


UNIVERSITÀ DEGLI STUDI DI BERGAMO

UNIVERSITÀ DEGLI STUDI DI PAVIA

Scuola di Alta formazione Dottorale

Corso di Dottorato in Scienze Linguistiche

Ciclo XXXVI

Settore scientifico disciplinare L-LIN/01 – Glottologia e linguistica

ANALYSING AND TEACHING THE INTERACTIONS BETWEEN ASPECT, *AKTIONSART* AND VERB POLYSEMY IN CROATIAN

Supervisore:
Prof.ssa Elisabetta Ježek

Candidata:
Costanza Marini

Matricola° 1073086

Anno accademico 2022/23

A mia nonna Clara,
che mi ha insegnato a scrivere

ACKNOWLEDGEMENTS

Now that I find myself at the very end of this long journey, I would like to thank the people that made it possible for me to get here.

First of all, I would like to thank my academic supervisor and mentor Elisabetta Ježek for the constant guidance, invaluable feedback and for having always supported my crazy idea to apply what I was learning from her about the lexicon and lexicography to Croatian. I feel extremely lucky.

A big thank you to the Centre for Croatian as a Second and Foreign Language of the University of Zagreb, especially to Sanda Lucija Udier, whose first-class teaching is greatly responsible for my Croatian proficiency and whose feedback as a reviewer was paramount, to Milvia Gulešić Machata, who kindly allowed me to witness her Croatian teaching practice while teaching absolute beginners, and Marica Čilaš Mikulić, who read this dissertation while it was being written, steering me in the right direction.

A special thank you to the Department Linguistics of the University of Zadar, especially to Marco Angster for allowing me to take part in his teaching activities during my stay in Zadar and for the precious feedback he gave as a reviewer, as well as to Mia Batinić Angster for the opportunity of organising a Conference together and for always making me feel welcome.

Another big thank you goes to the University Language Centre of the University of Pavia, especially to Cristina Mariotti for helping me turn the idea of a Croatian Language Lab into a reality, and to Antonella Meduri for her impeccable administrative support.

Thank you to Marco Passarotti for his suggestions and encouragement throughout the years, to Vítek Baisa for lending me his software expertise, and to Maria Grazia Cammarota for managing our PhD School through thick and thin.

A humongous thank you to Irina Knežević, the one responsible for introducing me to Croatia and making me fall in love with its language and culture. *Hvala ti, draga,* for always being available to give me your native speaker opinion at whatever hour of the day or night.

Hvala vama also to all the other Croatian native speakers who joined in the fun of answering my surveys and questionnaires: Domenika, Andrea, Marin, and so many others… thank you!

I would also like to thank my colleagues Tanja, Chiara, Lucrezia, Serena, and Caterina for sharing this roller-coaster of a journey with me, and all my students from the Croatian Language Lab for being such great students… you exceeded my wildest expectations and set a very high bar!

This being said, the biggest thanks goes to my family and friends for their emotional support during these trying years. Writing a PhD is no easy quest *per se*, but – coupled with a global pandemic – it could have become a serious challenge.

Quindi, grazie a Claudia, Gigi, Andrea, Roby, Sara, Silvia, Dandy, Leila, e Ale. You have been the closest to me during this time and I would definitely not have made it without you, guys!

Analysing and teaching the interactions between aspect, *Aktionsart* and verb polysemy in Croatian

INTRODUCTION

[…] progress in linguistics, as in other sciences,
depends on the cross-fertilization of developing theory and accumulating data.
Uriel Weinreich (1964)

When a native speaker of a Romance language embarks on the journey of learning a Slavic language, their first encounter with aspectual verb pairs is not dissimilar from an encounter with a hornets' nest.

While getting acquainted with the possibility of expressing the "different ways of viewing the internal temporal constituency of a situation" – as Comrie would say – by means of different lexical variants, one is often faced with differences in terminologies and few resources to resort to, especially in non-Slavic languages. This is the main reason why this work was written in English.

This Dissertation wishes to give a contribution to the study of Croatian verbal aspect and its interaction with *Aktionsart*, as well as to the teaching of this South Slavic language. In both instances, the Candidate has chosen verb polysemy as the starting point, specifically the CroaTPAS (*Croatian Typed Predicate Argument Structures*) resource, which is to be considered one of the deliverables of this research.

CroaTPAS is a digital semantic resource, tailor-made to portray verb polysemy, consisting in a collection of corpus-based predicate argument structures whose argument slots have been annotated using a set of Semantic Types labels. The resource was first created during the MA thesis of the Candidate and was finalised during her PhD. CroaTPAS currently contains 180 Croatian verbs for a total of 795 different verb senses and is accessible online at the address: https://croatpas.baisa.cz/.

Before turning to CroaTPAS and its applications, however, the first three Chapters of this Dissertation are devoted to introducing the theoretical concepts and background knowledge needed to further venture in these topics.

Chapter 1 is devoted to the Croatian language, providing not only the necessary introduction to it is sound and case system, but also an overview of its complex sociolinguistic history, which is paramount for whoever may approach the field.

On the other hand, Chapter 2 focuses on the two theoretical concepts of aspect and *Aktionsart*. After an attempt at clarifying their intertwined history and the rich terminology used to refer to them, the Chapter provides an overview of the current Western interpretations of these phenomena, before looking at the way they are systematised in Croatian literature, specifically Croatian grammars.

Chapter 3 is meant to introduce the concepts and terminology linked to verb valency theory, argument structures and verb polysemy and lays the grounds for Chapter 4, which is the devoted to the CroaTPAS resource and its evaluation procedure.

In Chapter 5, information is drawn from all previous chapters a comparison is carried out between the meaning inventories of a selection of symmetric and asymmetric CroaTPAS verb pairs, whose *Aktionsarten* are modelled using a customized version of the *event structures* formalism. The aim of the chapter was to investigate the interaction between aspect and *Aktionsart* and to test whether event structures were actually able to portray the relationship between perfective and imperfective Croatian aspectual variants.

Finally, Chapter 6 revolves around the Croatian language teaching experience that was part of the PhD curriculum of the Candidate. The experience took place at Language Centre of the University of Pavia in the second semester of the academic year 2022/23 and consisted in 30-hour Croatian Language Intensive Course for beginners (A2 level).

The Course was the perfect opportunity to apply language teaching principles, develop game-based activities and use the CroaTPAS resource to introduce verbal aspect to the class. A final test was administered to the class at the end of the course, whose results were analysed following both a quantitative approach (see section § 6.6) and a statistical one (see section § 6.7).

CHAPTER 1: THE CROATIAN LANGUAGE

Croatian (*hrvatski*) is the official language[1] of the Republic of Croatia (*Republika Hrvatska*), the youngest Member State of the European Union since 2013.

Croatia is a parliamentary democracy and, according to the latest 2023 *Eurostat* counts, it hosts a population of approximately 3.85 million inhabitants[2]. The country shares its borders with Slovenia to the northwest, Hungary to the north, Serbia and Bosnia-Herzegovina to the east and Montenegro to the south, while the Adriatic Sea washes its shores to the southwest. Croatia's capital city is Zagreb and euro has become its official currency since January 2023, after nearly thirty years of *hrvatska kuna* 'Croatian kuna'. The country is renowned for its coastline, which is faced by over 1000 islands (Krk, Brač, Hvar and Pag being the biggest ones), as well as for its national parks (like the Brijuni Islands, and Plitvice Lakes National Park).

Croatia, however, has not been an independent country for long, nor has its language. In this Chapter, we will therefore provide a brief sociolinguistic account of the history of Croatian (§ 1.1), as well as an overview of its phonological (§ 1.2), and case system (§1.3).

1.1 A SHORT HISTORY OF CROATIAN

From a sociolinguistic standpoint, Croatian is a South Slavic language characterised by three dialect groups, which are named after the word each of them has for the interrogative pronoun "what" (Sussex & Cubberley, 2006, p. 506; Udier, 2016, p. 149), namely:

- *Kaj* – in the Kajkavian dialects (*kajkavski*) spoken in north-west Croatia around the Zagreb area and near the Slovenian border.
- *Ča* – in the Čakavian dialects (*čakavski*) spoken in all of Istria, parts of Gorski Kotar and Dalmatia, as well as on most of the islands.
- *Što* – in the Štokavian dialects (*štokavski*) spoken in the rest of Croatia[3].

Figure 1 provides an overview of the above-mentioned tripartition: Kajkavian dialects are spoken in the green areas, Čakavian dialects in the yellow areas, and Štokavian dialects in the red and pink ones.

[1] Italian being the only other official language, but only in the Istria region.

[2] https://ec.europa.eu/eurostat/databrowser/view/tps00001/default/table?lang=en. Last visited on July 22nd, 2023.

[3] Štokavian has seven dialects, of which Croats use four, while the remaining three are part of the Serbian, Bosnian, and Montenegrin languages.

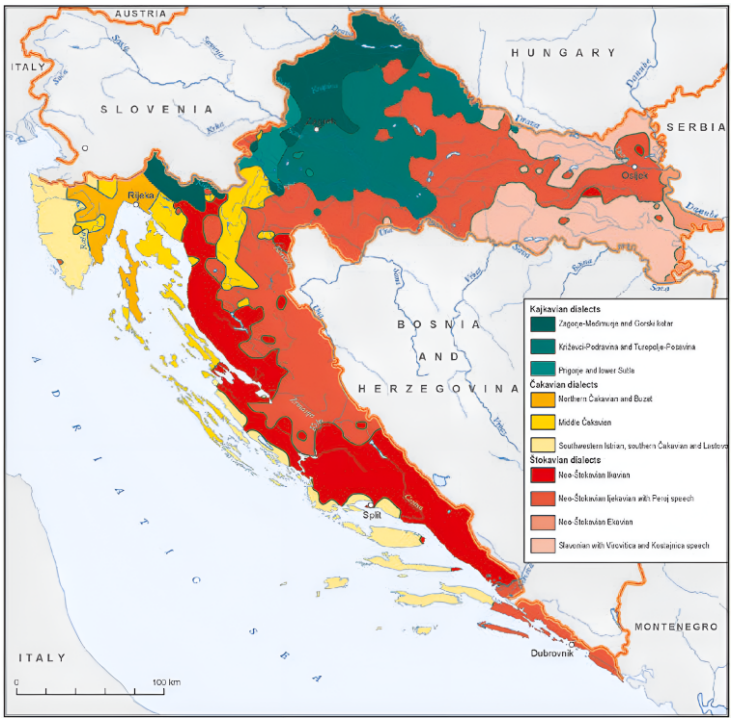


Figure 1 – Map of Croatian dialects, taken from Brozović (1997, p. 155).

Apart from this distinction, within the Čakavian and Štokavian dialect areas, we observe a further sub-division based on the diachronic developments of the Proto-Slavic low front vowel /ě/ (*jat*), also known as *jat* reflexes (Sussex & Cubberley, 2006, p. 111). This vowel evolved into:

- /e/ – in Ekavian dialects (*ekavski*) spoken in Serbia and Vojvodina.
- /i/ – in Ikavian dialect (*ikavski*) spoken in Dalmatia and the west of Bosnia .
- /ie/ – written *je* as a short and *ije* as a long reflex (Barić et al., 2005, p. 28)[4] – in Jekavian dialects (*jekavski*) spoken in the rest of Croatia, Bosnia, and Montenegro.

	***ě* in a LONG SYLLABLE**	***ě* in a SHORT SYLLABLE**
Proto-Slavic	*rěka*	*věra*
Ekavian dialects	*reka*	*vera*
Ikavian dialects	*rika*	*vira*
Jekavian dialects	*rijeka*	*vjera*
English translation	river	faith

Table 1 – The evolution of the Proto-Slavic *jat*

[4] The seminal work *Hrvatska Gramatika*, whose fourth edition was published by Školska knjiga in 2005, is the result of the fruitful collaboration of all its authors, who we would like to list here: Eugenija Barić, Mijo Lončarić, Dragica Malić, Slavko Pavešić, Mirko Peti, Vesna Zečević, and Marija Znika.

These notions are going to be paramount to understand the historic development of the language in the rest of this paragraph.

As a matter of fact, even though it gained official language status only with the establishment of an independent Republic of Croatia in 1991, having only been recognised as a national variant of Serbo-Croatian or Croato-Serbian (*srpsko-hrvatski* or *hrvatsko-srpski*) for most of the 19th and 20th century, Croatian's history starts well before that.

The most important Croatian document is without a doubt the *Bašćanska ploča* 'Baška Tablet', a stone tablet written around 1100 AD in Old Croatian using the Glagolitic alphabet. The tablet was found in 1851 in Jurandvor, near Baška, on the island of Krk, and is the oldest document mentioning a "Croatian people".

At the time, culture in Croatia was trilingual and trigraphic, that is to say, it was produced in three languages (Latin, Old Slavic, and Old Croatian) and written in three alphabets (Latin, Western Cyrillic and Glagolitic) (Udier, 2016, p. 152). Of these three alphabets, Glagolitic is the most relevant to Croatian culture, because it was the first one used by Croats.

Glagolitic was invented in the 9th century by the Byzantine evangeliser Cyril to provide Slavic people with an alphabet of their own. After his brother Methodius translated sacred books into Old Slavic using Glagolitic, the two were called to Moravia (Southern Czech Republic) to evangelise the local Slavic population. It was the later missionary work of their students in the South Slavic regions which resulted into the first contact between Croats and the Glagolitic alphabet.

After a first phase of rejection, during which the Roman Catholic Church opposed the spread of Slavic liturgy using Glagolitic, the alphabet was finally approved in the 13th century. This resulted in a great development of Glagolitic literacy, which reached its peak in the 14th and 15th century, when the Croatian language entered all areas of public life and literature (Malić, 2005, p. 14-16). During this time, works were rewritten in all alphabets and languages, and Croatian dialects experienced a time of reciprocal contamination.

In 1325, the term *jezik hrvacki* 'Croatian language' was recorded for the first time in a legal document regulating land boundaries. After the invention of the Gutenberg press, the first Croatian book was printed in Croatia in 1483. In the following years, several Croatian works were printed also in Venice, as in the case of *Judita* by Marko Marulić (1521), which is considered the first piece of Croatian secular literature.

During the 16th century, the use of the Glagolitic alphabet decreased in favour of the Latin one, and internal migration due to the Ottoman invasion led to even more linguistic contact between dialects, which in turn led to the development of a Croatian language based on a hybrid of the three Croatian dialects (Udier, 2016, p. 156).

Mirroring this reality, the first Croatian dictionaries and grammars include different dialect mixes. For instance, the first Croatian dictionary[5] by Faust Vrančić (1595) is a Čakavian dictionary with Kajkavian and Štokavian elements, the first Croatian grammar[6] by Bartol Kašić (1604) is based on Čakavian with Štokavian elements, while the dictionary *Blago jezika slovinskoga* 'Treasure of the Slovin Language' by Jakov Mikalja (1649-1651) recommends a Štokavian base for a Croatian literary language. By the middle of the 18th century, the processes leading to the formation of a standard Croatian language with a Neoštokavian base had actually started (Malić, 2005, p. 24).

In the 1830s, the need for a common language with a Štokavian foundation which could be understood by most Croats and unify all South Slavs became one of the goals of the Croatian political, literary, and cultural group called *Ilirski pokret* 'Illyrian movement', which fostered the development of a Croatian national consciousness and tried to oppose the Hungarian rule[7] (Udier, 2016, p. 158). The main figure of the movement was the linguist Ljudevit Gaj, who proposed a series of reforms to Croatian orthography, e.g., the addition of diacritic signs, as well as six new letters – which he took from Czech (*ž*, *š*, *č*, *lj* and *nj*) and Polish (*ć*) – thus creating a phonetic alphabet, in which each letter corresponds to an individual sound.

The efforts of the Illyrians paid off, as Croatian began to be taught at the Zagreb Academy in 1832, and the first parliamentary speech in Croatian was given in the Croatian Parliament in 1843 (until then only Latin had been spoken).

In the meantime, the political unrest in the Hapsburg Empire had reached its breaking point and, after the Budapest riots of 1848 (Sussex & Cubberley, 2006, pp. 73–75), Croatia was granted independence from the Kingdom of Hungary under the lead of Ban Josip Jelačić.

Following the fruitful collaboration that had taken place between Croats and Serbs during the riots, some of the most prominent Croatian and Serbian academics of the time[8] met in Vienna in 1850 for what came to be known as the *Bečki književni dogovor* 'Vienna Literary Agreement'.

---

[5] *Dictionarium quinque nobilissimarum Europae linguarum, Latinae, Italicae, Germanicae, Dalmaticae et Ungaricae* 'Dictionary of the five most noble European languages – Latin, Italian, German, Dalmatian and Hungarian'.

[6] *Institutionum linguae Illyricae libri duo* 'Basis of the Illyrian Language in Two Volumes'.

[7] Take notice that – thanks to the reforms of Maria Theresa – the territories making up modern Croatia had had a special administrative status within the Hapsburg Empire since the second half of the 18th century. When the Empire entered into crisis, Croatian lands were divided between Austrian and Hungarian rule.

[8] Among others, we recall the poet Ivan Mažuranić (1814-1890), the poet and novelist Dimitrija Demeter (1811-1872), the novelist and politician Ivan Kukuljević (1816-1889), and the linguist Vuk Stefanović Karadžić (1787-1864), who – according to the motto of the German philosopher Adelung *"piši kao govoriš"*, i.e., 'write as you speak' – had reformed Serbian Cyrillic following a phonetic script, just like Gaj had done for Croatian.

In line with the Panslavic motto "one language one people", the intention of the meeting was to unify the language of the Croats and the Serbs by creating a standard literary language: Serbo-Croatian or Croato-Serbian depending on the perspective. If certain resolutions were agreed on, such as the differentiation between alphabets[9], the party concluded that the common language should be created based on the *ije*-variant of the Neoštokavian dialect of East Herzegovina and refused several linguistic features put forth by the Croatian delegates adhering to the Illyrian Movement: *inter alia*, no suffix -*h* was to be written for the genitive plural of nouns, and no vowels were to accompany the syllabic *r* (see § 1.2.3).

After initial resistance, the Croatian linguist Ivan Broz published *Hrvatski pravopis* 'Croatian Orthography' (1892) in accordance with the principles of the Vienna Literary Agreement, followed, in 1899, by Tomislav Maretić, who published *Gramatika i stilistika hrvatskoga ili srpskoga jezika* 'Grammar and stylistics of the Croatian or Serbian language', which had a great influence on the standardisation process of the language.

From 1918 until 1929 Croatia became part of the Kingdom of Serbs, Croats, and Slovenes from the merger of the Kingdom of Serbia with the provisional State of Slovenes, Croats and Serbs formed from territories of the former Austro-Hungarian Empire. The Kingdom of Serbs, Croats, and Slovenes was later renamed Kingdom of Yugoslavia (1929-1941), which became the Federal People's Republic of Yugoslavia in 1945 and finally the Socialist Federal Republic of Yugoslavia in 1963.

If during WWII[10] many Croatian neologisms had been created and Serbian elements had been removed from Croatian (Udier, 2016, p. 162), after 1945, Serbo-Croatian became once again the *lingua franca* of the whole area and was taught throughout the entire Federation. It was the official language of four of the six Socialist Republics of Yugoslavia – i.e., Bosnia and Herzegovina, Croatia, Montenegro, and Serbia – while Slovenian was the official language of Slovenia and Macedonian the official language of Macedonia (nowadays North Macedonia).

Figure 2 provides an overview of the geographical make-up of Former Yugoslavia.

[9] Croatian Latin (*latinica*) for the Croats and Serbian Cyrillic (*ćirilica*) for the Serbs.

[10] The time of the Independent State of Croatia (1941-1945).

Figure 2 – Map of the Former Socialist Federal Republic of Yugoslavia

To provide some concrete figures, according to Eller & Hladczuk (1992, p. 454), 73% of the Yugoslav population in 1982 spoke Serbo-Croatian as their first language, while the remaining 27% spoke it as a second language[11].

In Croatia, however, several reactions to this linguistic unification process could be witnessed throughout the decades preceding the breakout of the Yugoslavian Wars, especially after the *Novosadski dogovor* 'Novi Sad Agreement' of 1954, which had officially declared that the language of the Serbs, Croats and Montenegrins (Bosnia was not even mentioned) was only one language, with either Ekavian or Jekavian pronunciation, and two equal alphabets (Malić, 2005, p. 35).

The Agreement led to a renewed attempt at unifying Croatian and Serbian, as well as to the publication of several dictionaries that tried to conceal the differences between the two languages, especially in terms of vocabulary. Croatian cultural institutions protested against these actions demanding for Croatian and Serbian to have an equal status within the Federation and to let Croatian develop independently from Serbian, according to Croatian tradition (Udier, 2016, p. 163). Their requests were condemned, but they sparked the publication of the first modern Croatian grammars and orthography manuals, which were of course banned by the Federal government.

Finally, in 1974, the Yugoslav constitution was amended and Croatian was declared the official language of Croatia, although the standard form of the language was still *hrvatski ili srpski* 'Croatian

[11] Quite interestingly, given its multilingual make-up, Yugoslavia allowed its citizens to use one of five languages in its federal institutions: Serbo-Croatian, Slovenian, Macedonian, Hungarian, or Albanian (Eller & Hladczuk, 1992, p. 455).

or Serbian' in Croatia and *srpski ili hrvatski* 'Serbian or Croatian' in Serbia (Piasevoli & Vranješ, 2014, p. 9). This was the starting point for the recognition of the status of Croatian as its own language.

In light of all the above, we can conclude that talking of Serbo-Croatian instead of naming the different independently standardized languages that have emerged since the end of the conflict, is not only "an anachronism from the political point of view" (Sussex & Cubberley, 2006, p. 4), but also a potentially dangerous political statement.

Indeed, stripping a language of its language status affects the sense of identity of the people who speak it, even if – from a strictly linguistic point of view – Bosnian, Croatian, Montenegrin, and Serbian belong to the same South Slavic language grouping and could be treated as different variants within a common geographical space, since the "linguistic differences between them are no greater than those between British and American English" (Naylor, 1980, p. 68).

This being said, this dissertation will focus on standard Croatian only.

1.2 THE CROATIAN SOUND SYSTEM

1.2.1 SOUND FORMATION VS. SYLLABIC ROLE

According to traditional literature, Croatian vowels and consonants are referred to using two different sets of labels, which depend on the criterion used to group them.

If we consider sound formation, phonemes can be divided into *otvornici* 'the open ones' and *zatvornici* 'the closed ones', depending respectively on whether the airflow is able to pass through the vocal tract without meeting any obstacle or not (Babić, 1995, p. 66; Jelaska, 2004, p. 33). On the other hand, a more phonologically oriented criterion focuses on the role sounds have in the syllable (Barić et al., 2005, p. 55; Jelaska, 2004, pp. 34–35): *samoglasnici* or *slogotvorni glasovi* are sounds that are able to form independent syllables, whereas *suglasnici* or *neslogotvorni glasovi* are not.

1.2.2 OTVORNICI

Following Babić (1995, p. 66) and Jelaska (2004, p. 38), the Croatian sound system is characterised by 6 *otvornici*, namely:

- 5 monophthongs: /i; e; a; o; u/
- 1 diphthong: /ie/

However, sporadically also the diphthongs /au/ and /ae/ can be found in words such as *auto* 'car' or *trinaest* 'thirteen' (Jelaska, 2004, p. 41). No nasal vowels are attested, only oral ones. Each vowel can be long or short, but more will be said on the topic in § 2.2.5.

The following vowel chart (Figure 3) offers a schematic representation of the Croatian vowel repertoire, where *visoki* means 'high', *srednji* 'medium' and *niski* 'low':

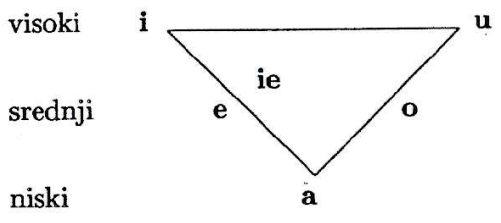


Figure 3 – The Croatian vowel chart according to Jelaska (2004, p. 39)

- /i/ is the close front unrounded vowel, as in *život* ['ʒiʋot] 'life'.
- /e/ is the mid front unrounded vowel, as in *leptir* ['lepti:r] 'butterfly'.
- /a/ is the open front unrounded vowel, as in *grad* ['gra:d] 'city'.

- /o/ is the mid back rounded vowel, as in *lopov* ['lo:poʊ] ‘thief’.
- /u/ is the close back rounded vowel, as in *put* ['pu:t] ‘way’.
- /ie/ has a close-mid collocation. As previously mentioned, it does not correspond to a single letter of the Croatian alphabet, but to either two letters, as in *rječnik* ['rje:tʃni:k] ‘vocabulary’, or three, as in *cvijet* ['tsʊje:t] ‘flower’.

Finally, after each separately pronounced consonant, Croatian always appends a *prateće šva* ‘supporting schwa’ [ə], i.e., a mid-central unrounded vowel, which does not have phonemic value, although it has syllabic potential (Barić et al., 2005, p. 55; Jelaska, 2004, p. 38).

1.2.3 SAMOGLASNICI

If we classify phonemes according to the role they have in the syllable, we can recognise that *otvornici* are the prototypical *samoglasnici*, i.e., the prototypical vowels. However, Croatian presents a certain number of *zatvornici* with syllabic potential in some of their allophonic realisations (Barić et al., 2005, p. 55; Brozović, 1991, p. 409), namely:

- /r/ as in *vrt* ['vr̩t] or *masakr* [ma'sakr̩], respectively ‘garden’ and ‘massacre’.
- /l/ usually in loanwords, e.g., *džentlmen* ['dʒentl̩men] or *bicikl* [bi'tsikl̩], respectively ‘gentleman’ and ‘bicycle’.
- /n/ in *Njutn* ['ɲutn̩], ‘Newton’.
- rarely /m/.

This being said, all of the above-mentioned phonemes are still more frequently realised as *suglasnici* than *samoglasnici*. Finally, as Jelaska (2004, p. 35) points out, some linguists talk of /r̩/ and /r/ as two distinct phonemes: a syllabic and a non-syllabic one.

1.2.4 ZATVORNICI

The consonant inventory of Croatian consists of 25 phonemes (Browne & Alt, 2004, p. 12; Jelaska, 2004, pp. 42–50), namely:

- 8 *zvonačnici* ‘sonorants’:
  - 3 *nosnici* ‘nasals’, namely /m, n, ɲ/;
  - 5 *približnici* ‘approximants’, namely /j; ʋ; r, l, ʎ/;
- 17 *šumnici* ‘obstruents’:
  - 6 *zapornici* ‘plosives’ or ‘stops’, namely /p, t, k, b, d, g/;
  - 6 *tjesnačnici* ‘fricatives’, namely /f, s, ʃ, x, z, ʒ/;

- 5 *slivenici* ‘affricates’, namely /ts, tʃ, tɕ, dʒ, dʑ/.

It is worth mentioning that the bilabial nasal /m/ has a labiodental realisation [ɱ] before /f, ʋ/, as in *tramvaj* ['traɱʋaj] ‘tram’, while the labiodental nasal /n/ becomes velar [ŋ] before the stops /k, g/, as in *stanka* ['staːŋka] ‘break’ (Landau et al., 1999, p. 68).

As for the alveolar vibrant /r/, depending on the number of tongue oscillations, Barić et al. (2005, p. 49) were able to differentiate when it is realised as a vowel [r̝̊] (3-5 oscillations) or as a consonant [r] (1-2 oscillations).

Unlike Italian, the Croatian sound system contains:

- the voiced labiodental approximant /ʋ/, as in *vjeverica* ['ʋjeʋeritsa] *squirrel*. Unlike in the realisation of the voiced labiodental fricative /v/, which is not present in Croatian, when producing /ʋ/ teeth touch the lower lip only very lightly (Jelaska, 2004, p. 49).
- the voiceless velar fricative /x/, as in *hvala* ['xʋaːla] ‘thanks’.
- the voiced alveo-palatal fricative /ʒ/, as in *žaba* ['ʒaba] ‘frog’, that in Italian appears only in regional variants (e.g. Tuscan *ragione* [ra'ʒone] ‘reason’) or French loanwords (*abat-jour* [aba'ʒuʁ] ‘abatjour’) (Basile et al., 2010, p. 86).
- the voiceless palatal affricate /tɕ/, as in *ćao* ['tɕao] ‘hi’, which can easily be mistaken by non-native speakers for the voiceless alveo-palatal affricate /tʃ/ of the Italian *ciao* ['tʃao] (Basile et al., 2010, p. 87).. Since in /tɕ/ the entire tongue is raised towards the palate, Browne & Alt (2004)’s suggestion for English speakers is to practice saying ‘cheap’ while smiling.
- the voiced palatal affricate /dʑ/, as in *anđeo* ['aːndʑeo], lit. ‘angel’, which could also be mistaken for the voiced alveo-palatal affricate /dʒ/ to be foud in *džemper* [dʒ'emper], lit. ‘jumper’. Browne & Alt (2004)’s tip for English speakers, in this case, is to practice saying ‘squeegee’ while smiling.

1.2.5 ACCENTS

Croatian is one of those languages that use accent as a meaning-distinctive feature, involving both pitch and syllable length (Jelaska, 2004, p. 193). Since pitch can be either rising or falling and accented syllables can contain either a long or a short vowel, Croatian differentiates between four accents, each with its own mark[12]:

[12] Nowadays, accents are usually not marked when writing Croatian, except in situations when there is a difference in meaning which cannot be contextually disambiguated, as in the case of sentences like *Trčim oko polja* ‘I run around the

- short falling (*kratkosilazni*), marked \\ as in *lȕk* ‘onion’;
- long falling (*dugosilazni*), marked ◠ as in *lûk* ‘bow’;
- short rising (*kratkoulazni*), marked \ as in *veličìna* ‘size’;
- long rising (*dugoulazni*), marked / as in *garáža* ‘garage’.

Falling accents occur almost exclusively on first syllables and can occur on monosyllables. Rising accents can occur on any syllable but the last, hence not on monosyllables. However, not all Croatian words have an accented syllable of their own (Browne & Alt, 2004, p. 15; Jelaska, 2004, p. 199):

- proclitics (e.g., some conjunctions, prepositions and the negation *ne*) hang onto the next word;
- enclitics (e.g., certain pronouns, verb forms and the question marker *li*) onto the previous one.

If a proclitic and an accented word with a falling accent form an accented unit, the accent can jump onto the proclitic, but only if it is on the first syllable (Browne & Alt, 2004, p. 15; Jelaska, 2004, p. 200), as we can see from the example (1):

(1) *ne znâm* → *nè_znām*
‘I do not know.’

1.2.6 SOUND CHANGES

Croatian displays a certain number of both consonant and vowel changes, which generally go under the name of *glasovne promjene* ‘sound changes’. Two of the most important ones when it comes to both case inflection and word-formation are palatalisation (§ 2.2.6.1) and sibilarisation (§ 2.2.6.2). Both were included in the syllabus of the Croatian language A2 intensive course, which will be dealt with in Chapter 4.

1.2.6.1 PALATALISATION

Velar palatalisation or *prva palatalizacija* is the process responsible for the transformation of

/k, g, x/ → /tʃ, ʒ, ʃ/

when followed by /e/ (Barić et al., 2005, p. 616; Browne & Alt, 2004, pp. 16–17; Ham, 2017, p. 18). In Croatian, this transformation take place, for instance:

1) In the vocative singular of masculine nouns:

field’ and *Trčim oko poljā* ‘I run around the fields’. Post-accentual length is also sometimes marked, as in example (1) above.

učenik ['utʃeniːk] + *-e* → *učeniče* ['utʃeniːtʃe] ‘oh, student’

Bog ['Boːg] + *-e* → *Bože* ['Boʒe] ‘oh, God’

duh ['duːx] + *-e* → *duše* ['duːʃe] ‘oh, spirit’

2) In the formation of the present tense of verbs like *pećі* ‘to bake’, since their stems end in /k/:

pek- ['pek] + *-em* → *pečem* ['petʃem] ‘I bake’

3) In the vocative singular and nominative plural of certain masculine nouns ending in /ts, z/:

stric ['striːts] + *-e* → *striče* ['striːtʃe] ‘oh, uncle’

knez ['kneːz] + *-evi* → *kneževi* ['kneːʒevɪ] ‘the princes’

Unlike other nouns, *stric* and *knez* result from the so-called *third palatalisation of velars* (Browne & Alt 2004: 17), which is responsible for their endings in /ts/ and /z/ from Proto-Slavic /k/ and /g/. This is why their endings still alternate with /tʃ/ and /ʒ/ when followed by /e/: it is a remnant of the old Proto-Slavic legacy.

