last visited [05/05/2020].

[4] http://croderiv.ffzg.hr/ Website last visited [05/05/2020].

[5] https://vidski-parnjaci.herokuapp.com/ Website last visited [05/05/2020].

the 5th eLex conference - Electronic lexicography in the 21st century. Leiden, Netherlands.
Brown, H.D., Lee, H. (2015). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Pearson.
Chan, T., Liou, H. (2005). Effects of Web-based Concordancing Instruction on EFL Students' Learning of Verb – Noun Collocations. In *Computer Assisted Language Learning*, 10(3), pp. 231-250.
Colman, L., Tiberius, C. (2018). A Good Match: a Dutch Collocation, Idiom and Pattern Dictionary Combined. In *Proceedings of the 8th EURALEX International Congress*. Ljubljana, Slovenia.
Filko, M., Šojat, K. & Štefanec, V. (2019). Redesign of the Croatian derivational lexicon. In *Proceedings of the 2nd International Workshop on Resources and Tools for Derivational Morphology*. Prague, Czech Republic.
Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. Cambridge: The MIT Press.
Hanks, P., Ježek, E. (2008). Shimmering lexical sets. In *Proceedings of the 13th EURALEX International Congress*.
Hanks, P., Pustejovsky, J. (2005). A Pattern Dictionary for Natural Language Processing. In *Revue française de linguistique appliquée*, 10 (2), pp. 63-82.
Hudeček, L., Mihaljević, M. (2017). The Croatian Web dictionary project - Mrežnik. In I. Kosem et al. (eds.) *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*. Brno, Czech Republic; Leiden, Netherlands.
Ježek, E. (2016). *The lexicon: An introduction*. Oxford: Oxford University Press.
Ježek, E. (2019). Sweetening Ontologies Cont'd: Aligning bottom-up with top-down ontologies. In *Proceedings of CREOL 2019*. Graz, Austria.
Ježek, E., Magnini, B., Feltracco, A., Bianchini, A. & Popescu, O. (2014). T-PAS: A resource of Typed Predicate Argument Structures for linguistic analysis and semantic processing. In *Proceedings of the 9th conference on International Language Resources and Evaluation (LREC)*. Reykjavik, Iceland.
Kilgarriff, A., Husák, M., McAdam, K., Rundell, M. & Rychlý, P. (2008). GDEX: automatically finding good dictionary examples in a corpus. In *Proceedings of the 13th EURALEX International Congress*. Barcelona, Spain.
Kilgarriff, A., Baisa, V., Busta, J., Jakubícek, M., Kovár, V., Michelfeit, J., Rychlý, P. & Suchomel, V. (2014). The Sketch Engine: ten years on. In *Lexicography*. 1(1), pp. 7-36.
Kilgarriff, A., Marcowitz, F., Smith, S. & Thomas, J. (2015). Corpora and Language Learning with the Sketch Engine and SKELL. In *Revue française de linguistique appliquée*, 20(1), pp. 61-80.
Kocijan, K., Šojat, K. & Poljak, D. (2018). Designing a Croatian Aspectual Derivatives Dictionary: Preliminary Stages. In *Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing*. Santa Fe, USA.
Lee, H., Warschauer, M. & Lee, J.H. (2018), The Effects of Corpus Use on Second Language Vocabulary Learning: A Multilevel Meta-analysis. In *Applied Linguistics*, pp. 1-34.
Liu, D., Jiang, P. (2009) Using a Corpus-Based Lexicogrammatical Approach to Grammar Instruction in EFL and ESL Contexts. In *The Modern Language Journal*, pp. 61-78.
Ljubešić, N., Erjavec, T. (2011). hrWaC and slWac: Compiling Web Corpora for Croatian and Slovene. In *Text, Speech and Dialogue*, Lecture Notes in Computer Science, Springer.
Marini, C., Ježek, E. (2019) CROATPAS: A Resource of Corpus-derived Typed Predicate Argument Structures for Croatian. In *Proceedings of the 6th Italian Conference on Computational Linguistics* (CLiC-it). Bari, Italy.
Marini, C., Ježek, E. (2020). Annotating Croatian Semantic Type Coercions in CROATPAS. In *Proceedings of the Sixteenth Joint ACL-ISO Workshop on Interoperable Semantic Annotation.* France, Marseille.
Mikelić Preradović, N., Boras, D. & Kišiček, S. (2009). CROVALLEX: Croatian Verb Valence Lexicon. In: *Proceedings of the 31st International Conference on Information Technology Interfaces*. Zagreb, Croatia.
Mikelić Preradović, N., Posavec, K. & Unić, D. (2019). Corpus-Supported Foreign Language Teaching of Less Commonly Taught Languages. In *International Journal of Instruction*, 12(4), pp. 335-352.
Nesselhauf, N. (2003), The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. In *Applied Linguistics*, 24(2), pp. 223-242.
Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge: The MIT Press.
Pustejovsky, J. (1998). The semantics of lexical underspecification. In: *Folia Linguistica* (32).
Pustejovsky, J., Ježek, E. (2008). Semantic Coercion in Language: Beyond Distributional Analysis. In *Italian Journal of Linguistics*, 20, pp. 181-214.
Šojat, K., Srebačić, M., Tadić, M. & Pavelic, T. (2014). CroDeriV: a new resource for processing Croatian morphology. *Proceedings of the 9th conference on International Language Resources and Evaluation (LREC)*. Reykjavik, Iceland.
Šojat, K., Kocijan, K. & Filko, M. (2018). Processing Croatian Aspectual Derivatives. *In Formalizing Natural Languages with NooJ 2018 and Its Natural Language Processing Applications. Communications in Computer and Information Science*.
Tadić, M., Brozović-Rončević, D. & Kapetanović, A. (2012). Hrvatski Jezik u Digitalnom Dobu – The Croatian Language in the Digital Age. In: *META-NET White Paper Series*, Rehm G. & Uszkoreit, H. (eds.), Springer: Heidelberg, New York, Dordrecht, London.
Ward, M. (2016). CALL and less commonly taught languages – still a way to go. In *CALL communities and culture – short papers from EUROCALL 2016*.
Croatian Valence Lexicon of Verbs. Accessed at: http://theta.ffzg.hr/crovallex/ [05/05/2020].
Croatian Derivational Lexicon. Accessed at: http://croderiv.ffzg.hr/ [05/05/2020].
E-Glava Valency Database of Croatian Verbs. Accessed at: http://valencije.ihjj.hr/ [05/05/2020].
Nooj Croatian Dictionary of Aspectual Derivatives. Accessed at: https://vidski-parnjaci.herokuapp.com/ [05/05/2020].
Woordcombinaties. Accessed at: http://woordcombinaties.ivdnt.org/ [05/05/2020].