1.2.6.2 SIBILARISATION

This process, also known as *druga palatalizacija* ‘second palatalisation’, is responsible for the transformation of /k, g, x/ → /ts, z, s/

when followed by /i/ (Barić et al., 2005, p. 616; Browne & Alt, 2004, pp. 16–17; Ham, 2017, p. 17). We can witness it in, *inter alia*:

1) In the formation of the imperative of verbs with stem-final /k/, /g/ or /x/:

rek- ['reːk] + *-i* → *reci* ['reːtsi] (from *reći* ‘to say’).

pomog- ['pomoːg] + *-i* → *pomozi* ['pomoːzi] (from *pomoći* ‘to help’).

vrh- ['vr̥x] + *-i* → *vrsi* ['vr̥si], (from *vrći* ‘to thresh’).

2) In masculine nominative plurals in *-i* and dative/locative/instrumental plural in *-ima*:

učenik ['utʃeniːk] + *-i* → *učenici* [utʃeniːtsi] ‘students’.

siromah ['siromax] + *-ima* → *siromasima* ['siromasima] ‘to the poor’.

3) In the dative and locative singular of feminine nouns of the *-a* declension:

ruka ['ruːka] + *-i* → *ruci* ['ruːtsi], ‘hand/arm’.

noga ['noːga] + *-i* → *nozi* ['noːzi], ‘foot/leg’.

1.3 THE CROATIAN CASE SYSTEM

According to Ham's definition (2017, p. 39) case endings "express the relationship of the noun with the other words in the sentence [so that] by changing its case the noun changes its relationship with the other words."[13] Similarly, Matovac defines cases as "the different forms of a noun indicating its different functions in a sentence" (2022, p. 19).

In Croatian, the case system consists of seven cases, namely nominative, genitive, dative, accusative, vocative, locative and instrumental. In addition to case, nominal endings also change according to gender and number, which results into Croatian featuring several different inflectional noun classes, as you can see in the following overview taken from Ham (2017, p. 35).

	živo	neživo		jednina			množina		
				ž. r.	m. r.	s. r.	ž. r.	m. r.	s. r.
nominativ	**tko?**	**što?**	stoji	srn-**a**	stol-**ø**	sel-**o**	srn-**e**	stolov-**i**	sel-**a**
genitiv	**koga?**	**čega?**	nema	srn-**e**	stol-**a**	sel-**a**	srn-**ā**	stolov-**ā**	sel-**ā**
dativ	**komu?**	**čemu?**	prilazim	srn-**i**	stol-**u**	sel-**u**	srn-**ama**	stolov-**ima**	sel-**ima**
akuzativ	**koga?**	**što?**	vidim	srn-**u**	stol-**ø**	sel-**o**	srn-**e**	stolov-**e**	sel-**a**
vokativ	**ej!**	**ej!**		srn-**o**	stol-**e**	sel-**o**	srn-**e**	stolov-**i**	sel-**a**
lokativ	**o kom?**	**o čem?**	govorim	o srn-**i**	stol-**u**	sel-**u**	srn-**ama**	stolov-**ima**	sel-**ima**
instrumental	**kim?**	**čim?**	upravljam	srn-**om**	stol-**om**	sel-**om**	srn-**ama**	stolov-**ima**	sel-**ima**

Figure 4 – The Croatian case system. Image taken from Ham (2017, p. 35)

For native speakers, the cases listed in the first column above are inextricably linked to the inflected forms of interrogative personal pronouns, which help them figure out which case ending is required by a given context without having to focus on the abstract grammatical category of case. The first column of inflected pronouns is used for animate referents and the second for inanimate ones. With the exception of vocative, which doesn't have a pronoun but only an exclamation (*ej!* 'hey!'), the other interrogative forms can be translated as follows:

- *tko/što?* 'who/what?'
- *koga/čega?* 'of whom/of what?'
- *komu/čemu?* 'to whom/to what?'
- *koga/što?* 'who/what?'
- *o komu/čemu?* 'about whom/about what?'
- *kim/čim?* 'by means of whom/by means of what?'

[13] "Padež izriče odnose imenice s ostalim riječima u rečenici i promjenom padeža imenica mijenja svoj odnos prema drugim riječima."

The fourth column provides a list of verbs whose valency structures require their arguments to be declined according to the cases to the left. However, as displayed by the last six columns, a *padežni nastavak* 'case ending' may change also according to number (*jednina*/*množina* 'singular/plural'), and gender (*ženski/muški/srednji rod* 'feminine/masculine/neuter gender'). Therefore, in the last six columns, we can see how three nouns belonging to the three different genders are declined both by case and number, namely the feminine noun *srna* 'doe', the masculine word *stol* 'table', and the neuter noun *selo* 'village'. In the following paragraphs, we will provide an overview of the inflectional paradigms and main functions of all Croatian cases.

1.3.1 NOMINATIVE

The nominative form is the citation form of a noun. The primary function of the nominative is that of expressing the subject of a sentence (Barić et al., 2005, p. 421; Ham, 2017, p. 126; Matovac, 2022, p. 42; Silić & Pranjković, 2007, p. 199), as in the following examples from Matovac (2022, p. 42):

(2) ***Marko*** *čita knjigu.*
'Marko is reading a book.'

(3) ***Televizor*** *ne radi.*
'The television does not work.'

In addition, nominatives can also express predicate nouns when they appear together with a linking verb such as *biti* 'to be', as in the following sentences also taken from Matovac (2022, p. 42):

(4) *Marko je* ***student****.*
'Marko is a student.'

(5) *Mercedes je* ***auto****.*
'Mercedes is a car.'

When it comes to inflection, as you can see from Table 2 (Čilaš-Mikulić et al., 2006, p. 27), we can identify three main inflectional noun classes depending on the gender of the nouns involved.

	Masculine	**Feminine**	**Neuter**	
singular	*poklon* 'gift'	*knjiga* 'book'	*stablo* 'tree'	*sunce* 'sun'
plural	*pokloni* 'gifts'	*knjige* 'books'	*stabla* 'trees'	*sunca* 'suns'

Table 2 – Singular and plural nominative forms of the three main inflectional noun classes. Table adapted from (Čilaš-Mikulić et al., 2006, p. 27).

As you can see in Table 2, the nominative singular of masculine nouns prototypically ends in a consonant (*poklon* 'gift'), but we can also find nominative singular forms ending in a vowel, especially when it comes to loanwords from languages that feature masculine vowel endings, such as *auto* 'car', *bendžo* 'banjo', *derbi* 'derby', or *intervju* 'interview'[14].
The nominative plural of masculine forms usually ends in *-i*, which is also the case for loanwords such as *auto, auti* 'car, cars', albeit with some exceptions.

A further subclass of monosyllabic masculine nouns can be identified. The nouns belonging to this group become longer in the plural by adding the morphemes *-ov-* or *-ev-* before the nominative plural ending *-i*, as in *lav, lavovi* 'lion, lions', and *noj, nojevi* 'ostrich, ostriches'. The two morphemes are actually allomorphs: the allomorph *-ev-* is required after palatals (*č, ć, dž, đ, j, lj, nj, š, ž*), the clusters *št*, *žd* and sometimes *c* or *r* for articulatory reasons (Ham, 2017, p. 47; Matovac, 2022, p. 39; Silić & Pranjković, 2007, p. 103).

When it comes to the formation of their nominative plural forms, another notable group of masculine nouns are those ending in *k, g,* or *h* in the nominative singular. As already mentioned in section 1.2.6.2., when these sounds are followed by an *-i* ending, the process of sibilarisation (see § 2.2.6.2) can take place, changing them into the sibilants *c, z, s*, as in *oblak, oblaci* 'cloud, clouds', *jastog, jastozi* 'lobster, lobsters', *orah, orasi* 'walnut, walnuts' (Ham, 2017, p. 47; Matovac, 2022, p. 40; Silić & Pranjković, 2007, p. 99).

Last but not least, another important morpho-phonological process that may take place, *inter alia*, when masculine nouns take on their plural form is the so-called *nepostojano a* 'mobile a', as in the case of *Čileanac, Čileanci* 'Chilean, Chileans', *policajac, policajci* 'policeman, policemen' or *pas, psi*[15] 'dog, dogs'. The Croatian label for this phenomenon actually means "fleeting a" since the morpheme is not always present in the paradigm of these nouns, except in the nominative singular (*policajac*) and the genitive plural (*policajaca*).

Going back to Table 2, the nominative singular of standards feminine nouns ends in *-a*, while their plural ends in *-e*, as in the case of *knjiga, knjige* 'book, books', or *žena, žene* 'woman, women'. There is, however, an additional closed group of feminine nouns ending in consonant, such as *obitelj* 'family', *kost* 'bone', and *krv* 'blood', that make up the so-called *i-vrsta* 'i type'. These nouns take on *-i* as their nominative plural ending and follow a completely different inflectional paradigm than other feminine nouns (Ham, 2017, p. 52; Matovac, 2022, p. 41; Silić & Pranjković, 2007, p. 111).

[14] Please, take notice of the phonological adaptation of loanwords in Croatian. As a matter of fact, since Croatian has a phonetic alphabet, loanwords are adapted sound by sound depending on the degree of similarity in the description of phonemes between the lending language and the borrowing language (Croatian). More on this in Filipović (1977).

[15] *Psi* 'dogs' is a further exception since, although its singular form *pas* 'dog' is monosyllabic, it does not become **psovi*.

Finally, the nominative singular of most neuter nouns ends in *-o*, as in the case of *stablo* 'tree', *pivo* 'beer' or *nebo* 'sky'. However, as in the case of the alternation between *-ov-* and *-ev-* in the long plurals for monosyllabic masculine nouns, also the nominative singular of neuter nouns takes an *-e* ending if preceded by one of the following palatal consonants (*č, ć, dž, đ, j, lj, nj, š, ž*), consonant clusters (*št*, *žd*) or sometimes letters *c* or *r*, as in the case of *sunce* 'sun' from Table 2.

Regardless of the ending, both neuter forms take on an *-a* ending for their nominative plural, as in the case of *stabla* 'trees' and *sunca* 'suns'.

1.3.2 ACCUSATIVE

The accusative case in Croatian is the case primarily used to express directs objects[16], i.e., those objects referring to persons or things directly involved by the action expressed by the verb (Barić et al., 2005, p. 432; Ham, 2017, p. 130; Matovac, 2022, p. 74; Silić & Pranjković, 2007, p. 223), as you can see in the following examples taken from Matovac (2022, p. 74):

(6) *Marko čita* ***knjigu***.
'Marko is reading a book.'

(7) *Imam* ***sestru*** *i* ***brata***.
'I have one brother and one sister.'

In terms of inflectional paradigms, Table 3 portrays the singular and plural accusative forms of the three inflectional classes that we have already identified for masculine, feminine and neuter nouns.

	Masculine		**Feminine**	**Neuter**	
singular	*otok* 'island'	*prijatelja* 'friend'	*kravatu* 'tie'	*stablo* 'tree'	*more* 'sun'
plural	*otoke* 'islands'	*prijatelje* 'friends'	*kravate* 'ties'	*stabla* 'trees'	*mora* 'suns'

Table 3 – Singular and plural accusative forms of the three main inflectional noun classes. Table adapted from (Čilaš-Mikulić et al., 2006, p. 74)

[16] As effectively summarized by Ham (2017, p. 132), objects can be divided into *izravni* 'direct' and *neizravni* 'indirect'. Direct objects are non-prepositional and can be expressed by the accusative, the partitive genitive, or the Slavic genitive, which is a term used to refer to those genitives used instead of the accusative in a negative sentence (Barić et al., 2005, p. 446). On the other hand, indirect objects can both be prepositional and non-prepositional. Non prepositional ones can be found in the genitive, dative and instrumental case, whereas prepositional objects accept all cases apart from nominative and vocative (genitive, dative, accusative, locative and instrumental). In light of this, we can say that the only case which always needs a preposition is the locative case.

The main characteristic of the accusative singular form of masculine nouns is that it differs depending on the animacy of the noun, that is to say, the ending changes depending on whether the noun refers to something animate or animate. If the noun refers to an animate entity, then the ending is *-a*, as in *prijatelja* 'friend'. On the other hand, if the entity is inanimate, the accusative singular form remains as the nominative singular form, as in the case of *otok* 'island'. The accusative plural of both animate and inanimate masculine nouns is the same and ends in *-e*.

When it comes to feminine nouns, the accusative singular form takes on the ending *-u*, as in the case of *kravatu* 'tie', while the accusative plural takes on the same *-e* ending as the masculine accusative plural.

Finally, the accusative singular and plural of neuter nouns remains the same as their singular and plural nominative counterparts.

As for the prepositional uses of the accusative[17], three of its most frequent ones are: with the preposition *za* 'for' to express the purpose of something, as in example (8); with the preposition *po* 'for' to express the meaning 'to go get something or pick up someone', as in example (9); and finally with verbs of movements such as *ići* 'to go' to express motion as in examples (10) and (11). All examples are taken from Matovac (2022, p. 74).

(8) *Marko ima poklon **za mamu**.*
'Marko has a gift for mom.'

(9) *Marko ide u trgovinu **po kruh i mlijeko**.*
'Marko is going to a shop to get bread and milk.'

(10) *Marko putuje **u Francusku**.*
'Marko is travelling to France.'

(11) *Marko ide **na koncert**.*
'Marko is going to a concert.'

When it comes to the last use, as you can see from examples (10) and (11), an alternation can be witness between two prepositions, *na* 'on' and *u* 'in', which in this context can be both translated with 'to'. The choice between the two depends on the type of location one is going to, i.e., the semantics of the noun imposes a selectional requirement on the preposition preceding it (Čilaš Mikulić et al., 2021, pp. 100–101; Čilaš-Mikulić et al., 2006, p. 93; Matovac, 2022, p. 78; Silić & Pranjković, 2007, p. 226)[18]. When the "goal of motion is a city, country, continent, or any kind of closed space, e.g., a

[17] For a more in-depth overview, see Silić & Pranjković (2007, pp. 224–230).

[18] The textbooks *Hrvatski za početnike 1: Udžbenik i rječnik (A1 i A2)* (2006) and *Razgovarajte s nama! Udžbenik*

house, a room or a building, [then the] preposition *u* 'in' will be used" (Matovac, 2022, p. 78). On the other hand, "if the goal of motion is an island, an open space, a location one climbs onto, a flat surface, [an activity, or an event, then the] preposition *na* 'on' will be used" (*ibidem*).

Indeed, from a cognitive linguistic perspective, *na* and *u* "are the two basic [Croatan] prepositions denoting the most central spatial relations – support and containment" (Šarić, 2008, p. 32) – which are the basis for meaning extensions in other domains.

Moreover, in addition to "support", *na* also implies "contiguity and coincidence" between the entities involved (Šarić, 2008, p. 40), which explains why it can be used with activities, since it allows for a conceptualisation of the place where the activity takes place and the activity itself as coincident, and, unlike *u*, it does not require clearly defined limits and boundaries (Šarić, 2008, p. 48).

The same alternation between *na* and *u* takes place in static contexts when they require the locative case (see section 1.3.5).

1.3.3 GENITIVE

The genitive case is the case with the broadest and most general meaning. Its basic function is that of specifying relationships between entities, be these relationships temporal, spatial or even logically more complex (Silić & Pranjković, 2007, p. 201)[19].

As for its inflectional paradigm, Table 4 provides an overview of all possible genitive endings.

	Masculine		**Feminine**		**Neuter**	
singular	*prijatelja* 'of the friend'	*studenta* 'of the student'	*mame* 'of the mum'	*djevojke* 'of the girl'	*stabla* 'of the tree'	*pismo* 'of the letter
plural	*prijatelja* 'of the friends'	*studenata* 'of the students'	*mama* 'of the mums'	*djevojaka* 'of the girls'	*stabla* 'of the trees'	*pisama* 'of the letters'

Table 4 – Singular and plural genitive forms of the three main inflectional noun classes. Table adapted from (Čilaš-Mikulić et al., 2006, p. 218)

hrvatskoga jezika za razine A1 i A2 (2021), published by Hrvatska Sveučilišna Naklada, are the result of the fruitful collaboration of all their authors, who we would like to list here: Marica Čilaš Mikulić, Milvia Gulešić Machata, Sanda Lucija Udier and Dinka Pasini (involved only in the writing of *Hrvatski za početnike 1*).

[19] "Genitiv je padež ticanja. Od svih kosih padeža on ima najšire i neopćenitije značenje. On znači da je kakav predmet (u sirem smislu) u odnosu s nekim drugim predmetom. Taj odnos može biti prostorni ili vremenski, a može biti i logički složeniji."

As you can see, apart from the genitive singular form of feminine nouns ending in *-e*, all other forms bear an *-a* ending. To help differentiate between genitive forms, the nouns ending in a consonant cluster in the nominative singular (like *student* 'student', *djevojka* 'girl', or *pismo* 'letter') take on an extra *-a-* in the genitive plural, which – when inserted in the consonant cluster – gives rise to the recognisable forms *studenata* 'of the students', *djevojaka* 'of the girls' and *pisama* 'of the letters'.

Going back to the functions of the genitive case, bare genitives can be used to specify other nouns (Matovac, 2022, p. 161), as in (12), and are also used with quantity words and numbers[20] (except for number 1 and all other numbers ending in 1 when pronounced) to specify the amount of entities we are talking about, as in examples (13) and (14).

(12) *Marko je student* ***matematike****.*
'Marko is a student of mathematics.'

(13) *Marko ima* ***tri sestre****.*
'Marko has three sisters.'

(14) *Marko ima* ***puno prijatelja***
'Marko has a lot of friends.'

Given its linking function, genitive combines with the most prepositions, taking on a wide array of meanings[21]. Among the many, we recall the following: *ispred* 'in front of', *iz* 'from', *pokraj* 'next to', *prije* 'before', *poslije* 'after', *zbog* 'because of', *osim* 'except', *bez* 'without', *protiv* 'against'.

1.3.4 DATIVE

The basic meaning of the dative case is that of directionality, which then gives rise to other extended meanings such as that of goal, of recipient, of beneficiary and so on (Barić et al., 2005, pp. 438–440; Matovac, 2022, pp. 124–126; Silić & Pranjković, 2007, pp. 219–222).

Table 5 portrays the singular and plural dative forms of masculine, feminine and neuter nouns, which coincide with the locative forms, both in the singular and in the plural (compare with Table 6).

[20] Numbers 2, 3, and 4 (and all numbers ending in 2, 3 or 4 when pronounced) are followed by nouns which are said to be in the genitive singular case (for didactic purposes). However, the case in question is actually paucal, which is the residue of what once was the Slavic dual case. All other numbers and quantity words like *puno* 'a lot of' are followed by nouns in the genitive plural (Matovac, 2022, pp. 185–187).

[21] For a complete overview of the prepositions combining with the genitive, see (Silić & Pranjković, 2007, pp. 203–219).

	Masculine	Feminine	Neuter
singular	*studentu* 'to the male student'	*studentici* 'to the female student'	*vremenu* 'to the time'
plural	*studentima* 'to the male students'	*studenticama* 'to the female students'	*vremenima* 'to the times'

Table 5 – Singular and plural dative forms of the three main inflectional noun classes. Table adapted from (Čilaš-Mikulić et al., 2006, p. 163)

As is the case with locative, sibilarisation (see § 2.2.6.2) occurs in the dative singular of feminine nouns and in the dative plural of masculine nouns, when *k*, *g*, or *h* are followed by the endings *-i* and *-ima*, respectively, as in the case of *vojnik, vojnicima* 'soldier, to the soldiers', *majka, majci* 'mother, to the mother' or *supruga, supruzi* 'wife, to the wife'. Sibilarisation does not occur in hypocoristic nouns indicating intimacy (e.g., *baka*, *baki* 'granny, to granny'), nor when *k*, *g*, or *h* are part of a consonant cluster (Matovac, 2022, p. 125).

In Croatian, dative can be used both independently and with a preposition. When used independently, it usually expresses a recipient, as in example (15), an experiencer, as in example (16), or it is required by the valency structure of a specific verb, as in (17).

(15) *Marko je dao poklon* ***mami****.*
'Marko gave a present to mum.'

(16) ***Marku*** *je hladno.*
'Marko is cold'.

(17) *Marko se veseli* ***pauzi.***
'Marko is looking forward to the break'.

When used with prepositions[22], it is most often used to express a direction, as in the case of both *prema* 'towards' (18) and *k* 'to, at' (19). All examples are taken from Matovac (2022, p. 125).

(18) *Marko putuje* ***prema moru****.*
'Marko is travelling towards the sea'.

(19) *Marko ide* ***k baki****.*
'Marko is going to his grandma's'.

[22] For a comprehensive account of all prepositional uses of the dative case, see Silić & Pranjković (2007, pp. 220–223).

1.3.5 LOCATIVE

The locative case is the only Croatian case that always requires a preposition. Its basic meaning is spatial and is linked to a static and motionless location (Silić & Pranjković, 2007, p. 230)[23].

Table 6 portrays the singular and plural locative forms of masculine, feminine and neuter nouns, which bear the same endings as their corresponding dative forms (compare with Table 5).

	Masculine	**Feminine**	**Neuter**
singular	*na stolu* 'on the table'	*u sobi* 'in the room'	*na računalu* 'on the computer'
plural	*na stolovima* 'on the tables'	*u sobama* 'in the rooms	*na računalu* 'on the computers'

Table 6 – Singular and plural locative forms of the three main inflectional noun classes. Table adapted from (Čilaš-Mikulić et al., 2006, p. 130)

Moreover, as in the case of the dative, sibilarisation (see § 2.2.6.2) occurs in the locative singular form of feminine nouns such as *ruka, u ruci* 'hand, in the hand', *noga, na nozi* 'leg, on the leg', as well as in the locative plural forms of masculine nouns, as in the case of *otok, na otocima* 'island, on the islands'.

As already mentioned in section 1.3.2, the locative case allows for the same alternation between the prepositions *na* 'on' and *u* 'in' allowed by the accusative case for the expression of a goal of motion, but this time when it comes to expressing static locations (Čilaš Mikulić et al., 2021, pp. 170–171; Čilaš-Mikulić et al., 2006, p. 130; Matovac, 2022, p. 90; Silić & Pranjković, 2007, pp. 230–233). Also in this case, *u* 'in' is used when the location is a closed or delimited space, as in example (20), while *na* 'on' is selected when open spaces, surfaces, or activities are involved, as in example (21) (Matovac, 2022, p. 92):

(20) *Marko je **u trgovini**.*
'Marko is in a shop.'

(21) *Marko pije kavu **na trgu**.*
'Marko is drinking coffee in the square'.

Other two prepositions frequently used with the locative case are *po* 'all over, for' and *o* 'about'. *Po* can take on two main meanings depending on the context: that of 'around, all over' when there is

[23] "Temeljno je značenje lokativa prostorno. Lokativ naime označuje mjesto, i to mjesto vezano uz mirovanje, statičnost, nepokretnost."

motion within a delimited surface (Matovac, 2014, pp. 13–15), as in example (22), or that of 'by, for' when it refers to the reason for somebody's fame, as in example (23). Finally, when paired with the preposition *o* 'about', the locative case is able to express the topic of a conversation, as in (24).

(22) *Marko šeta **po parku**.*
'Marko is walking around the park.'

(23) *Hrvatska je poznata **po moru**.*
'Croatia is famous for its sea.'

(24) *Marko i Ivana razgovaraju **o Zagrebu**.*
'Marko and Ivana are talking about Zagreb.'

1.3.6 VOCATIVE

Vocative is the case we use to call or address a person (Barić et al., 2005, p. 102; Matovac, 2022, p. 200; Silić & Pranjković, 2007, p. 200). Syntactically speaking, vocative forms are not part of the sentence, and their isolation is overtly marked by the use of a comma. Over time, this case is becoming less and less used, and nominative forms are taking its place.

In terms of case inflection, vocative and nominative endings already overlap in most of the paradigm, as you can see from Table 7.

	Masculine		**Feminine**		**Neuter**	
singular	*Ivane* 'Ivan!'	*mužu* 'husband!'	*ženo* 'wife!'	*profesorice* 'professor!'	*selo* 'village!'	*sunce* 'sun!'
plural	*muškarci* 'men!'	*muževi* 'husbands!'	*žene* 'wives!'	*profesorice* 'professors!'	*sela* 'villages!'	*sunca* 'suns!'

Table 7 – Singular and plural vocative forms of the three main inflectional noun classes. Table adapted from (Čilaš-Mikulić et al., 2006, pp. 276–277)

Plural vocative forms coincide with their nominative counterparts throughout the Table, as do singular neuter ones. The only difference are the masculine and feminine singular forms.

The vocative form of a masculine nouns typically ends in *-e*, as in the case of *Ivane* 'Ivan!'. However, if the ending is preceded by a palatal consonant (*č, ć, dž, đ, j, lj, nj, š, ž*), a consonant cluster (*št, žd*), or sometimes letters *c* or *r*, the ending becomes *-u*, as in *mužu* 'husband!' or *učitelju* 'teacher!' (Matovac, 2022, p. 201). On the other hand, when the nominative form of a masculine noun ends in *k*, *g*, *h* or *c*, palatalisation (§ 2.2.6.1) takes place because of the *-e* ending, which causes the consonants to change into *č, ž, š,* or *č*, respectively, as in *čovjek, čovječe* ('man, man!') or *stric, striče* ('uncle, uncle!').

As for feminine nouns, their vocative singular form typically ends in *-o*, as in *žena, ženo* 'wife, wife!'. Only feminine nouns ending in *-ica* take on an *-e* ending, as in the case of *ljepotica, ljepotice* 'beautiful, beautiful!'

1.3.7 INSTRUMENTAL

The instrumental case is primarily used to express the instrument or means an action is performed with or the people in whose company it has taken place (Matovac, 2022, p. 145; Silić & Pranjković, 2007, p. 234). Depending on its use, instrumental can appear both with and without prepositions, as shown by the different examples in Table 8, which portrays its inflection.

	Masculine		**Feminine**	**Neuter**
singular	*autom* 'by car'	*s prijateljem* 'with a friend'	*s mamom* 'with mum'	*računalom* 'with the computer'
plural	*autima* 'by car'	*s prijateljima* 'with friends'	*s mamama* 'with the mums'	*računalima* 'with the computers'

Table 8 – Singular and plural instrumental forms of the three main inflectional noun classes. Table adapted from (Čilaš-Mikulić et al., 2006, p. 204)

As shown above, the instrumental singular form of masculine, feminine, and neuter nouns typically ends in *-om*, while the plural form of masculine and neuter nouns ends in *-ima*, and the feminine one in *-ama*. Therefore, we can state that the plural dative, locative, and instrumental forms of Croatian nouns are a typical case of syncretism, which are disambiguated by context.

If we look at Table 8, we can also observe an alternation between *o* and *e* in the instrumental singular forms of masculine nouns, whose endings change from *-om* into *-em* after *č, ć, dž, đ, j, lj, nj, š, ž, št, žd* and sometimes *c* or *r*, as in examples (25) and (26) taken from Matovac (2022, p. 144).

(25) *Igram nogomet s **prijateljem**.*
'I am playing football with a friend.'

(26) *Vozim se **tramvajem**.*
'I am going around by tram.'

Last but not least, sibilarisation (see § 2.2.6.2) occurs in the instrumental plural form of masculine nouns when *k, g,* or *h* are followed by *-i*, thus changing into *c*, *z,* and *s*, as in *unuk, s unucima* 'grandson, with grandsons', and *arheolog, s arheolozima* 'archaeologist, with archaeologists'.

As for the main functions of the case, instrumental is primarily used by itself to express instruments, such as means of transport, as in example (26), and with the preposition *s* 'with' to

express complements of company as in example (25).

Among its many other functions, we recall that it can be used with prepositions *pod* ‘below’, *nad* ‘above’, *pred* ‘in front of’, *za* ‘behind’ and *među* ‘among’ to express locations, as in (27).

(27) *Marko je **pred fakultetom**.*
‘Marko is in front of the faculty.’

After having provided a comprehensive introduction to the sociolinguistic make-up (§ 1.1), as well as to the phonological (§ 1.2) and case system (§1.3) of the Croatian language, which we hope will allow even novices to this language to gain a good understanding of the subject matter, we now turn to Chapter 2 to focus on the two theoretical notions at the heart of this work, namely those of aspect and *Aktionsart*, that are central to the Croatian verb system and its aspectual pairs (§ 2.3.1).

CHAPTER 2: ASPECT AND *AKTIONSART*

> "Whatever tense, or the tenses, may be, speakers [of the familiar Western European languages] have some sort of notions about them: it is satisfying, for example, to consider the past tense to express past time. But aspect is not a traditional concept in the same way, and speakers of most European languages have no very clear notions concerning it." (Binnick, 1991, p. 135)

What is verbal aspect? What is the difference between grammatical and lexical aspect? What is meant by *Aktionsart*? Is there a difference between the term *aspect* and the term *vid* in the Slavic tradition?

Answering these questions is no trivial task, since the terminology concerning these categories was developed by different traditions focusing on different facets of the concepts involved, while at the same time influencing each other often resorting to calques.

In section 2.1, a short history of the terminology surrounding the topic is presented, followed by a focus on the current definitions of the involved phenomena in Western literature (§ 2.2), and a comparison with the Croatian perspective (§ 2.3). Croatian aspectual verb pairs (§ 2.3.1) will be introduced in this Chapter, as well as their morphology.

2.1 A SHORT HISTORY OF TERMINOLOGY

The verb and its categories have always drawn the interests of scholars since ancient times. The first ones to distinguish between the three main verbal tenses (past, present, and future) were Aristotle and Plato, soon followed by the Stoics who added *aorist* and *imperfect*, even though they did not ascribe their difference to an aspectual one. Similarly, Marcus Terentius Varro (116 – 27 BC) contributed to the development of Latin grammar distinguishing between *infectum* and *perfectum* and is considered by many the first one to identify the category of aspect (Binnick, 1991, p. 34), which will however be properly characterised only many centuries afterwards.

The first appearance of the term *vid* (Rus. 'vidy', from the Greek 'eidos', Eng. 'aspect') was made in the Russian grammars of the 15th and 16th century, where it was used as a synonym of "verbal type" (Gojmerac, 1980, p. 9). The term was then borrowed by M. Smotricki in his *Slavonic Grammar with Correct Syntax* (1619), where he uses it to divide verbs into primary and derived based on their morphological structure. Smotricki called primary verbs *soveršennyj vid*, a label which has lived on in Slavic grammars until now, despite its change in meaning (*svršeni glagoli* are 'perfective verbs' in Croatian). Another pivotal contribution to the development of the category of aspect was indeed made by M.V. Lomonosov, who in his 1755 Russian grammar identified ten verbal tenses divided into three main groups: *soveršennyj, neoperdelennyj*, and *odnokratny.* These groupings, however, had not much

to do with tenses, but rather with being derivationally related by prefixation from a base verb (*soveršennyj*), with being derived from the present tense (*neoperdelennyj*), or finally from the *preteritum* (*odnokratny*). When J. S. Vater learned of this categorization, he called these three morphological groupings: *perfectum*, *imperfectum*, and *simplex* (Čilaš Mikulić, 2012, p. 9). According to Gojmerac (1980, p. 11) it was indeed with this categorization that the category of aspect detached itself from that of verb tense, even though the term *vid* kept on being used also to refer to the different verb classes one can identify using lexical-semantic criteria[24], which is still a problem to this day.

The first time the term *vid* was translated with the term *aspect* was in the French translation of Greč's Russian Grammar of 1827 by Karl Philipp Reiff. After that, the term entered several Western European languages throughout the 19th century – e.g., English in 1853 (Novak Milić, 2010, p. 126) – and started to intertwine with the notion of *Aktionsart*, which was originally chosen as the German translation of *vid*, but soon began to be used to refer to both the perfective vs. imperfective distinction, as well as "the manner in which the event expressed by the verb unfolds"[25] – a definition provided by Brugmann (1904, pp. 492–493).

Meanwhile, Slavic aspectologists were trying to answer the same theorical questions using their own terminology. For instance, Vostokov (1831) distinguished between three *vida* ('aspects'), imperfective, perfective, and iterative, and several *podvidova* ('sub aspects'), i.e., verbal groupings made on the basis of semantic features, such as ingressive, egressive, durative, etc. Indeed, at that time, "the expression *podvid* meant what today we mean by *Aktionsart"*[26] (Gojmerac, 1980, p. 12). – or *način vršenja glagolske radnje*[27], if we want to use an equivalent label used in Croatian literature.

This being said – as Čilaš Mikulić points out – "it is not surprising that the mixing of the concepts of aspect and *Aktionsart* happened in Slavic linguistics, since both categories are encoded in the verb and [at the same time] the syntactic context in which the verb appears is relevant for building the meaning of both"[28] (2012, p. 9).

---

[24] "Ovime se glagolski *vid* definitivno odvojio od kategorije glagolskog vremena, ali i dalje je značio vrstu, tip, kategoriju u različitim podjelama glagola, uglavnom prema leksičko-semantičkim kriterijima."

[25] "Aktionsart ist, im Gegensatz zur Zeitstufe, die Art und Weise, wie di Handlung des Verbums vor sich geht."

[26] "Pojam podvid znači u to vrijeme otprilike isto što danas znači *Aktionsart*."

[27] Literally, 'manner of carrying out the verbal action'.

[28] "Smatramo da nije slučajnost da se miješanje pojmova vid i *aktionsart* dogodilo se u slavistici jer su i jedna i druga kategorija kodirani u samome glagolu [...], te je za određivanje značenja i jedne i druge relevantan sintaktički kontekst u kojem se pojavljuje promatrani glagol."

To add more colour to the plot, with the publication of Curtius's *Formation of the Tenses and Moods in Greek and Latin* (1846), it was assumed that Greek and Slavic aspect represented the same phenomenon. A consideration that was soon afterwards extended by Jacob Grimm to all Germanic languages (Binnick, 1991, p. 141).

However, it was not until the pivotal contribution *Aspektänderung und Aktionsartbildung beim polnischen Zeitworte* (1908) by the Swedish Slavicist Sigurd Agrell that a clear line was drawn between the concept of *Aktionsart* and aspect in their current understanding:

> "I do not use the label *Aktionsart* to refer to the two main Slavic verb categories, namely the unterminative and the terminative forms (the Imperfective and the Perfective) – those I call *aspects*. With the expression *Aktionsart*, I refer to the until now mostly unobserved – and even less classified – *meaning functions of affixed verbs* […], that better express how the action is carried out, the way its development unfolds".[29] (1908, p. 78)

The importance of this work cannot be stressed enough. According to Gojmerac (1980, p. 18), Agrell is the first one to define aspect as the category allowing for different ways of looking at the events expressed by verbs, which is absolutely in line with later definitions such as Comrie's, who famously defines aspect as the "different ways of viewing the internal temporal constituency of a situation" (1976, p. 3). Comrie also adds that aspect is the category encoding the difference between *he was reading* and *he read* (which English grammars insist on referring to as different tenses), and that "perfectivity indicates the view of a situation as a single whole, without distinction of the various separate phases that make [it] up; while the imperfective pays essential attention to the internal structure of the situation" (1976, p. 16).

Following Koschmieder's works in the 1930s, the two concepts of aspect and *Aktionsart* started not only to develop independently from one another, but also to become more and more associated with grammaticalization (the former) and lexicalisation (the latter), to the point that they sometimes go by the name of *grammatical aspect* and *lexical aspect*. Quoting Binnick:

[29] "Unter *Aktionsart* verstehe ich […] nicht die beiden Hauptkategorien des slavischen Zeitwortes, die unvollendete und die vollendete Handlungsform (das Imperfektivum und das Perfektivum) – diese nenne ich *Aspekte*. Mit dem Ausdruck *Aktionsart* bezeichne ich bisher fast gar nicht beachtete – geschweige denn klassifizierte – *Bedeutungsfunktionen der Verbalkomposita* […], die genauer ausdrücken, wie die Handlung vollbracht wird, die Art und Weise ihrer Ausführung markieren".

> "Aspect is a fully grammaticized, obligatory, systematic category of languages, operating with general oppositions such as that of perfective and non-perfective, while *Aktionsarten* are purely lexical categories, nongrammatical, optional, and unsystematic". (1991, p. 170)

Aspectual class, aspectual character, situation type, situation aspect, action, verb class, lexical aspect, inherent aspect, and eventuality type are among the many labels *Aktionsarten* can be referred to (Polančec, 2020, p. 41). In this work, the terms aspect and *vid* will therefore be used interchangeably since they both refer to grammatical aspect, whereas for lexical aspect, the terms *Aktionsart*, aspectual classes and event type will be used depending on the context.

2.2. CURRENT WESTERN DEFINITIONS

In the following paragraphs, definitions of both aspect and *Aktionsart* stemming from the most relevant international works dealing with aspectual questions will be provided, to shed some light on the current interpretation of these phenomena, before looking at the way they are systematised in Croatian literature (§ 2.3).

2.2.1 ASPECT

The most common definition of *aspect* is the one provided by Comrie (1976), which is completely in line with Agrell (1908). Comrie defines aspect as the "different ways of viewing the internal temporal constituency of a situation" (1976, p. 3) and adds that is the category encoding the difference between *he was reading* and *he read*, which English insists on referring to as different tenses, while in reality they are different aspects.

According to him, "perfectivity indicates the view of a situation as a single whole, without distinction of the various separate phases that make up that situation; while the imperfective pays essential attention to the internal structure of the situation" (1976, p. 16). That being said, he goes on providing examples – from both Slavic and Romance languages – that it is not the duration of an event that influences the choice of the perfective form over the imperfective one, but rather what we want to highlight when selecting one aspect over the other (1976, p. 17). For example, let us compare the following sentences featuring an Italian perfective tense (1.a) and an imperfective one (1.b).

(1) a. *Il re **regnò** per trent'anni*.
'The king reigned for thirty years'.

b. *Il re **regnava** da trent'anni quando finalmente ebbe una figlia.*
'The king had been reigning for thirty years when he finally had a daughter.'

In (1.a), by using the perfective tense called *Passato Remoto*, the whole period of thirty years is condensed into a single complete whole, while in (1.b) the use of the *Imperfetto* carries the meaning that at any point during those thirty years the king was indeed reigning. Moreover, since the Italian *Imperfetto* is used to express simultaneity with another past tense (Bertinetto, 1986), the form *regnava* serves as a background statement to introduce the event of the birth of the king's daughter.
That being said, Comrie goes onto criticizing the idea that the basic function of the perfective should be that of representing events as momentary or punctual. As an alternative, he points out that – rather than reducing an event to a single point – "a more helpful metaphor would perhaps be to say that the

perfective reduces a situation to a *blob*" (1976, p. 18). Indeed, being a three-dimensional object with potential internal complexity and clearly circumscribed limits makes a blob a better representation of the perfective view of a situation. This idea ties in quite nicely with the representation of perfectivity that is going to be put forth in this thesis using the formalism of event structures to model Croatian aspectual verb pairs (see Chapter 4).

Last but not least, Comrie clarifies that he prefers the term *complete* to *completed* when talking of perfective actions, since the latter puts "too much emphasis on the termination of the situation, whereas the use of the perfective puts no more emphasis, necessarily, on the end of a situation than on any other part of the situation, rather all parts of the situation are presented as a single whole" (1976, p. 18). The example he brings is from Russian, but it works just as well for Croatian, and consists in the future use of perfective verbs. Indeed, the fact that a perfective verb may be used to express a future action – which is for sure not completed yet – proves the point.

As for imperfectivity, Comrie defines it as making "explicit reference to the internal temporal structure of a situation, viewing a situation from within" (1976, p. 24). Figure 1 provides an overview of the different aspectual oppositions we find in language systems, with a focus on the different semantic distinctions that pertain to the realm of imperfectivity, which – depending on the language – may or may not correspond to a separate form or tense.

Table 1. ***Classification of aspectual oppositions***

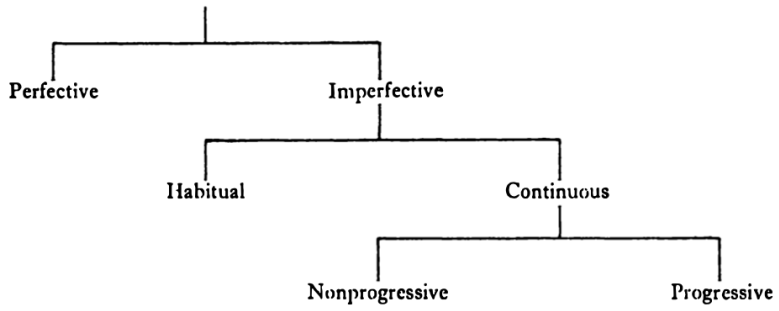


Figure 1 – Classification of aspectual oppositions according to Comrie (1976, p. 25).

Some languages, like English, have a separate habitual form (even if only in the past tense), as you can see in (2.a), and a separate progressive form, as in (2.b), while other languages, like Spanish, can express the progressive form both with a dedicated periphrastic form, as in (2.c), and an underspecified imperfective form (*Preterito Imperfecto*), as you can see in (2.d).

(2) a. *Mara* ***used to love*** *him.*

b. *Mara* ***was working*** *when she was killed.*

c. *Juan **estaba llegando.***
'Juan was arriving'

d. *Juan **llegaba**.*
'Juan was arriving'

After several examples, Comrie stresses the fact that – even if traditional grammars may distinguish between habitual, continuous, and all these semantic distinctions falling under the umbrella of imperfectivity – "these various subdivisions do in fact join together to form a single unified concept, as is suggested by the large number of languages that have a single category to express imperfectivity as a whole, irrespective of such subdivisions as habituality and continuousness" (1976, p. 26). This is for example the case of Croatian, whose imperfective lexical variants are able to express both a habitual and a progressive meaning.

The take-home message that we can draw from Comrie's work is that the concepts of perfectivity and imperfectivity that we find grammaticalized in the different verb tenses and constructions he quotes from non-Slavic languages such as English, Spanish or Italian, also apply to Slavic languages where such distinctions are lexicalised in different lexical variants (1976, p. 7) (see section § 2.3), as testified by the many examples he takes from Russian. In a way, his treatment of Slavic examples is in line with Dahl, who postulates that Slavic aspectual pairs may be regarded as "grammaticalized lexical categories" (1985, p. 89).

Of course, this stance is not shared by all. For instance, Bertinetto and Delfitto suggest that the perfective-imperfective pairs of Slavic languages do not belong to the realm of aspect proper but that of *actionality*, another term to refer to *Aktionsart* (2000, p. 189) (more on this in section § 2.2.2). They define aspect as "the specific perspective adopted by the speaker/writer [on an event, which] may be considered from a global or a partial point of view" (2000, p. 190) and recognise this the basis for the distinction between perfective and imperfective aspects. By actionality, on the other hand, they mean "the type of event, specified according to a limited number of relevant properties" (*ibidem*). For the sake of simplicity, they refer exclusively to Vendler's classes (1957), and identify the telicity of most perfective verbs making up aspectual pairs as the reason why the distinction between the two variants of said pairs should be considered a matter of *Aktionsart.* It is also worth mentioning that they suggest to employ different labels to refer to the different phenomena, namely "the pair *terminative/non-terminative* when referring to the aspectual domain proper, and *bounded/unbounded* when referring to the lexical oppositions available in the Slavic languages" (2000, p. 193), in order to avoid misunderstandings.

Smith's (1991) perspective on the matter, on the other hand, is in line with Comrie's, as she traces back the perfective-imperfective opposition in Slavic verb pairs to a matter of grammatical aspect and not *Aktionsart*, even though she uses a different terminology. In her PhD dissertation *The Parameter of Aspect*, she distinguishes two kinds of aspectual meaning a sentence may convey: *viewpoint aspect* and *situation aspect*.

Her starting point is that a situation is always presented from a certain perspective (or viewpoint), and it can be indirectly classified as an event of a certain type. In her framework, there are three main viewpoint types: perfective, imperfective, and neutral. *Perfective viewpoints* focus on events in their entirety, including both their initial and final endpoints. *Imperfective viewpoints* focus only part of an event and do not include their initial or final endpoints. *Neutral viewpoints* are flexible, include the initial endpoint and at least one internal stage.

For example, the following sentences present the same event, but differ in viewpoints, which are conveyed by the verb form, (3.a) having a perfective and (3.b) an imperfective viewpoint:

(3) a. *John and Mary **built** a rock garden last summer.*

b. *John and Mary **were building** a rock garden last summer.*

The *situation type* of a sentence, on the other hand, classifies the event talked about according to its temporal properties. Smith distinguishes five types of situation aspects (states, activities, accomplishments, semelfactives, and achievements), clearly building onto Vendler's (1957) classification of *Aktionsarten* (see section § 2.2.2).

Last but not least, Dowty defines aspect as the term usually used in linguistic terminology to "distinguish such things as whether the beginning, middle or end of an event is being referred to, whether the event is a single one or a repeated one, and whether the event is completed or possibly left incomplete" (1979: 52). He adds that aspect is usually "understood to refer to different inflectional affixes, tenses or other syntactic frames that verbs can acquire (*aspect markers*)" (*ibidem*) and that the only instances of pure aspect markers in English are the *progressive form* and the quasi-auxiliary *used to*, while Slavic languages are well known to possess the most extensive inventory of aspectual verb affixes.

Before turning to the main focus of his work – which is the classification of *Aktionsarten* – he adds a very interesting remark on the interaction between grammatical and lexical aspect, i.e., that in all languages, "the semantic differences inherent in the meanings of verbs themselves cause them to have differing interpretations when combined with these markers. […] It is because of this intricate interaction between classes of verbs and true aspect markers that the term aspect is justified in a wider sense to apply to the problem of understanding these classes of verbs as well" (*ibidem*).

He concludes suggesting that – if it really is necessary to distinguish the two uses of the term aspect – then we could distinguish the *aspectual class* of a verb (i.e., its Aristotelian class or *Aktionsart*) from its *aspectual form* (i.e., the specific aspect marker or markers it occurs with within a given context).

2.2.2 AKTIONSARTEN

The first traces of a discussion surrounding *aspectual verb classes* can be found in Aristotle, who in the *Metaphysics*, book 9, section 1048b distinguished between *kineseis* ("movements") and *energiai* ("actualities") to contrast actions that necessarily involve an end or result (such as that of *building*) with others which are complete in themselves (such as *being happy*).

The Oxford philosopher Gilbert Ryle (1949) later came back to Aristotle's classes and coined the term *achievement* for resultative verbs (such as *winning*, *unearthing* or *finding*) to be distinguished from irresultative activities (such as *listening* or *hunting*) that usually require a longer period of time to unfold. But it was Zeno Vendler, who put forth, in his paper *Verbs and Times* (1957), the traditional classification of verbs in four different lexical classes he calls *time schemata*, namely:

- Activity terms – *run, push a cart*
- Accomplishment terms – *run a mile, draw a circle*
- Achievement terms – *recognize, spot*
- State terms – *know, believe*

The three features Vendler bases his classification on are the following: (a) dynamicity (the presence of change), (b) duration and (c) telicity (the presence of a goal or *télos*[30]).

Verb classes	**Dynamicity**	**Duration**	**Telicity**	**Examples**
States	-	+	-	*know*
Activities	+	+	-	*run*
Accomplishments	+	+	+	*fix*
Achievements	+	-	+	*spot*

Table 1 – Vendler's verb classes

State terms, such as *know* in (4.a), are not considered dynamic since they contain no elements of change. They can be seen as homogeneous situations characterized by a given duration but lacking a specific goal. Activity terms, such as *run* in (4.b), on the other hand, express dynamic events lasting

[30] The term *telicity* was first introduced in linguistics by Garey (1957).

a certain amount of time, which still do not point at any *télos*. Accomplishment terms, such as *fix* in (4.c), express events which have a duration, are dynamic and are characterized by a gradual progression towards a precise goal which must be reached for the event to have occurred. Finally, achievement terms, such as *spot* in (4.d), express events without any duration culminating in an instant.

(4) a. *I **know** the truth.*
b. *Claire **runs** in the park.*
c. *I **fixed** the thermostat.*
d. *I **spotted** Mary in the crowd.*

Anthony Kenny (1963) also contributed to the development of this classification by adding more logical criteria to distinguish between the different classes and was the first one to observe that:

- if φ is an accomplishment, saying "A is (now) φing" implies that "A has not (yet) φed"
- activities and achievements can occur in progressive tenses, while states cannot
- the simple present of activities and achievements has a habitual meaning that states lack.

It is worth mentioning that the class of achievements contains several verbs (such as *cough*, *flash,* or *blink*) that are able to express not only single achievements, as in (5.a), but also iterated instantaneous events, as in (5.b). Verbs such as these are called semelfactives (from Latin *semel* 'once') or *points* (Moens & Steedman 1988; Smith 1991) and provide a prime example of *aspectual coercion*. As defined by Moens & Steedman (1988), aspectual coercions consist in the "change in the aspectual type of a proposition under the influence of modifiers like tenses, temporal adverbs and auxiliaries" (1988, p. 17). Aspectual coercions are also called *actional shifts* and the objects and other clausal elements contributing to the shifts are called *shifters* (Filip, 1999). In (5.b) for instance, it would be the combination of a punctual verb like *cough* with a durative modifier such as *for five minutes* to induce an aspectual mismatch and trigger an iterative interpretation. On the other hand, in a sentence like (5.c), it is the use of the progressive that coerces once again a semelfactive into an iterative process.

(5) a. *Blake **coughed** once to say yes.*
b. *Charlotte **coughed** for five minutes.*
c. *Sally **is coughing**.*

Another verb class worth noting is that of *degree achievements* (Dowty, 1979) – called *gradual completion verbs* in Bertinetto & Squartini (1995) and *multi-point scalar verbs* in Rappaport Hovav

(2008). It is the class verbs such as *shrink* or *cool down* belong to, namely verbs describing events in which participants can display different degrees of the property expressed by the verb. Consequently, these verbs are often characterised by two ambiguous readings, namely [BECOME *x*] and [BECOME MORE *x*]. For instance, a sentence such as 'The soup cooled down' without any further details leaves doubt on whether said soup is now cool enough to be eaten without any risk of burning one's tongue or whether maybe it is now too cold to be eaten at all. The two readings lexicalise two different states, an intermediate one and a possible end state.

Another verb class associated with incrementality is that of the so-called *incremental theme verbs* (such as *fill*, *build*, and *write*), which express events resulting into the gradual change of the entities playing the role of Theme. For instance, while it is 'being built', a house undergoes several stages of change. However, unlike in degree achievements, no ambiguity lies between intermediate and end state reading. If a 'house is built', the event of building it has necessarily come to an end.

2.2.2.1 VENDLER CLASS TESTING

In the following, we will list a few of the tests from Dowty (1979) to distinguish among aspectual verb classes in English.

- Only non-statives occur in the progressive, as in (6.b) and (6.c).
- Only non-statives can occur as imperatives, as in (7.b) and (7.c).
- Activities and achievements in the present tense have a habitual meaning – see (8.b) and (8.c) – that states lack, as in (8.a).
- Accomplishment verbs usually take *in*-adverbials, as in (9.a), while activities tend to allow *for*-adverbials, as in (9.c).
- Achievement verbs are usually infelicitous with *for*-adverbials, as in (10.b).

(6) a. **Sarah* ***is knowing***.
b. *Sarah* ***is running***.
c. *Sarah* ***is writing*** *a novel.*

(7) a. **Sarah,* ***know*** *the answer!*
b. *Sarah,* ***run****!*
c. *Sarah,* ***write*** *a novel!*

(8) a. *(?)Sarah* ***knows*** *the answer.*

b. *Sarah **runs**.*

c. *Sarah **writes** a novel.*

(9) a. *Sarah wrote a letter **in an hour**.*

b. *(?) Sarah wrote a letter **for an hour**.*[31]

c. *Sarah ran **for an hour**.*

d. **Sarah ran **in an hour**.*

(10) a. *Sarah spotted Mark **in a few seconds**.*

b. *(?) Sarah spotted Mark **for a few seconds**.*

2.2.2.2 PROBLEMS WITH VENDLER'S CLASSIFICATION

Vendler conceived his *time schemata* as lexical properties of verbs, since the way verbs "presuppose and involve the notion of time" was to him more crucial than other factors, such as "the presence or absence of an object" (Vendler, 1957, p. 143), in his classification of English verbs.

However, his position was criticised by many linguists, starting with Dowty (1979), who clearly stated that it is not just verbs but whole verb phrases that must be considered to distinguish aspectual classes. For instance, he realised that "an activity verb describing movement behaves like an accomplishment verb if it occurs with either a locative or destination or with an adverb of extent" (Dowty, 1979, p. 60), as in sentences (11.a) and (11.b), respectively. Moreover, he noticed that accomplishment verbs taking direct objects "unexpectedly behave like activities if an indefinite plural direct object or a mass-noun direct object is substituted for the definite one" (1979, p. 62) as in the case of (11.d).

(11) a. *John walked **to the park**.*

b. *John walked **a mile**.*

c. *John ate **the bag of popcorn**.*

d. *John ate **popcorn**.*

Another critic of Vendler's classification is the Dutch linguist Henk Verkuyl. Already in his work *On the Compositional Nature of Aspects* (1972), Verkuyl clearly advocates against the lexical nature of

[31] "Sarah wrote a letter for an hour" is not agrammatical as "*Sarah ran in an hour". However, it is not very communicative to underlines how much time one was busy doing something without hinting at whether the accomplishment is accomplished or not.

the opposition between durative and non-durative aspect, stating that the categorization must take place at a higher node than the VP since outside information must be taken into account. Afterwards, in *A Theory of Aspectuality* (1993), he specifically directs his criticism towards Vendler's classes, not only denying their usefulness since they ignore the basic principles of compositionality involved in building aspectual information, but also questioning whether the distinction between aspect and *Aktionsart* should be discarded altogether as "an ontological wolf [disguised] in linguistic sheep's clothing" (Verkuyl, 1993, p. 11). Consider, for example, the following sentences:

(12) a. *Judith ate **a sandwich**.*
b. *Judith ate **sandwiches**.*

Unlike Moens and Steedman (1988), who would explain sentence (12.b) as a case of *aspectual coercion* from accomplishment to activity due to the indefinite plural direct object, Verkuyl (1993) concentrates on the fact that no verb exists in isolation, thus challenging the very need to identify a kernel meaning to assign to a certain lexical aspectual class. In his opinion, in both sentences the given elements simply compositionally codetermine the aspectual value of the whole.

One way or the other, what can indeed be pointed out, as remarked by Ježek (2016, p. 125), is that "valency structure and *Aktionsart* are inextricably linked; arguments are those participants in the eventuality that are selected as syntactically relevant by the verb, and *Aktionsarten* define how the event denoted by the verb unfolds in time." Indeed, to determine the aspectual verb class of an event, the valency structure chosen to best fit the expressive needs of the event itself has a role to play.

In conclusion, lexical aspect is not actually merely lexical, but compositional. Verbs taken out of context may very well appear as belonging to a given aspectual class, but the moment they are put into an actual context of use, their *Aktionsart* potential is realised through the interaction with various contextual elements:

- on a clausal level, *adverbials*: iterative and frequency adverbials (*twice*, *sometimes*), point adverbials (*at that moment*), time-span (*in a month*) and durative (*for a week*) adverbials, directional adverbials (*to the park*) (Filip, 1999, pp. 63–68);
- on a phrasal level, the *direct objects* selected by the verb and their referential and quantification properties, as in (4), where the singular count noun a *sandwich* in (12.a) yields a telic interpretation which is lacking in (12.b);
- on a morphological level, the aspectual information supplied by the chosen verbal tenses or periphrases (Dowty's *aspect markers*). For Croatian and other Slavic languages, this translates in the choice of an aspectual lexical variant (perfective vs. imperfective).

2.3 THE CROATIAN PERSPECTIVE

After having taken into account the main international works devoted to aspect and *Aktionsart* and having found our bearings mainly thanks to English examples, we can now turn to Croatian literature and try to see how these phenomena are treated there.

Since most of the authors mentioned in this section have published mainly in Croatian, an effort has been made to translate all the definitions that have been deemed of interest (keeping the original Croatian excerpts in the footnotes) with the hope they may become more accessible to the Western reader interested in knowing more about this language.

2.3.1 CROATIAN ASPECTUAL VERB PAIRS

Quoting Gojmerac (1980, p. 21), "for both Koschmieder and Agrell, verbal aspect is a perspective on the event, which becomes grammaticalized in Slavic languages and represents an actual grammatical category of the verb. Two are the possible ways of looking, each of them corresponding to one of the members of an aspectual opposition[32]."

With these words, Gojmerac – who is in line with both Comrie (1976), Dowty (1979), Dahl (1985), and Smith (1991) – helps bringing together Comrie's definition of aspect and the reality of Slavic languages, including Croatian, where perfective and imperfective actions are lexicalised in two different forms, also called *aspectual pairs* (Babić, 2002, pp. 516–531; Barić et al., 2005; Browne, 1993, pp. 331–332; Kordić, 1997, p. 41; Matovac, 2022, pp. 225–231; Silić & Pranjković, 2007, p. 288).

For instance, let us look at the following examples taken from Silić & Pranjković (2007, p. 288) featuring the aspectual pair *prodavati* / *prodati* ('to sell', imperfective / perfective):

(13) a. *Susjed **je prodavao** kuću*.
'The neighbour was selling the house.'

b. *Susjed **je prodao** kuću.*
'The neighbour sold the house.'

[32] "Za Koschmieder je, kao i za Agrella, glagolski vid pogled na radnju, koji se u slavenskim jezicima gramatikalizirao i predstavlja objektivnu gramatičku kategoriju glagola. Moguća su dva načina gledanja, a svakom odgovara jedan član vidske opozicije".

> "In the first case, the focus is on an action in which only the process is emphasised, while in the second [the focus] is on a completed action, on the result of the process, on its being executed, on an action which is not characterised by processuality"[33] (Silić & Pranjković, 2007, p. 288).

In other words, Croatian aspectual pairs allow speakers to express "the different ways of viewing the internal temporal constituency of a situation" (Comrie, 1976, p. 3) by choosing between different (but morphologically related) lexical variants.

2.3.1.1 THE MORPHOLOGY OF ASPECTUAL PAIRS

At this point, one might wonder about the morphological processes behind the formation of Croatian aspectual pairs. According to Croatian literature, the members of an aspectual pair tend to be related to each other via prefixation, suffixation, suppletivity, or root allomorphy (Isačenko 1968; Silić 1978).

As pointed out by Barić et al. (2005, p. 227), prefixation is the morphological process most often involved in the perfectivisation (*perfektivizacija*) of an imperfective verb to create its perfective counterpart, as in the case of the following aspectual pairs taken from Appendix 1:

- *čitati/**pro**čitati* 'to read'
- *jesti/**po**jesti* 'to eat'
- *liječiti/**iz**liječiti* 'to heal'

Prefixes can also be added to perfective verbs to create other perfectives (*ibidem*). In these instances, prefixes tend to modify the semantics of the base verbs creating new lemmas, which are not aspectual variants, as in the case of *reći/**pro**reći* 'to say/to predict'.

Suffixation, on the other hand, is mostly involved in the creation of imperfective variants from perfective verbs (Barić et al., 2005, p. 228), but can also be responsible for the derivation of perfective verbs from imperfective ones, as in the case of the aspectual pair *dirati/dir**nu**ti* 'to touch', where the addition of the suffix ***-nu-*** to the imperfective verb *dirati* results in its perfective counterpart *dirnuti*.

Going back to imperfectivisation, among the most common suffixes used to create imperfective variants from perfective verbs, we can find the suffixes **-iva-** and **-ava-** (Čilaš Mikulić, 2012, p. 68), as in the case of the following pairs from Appendix 1:

- *dovršiti/dovrš**ava**ti* 'to complete'

[33] "U prvome slučaju riječ je o radnji u kojoj je naglašen sam proces, a u drugome o cjelovitoj radnji, o rezultatu procesa, o izvršenju, o radnji kojoj nije svojstvena procesualnost".

- *isključiti*/*isključ**iva**ti* ‘to exclude’
- *posjetiti/posje**ćiva**ti* ‘to visit’
- *spasiti/spa**šava**ti* ‘to save’

Sometimes, the addition of the suffix does not create any phonemic alternation in the root, sometimes it does, as in the case of *posjetiti/posje**ć**ivati* and *spasiti/spa**š**avati.*

Last but not least, let us look at suppletivity and root allomorphy. If suppletivity involves two forms that cannot be derived from one another by means of a phonological process – as in the case of the pair *dolaziti/doći* ‘to arrive’ (Barić et al., 2005, p. 229) –, in root allomorphy there is some common phonology between the two forms, as in the case of the following pairs from Appendix 1:

- *izlagati/izložiti* ‘to exhibit’
- *nastavljati/nastaviti* ‘to continue’

Finally, certain verbs can have both an imperfective and a perfective reading based on the context of use (Barić et al., 2005, p. 229). These verbs are called *dvovidni glagoli* ‘biaspectual verbs’ and many of them are loanwords. Among the ones contained in Appendix 1, we recall:

- *informirati* ‘to inform’
- *kontaktirati* ‘to contact’
- *čuti* ‘to hear’

Biaspectual verbs should not be confused with defective verbs, i.e., verbs that can only be used to express one aspect and do not have a related aspectual variant to express the other. This is the case, of the imperfective verbs *djelovati* ‘to act’ and *posjedovati* ‘to possess’ from Appendix 1.

2.3.2 *VID* AND *AKTIONSART* ACCORDING TO CROATIAN GRAMMARS

In this section, we will compare the description of the categories of *vid* and *Aktionsart* that can be found in two of the main Croatian reference grammars: the *Hrvatska Gramatika* (‘the Croatian Grammar’) by Barić et al. (2005), and the *Gramatika hrvatskoga jezika za gimnazije i visoke učilište* (‘the Grammar of the Croatian language for high schools and colleges’) by Silić and Pranjković (2007). These two works are amongst the most referenced sources when it comes to Croatian grammar, and will be compared with the new volume by Darko Matovac (2022), *Basic Croatian Grammar: For Croatian Language Learners*.

As we will see, the two Croatian grammars aimed at native speakers show a tendency towards introducing aspectual concepts in their morphological section, which does not help non-native readers, as pointed out by Čilaš Mikulić (2012, p. 24). Moreover, clear definitions of aspect and

Aktionsart tend to be missing – the latter replaced by lists of semantic groupings of verbs displaying similar actional traits called *glagolskovidska značenja* 'verbal aspectual meanings' or *načine vršenja glagolske radnje* 'manners of carrying out the verbal action'. Only when reaching the chapters on syntax, more clear definitions of aspect are provided. On the other hand, Matovac's (2022) work shows an entirely different attitude and structure, thus presenting itself as a much more appropriate option for non-native speakers, especially if not yet proficient.

2.3.2.1 BARIĆ ET AL. (1995)

In their paramount work *Hrvatska Gramatika* – also known colloquially as the *Siva Gramatika* 'the Grey Grammar' – the first time the authors mention the concept of *vid*, they do so to distinguish it from the several semantic groups of *Aktionsarten* identified in Croatian that they call *načine vršenja glagolske radnje* 'manners of carrying out the verbal action'. The grouping tend to have a similar morphological structure and to share the same prefixes (2005, p. 224), which is why their discussion takes place in the section devoted to morphology.

They identify the following groupings: *neprekidni/durativni glagoli* 'durative verbs', *učestali glagoli* 'iterative verbs', *trenutni glagoli* 'momentane verbs', *učinski/uzročni glagoli* 'causative verbs', *pantivni glagoli* 'gradual completion verbs', *totivni glagoli* 'total verbs', *početni glagoli* 'inchoative verbs', *završni glagoli* 'terminative verbs', *poredbeni verbs* 'comparative verbs', *intenzivni glagoli* 'intensive verbs', *sativni glagoli* (verbs expressing enough of an action), *glagoli subjektnoga razreda* 'subject class verbs' and *glagoli objektnoga razreda* 'object class verbs'.

After that, aspect is defined rather vaguely as "a category which is characteristic of Croatian as well as other Slavic languages"[34] (2005, p. 225), which divides verbs into two groups, namely *perfective* and *imperfective* ones:

> "Some verbs express actions that are being carried out, that are still in progress. Others, on the other hand, the completion of said actions, the carried-out process, while encompassing at the same time the unfolding of the action before its coming to an end. Verbs that express an action while in progress are called *imperfective verbs*. […] Verbs that express the completion of an action are called *perfective* verbs[35]." (1995, p. 225)

[34] „*Vid* ili *aspekt* glagolska je kategorija karakteristična za hrvatski jezik, kao i za druge slavenske jezike."

[35] „Neki glagoli izriču radnju koja se izvršava, u procesu je. Drugi pak izvršenost ostvarenost radnje, obavljeni proces, a istodobno i trajanje radnje prije izvršenja. Glagoli koji izriču radnju u vršenju se zovu *nesvršeni (imperfektivni)* ili *glagoli nesvršenoga (imperfektivnoga) vida*. […] Glagoli koji izriču izvršenost radnje zovu se *svršeni (perfektivni)* ili *glagoli svršenoga (perfektivnoga) vida*."

As already pointed out by Čilaš Mikulić (2012, p. 22), the definition can be mostly agreed with, but for the part on perfective verbs expressing the "unfolding of the action before its coming to an end", which is definitely not the case, since perfective verbs do not allow the speaker to think of the action expressed by the verb as in progress.

More thorough definitions of perfective and imperfective aspects are presented in the chapter devoted to syntax (Barić et al., 2005, p. 406), which are more in line with the actual phenomena:

> "The perfective aspect expresses an action, event, or state, regardless of the time in which it takes place. […] An action, event, or state expressed using the perfective form is therefore necessarily seen as a whole, regardless of its duration, and therefore it is not seen as distributed over a certain period. […] The imperfective aspect expresses an action, event or state considering the time in which it takes place. [...] The action, event or state expressed in the imperfective form is therefore not necessarily viewed as a whole, and its duration and distribution over time are important."[36] (2005, p. 406)

2.3.2.2 Silić & Pranjković (2007)

In their work *Gramatika hrvatskoga jezika za gimnazije i visoke učilište*, Silić and Pranjković define *glagolski vid* 'verbal aspect' as "the means to express the perfectivity or imperfectivity of a verbal action, state or event"[37] (2007, p. 48), but fail to shed more light on the concepts of perfectivity and imperfectivity.

On the other hand, since this definition is provided in the section devoted to morphology, their focus immediately shifts towards how to build perfective and imperfective verbs. After mentioning that "perfective verbs can be formed from imperfective ones either by changing the suffix or by adding a prefix to the imperfective base verb"[38] (*ibidem*) and providing a detailed overview of all the different suffixes involved in the inverse process of turning perfective verbs into imperfective ones, the authors change topic – leaving out a description of the process of perfectivisation – and turn to a

---

[36] "Svršenim se vidom izriče neka radnja, zbivanje ili stanje bez obzira na vrijeme u kojemu se odvija ili smješta. […] Radnja, zbivanje ili stanje izrečeni svršenim vidom promatraju stoga nužno kao cjelina, bez obzira na svoje trajanje, pa se stoga ne promatraju raspoređeni u vremenskom toku. […] Nesvršenim se vidom izriče neka radnja, zbivanje ili stanje s obzirom na vrijeme u kojemu se odvija ili u kojem se smjesta. […] Radnja, zbivanje ili stanje izrečeni nesvršenim vidom ne promatraju se stoga nužno kao cjelina, nego je važno njihovo trajanje i raspoređenost u vremenu."

[37] "Glagolski je vid sredstvo izražavanja svršenosti i nesvršenosti glagolske radnje, stanja ili zbivanja. Po njemu se glagoli dijele na svršene (perfektivne) i nesvršene (imperfektivne)."

[38] "Svršeni se glagoli tvore od nesvršenih bilo zamjenom sufiksalnoga morfema osnove nesvršenih glagola bilo pridruživanjem prefiksalnoga morfema osnovi nesvršenih glagola."

description of the so-called *vidskoznačenjske faze* 'phases of aspectual meaning' and *glagolskovidska značenja* 'verbal aspectual meanings' (2007, p. 56).

As Čilaš Mikulić clarifies, by *vidskoznačenjske faze* 'phases of aspectual meaning' we "refer to the ability Croatian verbs have to [repeatedly] change their aspect by prefixation or suffixation in three consecutive phases"[39] (2012, p. 26). In the first phase, we find imperfective verbs, in the second perfective ones (derived from imperfective verbs via prefixation), and in the third phase, once again imperfective verbs (derived from perfective verbs via suffixation) (Silić & Pranjković, 2007, p. 56). An example of such a derivational chain would be the triad *pisati* > *potpisati* > *potpisivati* ('to write' imperf. > 'to sign' perf. > 'to sign' imperf.). It is also possible to encounter such chains starting with a perfective verb in the firs phase already, as in the case of *uzeti* > *oduzeti* > *oduzimati* ('to take' perf. > 'to take away' perf. > 'to take away' imperf.) (*ibidem*).

At this point, the authors focus on describing what they call *glagolskovidska značenja* 'verbal aspectual meanings', which is a less conventional label used in Croatian literature to refer to *Aktionsart* (Čilaš Mikulić, 2012, p. 26). Verbs are then grouped into different classes according to semantic features they share, namely:

- *količina radnje* 'the amount of action', which is what supports the distinction between *sativno, intenzivno, augmentativno, majorativno,* and *deminutivno značenje.*
- their focus on a single phase of the event or on the event in its entirety (*istaknutost određenog dijela, faze, ili cjelovitosti radnje*), which allows for the distinction between *inkoativno, totivno*, and *finitivno značenje.*
- *raspodijeljenost radnje* 'the divisibility of the action', which is typical of the *distributivno značenje.*
- *stilističnost* 'the style', typical of those verbs having *pejorativno značenje.*

These groupings are connected to the use of certain prefixes and suffixes, which change the meaning of the base verbs, however no actual definition of *glagolskovidska značenja* is provided.

As for aspect, the definition of what *perfective* and *imperfective* mean is provided only by the end of the text in the section devoted to syntax, specifically to the grammar categories that characterise verbs. The category is called *kategorija vida ili aspekta* and the author define it as the category:

> "[establishing] the difference between imperfective and perfective actions – e.g., *prodavati* (*to sell*, imperfective) vs. *prodati* (*to sell*, perfective) – *cf. Susjed* ***je prodavao*** *kuću* (The neighbour was selling

[39] "Pod vidskoznačenjskim fazama podrazumijeva se tvorbena sposobnost glagola hrvatskoga jezika da se kroz tri stupnja prefiksacijom, odnosno sufiksacijom mijenja vid glagola."

the house) and *Susjed **je prodao** kuću* (The neighbour sold the house). In the first case, the focus is on an action in which only the process is emphasised, while in the second [the focus] is on a completed action, on the result of the process, on its execution, on an action which is not characterised by processuality."[40] (Silić & Pranjković, 2007, p. 288)

In conclusion, we can remark on the peculiarity of not having found a definition of aspect in the morphological section of the work, but only in the one devoted to syntax, as in Barić et al. (1995).

2.3.2.3 MATOVAC (2022)

In his *Basic Croatian Grammar: For Croatian Language Learners*, Matovac defines aspect as a "grammatical category that expresses how an action expressed by a verb extends over time" (2022, p. 225). He also underlines the fact that in Croatian, aspect is an inherent category of the verb, which means that it is part of the verb's meaning.

That being said, he adds that a verb may either be *perfective* or *imperfective* and provides definitions of both: "Perfective verbs denote an action which is completed and has a result, [while] imperfective verbs denote an action which is in duration and lasts for a longer period" (*ibidem*).

In order to clarify this distinction in the most efficient way, Matovac (2022, p. 226) provides a number of Croatian sentences together with their English translations using a variety of verb tenses:

(14) a. *Jučer **sam čitao** knjigu na plaži.*
'Yesterday I was reading a book on the beach.'

b. *Danas **čitam** knjigu na plaži.*
'Today I am reading a book on the beach.'

c. *Sutra **ću čitati** knjigu na plaži.*
'Tomorrow I will be reading a book on the beach.'

Featuring the progressive form, all three English translations stress the fact that the process of reading on the beach lasted, currently lasts, and will last for a longer period of time, hence the choice of the imperfective variant *čitati* and not of the perfective one *pročitati*, exemplified in the following sentences:

[40] "Tom se kategorijom naime uspostavlja razlika između nesvršenih (imperfektivnih) i svršenih (perfektivnih) radnji, npr. *prodavati* prema *prodati*, usp. *Susjed **je prodavao** kuću* prema *Susjed **je prodao** kuću*. U prvome slučaju riječ je o radnji u kojoj je naglašen sam proces, a u drugome o cjelovitoj radnji, o rezultatu procesa, o izvršenju, o radnji kojoj nije svojstvena procesualnost".

(15) a. *Jučer **sam pročitao** novine u kafiću.*
‘Yesterday I read the newspaper (through) at the bar.’
b. *Uvijek **pročitam** novine u kafiću.*
‘I always read the newspaper (through) at the bar.’
c. *Sutra **ću pročitati** novine u kafiću.*
‘Tomorrow I will read the newspaper (through) at the bar.’

In (15.a – c), it’s the addition of the particle “through” in each English translation which stresses the boundedness of the featured event and makes us understand that the original Croatian sentences did need a perfective verb to express that meaning.

After this overview, the author devotes a short paragraph to perfective verbs and the present tense, stressing the fact that – since perfective verbs have their focus on the completion of the action – “they cannot be used to describe that something is happening at the moment of speaking” (2022, p. 227).

(16) a. ***Pijem** kavu.*
‘I am drinking coffee.’
b. ***Kupujem** suvenir.*
‘I am buying a souvenir.’

On the other hand, they can be used in the present if they are in a complex sentence or in a sentence describing repeating or regular actions (2022, p. 228), as in the case of:

(17) a. *Uvijek **kupim** čokoladu bratu za rođendan.*
‘I always buy chocolate to my brother for his birthday’.
b. *Kad sam umoran, **popijem** kavu.*
‘When I am tired, I drink a coffee.

Last but not least, Matovac devotes a section of his chapter on aspect to *signal words*, i.e., “those adverbs and other words which can help to decide whether a perfective or imperfective verb should be used in a sentence” (*ibidem*). Imperfective verbs tend to be used in sentences containing words like: *sad* (‘now’), *dugo* (‘for a long time’), *stalno* (‘constantly’), *cijeli dan* (‘all day’), *cijelu godinu* (‘all year round’), *satima* (‘for hours’), *godinama* (‘for years’), *pet minuta* (‘for five minutes’), etc.

Perfective verbs, on the other hand, are used in sentences that contain words like *čim* (‘as soon as’), *odmah* (‘immediately’), *za pet minuta* (‘in five minutes’), etc. (2022, p. 229).

To sum up, in this Chapter we have first provided a short history of the development of the terminology used to discuss the topics of aspect and *Aktionsart* (§ 2.1). We have then turned to the current Western definitions of these phenomena (§ 2.2) and to their treatment in the Croatian tradition (§ 2.3). This theoretical introduction was necessary to be able to proceed to the modelling of Croatian aspectual verb pairs (§ 2.3.1) with the formalism of event structures (Chapter 5), and to understand the way they were explained in the Croatian intensive course discussed in Chapter 6. Now, we will however turn to the notions of verb argument structure and verb polysemy, which are paramount for semantic resources such as CroaTPAS (Chapter 4).

CHAPTER 3: ARGUMENT STRUCTURES AND VERB POLYSEMY

3.1 VALENCY THEORY AND ARGUMENT STRUCTURES

In chemistry, *atomic valency* provides a measure of the combining powers of an atom when it needs to form a stable molecule or compound, while – when it comes to the study of language – the term *verb valency* tells us how many obligatory complements a verb needs to fully express its meaning.

The first to introduce the concept of verb valency in linguistics was Lucien Tesnière in his 1959 *Éléments de syntaxe structurale*, where he identifies the *nœud verbal* 'verbal node' as the head of the sentence, around which a *petit drame* 'small play' is taking place. Like all plays, also this one requires *un procès* 'an event' to revolve around, *des acteurs* 'some actors' to play in it and *des circonstances* 'a set of [additional] circumstances' (Tesnière, 1959, p. 102).

If we transport this theatrical metaphor onto structural syntax, the linguistic counterparts of the elements he identifies would be (a) the verb itself, (b) one or more *actants*, and (c) some potential *circumstantials*. While Tesnière calls *actants* the obligatory complements needed to express a verb's meaning, *circumstantials* correspond to the optional ones, such as adverbs or adverbial expressions (Tesnière, 1959, p. 103). For instance, it is quite obvious from example (1.a) that the verb *to give* requires at least three actants, i.e., a giver (*Catherine*), something that is given (*a book*) and a recipient (*Fiona*), while the circumstantial *right now* is not necessary for the sentence to be well-formed. On the other hand, without even one of the actants, the sentence would become agrammatical, as in the case of example (1.b).

(1) a. *Catherine is giving the book to Fiona [right now].*
b. **Catherine is giving the book.*

As you can see from Figure 1, which is taken from Ježek (2016, p. 112), verbs can be divided into different classes according to the number of actants they require, i.e., their valency structure. One-slot, two-slot and three-slot verbs are the most common ones, but also four-slot verbs can be found, e.g., *translate*, which requires a *translator*$_1$ translating *a text*$_2$ from *one language*$_3$ into *another*$_4$.

Verb Classes	Verbs	Examples
One-slot Verbs	*be born*, *caught*	*A baby girl*$_1$ was born; *Paul*$_1$ coughed loudly
Two-slot Verbs	*rent*, *live in*	*Sara*$_1$ rented *a car*$_2$; *Mark*$_1$ lives *in Rome*$_2$
Three-slot Verbs	*put*, *dedicate*	*Lynda*$_1$ put *the keys*$_2$ *in the bag*$_3$; *Lisa*$_1$ dedicated *her book*$_2$ *to her father*$_3$

Figure 1 – Verb classes based on the number of actants. Figure taken from Ježek (2016, p. 112).

Another term used to refer to valency structure is *argument structure* (Grimshaw, 1990), which can be defined as "the number and type of event participants which are required for a successful use of the verb in question" (Ježek, 2016, p. 107).

Indeed, as also Simone points out, "a verb requires a number of virtual slots that can only be filled by noun phrases with specific meanings, called *arguments*; the set of arguments [required by] a verb make up its *argument structure*"[41] (Simone, 2013, p. 128).

One of the most comprehensive works devoted to the theories of how a verb's semantics can determine the morphosyntactic realization of its arguments is the survey by Levin & Rappaport Hovav (2005), which effectively provides a bridge between lexical-semantic and syntactic research. A recent contribution to the field is the work by Mereu (2020), which borrows the theoretical tools of cognitive linguistics and Construction Grammar to shed more light on a selection of Italian verb argument structures.

As it happens, all these works stress both the syntactic and semantic dimension of verb argument structures, meaning that to correctly identify an argument structure, one has to consider:

- how verb arguments are syntactically expressed (e.g., as direct objects, indirect objects, clausals, etc.), which in a Slavic language such as Croatian translates into the use of a case system (see § 1.3).
- what type of entity is expected to fill each argument position (e.g., an animate entity, an inanimate entity, etc.), i.e., the verb's *semantic selection*.

Both these notions are paramount to understanding resources such as CroaTPAS, whose focus lies precisely on portraying both the syntactic subcategorisation of Croatian arguments (see § 4.3.2.4) and their semantics (see § 4.3.2.1 on Semantic Type labels) to best frame verb polysemy.

[41] "Il verbo determina alcuni posti virtuali che possono essere riempiti solo da sintagmi dotati di significati specifici, denominati *argomenti*; l'insieme degli argomenti di un verbo costituisce la cosiddetta struttura argomentale."

3.1.1 DISAMBIGUATING TERMINOLOGY: ACTANTS, ARGUMENTS OR COMPLEMENTS?

As already pointed out, Tesnière (1959) calls obligatory sentence components *actants*, but we have also been using the labels *arguments* and *obligatory complements*. As always, the difference lies in the theoretical framework of reference.

When we talk of *complements*, we are using the terminology of traditional grammars, which differentiate between subject, predicate, and complements, albeit without further specifying between obligatory and optional ones. According to Graffi (2019, p. 271), the main consequence of valency theory on traditional syntactic roles is the subject's loss of status. Indeed, with the predicate taking on the pivotal role in the clause, the subject becomes just another *actant*.

On the other hand, when talking of *arguments*, we are using a terminology which stems from logic (Frege, 1892) and which was eventually introduced to linguistics in the 1980s by generative linguists working on Government and Binding Theory. According to this perspective, the definition of *argument* basically overlaps with that of *actant* as the obligatory sentence component. However, the subject maintains its role as the most prominent argument in the clause, taking on the name of *external argument*, in order not to be mistaken with direct and indirect objects, which are referred to as *internal arguments*. Within this framework, circumstantials go by the name of *adjuncts*. Figure 2 provides a comprehensive overview of all these terminologies.

Traditional grammar	subject	predicate	complements	
Valency grammar	actant	predicate	actant(s)	circumstantial(s)
Generative grammar	external argument	predicate	internal argument	adjunct(s)

Figure 2 – Terminological differences in clause roles. Figure taken from Ježek (2016, p. 113).

3.1.2 IDENTIFYING VERB ARGUMENT STRUCTURES

Before even starting to discuss how to identify verb argument structures, a preliminary difference needs to be drawn between the *participants* to an event and their actual realisations as *arguments* selected by a verb (Ježek, 2016, p. 115).

(2) a. *Mark is eating spaghetti*.
b. *Mark is phoning Sarah*.
c. *Mark is phoning Sarah with his mobile.*

For instance, if we compare the previous examples, we can notice that all the participants in the event taking place in sentence (2.a) are syntactically expressed as arguments (i.e., *Mark* and the *spaghetti*),

while, in sentence (2.b), the instrument *Mark* uses to phone *Sarah* is not overtly expressed, but rather incorporated in the verb meaning. This kind of unexpressed arguments are called *shadow arguments* (Pustejovsky, 1995) and according to Ježek they can be overtly expressed only if further information is provided about them (2018, p. 59), as in sentence (2.c).

This being said, already the mere existence of shadow arguments points to the fact that we should distinguish between two different *valencies*: a *semantic valency* on the conceptual level, which refers to the obligatory participants in an event, and a *syntactic valency* on the linguistic level, which refers to the syntactic realisation of those participants (Ježek, 2016, p. 114). When talking of argument structures in this dissertation, we will always refer to the latter.

This being said, when trying to identify a verb argument structure, how does one distinguish between arguments and adjuncts? The easiest way to tell them apart is to progressively subtract candidate arguments from the sentence and evaluate whether the result becomes agrammatical. If it does, then we are dealing with a proper argument, as in the case of example (3.b), which is not an acceptable sentence in Italian, since the verb *abitare* 'to live' requires a location, as in (3.a).

(3) a. *Marco abita a Roma.*
'Marco lives in Rome.'
b. **Mark abita.*
'Marco lives.'

Conversely, in sentence (4.a), we can see that the same prepositional phrase (*a Roma*) that was acting as an argument in (3.a) can also take an adjunct role, since it can be easily omitted without the sentence becoming agrammatical, as demonstrated by example (4.b).

(4) a. *Marco ha incontrato Maria a Roma.*
'Marco met Maria in Rome.'
b. *Marco ha incontrato Maria.*
'Marco met Maria.'

However, becoming agrammatical or not is not all that can happen to a sentence when an argument is subtracted from it, as you can see from examples (5.a) and (5.b).

(5) a. *Marco beve dell'acqua.*
'Marco is drinking water.'
b. *Marco beve.*
'Marco is drinking [some kind of liquid].' / 'Marco habitually drinks alcohol.'

Indeed, the omission of the beverage that is being drunk gives room to two different interpretations of sentence (5.b). On the one hand, *Marco beve* may still be translated as 'Marco is drinking [some kind of liquid]', where [some kind of liquid] could be treated as a *default argument*, namely an argument that can be linguistically omitted while still being necessary to correctly interpret the sentence meaning (Ježek, 2018, p. 59). This would be the same case as example (6) taken from Ježek (*ibidem*), where it is obvious that [some kind of food] must have been eaten by Luca, even if has not been explicitly mentioned.

(6) *Luca ha mangiato alle 7.*
'Luca ate [some kind of food] at 7.'

On the other hand, the other interpretation of (5.b) requires the Italian verb form *beve* to be interpreted as a habitual use of the present tense, which would then leave room for the sentence to be translated as 'Marco habitually drinks alcohol' or 'Marco is not a teetotaller'.

If we accept this second interpretation, we are entering the realm of verb *polysemy*, i.e., the ability of the same verb (in this case *bere*, 'to drink') to convey multiple meanings, which are usually associated with multiple verb argument structures. To put it in Ježek's words, "change in argument structure often goes together with change in meaning" (2016, p. 114), which is the basic idea on which all semantically Typed Predicate Argument Structure resources rest on, including CroaTPAS.

3.2 VERB POLYSEMY

According to the *Merriam-Webster Dictionary*[42], the word *polysemy* comes from the Late Latin adjective *polysemus*, which in turns comes the Greek adjective *polysēmos*, the union of *poly-* 'many' and *sēma* 'sign'. In line with this etymology, the Oxford English Dictionary[43] defines *polysemy* as "the fact of having several meanings; the possession of multiple meanings, senses or connotations."

Before delving into the topic, a couple of definitions are necessary, namely those of *lexicon* and *lexeme*: the first can be defined as "the set of words of a language" (Ježek, 2016, p. 1), while the latter "identifies the unit of the lexicon that constitute the base form [of a word] to which all the inflected forms [of that word] are ascribed" (Ježek, 2016, p. 22).

After presenting an overview of both traditional and innovative approaches to polysemy (see section § 3.2.1), we will focus on the one put forth by Generative Lexicon Theory (section § 3.2.2), which is the framework adopted in the CroaTPAS resource, which will be presented in Chapter 4.

3.2.1 TRADITIONAL VS. INNOVATIVE APPROACHES TO POLYSEMY

In traditional approaches to the lexicon, polysemy is conceived as a list of pre-defined senses stored in the lexeme that get selected according to the context of use. These approaches rely on what Fillmore called *checklist theories of lexical meaning* (1975), a similar concept to *sense enumeration lexica* (Pustejovsky, 1995). This is the standard way traditional dictionaries, as well as resources used in Natural Language Processing tasks are put together, e.g., WordNet (Miller, 1995), which is used for word sense disambiguation tasks.

However, according to Ježek (2016, p. 77), this approach to polysemy has proven to be:

- *uneconomic*, since resources built this way require long lists of meanings for each word.
- *incomplete*, since, depending on the context, words can express a potentially infinite number of meanings, which cannot be covered by a checklist.
- *inadequate*, given that the boundaries between word senses are not rigid, thus creating problems in telling similar word senses apart[44].

In his famous contribution *I don't believe in word senses* (1997), Kilgarriff comes to the conclusion that the concept of *word sense* is not "sufficiently well-defined to be a workable basic unit of meaning" (1997, p. 108) and that only the contextual corpus occurrences of a word should be used to

[42] https://www.merriam-webster.com/dictionary/polysemous. Website last visited on August 23rd, 2023.

[43] https://www.oed.com/search/dictionary/?scope=Entries&q=polysemy. Website last visited on August 23rd, 2023.

[44] Kilgarriff sums up this problem saying that words senses are "very slippery entities" (1997, p. 1).

identify its different meanings, specifically by grouping corpus occurrences into different clusters representing different word senses. Another non-traditional approach to polysemy is that formulated by Hanks (2013), who believes that "when [words] are isolated from context, [they] have *meaning potential* rather than meaning as such. Meanings are contextually bound" (2013, p. 66), i.e., only when words are contextualized, can meanings be identified.

According to similar innovative approaches to polysemy, words are believed to possess a *core meaning*, which varies as a function of the interaction with the other words it combines with, as well as of the situational context it is uttered in (Pustejovsky, 1995; Recanati, 2012). The outcome of said interaction generates the meaning of the sentence. Therefore, if we accept this perspective, polysemy is the result of the contextual adjustments and modulations of the core meaning of a word in semantic composition.

In the next section, we will explore a framework that relies and further elaborates on this compositional principle, namely that of Generative Lexicon Theory (Pustejovsky, 1995; Pustejovsky & Ježek, 2008), which is the framework the CroaTPAS resource is based on.

3.2.2 GENERATIVE LEXICON THEORY

Generative Lexicon Theory is a compositional theory of lexical semantics developed by the computational linguist James Pustejovsky (1995; Pustejovsky & Ježek, 2008), which claims that lexical meaning is not an intrinsic feature of lexical items but is *contextually generated* by means of their interaction with one another according to specific generative devices.

As effectively summarised by Ježek (2016, p. 78), these principles operate at a sub-lexical level targeting specific aspects of word meaning and are able to provide a thorough interpretation for a wide range of lexical phenomena, such as verb polysemy or metonymic shifts in verb arguments.

The computational apparatus of this theory rests on four pivotal linguistic structures and three generative devices. The structures, also known as *levels of representation* (Pustejovsky, 1995, p. 60), are the following:

I. ARGUMENT STRUCTURE – the structure specifying the number and nature of the arguments required by a verb (see section § 3.1).
II. EVENT STRUCTURE – the structure which defines the type of event (state, activity, or transition) expressed by a verb (see Chapter 4).
III. QUALIA STRUCTURE – the structure portraying the four fundamental semantic properties of any lexical item, namely its *Formal*, *Constitutive*, *Telic* and *Agentive quale* (Pustejovsky & Ježek, 2012) (see § 3.2.2.1).

IV. LEXICAL INHERITANCE STRUCTURE – the structure able to assign an appropriate Semantic Type within a Semantic Type System (Pustejovsky & Ježek, 2008, p. 183) to any lexical item. As underlined by Ježek, "Semantic Types are not abstract categories, but semantic classes discovered by generalizing over statistically relevant lists of verb collocates, which take the name of *lexical sets*" (2014, p. 890).

As for the generative devices at play, which are also called *principles for strong compositionality,* they are the following:

I. The PRINCIPLE OF CO-COMPOSITION, which we will see at play when it comes to verb polysemy (see § 3.2.2.2).
II. SEMANTIC TYPE COERCION, which plays a role in metonymic shifts (see § 3.2.2.3).
III. SELECTIVE BINDING, which is able to account for adjectival polysemy, which we will not focus on in this contribution.

3.2.2.1 QUALIA STRUCTURE

Although this contribution will mainly focus on verb argument structures, we also need to shed some light on *qualia* structures, which are paramount to the understanding of Semantic Type coercions (see § 3.2.2.3). As you can see from Figure 3, *qualia* – plural of the Latin *quale* 'what kind?' – are semantic properties of lexical items, in this case of the noun *sandwich*, and are grouped within square brackets.

$$\begin{bmatrix} \textbf{sandwich(x)} \\ \text{CONST} = \{\textbf{bread,...}\} \\ \text{FORMAL} = \textbf{physform(x)} \\ \text{TELIC} = \textbf{eat(P,w,x)} \\ \text{AGENTIVE} = \textbf{make\_activity(z,x)} \end{bmatrix}$$

Figure 3 – The qualia structure of the noun *sandwich* (Pustejovsky & Ježek, 2008, p. 185)

The *constitutive quale* (CONST) is a list of all the parts that make up the meaning of the lexical entity, which in our case are all the sandwich's ingredients, namely, {bread, ham, etc.}. The *formal quale* (FORMAL) answers the question "What type of thing is this?", which in this case would be a [Physical Entity]. The *telic quale* (TELIC), from the Greek word *télos* 'end', expresses the goal of our lexical item, which in this case – since we are dealing with a sandwich – is *eating*. Last but not least, the *agentive quale* (AGENTIVE) identifies the origin of the entity denoted by the lexical item, which in this case has been *made*. However, not every word possesses all four *qualia*: for example, natural entities

usually are not equipped with the telic one, since they have not been created for a purpose (Hanks, 2017).

3.2.2.2 PRINCIPLE OF CO-COMPOSITION

According to Pustejovsky (1995), meanings do not only compose, they *co-compose*, that is to say, they influence each other in meaning composition.

According to this principle, verb polysemy can be seen as the result of the influence of the Semantic Type of the arguments on the core meaning of the verb. In other words, the variation in argument semantics co-causes variation in verb meaning.

In order to illustrate this principle, Pustejovsky focuses on the polysemy of *to bake* (1995, p. 122), a verb having both a *change of state* (7.a) and a *creation* meaning (7.b).

(7) a. *John baked the potato.*
b. *John baked the cake.*

Saying that these different senses are contextually generated by the verb itself and the semantics of the arguments equals to saying that it is the presence of *potato* as a direct object in sentence (7.a) that causes the meaning of *to bake* to shift towards "making food undergo a change of state in order to be eaten", whereas in sentence (7.b) the verb takes on its *creation* sense because *cakes* do not undergo any change of state, but are "the result of baking".

This being said, if we consider lexical items expressing verb arguments to be as semantically active as the verb itself in generating verb meaning, clustering them in semantic classes and assigning them Semantic Type labels becomes a useful representation strategy to better portray their role in co-composition.

For instance, let us consider the following examples taken from the Croatian Web as Corpus (Ljubešić & Klubička, 2014) featuring two of the possible senses of the verb *piti* 'to drink', i.e., that of "drinking a liquid" (8.a) and that of "taking, ingesting a medicine" (8.b):

(8) a. *Djeca ne piju kavu.*
'Children do not drink coffee.'
b. *Većina ljudi pije antibiotike na svoju ruku.*
'Most people take antibiotics on their own initiative.'

Following the same logic as in the previous examples, given that both (8.a) and (8.b) feature the same verb and a human subject (*djeca* 'children' and *većina ljudi* 'most people', respectively), it is mainly the semantics of the direct object to contribute to the different meaning co-composition. In the case

of (8.a), it is the presence of a beverage (*kavu* 'coffee') in the accusative case that allows us to interpret the verb meaning as "drinking a liquid", while in (8.b), it is the presence of direct object denoting a drug (*antibiotike* 'antibiotics') that makes verb meaning sway towards "taking a medicine". As you can see, already while providing these simple definitions, one has to resort to general labels, such as *beverage*, *liquid*, *drug,* or *medicine*, to expressing argument semantics. This is precisely the function of Semantic Types, which are not to be understood as "abstract categories, but semantic classes discovered by generalizing over statistically relevant lists of verb collocates, which take the name of *lexical sets*" (Ježek et al., 2014, p. 890).

In the CroaTPAS resource, the Semantic Types that were used to create Typed Predicate Argument Structures (henceforth, *patterns*), representing the different Croatian verb senses, were taken from the *System of Semantic Types* (Ježek, 2019), an ontology containing 180 Types used also for the T-PAS resource (Ježek et al., 2014), which was inspired by the *Brandeis Shallow Ontology* (Pustejovsky et al., 2004) which originally contained 65 Types.

Within Generative Lexicon, Semantic Types are represented between square brackets, both in isolation and when they appear in a pattern, as you can see in Figure 4, which shows the two patterns corresponding to the two senses of the verb *piti* 'to drink' discussed in (8.a) and (8.b), respectively.

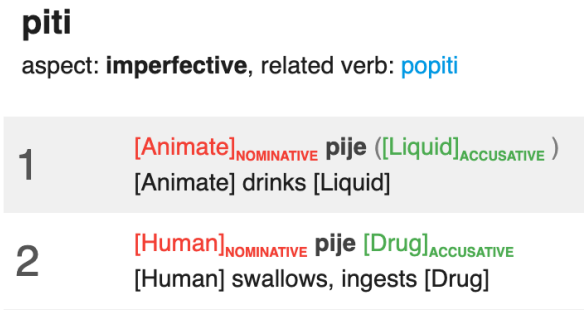


Figure 4 – Two of the senses of the verb *piti* 'to drink' in CroaTPAS

Notice that, while the Semantic Type chosen to represent the semantics of the subject of pattern 1 "drinking a [Liquid]" is [Animate], the Semantic Type chosen to portray the semantics of the subject of pattern 2 "swallowing, ingesting a [Drug]" is [Human]. This is because, while both humans and animals intentionally drink liquids, usually medicines are not intentionally ingested by animals, but only by humans. This called for the use of the Semantic Type [Animate] in pattern 1, which is a hypernym of both [Human] and [Animal] in the hierarchy of the System of Semantic Types[45].

[45] More on Semantic Types and the hierarchical organization of the System of Semantic Types in section 3.3.

3.2.2.3 Semantic Type coercion

Within the Generative Lexicon, we talk of *pure selection* when the Semantic Type requirements of a given verb sense are satisfied by its arguments (Pustejovsky & Ježek, 2008, p. 186). However, when there is a mismatch between the Semantic Type that a specific verb sense should require and the actual Semantic Type of one (or more) of its arguments, a metonymic shift takes place, which in this framework takes the name of *Semantic Type coercion*.

To provide an example, Pustejovsky (1995, pp. 115–116) offers the following examples:

(9) a. *John began reading a book.*
b. *John began a book.*

In sentence (9.a), the Semantic Type that we would assign to the second argument is that of an event, namely the [Activity] of "reading a book". However, in sentence (9.b), despite there being no change in meaning, the Semantic Type of the second argument does not satisfy the verb's selectional requirements. How is this possible? According to Pustejovsky, a Semantic Type coercion "applies to reconstruct the semantics of the complement" (*ibidem*), i.e., the verb forces the noun phrase *a book* into an event denotation by exploiting the value of the *telic quale* associated with it (see § 3.2.2.1), namely, *reading.* This is why, Semantic Type coercion is also referred to as Qualia Exploitation (Pustejovsky & Ježek, 2008, p. 195).

Both the T-PAS (Ježek et al., 2014) and CroaTPAS (Marini, 2022a; Marini & Ježek, 2019) resources keep trace of metonymic shifts in the argument structures of their verb senses. The first Semantic Type coercion annotation exercise for Italian was carried out by Ježek & Quochi "to provide a reliable […] corpus annotated for coercions to be used as training and test set […] for metonymy recognition [tasks] and/or processing figurative language" (2010, p. 1464). In their study, they annotated over 4,000 metonymic instances focusing on a sample of 26 T-PAS verbs and obtained a good Inter Annotator Agreement score.

The CroaTPAS verb inventory includes the Croatian translational equivalents of all the Italian *coercive verbs* from Ježek & Quochi (2010), but – unlike in the case of T-PAS – Semantic Type coercions were annotated for all its Croatian verb entries.

Sentence (10) provides a prime example of Semantic Type coercion featuring the perfective variant of the Croatian verb *piti* 'to drink', namely *popiti*.

(10) *Stipe je popio čašu.*
'Stipe drank a glass.'

As in the case of the fist pattern of its imperfective variant (Figure 4), also the verb *popiti* 'to drink' usually selects direct objects semantically typed as [Liquid] to convey the meaning of "drinking" (see Figure 5). However, in sentence (10), we can see it combines with the direct object *čašu* 'glass', which we could label as [Container], without a corresponding change in verb meaning. What happens it that, since the *telic quale* of such a [Container] is "holding a liquid" (Pustejovsky & Ježek, 2008, p. 200), the verb is able to instantiate a metonymic shift *Container* → *Content* by exploiting said *telic quale*, thus enabling us to interpret the [Container] as the [Liquid] it contains.

As you can see from Figure 5, Semantic Type coercions are encoded and displayed as metonymic sub patterns in CroaTPAS (in 1.1.m, "m" stands for metonymic). Moreover, we can notice how the sense description of the sub pattern has been adapted to make the metonymic relationship between [Liquid] and [Container] easy to decode.

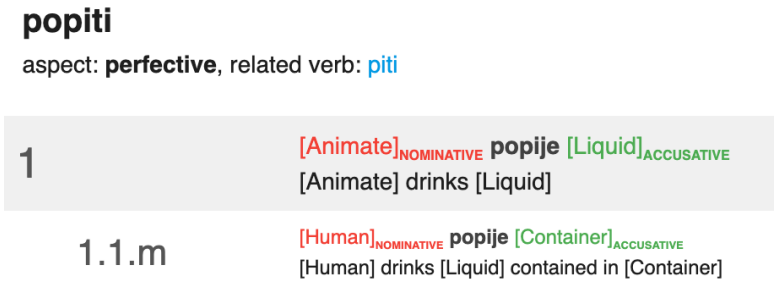


Figure 5 – Pattern 1 of the verb *popiti* 'to drink' and its metonymic sub pattern 1.1.m encoded in CroaTPAS

In this Chapter, we have introduced the notion of verb argument structure (§ 3.1) and provided an overview of the existing approaches to verb polysemy (§ 3.2), focusing specifically on Generative Lexicon Theory (section § 3.2.2). As we will see in Chapter 4, this is in fact the theoretical framework adopted in the CroaTPAS resource, which will be presented hereafter.

CHAPTER 4: THE CROATPAS RESOURCE

The CroaTPAS (*Croatian Typed Predicate Argument Structures*) resource is particularly relevant to this PhD research project, not only because it can be considered one of its concrete deliverables, but also because its corpus-based data have been used as the starting point for the analysis in Chapter 5, and to introduce the topic of aspectual verb pairs to the students of the Croatian language intensive course that will be the subject of Chapter 6.

CroaTPAS (Marini, 2022a; Marini & Ježek, 2019) is a digital lexicographic resource for Croatian verbs able to frame verb polysemy (see § 3.2) and metonymic shifts (see § 3.2.2.3). The resource relies on the theoretical framework of Generative Lexicon Theory (see § 3.2.2) and a lexicographic methodology inspired by Corpus Pattern Analysis (see § 4.2.2).

Just like in the case of its Italian sister resource T-PAS (Ježek et al., 2014), CroaTPAS consists in a collection of corpus-derived predicate argument structures (called *patterns*) whose argument slots have been semantically annotated using a set of labels called Semantic Types (see § 4.2.1). The resource currently contains 180 Croatian verbs for a total of 795 different verb senses.

The editing environment for said patterns, as well as the online resource, have been developed thanks to the technical support of *Lexical Computing Ltd.*[46] in the person of Vítek Baisa. CroaTPAS is currently accessible online at the address: https://croatpas.baisa.cz/.

In the following sections, more details will be provided regarding the selection of Croatian verbs available in CroaTPAS (see § 4.1), the methodology that was followed to create the resource and its online interface (see § 4.2), and the resource's evaluation process (see § 4.3).

4.1 VERB SELECTION

The CroaTPAS resource contains 81 Croatian verb pairings, the majority of which are true aspectual pairs (see section § 2.3.1), 10 biaspectual verbs, 7 unpaired imperfective verbs and 2 unpaired perfective verbs, for a total of 180 separate entries[47].

The reason for this diverse inventory is that the Croatian verb entries for the CroaTPAS resource were selected following an Italian-driven procedure, that is to say, it was decided to prioritise

[46] *Lexical Computing Ltd*. is the company that has developed and owns the Sketch Engine corpus management platform.

[47] The total number of entries is 180 and not 181, because the verb *pozvati* is part of two pairings: *zvati/pozvati* 'to call' and *pozivati/pozvati* 'to invite'.

Croatian verbs whose Italian translational equivalents were already available in the T-PAS resource in order to facilitate future multilingual linking between the two.

The first 37 entries that were added to the resource are the Croatian translational equivalents of a sample of 24 Italian verbs taken from Ježek & Quochi (2010). These verbs are known as *coercive* verbs[48], i.e., verbs that instantiate metonymic shifts in their arguments structures (see section § 3.2.2.3 on Semantic Type Coercions).

The remaining 143 verb entries are the Croatian translational equivalents of a sample of 74 verbs belonging to the *Vocabolario Fondamentale* 'fundamental vocabulary' of the Italian language, a group of approximately 2,000 words with the highest frequency counts covering about 90% of all Italian written and spoken text (Chiari & De Mauro, 2014, p. 113). These 2,000 lexemes, together with about 3,000 lexemes pertaining to the *Vocabolario d'Alto Uso* 'high usage vocabulary' and 2,400 lexemes that are part of the *Vocabolario di Alta Disponibilità* 'high availability vocabulary', constitute the so-called *Vocabolario di Base della lingua italiana* 'basic vocabulary of the Italian language'[49] (De Mauro, 2016).

The Italian corpus we used to extract the Italian verbs which we later translated into Croatian, is a reduced version of the Italian Web as Corpus (Baroni & Kilgarriff, 2006), which is freely available on the *Sketch Engine* corpus management platform (Kilgarriff et al., 2014). After extracting the frequency counts for all the De Mauro *fundamental verbs* present in T-PAS and taking out the first and the last 20 verbs on the list, we selected 74 Italian candidates around the median frequency value. Once the Italian verb list was made up, the Croatian translational equivalents were selected from the Italian/Croatian bilingual dictionary by Špikić (2017). Since Špikić usually offered multiple translation options, some of the Croatian verb pairings have been selected to evaluate the semantic impact of prefixes on base verbs (e.g., *vući/izvući*) rather than based on pure aspectual opposition.

See Appendix 1 for a complete list of all CroaTPAS verbs, together with their T-PAS counterparts and English equivalents. Please, take notice that, in the list of Italian T-PAS verbs, the verbs *sentire* and *guidare* appear twice because, since they both display a highly polysemous behaviour, we decided to create entries for more than one of their Croatian translational equivalents, namely *čuti* 'to hear' and *osjećati/osjetiti* 'to feel' for the first, *voditi/provoditi* 'to lead' and *voziti* 'to drive' for the second.

[48] The Italian *coercive* verbs from Ježek & Quochi (2010) are underlined in Appendix 1.

[49] The *Nuovo Vocabolario di Base della lingua italiana* 'new basic vocabulary of the Italian language' first appeared as an annex to the volume *Guida all'uso delle parole* (De Mauro, 1980).

4.2 Methodology

Just like its Italian sister resource T-PAS (Ježek et al., 2014), also the CroaTPAS resource rests on five key-components, namely:

I. a representative corpus for Croatian
II. a shallow ontology of Semantic Types (see § 4.2.1)
III. a sound lexicographic methodology (see § 4.2.2)
IV. an adequate pattern editing environment (see § 4.2.3)
V. a devoted online interface (see § 4.2.4)

As our first component, we selected the RELDI PoS-tagged version of the Croatian Web as Corpus (Ljubešić & Klubička, 2014), a corpus containing over 1.2 billion web-scraped tokens, which gives a reliable picture of contemporary Croatian with enough variety of registers and contexts of use.

Since the Croatian Web as Corpus was already available on the *Sketch Engine* platform, it was quite convenient for Lexical Computing Ltd. to integrate the editing environment on the platform itself and then connect it to the corpus itself. Moreover, by choosing to work with the Croatian Web as Corpus as our reference corpus, we maintained the highest level of compatibility with T-PAS, since also the Italian resource uses a corpus from the Web as Corpus family, namely a reduced version of the Italian Web as Corpus (Baroni & Kilgarriff, 2006), which was expressly created for this purpose.

In the following sections, more will be said also about the other components, starting with the shallow ontology we used for the semantic typing of CroaTPAS patterns, called System of Semantic Types (see § 4.2.1).

4.2.1 The System of Semantic Types

In recent years, the term *ontology* has become a buzzword in computational linguistics and other related research fields, even though it can refer to very different concepts.

Philosophical ontologies are structured systems of entities and relations which enable us to study reality (Gaio et al., 2010, p. 107), while, in Information Technology, *computational ontologies* are data architectures providing a robust formulation of the properties and relationships between the entities belonging to a certain domain (*ivi*, 108). However, in a wider sense, an ontology can also be understood as the list of "all the words and expressions in a particular language that denote concepts, both abstract and concrete" (Hanks & Ježek, 2008, p. 392): this last definition is the one we are going to build upon.

The *System of Semantic Types* (Ježek, 2019) is a corpus-based ontology, shared by both the T-PAS and CroaTPAS resources, which contains 180 Semantic Type labels obtained from manual clustering of lexical items found in the argument positions of verbal structures from large Italian corpora. Therefore, despite the fact that these Semantic Types "look very much like ontological categories, [what they actually do is] reflecting the way in which humans talk about events and states of entities through language" (*ivi*, 2) – and specifically the way Italian people do – which justifies, e.g., the presence of the Semantic Type [Wine] in the System, due to its relevance for very specific, but widely used, wine-related verbs such as *invecchiare* 'to age' and *maturare* 'to mature'. Figure 6 offers an outline of the top-level of the T-PAS ontology (Ježek, 2019).

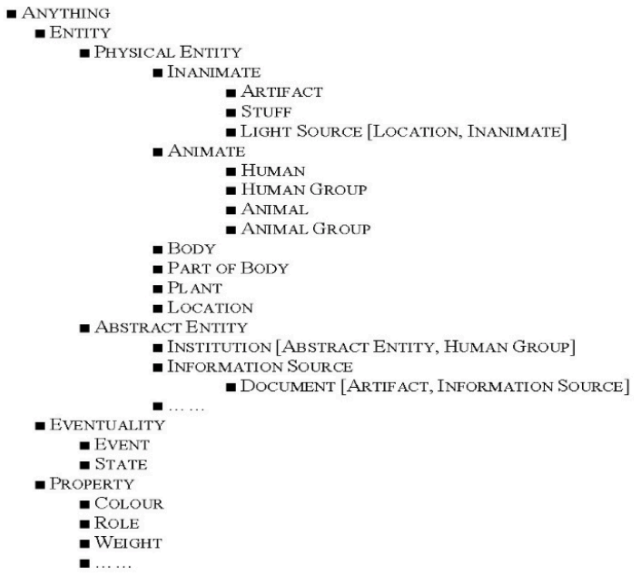


Figure 6 – Top-level of the T-PAS Semantic Type System taken from Ježek (2019).

As already mentioned, Semantic Types should be understood "semantic classes that were discovered by generalizing over statistically relevant lists of verb collocates, which take the name of *lexical sets*" (Ježek et al., 2014, p. 890).

According to both Hanks (1996), and Pustejovsky & Ježek (2008, p. 187), a lexical set is the set of nouns which "typically co-occur with [a given verb] in a specified grammatical relation", thus providing evidence to *populate* a node in the ontology, i.e., a Semantic Type.

It is, however, important to stress the fact that lexical sets do not behave as homogeneous structures and cannot be therefore mapped onto Semantic Types. As Hanks & Ježek (2008, p. 399) put it, "lexical sets tend to *shimmer* […] from verb to verb: some words drop out while others come in". For example, in spite of the fact that both the verb *to wash* and the verb *to amputate* select direct objects typed as [Part of Body], one will hardly find examples of a *face* or *hair* being amputated,

while washing them remains quite the prototypical thing to do: this is the difference between *core* and *satellite members* of a lexical set (Pustejovsky & Ježek, 2008, p. 189).

In the CroaTPAS resource, lexical sets are manually annotated and displayed next to the Semantic Type they populate in curly brackets. Figure 7, for instance, showcases the Croatian lexical set of the Semantic Type [Liquid] to be found on the direct object of the first pattern of the verb *piti* 'to drink'. The lexical set contains, among others, the words *vodu* 'water', *kavu* 'coffee', *vino* 'wine', and *čaj* 'tea'.

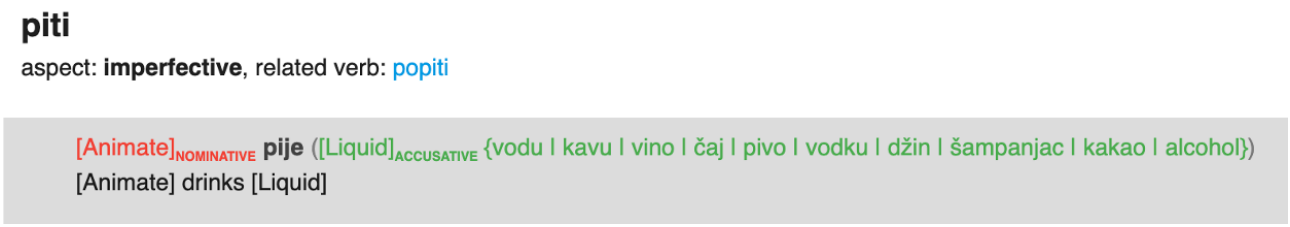


Figure 6 – Top-level of the T-PAS Semantic Type System taken from Ježek (2019).

4.2.2 Corpus Pattern Analysis

Corpus Pattern Analysis (henceforth, CPA) (Hanks, 2012, 2004) is a corpus-based lexicographic methodology that has been applied in a number of resources[50]: the Pattern Dictionary of English Verbs, PDEV[51] (Hanks & Pustejovsky, 2005), the Italian Typed Predicate Argument Structure resource, T-PAS[52] (Ježek et al., 2014), the Pattern Dictionary of Spanish Verbs, PDSV[53] (Renau & Nazar, 2011), and, of course, CroaTPAS.

While being rooted in Sinclair's pioneering work in corpus linguistics (Sinclair, 1991), CPA relies heavily on Pustejovsky's Generative Lexicon Theory (see section § 3.2.2), thus bridging the gap between the two fields.

From a theoretical standpoint, CPA also relies on the Theory of Norms and Exploitations (Hanks, 2013), which distinguishes between conventional word uses (also called *norms*) and deviations from these norms (*exploitations*). CPA aims to identify all normal patterns of word usage and separate them from creative exploitations by means of extensive lexicographic analysis (Hanks, 2013, p. 404). Its potential lies in the fact that it focuses on identifying word meaning in its

[50] For an overview of CPA-based resources, see Marini (2020).

[51] https://pdev.org.uk/. Last visited on September 1st, 2023.

[52] https://tpas.unipv.it/. Last visited on September 1st, 2023.

[53] http://www.tecling.com/verbario/. Last visited on September 1st, 2023.

prototypical contexts, i.e., in their syntagmatic patterns of use, which – in the case of verb senses – consist in their valency structures.

CroaTPAS's pattern annotation methodology is a customized version of CPA, which can be entirely carried out on the Sketch Engine platform and requires the following steps:

I. 250 corpus lines are randomly sampled for each Croatian verb entry from the Croatian Web as Corpus and displayed in the Key Word In Context (KWIC) view.
II. the different verb senses are identified by the lexicographer going through the corpus lines looking for regular patterns of use.
III. pattern strings are created in an appropriate pattern editing environment (see § 4.2.3), integrated in the Sketch Engine, by labelling argument slots with the right Semantic Types, which were identified generalising over recurring lexical sets.
IV. numbers are assigned to the corpus lines exemplifying each pattern, so that each semantically tagged valency structure is justified by corpus evidence. The same number is assigned to the corpus lines instantiating a pattern and the pattern itself.

Up until this fourth step, CroaTPAS and T-PAS share the same CPA-inspired annotation methodology, which is visually summarized in Figure 7 taking the Italian verb *bere* 'to drink' as an example.

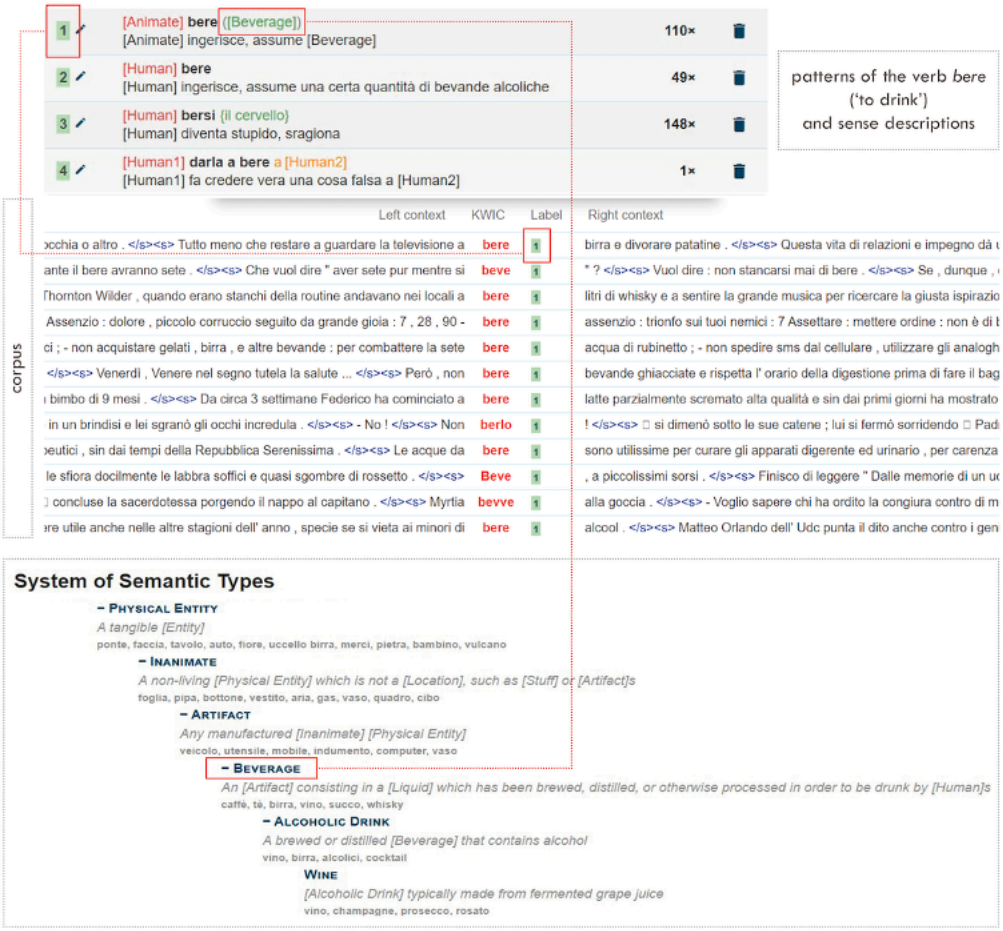


Figure 7 – Verb patterns, concordances, and an excerpt of the System of Semantic Types (T-PAS data).

Unlike T-PAS, however, CroaTPAS's customized CPA methodology includes a final fifth step due to the existence of aspectual verb pairs in Slavic languages (see section § 2.3.1).

As a matter of fact, a Croatian verb sense attested in a conspicuous number of concordances within the 250-line sample of a given entry may sometimes be justified only by a few corpus instances in the 250-line sample of its related aspectual variant, or it may even be found only outside of the random sample. This difference in the frequency of use of the same verb sense across related aspectual variants is easily explained, since not all the "different ways of viewing the internal temporal constituency of a situation" (Comrie, 1976, p. 3) are equally represented (i.e., not equally useful) in language use. Therefore, when a verb sense is well attested in a variant but lacking in its partner, to assess whether its absence is due to distributional reasons or whether it is actually beyond the expressive power of that variant, it was decided to expand the search beyond the 250-lines sample and to create patterns also for those verb senses entirely justified by concordance lines found outside the sample.

For instance, let us focus on the Croatian aspectual pair *padati/pasti* 'to fall' and specifically on one of the senses that was retrieved in the 250-line samples of both the imperfective and perfective variant, namely that of "raining", which is encoded in pattern 2 of Figure 8.

Figure 8 – The first 4 patterns of the Croatian verb *padati* 'to fall' encoded in CroaTPAS

In the 250-line sample of the imperfective entry *padati*, this sense is attested in 20 concordance lines, whereas, in the sample of the perfective entry *pasti*, only in 5. This disparity is due to the different contexts of use: since rain – *kiša* in Croatian – is a weather event that usually lasts for a certain period of time, it is more common to talk and write about it using the imperfective variant *padati*, as you can see from the examples (11.a), (11.b), and (11.c), taken from pattern 2 in Figure 8:

(11) a. *Više voliš kada pada kiša ili snijeg?*
'Do you like it better when it rains or when it snows?'

b. *Ova kiša pada već danima.*
'This rain has already been falling for days.'

c. *U Dalmaciji je bilo sunčano dok je na kontinentu cijeli dan padala kiša.*
'In Dalmatia it was sunny, while in the continental part it has been raining all day long.'

However, the possibility of talking about rain perfectively exists and, even though few examples were found within the 250-line sample, several good ones were found once we started looking outside[54],

[54] More on the concept of *good example* in section § 4.2.4.1.

as you can see from sentences (12.a), (12.b), and (12.c) from pattern 2 of the CroaTPAS perfective entry *pasti* (Figure 9).

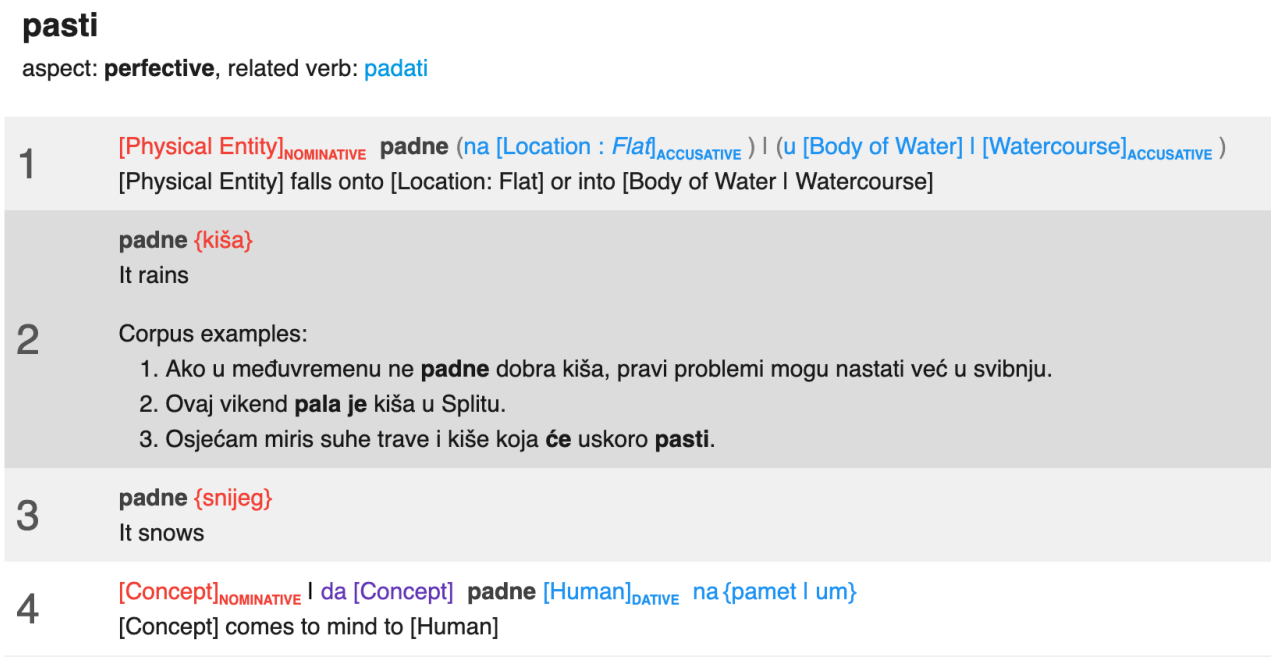


Figure 9 – The first 4 patterns of the Croatian verb *pasti* 'to fall' encoded in CroaTPAS

(12) a. *Ako u međuvremenu ne padne dobra kiša, pravi problemi mogu nastati već u svibnju.*
'If in the meantime it does not rain, big problems may come up already in May.'

b. *Ovaj vikend pala je kiša u Splitu.*
'This weekend, it rained in Split.'

c. *Osjećam miris suhe trave i kiše koja će uskoro pasti.*
'I can recognise the smell of dry grass and rain, that will fall soon.'

In (12.a) and (12.c), the perfective variant was chosen because the focus is on when it will rain or start to rain, but there is no emphasis on the process, while in (12.b) the point that the author of the sentence wants to get across is that, in Split, this weekend, it rained. As Silić and Pranjković would say, "[the focus] is on the result of the process, on its being executed, on an action which is not characterised by processuality" (2007, p. 288). We do not know for how long, but we do know that it was not sunny all day, because it *rained*. In sentences (11.b) and (11.c), on the other hand, the events expressed by the verb *padati* were framed imperfectively, looking at their internal structure, at how long they lasted – a choice underlined even by the choice of adverbs: *danima* 'for days' and *cijeli dan* 'all day long'.

More will be said on how Croatian patterns are created using the pattern editor developed for CroaTPAS in the following paragraph.

4.2.3 SKEMA

Skema (Baisa et al., 2020) is the pattern editing environment developed by *Lexical Computing Ltd.* in collaboration with the University of Pavia for the T-PAS resource, which was later adapted to suit the needs of CroaTPAS. The software is currently integrated in the Sketch Engine platform in order for patterns to be linked to the corpus lines justifying them.

As all Slavic languages, Croatian displays several language-specific features which were taken into account while customising *Skema*, namely: verbal aspect (see section § 2.3.1), the Croatian case system (see section § 1.3), and its different argument categorisation.

However, since the customization of *Skema* has been dealt with in detail in the Master thesis (Marini, 2020), in this dissertation, we will focus on the development of the online interface of the resource instead.

4.2.4 CROATPAS ONLINE INTERFACE

The CroaTPAS online interface was designed in order to provide a user-friendly access to its pattern inventory. As you can see from Figure 10, CroaTPAS presents itself as a list of its verb entries in alphabetical order. Each entry is followed by the number of its identified senses, the verb's corpus frequency on the Croatian Web as Corpus (Ljubešić & Klubička, 2014) and its normalised per-million frequency. If you look at the number of verb senses (also called *pattern number*, since each sense is encoded in a pattern), you may notice that sometimes the verbs that are part of a proper aspectual pair (see section § 2.3.1) have the same pattern number, as in the case of the pair *bacati/baciti* 'to throw'. In other aspectual pairs, however, this is not the case, as for the verbs *dirati/dirnuti* 'to touch', where one verb sense goes missing in the perfective variant *dirnuti*. These meaning asymmetries are going to be further explored in Chapter 5.

If users want to access a specific verb entry, they can search for it on the Filter search bar. However, if they are interested only in the patterns containing a specific Semantic Type, they can exploit the Pattern Search function to the top right of the resource homepage (Figure 10).

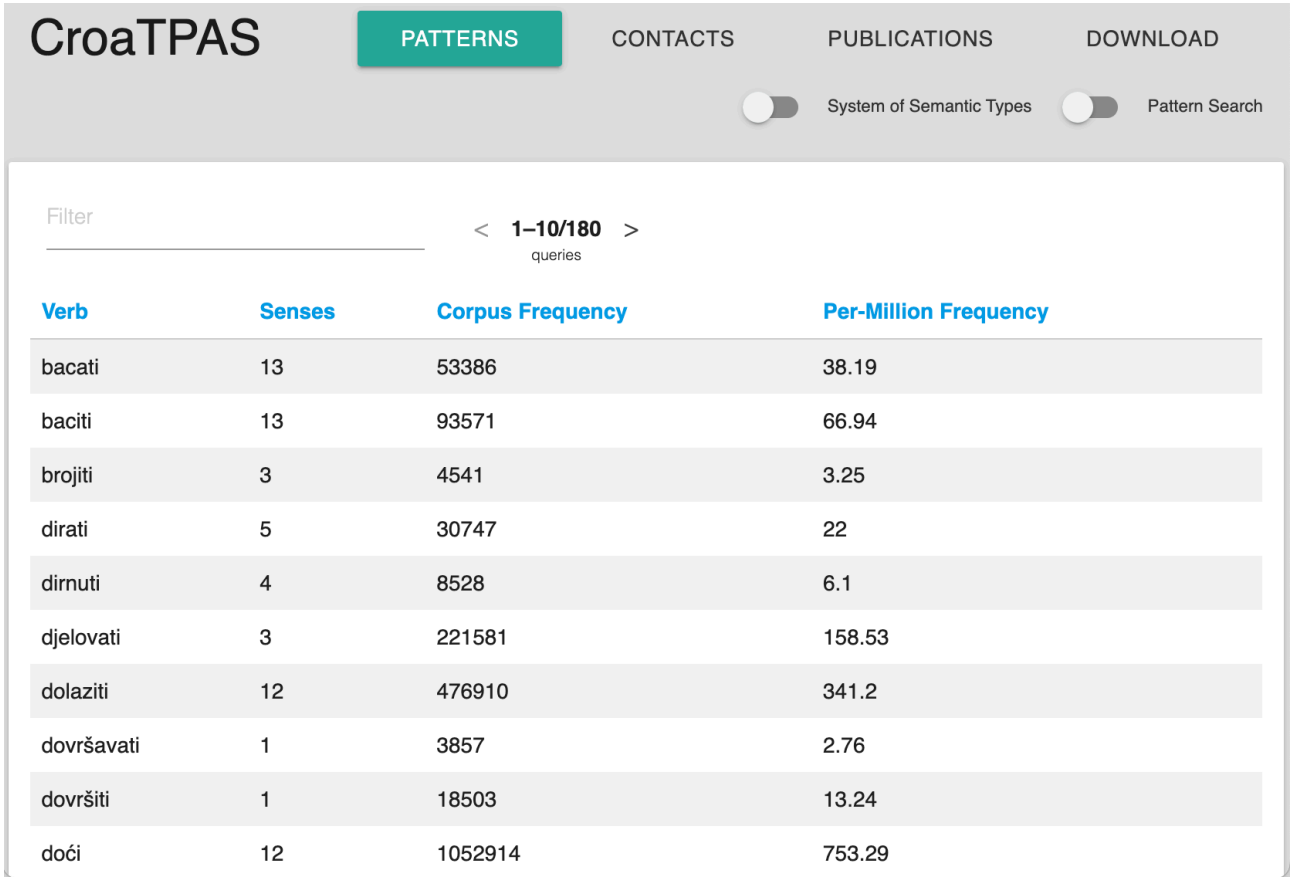


Figure 10 – The homepage of CroaTPAS online interface

The Pattern Search option allows to query the CroaTPAS database by Semantic Type and argument slot, and it returns the verb entries that satisfy those criteria specifying which of their patterns match the selected requirements. If we look at Figure 11, for example, the search is focused on all the CroaTPAS patterns where the Semantic Type [Relationship] is used as a direct object. As we can see, this search returns only two aspectually related verbs, *spasiti* and *spašavati* 'to save', both featuring [Relationship] on the direct object slot of their second pattern.

Given that, in order to carry out a proper Pattern Search, users have to be at least familiar with Semantic Type labels, full access was given to the System of Semantic Types[55]. One only needs to activate the *System of Semantic Types* button to the top right of the homepage, next to Pattern Search, and the whole hierarchy will unfold at the bottom of the screen. This is particularly helpful for two reasons: on the one hand, thanks to the English descriptions below each Semantic Type label, users will have a deeper insight of argument semantics; on the other hand, seeing Semantic Types in the context of their ontology will help them better understand the annotation choices made when creating CroaTPAS patterns.

[55] The System can also be downloaded, together with the whole resource, by clicking on the Download button.

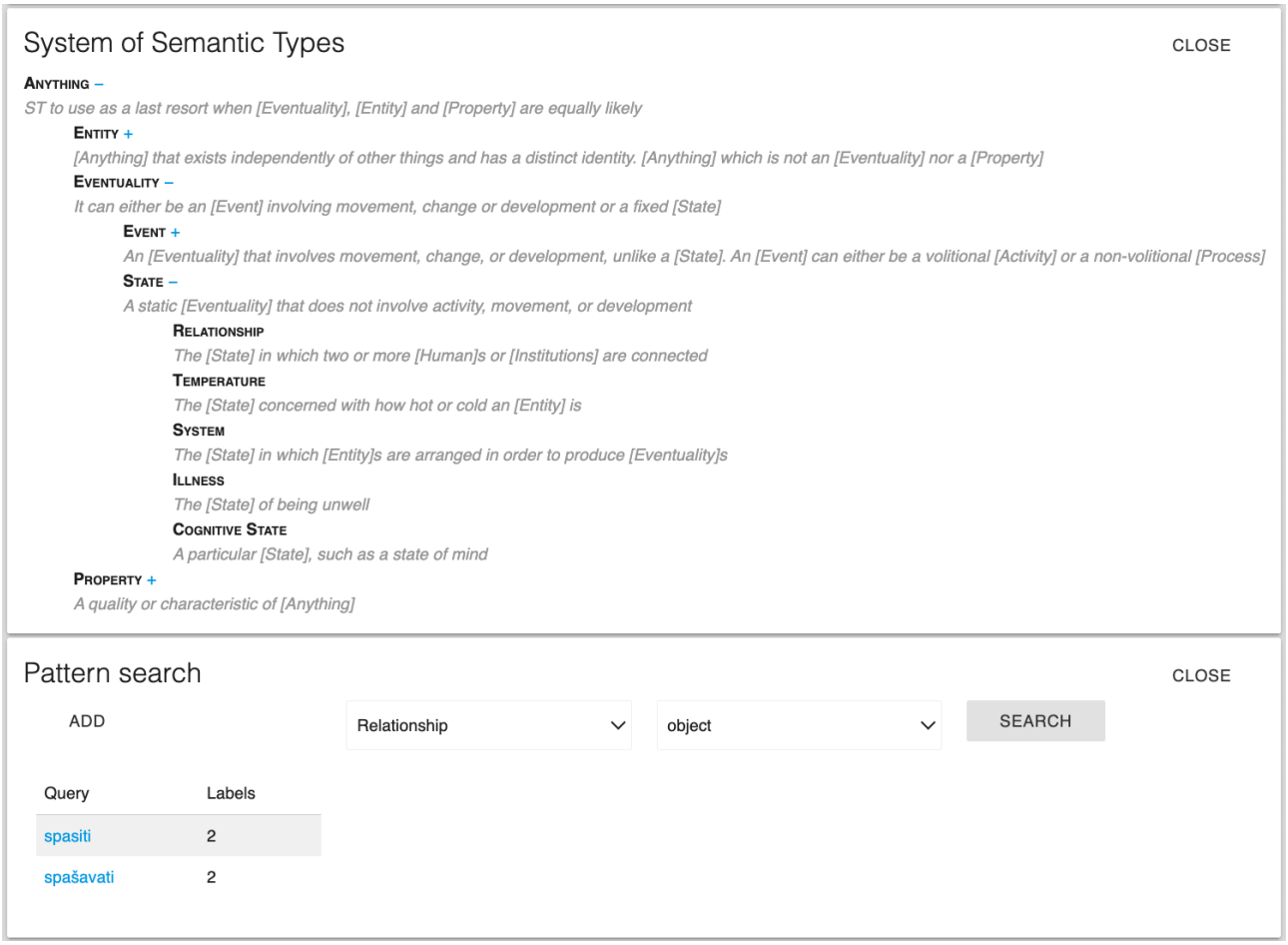
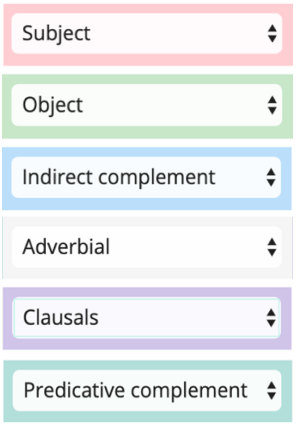


Figure 11 – Pattern Search of the Semantic Type [Relationship] in object position on CroaTPAS

Since Slavic languages like Croatian lack a fixed word order, express grammatical relations between sentence components with case markings and have a variety of case/preposition combinations to express the same arguments, already when adapting the *Skema* editor to Croatian (Marini, 2020), we realized we had to find an effective way to display the morpho-syntactic information provided by case in the context of our patterns.

The same strategy was also applied to CroaTPAS's online interface, namely bottom-right indexes expressing case markings were combined with an argument colour-coding strategy. In this way, patterns become much more user-friendly and accessible to potential learners.

Figure 12 provides an overview of the strategy, which includes the following argument-colour pairings: *Subject* (red); *Object* (green); *Indirect Complement* (blue); *Adverbial* (grey); *Clausal* (purple); *Predicative Complement* (turquoise). To the right of each argument slot, you can see its mapping onto the corresponding complements recognised by Croatian grammar, which are governed by the presence of cases and prepositions (see section § 1.3).



nominative and dative subjects

direct objects in the accusative case, Slavic genitives, and partitive genitives

indirect objects in the genitive, dative or instrumental case and prepositional objects[56]

obligatory complements expressed by adverbs

clausal objects and subjects

subject and object predicative complements

Figure 12 – CroaTPAS's argument colour-coding strategy and its mapping with traditional Croatian complements

Since CroaTPAS is first and foremost a semantic resource, the same Semantic Type can have multiple sentential realisations, like in the case of the Semantic Type [Activity] in the first pattern of the verb *preporučiti* 'to recommend' (Figure 13), which appears both as a green accusative Object and a purple Clausal Object introduced by the complementizer *da*, the most common Croatian equivalent of *that*.

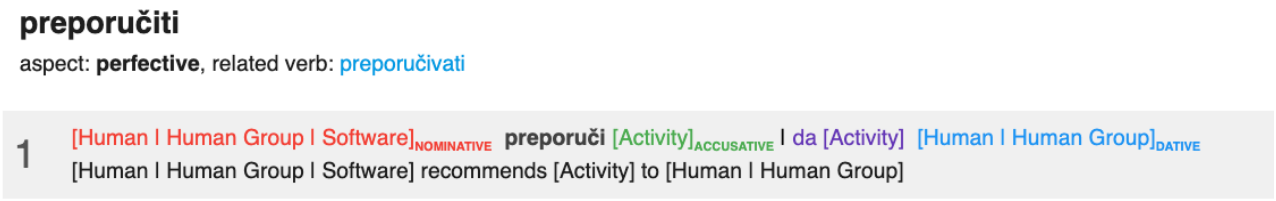


Figure 13 – CroaTPAS's argument colour-coding strategy and its mapping with traditional Croatian complements

Figure 14 offers a further example of CroaTPAS verb entry. As you can see, below the entry label (*spasiti*), the aspectual value of the verb is provided ('perfective'), as are any existing links to other related verbs available in the resource (*spašavati*).

The verb features six different patterns, one for each of its senses. Their meanings are encoded in English in the line below each pattern, called *sense descriptions*. This choice was made to make CroaTPAS a fully-fledged bilingual resource accessible also to Croatian language learners as well as to linguists interested in Croatian semantics.

[56] Croatian grammar makes a distinction between *non-prepositional* and *prepositional complements* (Barić et al., 2005, p. 443): subjects, direct objects (bare accusatives, Slavic genitives and partitive genitives) and indirect objects (bare datives, genitives and instrumentals) all fall within the first category, whereas accusative, instrumental and locative complements which require a preposition are the prepositional ones.

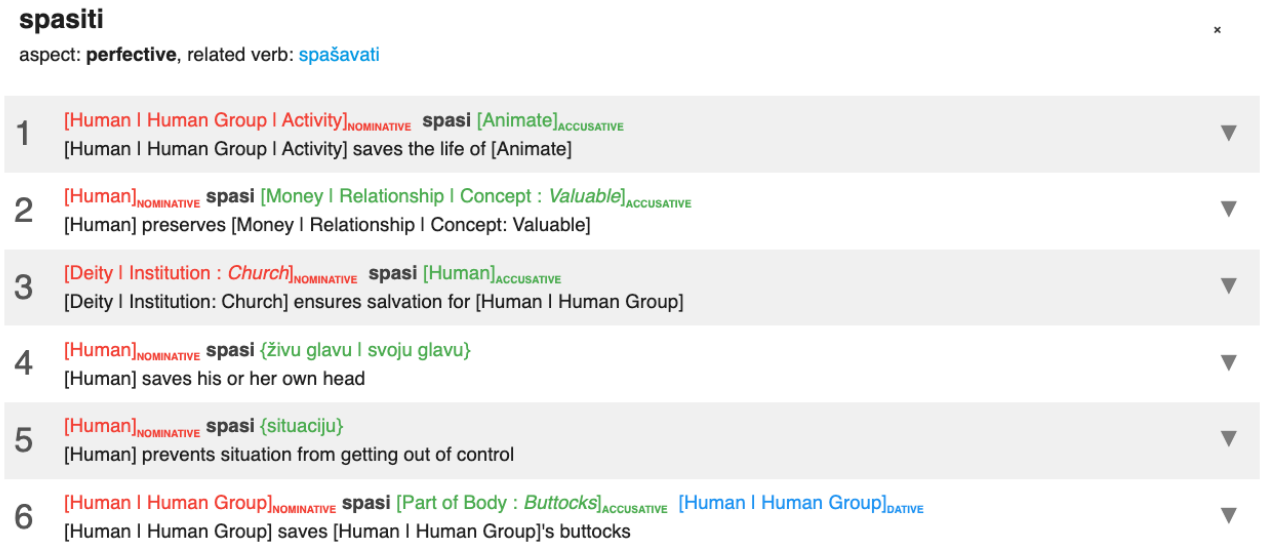


Figure 14 – The patterns inventory of the CroaTPAS entry *spasiti* 'to save'

As you can see from the inventory, a difference was drawn between the meanings of pattern 1 "saving the life of an [Animate]" and pattern 2 "preserving [Money | Relationship | Concept: Valuable]". In the second pattern, we find the Type [Relationship] in alternation with other Semantic Types on the same direct object slot marked in the accusative case. This phenomenon is called Semantic Type Alternation (Ježek et al., *forthcoming*) and is the result of an annotation choice.

As a matter of fact, as pointed out by Hanks (2013, p. 92), grouping together semantically similar corpus lines to justify a verb meaning is no trivial task. In some instances, it is practically impossible to establish "where the boundary between senses should be drawn, or indeed whether a distinction should be made at all" (Hanks & Ježek, 2008, p. 397). These two different approaches are usually referred to, in lexicography, as *splitting* and *lumping*. In this case, lumping together all those corpus lines referring to saving money, relationships and other valuable concepts was preferred, in order to maintain the semantic connection between them. The meaning encoded in the first pattern, on the other hand, was deemed different enough to warrant the splitting.

Going back to CroaTPAS's interface functions, if we click on any pattern from any verb entry, an extended pop-up version of said pattern will appear (see Figure 15 below), revealing a selection of the lexical items making up the lexical sets of its Semantic Types (see § 4.2.1) and at least three corpus examples extracted from the Croatian Web As Corpus via the GDEX algorithm (Kilgarriff et al., 2008) (see § 4.2.4.1), which were deemed representative of their respective verb senses.

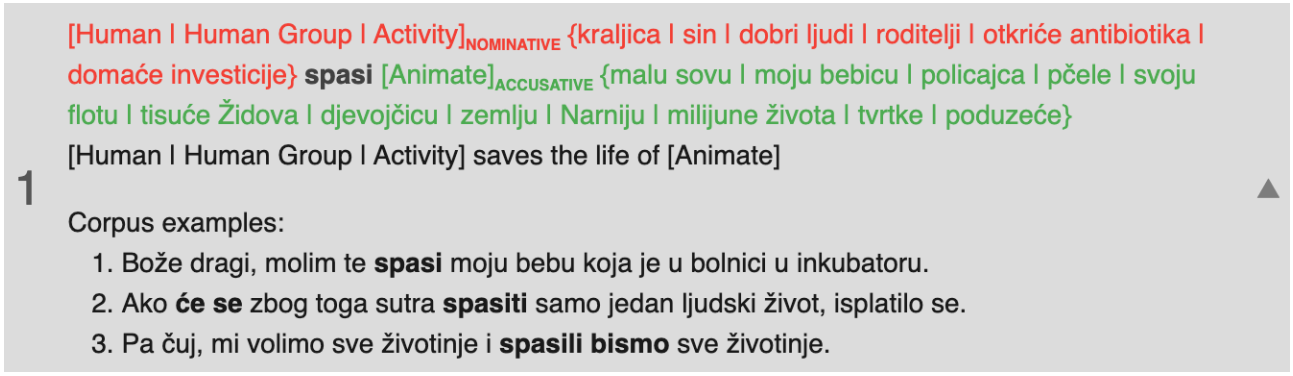


Figure 15 – Extended pop-up version of pattern 1 of the CroaTPAS entry *spasiti* 'to save' with GDEX examples

All the lexical items that appear in curly brackets after the Semantic Types [Human | Human Group | Activity] alternating on the Subject slot and after [Animate] on the Object slot were manually selected from the corpus instances justifying the pattern. Some of them were deemed representative of the Semantic Type itself, while others were considered insightful and potentially stimulating for Croatian language learners. For instance, {*kraljica* 'queen' and *sin* 'son'} are both prime examples of [Human], while {*malu sovu* 'small owl'} is an interesting example of [Animate].

More will be said on GDEX examples in the next section.

4.2.4.1 Good Dictionary EXamples (GDEX)

GDEX is the acronym of Good Dictionary EXamples, and is an algorithm used to evaluate sentences with respect to their suitability to serve as dictionary examples (Kilgarriff et al., 2008). It identifies such sentences automatically so that so that good examples do not need to be identified manually by a concordance search, which is especially useful in multi-billion-word corpora where concordances return hundreds or thousands of hits.

GDEX assigns concordances a score ranging from 0 to 1 based on their lexical and syntactic complexity, sorts them, and places the ones with the highest score at the top. In this way, poor candidates are ruled out and the lexicographer is offered a preselected set of sentences with a higher chance of being good examples.

GDEX was originally developed by the engineers and linguists at *Lexical Computing Ltd.* (the company behind the *Sketch Engine* corpus management system (Kilgarriff et al., 2014) for the electronic version of the *Macmillan English Dictionary for Advanced Learners* (Rundell, 2002), but it has long been integrated in the platform.

Figure 16 provides an example of concordance search to with the GDEX algorithm has been applied. The image shows the seven best dictionary example candidates containing an inflected form of the Croatian verb *piti* 'to drink', all of them scoring 0.95 on the GDEX scale.

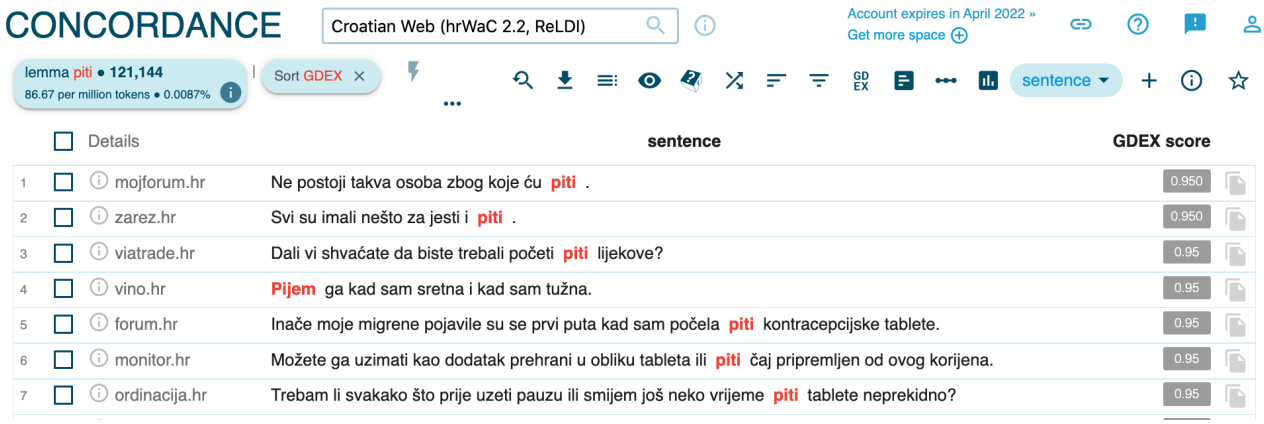


Figure 16 – Concordance lines containing the verb *piti* 'to drink' to which the GDEX algorithm has been applied

But what makes a good example and how does it translate in a GDEX score? According to Kilgarriff (2008), a good example must be:

- *typical*, i.e., it should exhibit frequent patterns of usage.
- *informative*, i.e., it should help to elucidate its corresponding definition.
- *intelligible to learners*, i.e., it should avoid gratuitously difficult language, puzzling names, anaphoric references, or other deictic expressions which cannot be understood without the wider context. This also goes by the name of "readability".

To translate these requirements into measurable features, they looked at:

- *Sentence length*: preference was given to sentences between 10 and 25 word-long, while longer and shorter ones were penalized.
- *Word frequencies*: sentences containing words that were not amongst the commonest 17,000 words in the language were penalized.
- *Pronouns and anaphors*: sentences containing expressions like *this*, *that*, *it*, and that need further context to be understood were penalized.
- *Main clause*: sentences where the target collocation is in the main clause were preferred.

To populate CroaTPAS's online interface with sound corpus examples that could well represent the 795 patterns available in the resource, an average of 3 GDEX examples per pattern were extracted via the *Sketch Engine* platform from the annotated corpus lines of the Croatian Web as Corpus (Ljubešić & Klubička, 2014) linked to CroaTPAS patterns via the *Skema* editor (Baisa et al., 2020) – for a total of nearly 2,400 GDEX examples.

Corpus examples with a GDEX score higher that 0.8 were given priority, but for verb senses instantiated only by a few corpus examples, also lower scores were accepted, and sentences were occasionally simplified. The Master students in Linguistics attending the course "Doing quantitative

research in linguistics" offered by the Department of Linguistics[57] of the University of Zadar took part in this semi-automated extraction. Before being uploaded to the resource, all GDEX examples were checked by a Croatian native speaker and language consultant.

Apart from their GDEX score, the other element that was considered while selecting sentences was their verb tense. Priority was given sentences featuring different verb tenses, in order to be able to show Croatian language learners the impact of verb tense on meaning and to contrastively compare the GDEX examples of related aspectual variants in different tenses. The following sentences are taken from Figure 15.

(13) a. *Bože dragi, molim te **spasi** moju bebu koja je u bolnici u inkubatoru.*
'Dear God, please save my baby who is in the hospital in an incubator.'

b. *Ako **će se** zbog toga sutra **spasiti** samo jedan ljudski život, isplatilo se.*
'If tomorrow, thanks to that, even one life will be saved, it was worth it.'

c. *Pa čuj, mi volimo sve životinje i **spasili bismo** sve životinje.*
'Listen, we love all animals and would save all animals.'

In (13.a), we can see the use of the imperative form *spasi* 'save', which can only be used in the perfective, since no imperfective view of a command or wish is allowed in Croatian. In (13.b), on the other hand, we can see an impersonal simple future, which leaves rooms for discussion, for instance, regarding the use of the particle *se* 'oneself' in Croatian. Last but not least, in sentence (13.c), we can witness the conditional form *spasili bismo* 'we would save'.

[57] A special thank you to Professor Marco Angster who allowed me to take part in his teaching activities in Zadar and involve his students in my PhD project.

4.3 CROATPAS'S RESOURCE EVALUATION

Since the identification of CroaTPAS's verb sense distinctions and subsequent pattern creation has been single-handedly carried out only by one annotator[58], who is not a native speaker of Croatian, an evaluation procedure was deemed necessary to give credibility to the resource.

It was decided that recruiting Croatian native speakers to perform an additional annotation task over a subset of CroaTPAS verbs to be able to calculate a measure of Inter Annotator Agreement (Pustejovsky & Stubbs, 2012) – as it has been done by (Cinkova et al., 2012) with a subset of 30 English PDEV verbs (Hanks & Pustejovsky, 2005) – would have been unpractical given the time-consuming nature of such a task and the difficulties in finding trained annotators.

Therefore, to evaluate the overall goodness of the identified verb senses stored in CroaTPAS, as well as the English sense descriptions elucidating them, an online questionnaire based on a multiple-choice sense disambiguation task was devised, pilot tested and subsequently distributed to native speakers of Croatian with a good command of English, as well as to non-native respondents with native-like Croatian proficiency, following a snowball sampling methodology.

Answers were collected and compared against a yardstick answer sheet containing the actual sense distinctions identified in CroaTPAS. The *Jaccard similarity index* was used as a measure of agreement between respondents and the resource's original annotator.

Since the verbs included in the survey are a representative selection of the verbs contained in CroaTPAS in terms of polysemy and aspect, they allowed for a generalization of the results to the whole of the resource.

This chapter will address the choices made regarding population sampling (§ 4.3.1), questionnaire design (§ 4.3.2), item choice (§ 4.3.3), verb selection (§ 4.3.4), and pilot testing (§ 4.3.5). The last section of the chapter (§ 4.3.6) will be devoted to the analysis of the survey answers. The results have already been published in Marini (2022b) in the *Proceedings of the18th Joint ACL - ISO Workshop on Interoperable Semantic Annotation (ISA-18)*.

4.3.1 POPULATION SAMPLING

According to Tony Johnstone Young's definition, a survey "aims to make inferences about a population by examining a sample from that population" (2016, p. 165). Since "a good sample is very similar to the target population in its most important general characteristics" (Dörnyei, 2007, p. 96), the aim should be that of obtaining a representative subset of the whole population.

[58] The Candidate.

However, if we were to apply this reasoning to CroaTPAS's evaluation, we would have needed to identify a representative subset of the global Croatian-speaking population with a good command of the English language through probability sampling (either random or stratified), which is beyond the practical means at our disposal.

Therefore, we opted for a non-probability sampling strategy, namely *snowball sampling* (Dörnyei, 2007, p. 98; Johnstone Young, 2016, p. 169), which involves the researcher contacting a small group of good candidate participants, who are then asked to identify further appropriate respondents among their contacts and so on, thus generating a sort of chain reaction. In this way, the sample cannot be said to be random since its relationship to the general population is not clear, but it does have the advantage of being a quick and convenient way of reaching respondents.

4.3.2 QUESTIONNAIRE DESIGN

Surveys are usually conducted using questionnaires, which are defined as "any text-based instrument that give survey participants a series of questions to answer or statements to respond to, either by indicating a response – by marking a page, writing a number, or checking a box on paper or online" (Johnstone Young, 2016, p. 167).

In the case of the CroaTPAS evaluation survey if was decided to opt for a Google Form survey, i.e., a free online survey which can be built using Google Suite and easily forwarded via link, in order to be able to proceed with the *snowball sampling* in the easiest way.

Bearing in mind that "most questionnaires consist of a title, some general information about the research project; instructions or guidance for completion, with examples; the items; and final, additional information" (Johnstone Young, 2016, p. 172), the CroaTPAS evaluation questionnaire was divided in four sections, structured as follows:

I. Title & General information about the research project:

Meaning in Context

Hello! My name is Costanza Marini, and I am a PhD candidate in Linguistics from the Universities of Pavia and Bergamo (Italy). My PhD project is focused on developing a digital bilingual dictionary of Croatian verbs, which is going be freely accessible online.

If you are a Croatian native speaker (or have native-like proficiency) with a good command of the English language, you can help me evaluate my work by filling out this survey.

The survey is made up of multiple-choice questions and is going to take you around 20 minutes to complete.

At the end, you are going to be asked some background information (age, education...) to help us interpret the data. Your responses are going to be completely anonymous and confidential.

If you would like to know more about the project before taking part in it, feel free to get in touch with me at the following address: costanza.marini@unibg.it
Thank you so much for your time and let us begin!

Already in this first section of the questionnaire, the bilingual nature of the resource is quite apparent. As a matter of fact, all instructions are given in English to ensure that respondents realize they do need to be proficient not only in Croatian but also in English to be able to carry out the verb sense disambiguation task that constitutes the main bulk of the survey.

The choice in title – namely, *Meaning in Context* – also reflects this line of reasoning, while keeping an eye on not overloading participants with technical information about verb semantics. As a matter of fact, it was decided not to mention the name of the resource in any part of the survey either nor to go into details when it comes to verb polysemy, so as not to distract respondents from the task at hand.

As recommended by Johnstone Young (2016, p. 172), it was specified that only one answer was acceptable for each item, even though the online nature of the form made it possible to override the problem by not allowing respondents to provide multiple answers. Having answered all questions was also made a prerequisite to be able to submit the form.

Special attention was devoted to thanking and reassuring respondents of the confidentiality and anonymity of their answers, as well as of the availability of the author to answer any possible question concerning the project.

II. Instructions with an example, followed by the items:

You are now going to start the multiple-choice section of the survey.
You are going to be presented with Croatian sentences like the one below, each followed by two or more English alternatives to choose from.

*( ) Kako *podnijeti* ljetne vrućine, a osjećati se ugodno?*[59]

o [Human | Human Group] can stand, endures [Anything: Negative]
o [Human | Institution] submits, files [Document | Request | Offer]

Each Croatian sentence contains a *word* between asterisks (like *podnijeti* in the example above).
Focus on the meaning of that *word* in the context of the sentence and choose the English description that you think better fits that meaning from the list of possible alternatives.

[59] Respondents were not provided a translation of the example sentence above, but it means 'How to endure the summer heat, while feeling comfortable?'.

In this second section of the questionnaire, instructions are given to respondents on how to carry out the sense disambiguation task. They are shown an example of a multiple-choice item, in order to prepare them to what they are expected to do before starting the real task. The verb that they are called to disambiguate is provided between asterisks in order to be easily identifiable. Also in this case, the instructions are given in English and references to technical terms are avoided.

III. Background information:

You are now going to be asked some background information to better interpret your responses.
Everything you say is going to be completely anonymous and confidential.

- How old are you?
- What is your gender?
  - female
  - male
  - other (non-binary)
- What is the highest level of education you have completed?
  - primary school
  - secondary school (high school, gymnasium…)
  - university (Bachelor, Master's, PhD…)
- If you attended (or are currently attending) university, what was (or is) your field of interest)?
- Which country did you grow up in? If more than one, list them all.
- Which country do you live in now?
- What is your native language? If you have more than one, list them all.
- Your English language skills are…
  basic < 1 < 2 < 3 < 4 < 5 < excellent
- Your Croatian language skills are…
  basic < 1 < 2 < 3 < 4 < 5 < excellent
- According to you, the English alternatives you could choose from in the multiple-choice section of the survey were…
  inadequate < 1 < 2 < 3 < 4 < 5 < fitting
- Optional comment

In this third section of the questionnaire, respondents were asked to provide background information. Since there is evidence that asking for demographic information at the beginning of a questionnaire can be off-putting (Fife-Schaw, 2006), those questions were asked after the actual survey. Questions included both open questions, multiple-choice items, as well as three sentence completion items involving semantic differential scales (Dörnyei, 2007, p. 105).

The first two scales were designed to let respondents complete statements concerning their language proficiency in English and Croatian by marking a 5-step continuum between two polar adjectives, namely *basic* and *excellent*, in order "to elicit a more meaningful answer than a simple question" (*ibidem*, 107). The last item involving a semantic differential scale, on the other hand, aimed at getting honest feedback on the perceived adequacy of the English sense descriptions respondents were presented with in the multiple-choice section of the survey.

IV. Thanks, and informed consent:

> Once again, thank you so much for your contribution!
> This could not have been done without your help.
> By submitting this form, you are accepting the use of your anonymous answers for research purposes.
> If you wish to be kept informed on the developments of my research, you can email me at:
> costanza.marini@unibg.it

In this fourth and last section of the survey, respondents are once again thanked for their contribution, reassured of the anonymity of their answers, as well as of the author availability for answering any project-related question.

Moreover, in order to be in compliance with the ethical principle of informed consent, respondents were explicitly asked to submit the form only if they accepted that their anonymous answers were going to be used for research purposes.

4.3.3 ITEM CHOICE

Since research has shown the existence of a bias making respondents more likely to lean towards middle-of-the-road and affirmative responses (Hui & Triandis, 1989; Johnstone Young, 2016, p. 168), it was decided not to opt for true-false items nor scales with mid items, but rather multiple-choice items as the one below.

*( ) Kako *podnijeti* ljetne vrućine, a osjećati se ugodno?*

o [Human | Human Group] can stand, endures [Anything: Negative]
o [Human | Institution] submits, files [Document | Request | Offer]

Each item consists of a Croatian sentence taken from the Croatian Web as Corpus (Ljubešić & Klubička, 2014) containing one of the verbs under scrutiny (see the following section § 3.4.4 on verb selection), as well as the English sense descriptions of the identified patterns for that given verb.
Since the corpus is available on the *Sketch Engine* platform, appropriate sentences were chosen based on their GDEX values (Kilgarriff et al., 2008). GDEX is an algorithm able to rate the suitability of a

sentence to serve as a Good Dictionary EXample. It takes into account several measurable parameters, such as sentence length, word frequency, presence of pronouns and anaphors, as well as presence of target collocations, and it returns a value between 0 and 1. For our study, we decided to stick to sentences that were rated over a score of 0.8. For more on the subject see section § 4.2.4.1 of this Chapter.

As for the English sense descriptions, they are presented to respondents as alternatives to choose from to better explain the meaning of the verb between asterisks in the example above them. It was decided to leave them as they will be presented in the online resource, namely containing Semantic Types between square brackets. This was deemed as a good way to indirectly verify whether this technical way of representing argument semantics was detrimental to the user understanding or not, based on the survey's results.

In order not to make the task too automatic, it was decided that the number of multiple-choice items for each chosen verb should be higher than the actual number of senses displayed by that verb. For example, 6 items containing 6 different corpus examples were set up to evaluate the verbs *misliti* 'to think' and *ubiti* 'to kill', which are both 4-senses verbs.

4.3.4 VERB SELECTION

Out of the 180 verbs in CroaTPAS, 32 were excluded since they only feature one sense and could thus not be used in a sense disambiguation task such as the one devised for the survey.

In order to provide respondents with corpus examples from a representative selection of verbs, the 148 remaining CroaTPAS entries were divided by pattern number as well as aspect, and percentages were subsequently calculated for each group of verbs (see Table 1 below).

Perf.	**N**	**%**	**Imperf.**	**N**	**%**	**Bias.**	**N**	**%**
2P	18	12	**2P**	18	12	**2P**	2	1
3P	14	9	**3P**	15	10	**3P**	1	1
4P	14	9	**4P**	11	7	**4P**	0	0
5P	7	5	**5P**	11	7	**5P**	1	1
6P	8	5	**6P**	8	5	**6P**	0	0
7P	2	1	**7P**	4	3	**7P**	0	0
8P	1	1	**8P**	1	1	**8P**	0	0
11P	3	2	**11P**	4	3	**11P**	0	0
12P	1	1	**12P**	2	1	**12P**	0	0
13P	1	1	**13P**	1	1	**13P**	0	0

Table 1 – CroaTPAS verbs divided by pattern number and aspect

The Table reports the number (N) of perfective, imperfective and biaspectual verbs that display two verb senses and therefore 2 patterns (2P), three verb senses (3P), found verb senses (4P), and so on. To guarantee a verb selection representative of the resource as a whole, we decided to keep the percentages from Table 1 fixed and determine how many verbs (N) would have to be chosen for each aspect/patterns grouping (2P perfective, 5P biaspectual, etc.) given an arbitrary number of verbs to be included.

Considering the structure of the survey and the fact that an item would have to be added to the sense disambiguation task for each of the senses of the selected verbs, it was decided to fix the overall number of verbs at 20.

Consequently, as you can see in Table 2 below, the number of verbs (N) for each grouping was calculated by multiplying the fixed percentages from Table 1 by the arbitrary size of our selection of verbs (20).

Perf.	**N**	**%**	**Imperf.**	**N**	**%**	**Bias.**	**N**	**%**
2P	2.4	12	**2P**	2.4	12	**2P**	0.3	1
3P	1.9	9	**3P**	2.0	10	**3P**	0.1	1
4P	1.9	9	**4P**	1.5	7	**4P**	0.0	0
5P	0.9	5	**5P**	1.5	7	**5P**	0.1	1
6P	1.1	5	**6P**	1.1	5	**6P**	0.0	0
7P	0.3	1	**7P**	0.5	3	**7P**	0.0	0
8P	0.1	1	**8P**	0.1	1	**8P**	0.0	0
11P	0.4	2	**11P**	0.5	3	**11P**	0.0	0
12P	0.1	1	**12P**	0.3	1	**12P**	0.0	0
13P	0.1	1	**13P**	0.1	1	**13P**	0.0	0

Table 2 – Distribution of CroaTPAS verbs to be included

Obviously, the results of the above-mentioned calculation had to be transformed into integers, since verbs are discrete entities. Therefore, the floats from Table 2 with a decimal value ≥ 0.5 were rounded up to the nearest integer, whereas those with a decimal value < 0.5 were rounded down.

Given that the only biaspectual verbs in Table 2 all had N values < 0.5 and that we would have ended up with a survey without biaspectual verbs hadn't we amended the selection criteria, it was decided to include 3 biaspectual verbs (one for each grouping) to the sample regardless, in order to guarantee their evaluation.

Table 3 provides a complete overview of the verbs that were first selected for the survey, whose senses were evaluated in the pilot testing phase.

Perf.	N	verbs	Imperf.	N	verbs	Bias.	N	verbs
2P	2	podnijeti, prekinuti	**2P**	2	gostiti, željeti	**2P**	1	informirati
3P	2	isključiti, sletjeti	**3P**	2	čitati, kupovati	**3P**	1	napredovati
4P	2	otkriti, ubiti	**4P**	2	osnovati, slati	**5P**	1	kontrolirati
5P	1	prodati	**5P**	2	voziti , pripadati			
6P	1	popiti	**6P**	1	žderati			
			7P	1	odgovarati			
			11P	1	pružati			

Table 3 – The selection of verbs included in CroaTPAS's evaluation survey prior to pilot testing

4.3.5 PILOT TESTING

Following Johnstone Young's recommendation (2016, p. 176), the questionnaire was piloted before beginning with data collection. The pilot group was a small one, namely two respondents, who were asked to complete the draft survey and reflect on its design, the wording of items and the clarity of the example sentences.

The items from section III (*Background information*) were deemed clear and the questions about their language history were able to capture the background of both. Neither one felt that items should be rephrased, and both appreciated the presence of a non-binary gender option in the multiple-choice question about gender. Moreover, they both reported that asking for participants who had a native-like proficiency of Croatian was a good way to include not only foreigners, but also Serbian, Bosnian and Montenegrin native speakers.

The most important amendment made after the pilot testing phase was that of eliminating all the multiple-choice items based on verbs with 7 and 11 senses. As a matter of fact, both participants found that skimming through multiple-choice lists containing that many senses took too hard a toll on their attention levels. Moreover, since the survey was deemed quite long, it was decided to remove the items for one of the 5-pattern imperfective verbs, too.

The following Table offers an overview of the final 19 verbs included in CroaTPAS's evaluation survey after pilot testing, corresponding to a total of 91 items.

Perf.	N	verbs	Imperf.	N	verbs	Bias.	N	verbs
2P	2	podnijeti, prekinuti	**2P**	2	gostiti, željeti	**2P**	1	informirati
3P	2	isključiti, sletjeti	**3P**	2	čitati, kupovati	**3P**	1	napredovati
4P	2	otkriti, ubiti	**4P**	2	osnovati, slati	**5P**	1	kontrolirati
5P	1	prodati	**5P**	1	voziti			
6P	1	popiti	**6P**	1	žderati			

Table 4 – The final selection of verbs included in CroaTPAS's evaluation survey after pilot testing

Following the respondents' feedback, several of the sentences included in the multiple-choice items were also discarded and replaced with shorter and simpler sentenced.

Despite their high GDEX scores, these sentences were identified as problematic since they:

- contained anaphoric pronouns pointing at referents outside sentence limits, thus taking away the readers' focus from verb meaning, as in the case of (14), where the absence of subject pronouns like *she* and *he* (see translation below), due to the fact that Croatian is a pro-drop language, makes it even more unclear to understand who we are referring to.

(14) *Santos govori da je očito da više ne voli Jose i da bi bilo bolje da *prekine* s **njom**.*
'Santos says that it's clear that [**she**] doesn't love Jose anymore and that it would be better if [**he**] *broke up* with **her.**'

- were deemed too long or too syntactically too complex, as in the case of sentences (15) and (16), where the meaning that needs to be disambiguated is in subordinate clauses introduced by semantically complex connectives.

(15) *Automobilske gume najvidljiviji su dio automobila **čija** se potrošnja može *kontrolirati* uvijek i na svakom mjestu.*
'Car tires are the most visible part of a car, **whose** consumption can be checked always and everywhere.'

(16) *Donosimo nekoliko metoda **pomoću kojih** možete *kontrolirati* strahove koji dolaze s ulogom primarnog skrbnika za dijete.*
'We provide several methods **with the help of which** you can *control* the fears that come with being the primary care giver of a child.'

- featured the lexically loaded part of the verb only at the end of the sentence, as in (17), thus making it difficult to disambiguate its meaning because of the demanding processing required to make sense of all the previous context.

(17) *I dakako, većina u Hrvatskoj pojma ne bi imala što to znači da nas mediji *nisu* pravodobno ***informirali**.*
'Therefore, most people in Croatian would not have any idea what it means that media *do not **inform*** us correctly.'

Some slight changes were made in case of misspelling of single words in otherwise well-formed sentences.

4.3.6 RESULTS

In a period of approximately two months, we were able to collect answers from 30 respondents.

4.3.6.1 THE RESPONDENTS

The average age of our 30 respondents is 35.4 years: 12 (40%) are in their 20s, 7 (23.3%) in their 30s, 7 (23.3%) in their 40s, 3 (10%) in their 50s and one (3.3%) in her 60s. Gender-wise, 20 respondents identify as female, 9 as male and one as non-binary.

For what concerns the educational level, 16.7% of the participants holds a secondary school diploma, while 83.3% has attended or is still attending university. Of the latter group, 87% have a Humanities background (Croatian Studies, Linguistics, Foreign Languages, Social studies, Political Sciences, Theatre), while 13% are into Hard sciences (Informatics, Chemistry, Engineering).

All respondents except one consider Croatian as their native language. All of them grew up in either Croatia or Bosnia and Herzegovina, except for four, who were either born in or moved to an English-speaking country in their childhood. Two of these still live abroad, while the rest lives in Croatia. All respondents declared to have a sufficiently high level of English to fill in the survey.

4.3.6.2 THE JACCARD INDEX OF SIMILARITY

For what concerns the multiple-choice section of the survey, each one of the 30 collected answer sheets was compared against a yardstick set of answers representing CroaTPAS's sense distinctions. To provide a measure of how similar each answer sheet was to the yardstick, we calculated the Jaccard *index of similarity*.

The Jaccard similarity between two sets A and B is defined as the "the ratio of the number of elements in the intersection of A and B over the number of elements in the union of A and B" (Zumel & Mount, 2014, p. 184). Since respondents were presented with 91 multiple-choice items, each of the 30 survey answers was assigned a similarity score ranging from 0 to 91 depending on the number of multiple-choice answers in line with CroaTPAS's yardstick. That number was then divided by 91 and multiplied by 100, thus returning a normalised Jaccard index expressing the similarity score (%) between each answer and CroaTPAS's annotation.

As you can see from Figure 17, all survey answers but one range between 100% and 83.51% similarity. The only answer with a lower similarity score (64.84%) was identified by Rosner's Test (Rosner, 1983) [60] as a possible outlier both at 5% and 1% significance and thus discarded.

[60] Rosner's Test was run using the statistical software ProUCL 5.0.

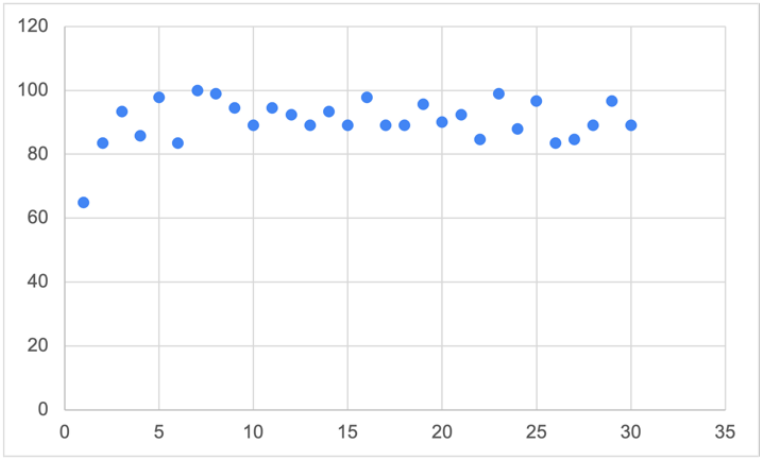


Figure 17 – Dispersion plot of the similarity scores (%) between survey answers and the yardstick set of answers in line with CroaTPAS (outlier included)

Therefore, since the mean similarity score of the remaining 29 survey answers stands at 91.36% (± 5.12) and data sets with a Jaccard similarity above 85% are considered highly stable (Zumel & Mount, 2014, p. 184), we can conclude that the collected answers form a proper cluster showing a high level of agreement with CroaTPAS's sense distinctions, thus corroborating its annotation.

Moreover, according to the Kolmogorov-Smirnov Test of Normality, the distribution of our 29 data appears not to differ significantly from a normal distribution, with a K-S test value (D) of 1.6228, a *p*-value of 0.38796, skewness at 0.03541 and kurtosis at -1.137667.

4.3.6.3 Similarity scores and polysemy

After assessing the overall similarity scores of the survey answer sheets with CroaTPAS's yardstick, the answers to the multiple-choice items were grouped together according to the degree of polysemy of the verb they feature (2P, 3P, 4P, 5P and 6P). Five new sets of similarity scores were thus calculated for these five answer classes and compared to their equivalent sets of yardstick answers representative of CroaTPAS's sense distinctions.

As you can see from the boxplot below (Figure 18), similarity scores appear to vary according to the number of senses expressed by the verbs. Respondents tend to be more in line with CroaTPAS's annotation when it comes to less polysemous verbs, scoring a mean similarity value of 95.5% (± 5.94) in the 2P answer class and of 95.89% (± 5.2) in the 3P answer class. Conversely, the mean similarity score decreases to 91.38% (± 6.8) for the answers to items on four-sense verbs, to 83.3% (±11.43) when it comes to five-sense verbs, and to 91.13% (± 8.89) for those to items containing six-sense verbs.

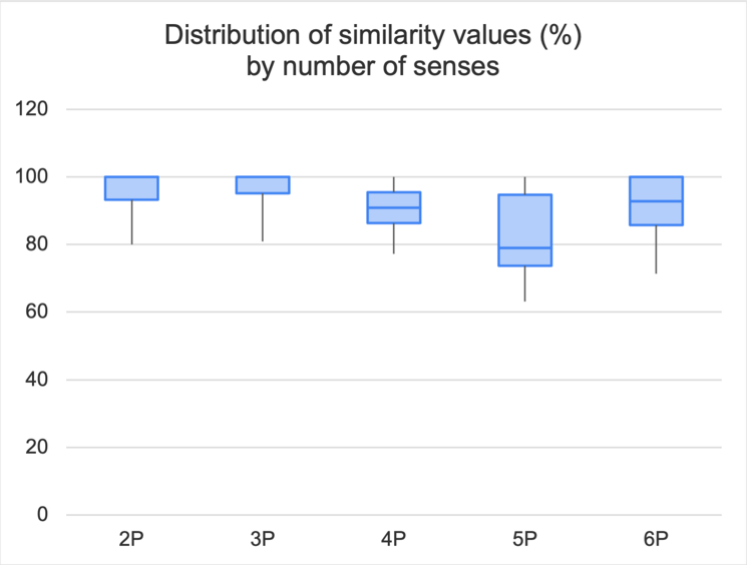


Figure 18 – Box plots showing the different distribution of similarity scores (%) in survey answers referring to verbs with a different degree of polysemy.

Even though the mean similarity scores for all answer classes are still higher than 80%, thus showing – regardless of verb polysemy – a high level of agreement with CroaTPAS's yardstick annotation, we might trace this difference back to the fact that a meaning disambiguation task becomes more demanding the more options one has to choose from.

4.3.6.4 Similarity scores and self-assessed level of English

To provide further support to CroaTPAS's evaluation, we divided the similarity scores of the survey answers by self-assessed level of English language proficiency.

The box plots in Figure 19 show the distribution of similarity scores according to the levels of English respondents declare to possess. Only one participant gave themselves a 1 out of 5 on the semantic differential scale provided in the online survey (see section § 3.4.2) and was thus discarded as an outlier. The remaining 28 respondents are scattered on three different language levels: 5 respondents on level 3, 12 on level 4, and 11 on level 5. The mean similarity scores for the three levels are all high, standing, respectively, at 92.53% (± 5.46), 91.21% (± 5.62) and 91.71% (± 4.46).

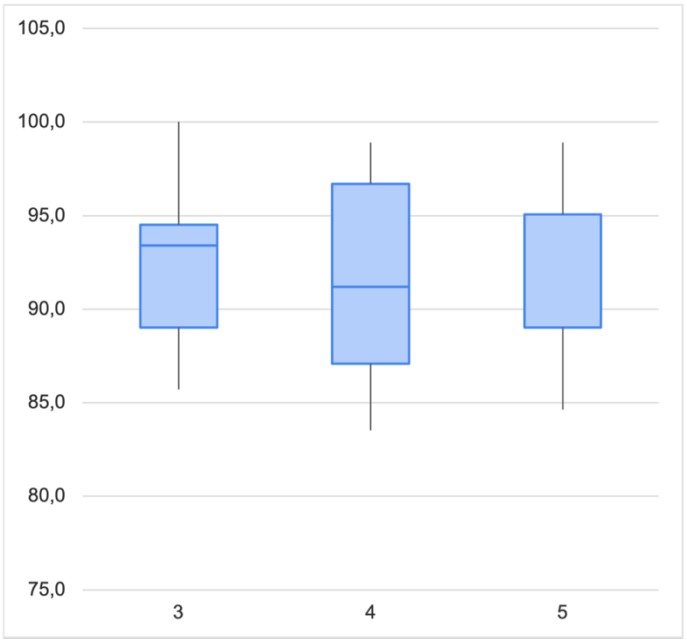


Figure 19 – Box plots showing the different distribution of similarity scores (%) by self-assessed level of English

After making sure that all populations are indeed normally distributed, t-Tests were run between the similarity scores associated to levels 3 and 4, 3 and 5, and 5 and 4, which returned the following t-values: 0.44395, 0.31856, and 0.23454. The corresponding critical values of t for 15, 14 and 21 degrees of freedom at 10% level of significance (5% in each tail) are 2.1314, 3.1448, and 2.0796. Since in all three cases, computed t-values are well below the corresponding critical values, we can conclude there is no statistically significant influence of the self-assessed English proficiency levels possessed by respondents on their sense-disambiguation capability.

This may either mean that asking respondents to self-evaluate their language skills is not a good indicator of their actual linguistic knowledge or that the English sense descriptions provided as multiple-choice options in the sense-disambiguation task were clear enough to guarantee an effective meaning disambiguation regardless of the respondents' English language proficiency.

4.3.6.5 Similarity scores and gender

As in the case of English language proficiency, it was decided to investigate the possible influence of gender on the recorded similarity scores. Given that only one respondent declared to be non-binary, only two similarity score distributions were compared, namely the one of the answers given by female respondents (19) and the other corresponding to the answers given by male participants (9). Two box plots were drawn to represent the two populations (Figure 20).

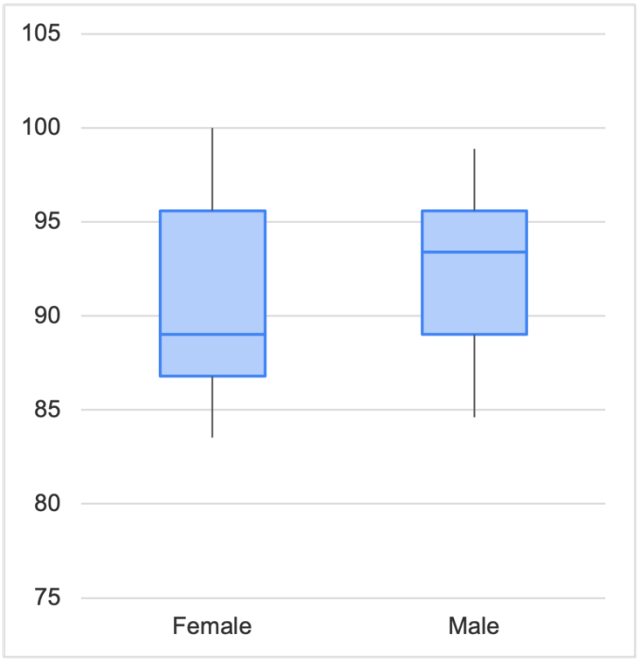


Figure 20 – Box plots showing the different distribution of similarity scores (%) in survey answers by gender

As you can see, the two box plots are quite similar: the mean similarity score for survey answers provided by women is 90.80 % (± 5.46), while the mean similarity score of men 92.80 % (± 4.60). However, to assess the possible presence of a statistically significant gender bias, after making sure that both populations qualify as normally distributed, we ran a t-Test for two independent means.

Also in this case, no significant difference was found between the two populations, since the computed t-value is 1.02785, which is lower than 2.0555, the critical value for 26 degrees of freedom and 10% level of significance (5% in each tail).

The findings of the last two sections seem to rule out any gender or English language proficiency bias in the collected survey answers, which have proven highly consistent with CroaTPAS's annotation, thus supporting its overall goodness.

In this Chapter, we have thoroughly introduced the CroaTPAS resource, explained its corpus-based annotation methodology (§ 4.2.2), online interface (§ 4.2.4), and evaluation strategy (§ 4.3), thus allowing us to proceed to Chapter 5, where CroaTPAS's data will be used to analyze the meaning inventories of a selection of symmetric and asymmetric Croatian verb pairs, whose *Aktionsarten* will be modelled using the event structures formalism.

CHAPTER 5: THE EVENT STRUCTURES OF CROATIAN VERB PAIRS

As already thoroughly explained in Chapter 2, Croatian aspectual verb pairs (see section § 2.3.1) allow to express "the different ways of viewing the internal temporal constituency of a situation" (Comrie, 1976, p. 3) by choosing between a perfective and an imperfective lexical variant (Babić, 2002, pp. 516–531; Barić et al., 2005; Browne, 1993, pp. 331–332; Kordić, 1997, p. 41; Matovac, 2022, pp. 225–231; Silić & Pranjković, 2007). However, not all the verb senses of one variant are always present in the meaning inventory of its aspectual counterpart. Polančec (2022, pp. 127–129) calls these aspectual pairs *asymmetrical* (see section § 5.2) and believes their asymmetry to be caused by the *Aktionsart* of the involved meanings (Kolaković, 2018, p. 95; Polančec, 2020, p. 160). In this Chapter, we will carry out a comparison between the meaning inventories of a selection of symmetric and asymmetric CroaTPAS verb pairs, whose *Aktionsarten* will be modelled using a customized version of the event structures formalism (see section § 5.4) to investigate the interaction between lexical and grammatical aspect in Croatian.

5.1 REPRESENTING EVENTS

Aktionsarten, also called event types (see section § 2.2.2), can be represented in many ways. Here we will provide a short overview of some relevant decomposition paradigms, starting with the one proposed by Van Valin & LaPolla (1997, p. 93), which relies on the three binary features [±static], [±telic], [±punctual] to represent the four Vendler (1957) verb classes: *states*, *activities*, *accomplishments* and *achievements* (see Table 1).

Verb class	**Feature decomposition representations**
State	[+static], [-telic], [-punctual]
Activity	[-static], [-telic], [-punctual]
Accomplishment	[-static], [+telic], [-punctual]
Achievement	[-static], [+telic], [+punctual]

Table 1 – Feature decomposition of Vendler's verb classes from Van Valin & LaPolla (1997, p. 93) with modifications.

Another decomposition approach is that of semantic primitives. Following Dowty (1979), Van Valin & LaPolla (1997, p. 102) put forth their *logical structures*, i.e., decomposed lexical representations of sentences relying on constants (e.g., *do*), modifiers (e.g., BECOME,) and variables. Table 2 provides an example sentence for each of Vendler's classes and their respective logical structures.

Verb class	Example	Logical Structure
State	Sarah is beautiful.	be' (Sarah, [beautiful'])
Activity	Mark is singing.	do' (Mark, [sing' (Mark)])
Accomplishment	Ava baked a cake.	BECOME baked' (Ava, cake)
Achievement	The window broke.	INGR broken' (window)

Table 2 – Logical structures representing Vendler's verb classes from Van Valin & LaPolla (1997, p. 102) modified.

As you can see, while achievements require the modifier INGR, which stands for *ingressive* and encodes instantaneous change, accomplishments need the BECOME modifier, which encodes gradual change over a certain time span. For a list of all the abbreviations, see Appendix 2.

Similarly, Rappaport Hovav & Levin (1998, p. 107) came up with *lexical semantic templates*, also called *event structure templates*. In their framework, predicate decomposition requires two major components: *constants* specify what is idiosyncratic to a particular verb, i.e., its core meaning (e.g., STATE), while *primitives* provide the structural aspect of verb meaning (e.g., CAUSE, BECOME). Table 3 provides their four basic event structure templates based on Vendler's verb classes.

Verb class	Event structure template
State	[x <STATE>]
Activity	[x ACT <MANNER>]
Accomplishment	[[x ACT<MANNER>] CAUSE [BECOME [y <STATE>]]]
Achievement	[BECOME [y <STATE>]]]

Table 3 – Event structure templates of Vendler's classes Rappaport Hovav & Levin (1998, p. 108) with modifications.

Next to these decomposition models, scholars have also tried to model Vendler's verb classes using graphical formalisms focused on portraying the internal make-up of events, as in the case of Moens & Steedman's *event nucleus* (1988) and Pustejovsky's *event structure* (1991) (see section § 4.2).

According to Moens & Steedman (1988), an event nucleus entails a preparatory process, a culmination point, and a consequent state (Figure 1). Activities (like *climbing*) refer to the preparatory phase only, states (like *being at the top*) only to the consequent phase, while achievements (like *reaching the top*) to the culmination point and to the consequent phase. Finally, accomplishments (like *climbing a mountain*) refer to all three parts of the event nucleus.

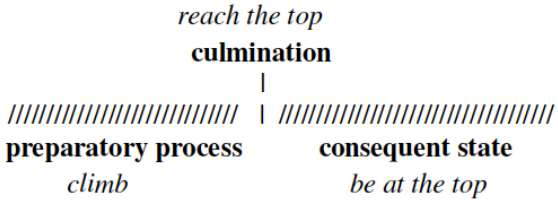


Figure 1 – An example of Moens & Steedman's event nucleus and its sub phases.

5.1.1 EVENT STRUCTURES

First introduced by Pustejovsky (1991), event structures represent a well-attested formalism aimed at modelling the *Aktionsart* of predicates by decomposing the events they denote into smaller subevents. In line with Vendler's classification, Pustejovsky divides events into *states* (S), *processes*[61] (P), and *transitions* (T), further subdivided into *achievements*, and *accomplishments*.

From a structural point of view, *states* are the only event type consisting of a single static, atelic, and lasting event *e* (Figure 2, to the left), while *processes* consist of a dynamic series of identical and subsequent subevents (Figure 2, to the right).

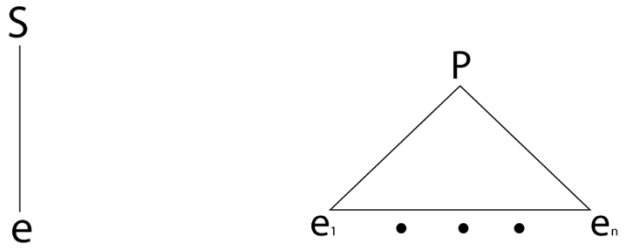


Figure 2 – Event structures for *states* (S) and *processes* (P) (Pustejovsky, 1991, p. 40) with modifications.

Language-specific tests have been developed to identify the *Aktionsart* encoded in predicates for several languages, including Croatian, but – to our knowledge – event structures have never been used to provide a graphical representation of the phenomenon for this language.

For instance, in Croatian, *states* are tested using the question *Što se događa?* 'What is happening?', which should elicit a negative answer, because nothing is happening if the situation is stationary (Polančec, 2015, p. 79). *Activities*, on the other hand, can be tested using the adverbial *za jedan sat* 'in an hour', which they do not allow since they entail no internal completion point (Polančec, 2015, p. 82).

The highlighted verbs in sentences (1.a) and (1.b) have been tested for their *Aktionsart* by Polančec (2015, pp. 81–82) and can thus be modelled using the event structures from Figure 2: the event of *hoping* from (1.a) can be modelled as a *state*, while that of *being involved* in a festival from (1.b) can be modelled as a *process*. Both are atelic and lasting, but the second is inherently dynamic.

(1) a. ***Nadamo se*** *pozitivnoj odluci Vijeća.*
'We hope for a positive decision from the Council.'

b. *Festivalima* ***se bave*** *samo kulturnjaci.*
'Only cultural people are involved in festivals.'

[61] What Vendler (1957) called *activities*.

On the other hand, *transitions* require the presence of two subsequent events (Figure 3). According to Pustejovsky , when it's a *process* that brings about the change, we are dealing with an *accomplishment*, whereas *achievements* require a change of state from S_1 to S_2.

Figure 3 – Event structures for *accomplishments* and *achievements* (Pustejovsky, 1991, p. 40) with modifications.

In terms of Croatian-specific tests for *Aktionsart*, both *accomplishments* and *achievements* pass the test of telicity (Polančec, 2015, pp. 82–83), but while the former accept adverbials with a longer time span (e.g., *za jedan sat* 'in an hour'), the latter require shorter ones (e.g., *u minutu* 'in a minute').

This is the case of sentences (2.a) and (2.b), respectively, which were also taken from Polančec (2015, pp. 82–83). The event in (2.a) can be modelled as an *accomplishment* (Figure 3 to the left) since, in order to learn something, a certain amount of time (in this case, one hour) must be devoted to the task. On the other hand, the event expressed in (2.b) encodes a sudden change of state, which can be modelled as an *achievement* (Figure 3 to the right).

(2) a. ***Naučio je*** *sve u jedan sat.*
'He learnt everything in an hour.'

b. *A sad* ***će*** *me ovaj led* ***baciti*** *nekoliko godina unatrag.*
'But now this ice will set me back several years.'

Let us now focus for a moment on the fact that, as we have learnt in Chapter 2 (see § 2.3.1), Croatian verbs actually come in aspectual pairs. Indeed, as Polančec points out, most of the verbs that pass the accomplishment test are perfective verbs whose imperfective counterparts qualify as processes (2015, pp. 82–83). Now, could the link between aspectual variants be also represented using the formalism of event structures?

We will answer this question in the section devoted to the Analysis (see § 5.4), taking into account both symmetrical and *asymmetrical aspectual pairs* (Polančec, 2022, pp. 127–129), i.e., pairs whose variants do not share the same meaning inventory. More will be said on this topic in the following paragraph.

5.2 Asymmetrical aspectual verb pairs

According to Croatian literature, the members of an aspectual pair tend to be related to each other via prefixation, suffixation or suppletivity (Isačenko, 1968; Silić, 1978). For what concerns prefixation, prefixes are believed to add a new meaning component to that of the base verb (Isačenko, 1968, p. 363), while aspectual pairs related via other derivational strategies should display the same meaning inventory (Łazorczyk, 2010, p. 105; Silić, 1978, p. 46).

However, sometimes this does not happen, as in the case of the pair *odgovarati/odgovoriti* 'answer' (Cvikić & Jelaska, 2007, p. 198; Kolaković, 2020, p. 122). Despite not being derived from one another via prefixation and sharing several meanings, these verbs do not share the meaning "to suit one's needs" represented in sentence (3), which can be only expressed by the imperfective variant *odgovarati*.

(3) *Termin mi ne **odgovara**.*
'The appointment does not suit me.'

Polančec observes this phenomenon in several verb pairs containing what he calls *osamostaljeni izvedeni nesvršeni glagoli* 'orphan secondary imperfectives' – i.e., "verbs that have the morphological structure of a derived imperfective verb, but lack their perfective partner in one or all of their meaning (Polančec, 2018, p. 114)"[62] – and provides a new label for these pairs namely *asimetrični vidski parovi* 'asymmetrical aspectual pairs'.

But why do these asymmetries take place? According to several scholars, it could be due to the *Aktionsart* of the involved verb meanings, which might not allow for a different viewing of the internal temporal constituency of the event by using a different aspectual variant.

Indeed, Kolaković states that "the possibility of realising an aspectual opposition, namely an aspectual partner, depends also on the lexico-actional functions [of the verb meaning]"[63] (2018, p. 95), while Polančec points out that:

> «The main reason for the importance of grammatical aspect in aspect languages lies in the fact that […] grammatical aspect is an obligatory, paradigmatic category. This entails that, in many contexts where a verb occurs, grammatical aspect marking is present, and actionality is bound to be perceived via aspect morphology» (2020, p. 160).

[62] "[Glagoli] koji imaju morfološku strukturu izvedenog nesvršenog glagola, ali nemaju svršeni parnjak bilo u jednom od svojih značenja ili pak u svim svojim značenjima".

[63] "O leksičko-akcionalnim funkcijama ovisi i mogućnost ostvarivanja aspektne opreke, odnosno aspektnoga partnera".

In other words, since multiple verb meanings can be associated with multiple event types (Breu, 1994; Rappaport Hovav & Levin, 1998), the *Aktionsart* of those meanings that do not appear in the sense inventory of their aspectual counterpart (henceforth called *asymmetrical meanings*) poses a constraint on the realisation of the aspectual counterpart itself.

To provide further evidence of the intertwined relationship between aspect and *Aktionsart* in Croatian verb pairs, a selection of examples featuring both symmetric and asymmetric verb meanings will be analysed in the following sections and their *Aktionsart* modelled using the formalism of event structures.

By modelling asymmetrical senses, we hope to find recurring traits in their *Aktionsarten* to justify their asymmetry, while symmetrical senses will be modelled to provide a graphical portrayal of the relationship between imperfective and perfective variants.

5.3 METHODOLOGY

Before modelling the *Aktionsarten* of symmetrical and asymmetrical Croatian verb meanings using event structures, said meanings had to be identified and a selection of appropriate sentences portraying them had to be retrieved. Therefore, we followed the following procedure:

- Asymmetrical verb pairs were identified within the CroaTPAS resource (Chapter 3) (§ 5.3.1).
- Corpus examples featuring both the asymmetrical and symmetrical meanings of the selected verb pairs were extracted from the Croatian Web as Corpus and validated by Croatian native speakers (§ 5.3.2).
- *Ad-hoc* examples where asymmetrical meanings were used in the syntactic context of their non-retrieved aspectual counterparts were created and evaluated by native speakers to confirm the asymmetries (§ 5.3.3).

5.3.1 VERB SELECTION

CroaTPAS (Marini, 2022a; Marini & Ježek, 2019) contains a total of 17 asymmetrical verb pairings. However, not all verb pairings in CroaTPAS are aspectual verb pairs (Table 4).

N.	**CroaTPAS verb pairings**	**Italian equivalent**
1	*brojiti/izbrojiti*	*contare* 'to count'
2	*dirati/dirnuti*	*toccare* 'to touch'
3	*glasati/izglasati*	*votare* 'to vote'
4	*gostiti/ugostiti*	*ospitare* 'to host'
5	*jesti/pojesti*	*mangiare* 'to eat'
6	*odgovarati/odgovoriti*	*corrispondere* 'to correspond'
7	*raditi/uraditi*	*funzionare* 'to work'
8	*silaziti/sići*	*scendere* 'to descend'
9	*štovati/poštovati*	*rispettare* 'to respect'
10	*tumačiti/protumačiti*	*interpretare* 'to interpret'
11	*tužiti/optužiti*	*accusare* 'to accuse'
12	*uputiti/upućivati*	*avviare* 'to start'
13	*učiti/naučiti*	*imparare* 'to learn'
14	*voditi/provoditi*	*guidare* 'to lead'
15	*vući/izvući*	*trarre* 'to draw'
16	*zamisliti/zamišljati*	*immaginare* 'to imagine'
17	*zvati/pozvati*	*chiamare* 'to call'

Table 4 – Asymmetrical verb pairings in CroaTPAS.

Indeed, the choice of Croatian verbs to include in the resource was made following an Italian-driven procedure (see section § 3.3.1) using the bilingual dictionary by Špikić (2017). Since Špikić usually offered multiple translation options, some of the Croatian verb pairings have been selected to evaluate the semantic impact of prefixes on base verbs (e.g., *vući/izvući*) rather than to create aspectual oppositions (e.g., *izvući/izvlačiti*).

However, since in this Chapter we are only interested in actual aspectual verb pairs – and specifically in asymmetrical ones not derivationally related by prefixation – the only CroaTPAS verbs that will be analysed are those from Table 5, whose English translations have now been adapted to fit their main Croatian meaning.

N.	**CroaTPAS aspectual verb pairs**	**English translation**
1	*dirati/dirnuti*	'to touch'
2	*odgovarati/odgovoriti*	'to answer'
3	*silaziti/sići*	'to descend'
4	*uputiti/upućivati*	'to send'
5	*zamisliti/zamišljati*	'to imagine'

Table 5 – Asymmetrical aspectual verb pairs in CroaTPAS.

5.3.2 GDEX EXAMPLES EXTRACTION

To conduct our analysis, a selection of representative corpus examples of all the attested symmetrical and asymmetrical meanings of the verbs from Table 5 were extracted from the Croatian Web as Corpus (Ljubešić & Klubička, 2014) and validated by Croatian native speakers.

Indeed, even though the evaluation of the CroaTPAS resource has already been successfully carried out (see § 4.3), Croatian informants were called upon to confirm the asymmetry of the verb pairs (see section § 5.3.3), as well as the goodness of the corpus examples representing them.

To provide respondents with an array of well-fitted examples, only corpus citations scoring at least 0.8 on the GDEX algorithm (see section § 4.2.4.1) were extracted. Table 6 provides an overview of all the validated examples, including their aspectual value (column 2), GDEX score (column 4) and whether they are symmetric (S) or asymmetric (A) (column 5).

Meaning	**Aspect**	**Example**	**GDEX**	**S/A**
dirati_1	*imperf.*	*Ne diraj me, ne **diram** te.* 'Don't touch me, and I won't be touching you.'	*0.871*	*S*
dirnuti_1	*perf.*	*Koliko **ste** puta loptu **dirnuli** u tih dvadesetak minuta na travnjaku?* 'How many times have you touched the ball in those twenty minutes on the field?'	*0.894*	*S*

dirati_se_5	*imperf.*	***Diraju se** bez ikakvog rizika od trudnoće.* 'They masturbate without any risk of pregnancy.'	*0.846*	*A*
odgovarati_1	*imperf.*	*Ne želim tebi ovdje ništa sugerirati, samo ti **odgovaram** na pitanje.* 'I do not want to suggest anything here, I am just answering the question.'	*0.924*	*S*
odgovoriti_1	*perf.*	*Mene si citirao, pa **ću** ti i **odgovoriti.*** 'You mentioned me, so I am also going to give you an answer.'	*0.925*	*S*
odgovarati_2	*imperf.*	*Firma ne **odgovara** za neuspjele rezervacije.* 'The company does not answer for unsuccessful bookings.'	*0.838*	*A*
odgovarati_3	*imperf.*	*Ako ti to ne odgovara - ne **odgovara** ti ni posao u bankarskom sektoru.* 'If this does not suit you – then also a bank job will not suit you.'	*0.925*	*A*
odgovarati_4	*imperf.*	*Niti jedan od naših postojećih dizajna ne **odgovara** Vašim željama?* 'Not even one of our existing designs corresponds to your wishes?'	*0.925*	*A*
odgovarati_5	*imperf.*	*Glavni tajnik za svoj rad **odgovara** Predsjedniku.* 'The Secretary General answers to the President for his work.'	*0.825*	*A*
odgovarati_6	*imperf.*	*Sutra pišem biologiju i **odgovaram** filozofiju.* 'Tomorrow, I have to sit a biology exam and to be orally tested in philosophy.'	*0.837*	*A*
silaziti_1	*imperf.*	***Silazim** s busa iz Botinca.* 'I am getting off of the bus from Botinac.'	*0.830*	*S*
sići_1	*perf.*	*Obojica **su sišla** s traktora i počeli se tući.* 'The two got off the tractor and started to hit each other.'	*0.905*	*S*
silaziti_6	*imperf.*	*Smiješak joj ne **silazi** s lijepog lica.* 'The smile does not get off of her face.'	*0.893*	*A*
upućivati_1	*imperf.*	*Poseban poziv **upućujemo** bivšim članovima zbora.* 'We send a special invitation to former choir members.'	*0.95*	*S*
uputiti_1	*perf.*	*Organizator **će** svim klubovima **uputiti** pismo zahvale.* 'The organiser will send thank you letters to all clubs.'	*0.95*	*S*
upućivati_4	*imperf.*	*Taj naziv **upućuje** na japodsku tradiciju.* 'That term refers to a Japanese tradition'	*0.902*	*A*
zamišljati_1	*imperf.*	*Naravno da ponekad u mašti **zamišljam** one grube situacije.* 'Of course, sometimes in my mind I imagine awful situations.'	*0.9*	*S*
zamisliti_1	*perf.*	***Zamislite** svijet gdje moraš raditi samo što ti kažu.* 'Imagine a world in which you have to do only what you are told.'	*0.95*	*S*
zamisliti_se_2	*perf.*	*Čovjek **se zamisli** nad svakom sitnicom koja mu padne na pamet.*	*0.95*	*A*

		‘The guy becomes pensive because of each small thing that comes to his mind’		

Table 6 – Validated GDEX examples extracted from the Croatian Web as Corpus.

5.3.3 *AD-HOC* EXAMPLE CREATION

In the case of asymmetric meanings, we had to make sure that the absence of a verb’s meaning in its aspectual counterpart was not due to insufficient corpus evidence but was actually systemic.

Therefore, *ad-hoc* examples featuring syntactic contexts that should trigger the use of the non-retrieved aspectual variants were created and submitted to the evaluation of three Croatian native speakers. Table 7 contains all the *ad-hoc* examples that were created to confirm the asymmetric meanings from Table 6.

N.	**Example**	**Options**	**Syntactic cues**
I	*Nakon što **su se** _____, vodili su ljubav.* ‘After engaging in masturbation, they made love.’	**dirali**/dirnuli	*nakon što, perfective-oriented*
II	*Čim mi posao _____, izgubim interes.* ‘As soon as a job suits me, I lose interest’.	**odgovara**/odgovori	*čim, perfective-oriented*
III	*Čim dizajn **bude** _____ Vašim željama, počet ćemo graditi zgradu.* ‘As soon as the design corresponds to your wishes, we will start building.’	**odgovarao**/odgovorio	*čim, perfective-oriented*
IV	*Prije nego što **je** Glavni tajnik za svoj rad _____ Predsjedniku, odgovarao je samo Upravnom odboru.* ‘Before the Secretary General responded for his work to the President, he answered only to the Steering Committee.’	**odgovarao**/odgovorio	*prije nego što, perfective-oriented*
V	*Nakon što **sam** _____ filozofiju, išla sam kući.* ‘After being orally tested in philosophy, I went back home’	**odgovarala**/odgovorila	*nakon što, perfective*
VI	*Čim diktator **bude** _____ za svoje zločine, prihvatit ćemo Vaše uvjete.* ‘As soon as the dictator answers for his crimes, we will accept your conditions.’	**odgovarao**/odgovorio	*čim, perfective-oriented*
VII	*Čim joj osmijeh _____s lica, rasplakat će se.* ‘As soon as the smile leaves her face, she will start crying.’	**side**/silazi	*čim, perfective-oriented*
VIII	*Prije nego što **je** ime na to _____, ne znamo što je značilo.* ‘Before the meaning referred to that, we don’t know what it meant.’	**upućivalo**/uputilo	*prije nego što, perfective-oriented*
IX	*Danima se umjetnik _____ .* ‘The artist was pensive for days.’	zamislio/zamišljao	*danima, imperfective-oriented*

Table 7 – *Ad-hoc* examples used to confirm the asymmetric meanings from Table 7. Grammatical option in bold.

As you can see, when the missing aspectual variant is the perfective one, the *ad-hoc* example is created starting off with a subordinative conjunction that requires a perfective afterwards, e.g., *čim* 'as soon as', *prije nego što* 'before', or *nakon što* 'after' (Čilaš Mikulić et al., 2015, p. 59; Matovac, 2022, p. 229).

This is the case of the example (4) (example II in Table 7), which was created to ascertain the asymmetry of *odgovarati_3* 'to suit somebody' (Table 6):

(4) *Čim mi posao _____, izgubim interes.* (**odgovara**/odgovori)
'As soon as a job suits me, I lose interest.'

Respondents were presented with a fill-in-the-blank task where they had to choose between the perfective or imperfective variant of each verb, in this case *odgovori* or *odgovara*, respectively. The option that the respondents agreed on is highlighted. The fact that, in this case, the imperfective *odgovara* was still necessary, even in a perfective-oriented scenario, gives us proof of the asymmetry of the meaning.

Please, consider that, although in sentences III and VI (Table 7) the perfective auxiliar *bude* was used, the choice of the lexical verb still falls, in both cases, on the imperfective variant *odgovarati* and not *odgovoriti*.

On the other hand, when the potentially asymmetrical meaning was a perfective one, an imperfective-oriented environment was created using the adverb *danima* 'for days' (Čilaš Mikulić et al., 2015, p. 59), as in example 5 (IX from Table 8), which was created to assess the asymmetry of the reflexive meaning *zamisliti_se_2* 'to become pensive' (Table 6):

(5) *Danima se umjetnik _____*. (zamislio/zamišljao)
'The artist was pensive for days.'

As you can observe, neither one of the answers is highlighted in bold, because the respondents found the use of both the perfective and imperfective variant agrammatical. More will be said on this in the following section.

5.4 Analysis

In this section, we will analyse the interaction between aspect and *Aktionsart* in Croatian aspectual verb pairs by modelling both the symmetric (§ 5.4.1) and asymmetric verb senses (§ 5.4.2) from Table 6 using the formalism of event structures (§ 5.1.1).

We will show how event structures can mirror the lexical link between the perfective and imperfective variant in aspectual pairs, while at the same time allowing for a view of perfectivity in line with Comrie's *blob* metaphor (1976, p. 18). Moreover, event structures will also turn out to be useful for the understanding of asymmetrical verb meanings.

The five CroaTPAS verb pairs at the centre of this analysis are the ones from Table 5, i.e., *dirati/dirnuti* 'to touch', *odgovarati/odgovoriti* 'to answer', *silaziti/sići* 'to descend', *uputiti/upućivati* 'to send', and *zamisliti/zamišljati* 'to imagine'.

5.4.1 Modelling symmetric verb meanings

The following examples are taken from Table 6 and are all cases of symmetric meanings, i.e., meanings that can be expressed using both the perfective and imperfective variant of the same aspectual pair. For instance, sentences (6.a) and (6.b) feature the meaning 'touching a physical entity', which can be expressed both perfectively by *dirnuti* 'to touch' and imperfectively by *dirati* 'to touch'.

(6) a. *Ne diraj me, ne **diram** te.*
'Don't touch me, and I won't touch you.'

b. *Koliko **ste** puta loptu **dirnuli** u tih dvadesetak minuta na travnjaku?*
'How many times have you touched the ball in those twenty minutes on the field?'

In (6.a), the imperfective variant is required because the focus is not on the completed action, but rather on a future scenario where the action will be carried out. The sentence could also be translated as 'don't touch me, and there will be no touching on my part'. On the other hand, in (6.b), although there is explicit mention of a number of times the ball has been touched, each touching event is seen as completed, which is why the perfective variant was chosen.

In light of this, the best event structures to model both highlighted events are *transitions*, specifically *accomplishments* (T_{acc}), like the ones in Figure 4 below.

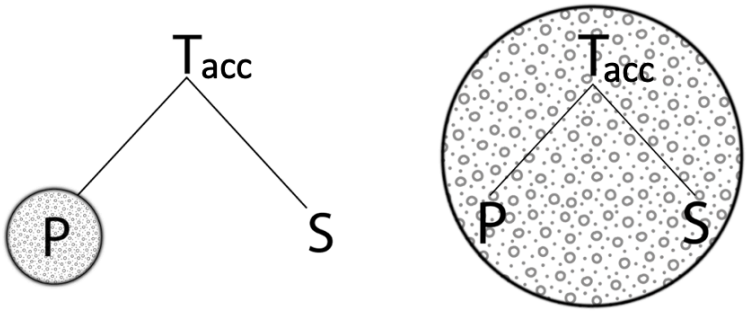


Figure 4 – Accomplishments event structures modelling symmetric meanings in Croatian aspectual pairs.

As you can see, the imperfective *touching* event in sentence (6.a) can be seen as the *process* phase of a T_{acc} event structure (Figure 4, to the left), while the structure to the right allows us to consider the *transition* in its whole including the resulting *state* of the ball being touched, as in example (6.b).

We believe T_{acc} event structures to be uniquely able to model symmetric verb senses, because the link between their branches mirrors the link between aspectual variants. As a matter of fact, when we use the imperfective variant, the focus is on the P branch, but the rest of the structure still exists – just like the possibility of using the same verb sense perfectively. On the other hand, when the perfective variant is used, the focus is on the T_{acc} structure as a whole, but we can still catch a glimpse of the P phase within the overall event.

To further implement this idea, *patterned focus circles* were added to show which event phase to focus on. This strategy is particularly helpful to represent the perfective view, since in this way the internal complexity of the event is maintained and not reduced to a single point, which is in line with Comrie's *blob* metaphor (1976, p. 18).

The next aspectual pair we will focus on is *odgovarati/odgovoriti*, specifically its symmetric meaning of 'providing an answer', portrayed by corpus examples (7.a) and (7.b):

(7) a. *Ne želim tebi ovdje ništa sugerirati, samo ti* ***odgovaram*** *na pitanje.*
'I do not want to suggest anything here, I am just answering the question.'

b. *Mene si citirao, pa* ***ću*** *ti i* ***odgovoriti.***
'You mentioned me, so I am also going to give you an answer.'

Also in this case, the events can be modelled using a T_{acc} event structure (Figure 4), where we can identify a *process* phase (the answering) resulting in a new *state* (the existence of an answer). Please, take notice of the choice of verb tenses used in the English translations: a progressive form to render the imperfective view, and the actual lexical presence of an end state (*the answer*) to provide the same perspective given by the perfective variant.

Indeed, the structure mirrors the link between the two aspectual variants since, although the imperfective one *odgovarati* allows for a view of the event as in progress, the rest of the structure does not disappear. Answering still requires a resulting state. Conversely, in the structure to the right of figure 4, the focus is on the transition as a whole, as expressed by the perfective variant *odgovoriti.*

The same goes for examples (8.a-b), (9.a-b), and (10.a-b):

(8) a. ***Silazim*** *s busa iz Botinca.*
'I am getting off of the bus from Botinac.'

b. *Obojica* ***su sišla*** *s traktora i počeli se tući.*
'The two got off the tractor and started to hit each other.'

(9) a. *Poseban poziv* ***upućujemo*** *bivšim članovima zbora.*
'We are sending a special invitation to former choir members.'

b. *Organizator* ***će*** *svim klubovima* ***uputiti*** *pismo zahvale.*
'The organiser will send thank you letters to all clubs.'

(10) a. *Naravno da ponekad u mašti* ***zamišljam*** *one grube situacije.*
'Of course, sometimes in my mind I imagine awful situations.'

b. ***Zamislite*** *svijet gdje moraš raditi samo što ti kažu.*
'Imagine a world in which you have to do only what you are told.'

In (8.a), (9.a), and (10.a), we are presented with imperfective views of the events of 'getting off of a bus', 'sending an invitation', and 'imagining something. However, all these meanings are symmetric, and can also be expressed using perfective aspectual variants, as in (8.b), (9.b), and (10.b).

All the events expressed by the imperfective variants can be associated with actions that are either portrayed as taking place at the moment of speaking, as in (8.a) and (9.a), or as habitual, as in (10.a). In all these cases, the events can be modelled as the *process* phase of a T_{acc} event structure (Figure 4, to the left).

On the other hand, all the events expressed by the perfective variants are presented as a whole – despite the change in verbal tenses: in (8.b) we have a past tense (*su sišla,* '[they] got off'), in (9.b) a future (*će uputiti*, '[he] will send'), and in (10.b) an imperative (*zamislite*, 'imagine!'). In all these cases, the events are considered globally and can thus be modelled as the T_{acc} event structure to the right of Figure 4, where the focus circle helps visualising them as a blob.

Before turning to the asymmetric meanings, let us consider that the fact that symmetric meanings can be modelled as the different phases of a T_{acc} event structure supports Polančec (2015, pp. 82–83),

who states that most of the Croatian verbs that pass the *accomplishment* test are perfective verbs belonging to aspectual pairs whose imperfective counterparts behave as *activities*.

5.4.2 MODELLING ASYMMETRIC VERB MEANINGS

In the following section, we will encounter only asymmetric verb meanings – i.e., meanings that are absent from the inventory of one of the aspectual variants making up a verb pair. To make sure they were indeed asymmetric, *ad-hoc* examples with syntactic contexts that require the use of the non-retrieved aspectual variant were created and validated by native speakers (see Table 7).

The *Aktionsart* of these validated asymmetrical meanings will now be modelled using the formalism of event structures to provide a graphical representation that may help us better understand their asymmetry.

The first meaning we will focus on belongs to the imperfective verb *odgovarati* and goes missing in its perfective counterpart *odgovoriti*. It is the meaning of 'suiting somebody' that we find in the corpus-extracted sentence (11.a):

(11) a. *Ako ti to ne odgovara – ne **odgovara** ti ni posao u bankarskom sektoru.*
'If this does not suit you – then also a bank job will not [suit you].'

b. **Čim mi nešto **odgovori**, izgubim interes.*
'As soon as something suits me, I lose interest.'

c. *Čim mi nešto **odgovara**, izgubim interes.*
'As soon as something suits me, I lose interest.'

As shown in (11.b), the perfective variant *odgovoriti* cannot be used to express this meaning despite the presence of the conjunction *čim* 'as soon as', which typically favors a perfective view on events. Our respondents opted for the use of the variant *odgovarati* both by itself (11.c) and introduced by the phase verb *početi* 'to start' as in *Čim mi nešto **počne odgovarati**, izgubim interes*. ('As soon as something starts to suit me, I lose interest').

This last solution is particularly interesting since it provides some useful insights on the way this event is conceptualized in Croatian.

Indeed, if we assume that the meaning of *Čim mi nešto **odgovara*** 'As soon as something suits me' is the same as that of *Čim mi nešto **počne odgovarati*** 'As soon as something starts to suit me', then (11.a) can be seen as an example of *aspectual coercion* due to the presence of *čim*.

As already mentioned in Chapter 2 (section § 2.2.2), aspectual coercions are changes in the *Aktionsart* of a proposition due to modifiers like tenses, temporal adverbs and auxiliaries (Moens & Steedman, 1988, p. 17). In this case, it is the presence of *čim* 'as soon as' that coerces the *state* meaning of 'suiting somebody' into the *achievement* reading 'starting to suit somebody'.

This still leaves one open question: why wasn't the perfective aspectual variant *odgovoriti* used to express the achievement meaning, since it is at least telic? Given that the variant *odgovoriti* is normally used to express accomplishments, as in the case of (7.b), and that phase verbs such as *početi* 'to start' need to be followed by imperfectives (Čilaš Mikulić et al., 2015, p. 59), we believe that coercing the imperfective variant is simply the most economical solution from a cognitive point of view. After all, *početi* identifies the beginning of a state, which is the *Aktionsart* of *odgovarati*.

We can therefore state that the *Aktionsart* of the meaning 'to suit somebody' poses a constraint to the use of the perfective variant of the verb to express that meaning, which makes the meaning asymmetric, in line with Kolaković (2018, p. 95) and Polančec (2018, p. 114).

In terms of event structure, it possible to model the meanings expressed in (11.a) and (11.c) using an achievement (T_{ach}) event structure as in Figure 5.

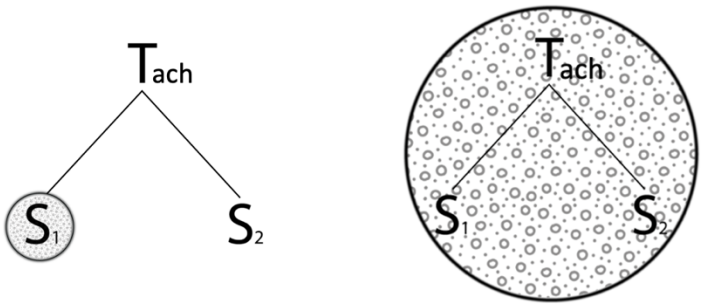


Figure 5 – Achievement event structures modelling asymmetric state meanings.

The structure to the left, focuses only on S1, which corresponds to the state expressed by *odgovarati* in (11.a), namely that state of not being suited for a bank job. On the other hand, even though there is no perfective lexical variant able to express the change of state, the structure to the right portrays example (11.c) and its equivalent featuring the periphrasis *početi* + imperfective. Indeed, as soon as something starts to suit somebody, an S1 ends and an S2 begins.

Examples (12.a-c) paint a similar picture, still relating to the pair *odgovarati/odgovoriti*. The asymmetric sense this time is that of 'corresponding to something', which can be expressed only with the imperfective variant *odgovarati*.

(12) a. *Niti jedan od naših postojećih dizajna ne **odgovara** Vašim željama?*
'Not even one of our existing designs corresponds to your wishes?'

b. **Čim dizajn **bude odgovorio** Vašim željama, početi ćemo graditi zgradu.*
'As soon as the design corresponds to your wishes, we will start building.'

c. *Čim dizajn **bude odgovarao** Vašim željama, počet ćemo graditi zgradu.*
'As soon as the design corresponds to your wishes, we will start building.'

In this case, the first part of (12.c) could be better translated in English as "as soon as the design meets your wishes", which makes it more obvious that an *achievement* reading is being coerced out of an otherwise *stative* meaning by means of the conjunction *čim* 'as soon as'.

Despite the conjunction trying to elicit the use of the perfective variant *odgovoriti* (12.b), only the perfective auxiliary *bude* aligns with it, but not the verb on a lexical level.

Why can't the perfective aspectual variant *odgovoriti* be used to express the change of state in (12.c)? We believe there are several reasons: on the one hand, *odgovoriti* is normally used to express accomplishments, as in (7.b); on the other, the presence of the perfective auxiliary *bude* is enough to convey telicity; finally, also in this case, an achievement marks the beginning of a new state, which is the *Aktionsart* of *odgovarati*. Therefore, to put it in Kolaković's words, it is once again the actional properties of the verb meaning that are preventing the realisation of its aspectual counterpart (2018, p. 95).

Regardless of the asymmetry, the event structures from Figure 5 prove adequate to frame both the state of not corresponding to one's wishes from (12.a) (event structure to the left) and the change of state of starting to expressed in sentence (12.c) (event structure to the right).

The same goes for examples (13.a-c), (14.a-c), as well as (15.a-c). In all these cases, the *Aktionsart* of the asymmetric meanings is that of *state* and the subordinative conjunctions *čim* 'as soon as' and *prije nego što* 'before' coerce an *achievement* reading in an imperfective variant without triggering the use of a perfective counterpart because of the actional properties of the meanings involved.

(13) a. *Glavni tajnik za svoj rad **odgovara** Predsjedniku.*
'The Secretary General answers to the President for his work.'

b. **Prije nego što **je** Glavni tajnik za svoj rad **odgovorio** Predsjedniku, odgovarao je samo Upravnom odboru.*
'Before the Secretary General answered to the President for his work, he answered only to the Steering Committee.'

c. *Prije nego što **je** Glavni tajnik za svoj rad **odgovarao** Predsjedniku, odgovarao je samo Upravnom odboru.*
'Before the Secretary General answered to the President for his work, he answered only to the Steering Committee.'

(14) a. *Firma ne **odgovara** za neuspjele rezervacije.*
'The company does not answer for unsuccessful bookings.'
b. **Čim diktator **bude odgovorio** za svoje zločine, prihvatit ćemo Vaše uvjete.*
'As soon as the dictator answers for his crimes, we will accept your conditions.'
c. *Čim diktator **bude odgovarao** za svoje zločine, prihvatit ćemo Vaše uvjete.*
'As soon as the dictator answers for his crimes, we will accept your conditions.'

(15) a. *Taj naziv **upućuje** na japodsku tradiciju.*
'That term refers to a Japanese tradition.'
b. **Prije nego što **je** ime na to **uputilo**, ne znamo što je značilo.*
'Before the meaning referred to that, we don't know what it meant.'
c. *Prije nego što **je** ime na to **upućivalo**, ne znamo što je značilo.*
'Before the meaning referred to that, we don't know what it meant.'

On the other hand, something quite different happens in examples (16.a-c):

(16) a. *Sutra pišem biologiju i **odgovaram** filozofiju.*
'Tomorrow, I have to sit a biology exam and to be orally tested in philosophy.'
b. **Nakon što **sam odgovorila** filozofiju, išla sam kući.*
'After being orally tested in philosophy, I went back home.
c. *Nakon što **sam odgovarala** filozofiju, išla sam kući.*
'After being orally tested in philosophy, I went back home.

In (16.b), the meaning that goes amiss in the perfective variant *odgovoriti* is that of 'being orally tested, undergoing an oral test', which is present in the sense inventory of the imperfective variant *odgovarati*. What justifies this asymmetry is the fact that a perfective view on the event would somehow undermine the main actional feature of the meaning itself, which is that of processuality. Being tested is an event which has a certain duration and dynamicity (not all its moments are the same), but who is being tested has no control over the end of the activity. Indeed, we can even test whether this meaning qualifies as a *process* (or *activity*) using a durativity test:

(17) a. *Odgovarala sam filozofiju **sat vremena**.*
'I have been orally tested in philosophy for an hour.'

This time the actional constraint which does not allow for the use of the perfective variant *odgovoriti* is not the *state* character of the event, but its being a *process*, which is being coerced into an *accomplishment* reading only by means of the subordinative conjunction *nakon što* 'after' (16.c).

The event structures in Figure 6 can mirror both views, even though they do not correspond to two related aspectual variants. The structure to the left focuses on the process phase of being tested exemplified by (16.a), while the structure to the right focuses on having been tested, that is to the say on the accomplished deed in its entirety.

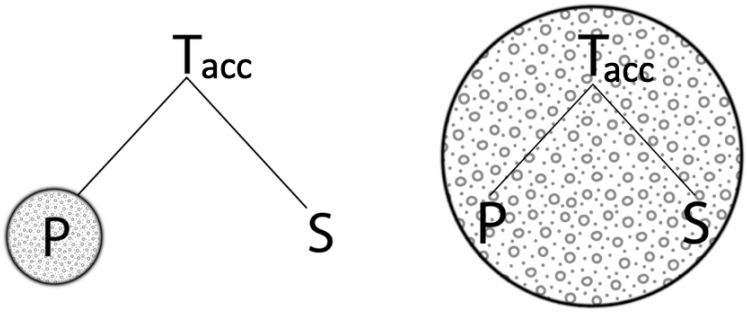


Figure 6 – Accomplishment event structures modelling asymmetric process meanings.

The same goes for the meaning exemplified in (18.a) and (18.c), namely that of 'engaging in masturbation'. Also in this case, we find ourselves facing an asymmetric meaning (a reflexive one this time), which cannot be expressed by the perfective variant *dirnuti se*, but only by the imperfective one *dirati se*.

(18) a. ***Diraju se*** *bez ikakvog rizika od trudnoće.*
'They masturbate without any risk of pregnancy.'

b. *Nakon što* ***su se dirnuli****, vodili su ljubav.*
'After having touched each other, they made love.'

c. *Nakon što* ***su se dirali****, vodili su ljubav.*
'After engaging in masturbation, they made love.'

As a matter of fact, if we used *dirnuti se* as in (18.b), the sentence would not be agrammatical, but the sexual meaning would be lost and the new sentence would read something like 'After having touched each other (with their hands), they made love'. In other words, the durativity of the meaning needs to be underlined by the use of the imperfective variant also in an *accomplishment* context, like the one coerced by the conjunction *nakon što* 'after'.

In terms of event structures, the best structure to model (18.a) and (18.c) are the same as is the same as (16.a) and (16.c) portrayed in Figure 6.

Let us now turn to the only asymmetric reflexive meaning *zamisliti se* 'becoming pensive', which lacks an imperfective counterpart. This prompted us to create an imperfective-oriented context using the adverb *danima* 'for days', to try to elicit the use of an imperfective form (Čilaš Mikulić et al., 2015, p. 59), as you can see in sentences (19.b-c).

(19) a. *Čovjek **se zamisli** nad svakom sitnicom koja mu padne na pamet.*
'The guy becomes pensive over every little thing that comes to his mind.'

b. **Danima **se** umjetnik **zamišljao**.*
'The artist was pensive for days.'

c. **Danima **se** umjetnik **zamislio**.*
'The artist was pensive for days.'

d. *Danima **je** umjetnik **bio zamišljen**.*
'The artist was pensive for days.'

'Becoming pensive' is a classic example of a verb expressing a change of state (an *achievement* in Vendler's terminology), i.e., an event involving two subsequent states (S_1 and S_2): the first being a non-pensive state of mind and the second a pensive one (Figure 7, to the left).

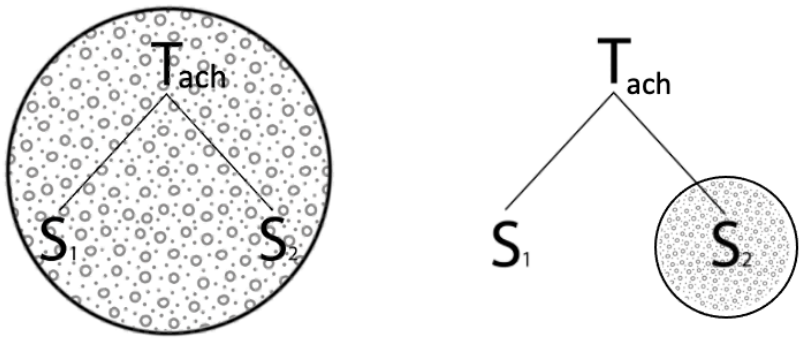


Figure 7 – Event structures modelling the meanings (19.a) and (19.d), respectively.

According to our respondents, neither the imperfective variant *zamišljati (se)* nor the perfective one *zamisliti (se)* can express the beginning state (S_1) nor the end state (S_2), as you can see from (19.b-c). However, other verbal expressions can be used, such as *biti zamišljen* 'being pensive', as you can see from example (19.d) *Danima **je** umjetnik **bio zamišljen*** 'The artist was pensive for days.' The event structure to the right of Figure 7 tries to model this situation.

Finally, the last meaning that we are going to consider is the one from sentence (20.a), namely the idiomatic meaning *smiješak nekome **ne silazi** s lica* 'somebody cannot get a smile off of their face'. In this case, respondents didn't find this meaning to be asymmetric, since it was enough to leave out

the negation from the perfective-oriented scenario to allow the meaning to be used perfectively with the variant *sići*, as you can see from (20.c).

(20) a. *Smiješak joj **ne silazi** s lijepog lica.*
'The smile does not get off of her face.'

b. **Čim joj osmijeh **silazi** s lica, rasplakat će se.*
'As soon as the smile leaves her face, she will start crying.'

c. *Čim joj osmijeh **siđe** s lica, rasplakat će se.*
'As soon as the smile leaves her face, she will start crying.'

In terms of *Aktionsarten*, the meaning in (20.a) can be modelled as a state, because "nothing is happening" if a "smile does not get off of somebody's face". On the other hand, example (20.c) expresses a transition, a change of state, from an S_1 (smile is there) to an S_2 (smile has left). The two meanings can therefore be modelled as the beginning state of an achievement (Figure 8, to the left) and as the achievement itself, considered in its entirety, as highlighted by the focus circle around the event structure on the right of Figure 8.

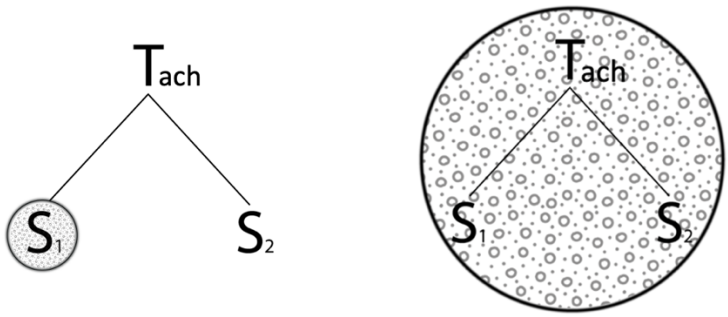


Figure 8 – Event structures modelling the symmetric meanings in examples (20.a) and (20.c) respectively.

What can be considered asymmetric is the idiomaticy of the imperfective meaning in (20.a), which is way more commonly used than the perfective one from (20.c).

After this analysis, which shed more light on the intricate relationship between aspect and *Aktionsart* in Croatian aspectual verb pairs, we now turn to the last Chapter of this Dissertation, namely Chapter 6, which is devoted to the Croatian teaching experience that the Candidate had the opportunity to carry out and where she could actually apply the CroaTPAS resource for the first time.

CHAPTER 6: TEACHING CROATIAN USING CROATPAS

In this Chapter, we will discuss the Croatian language teaching experience that the Candidate had the opportunity to carry out by organising a 30-hour A2 Croatian Language Intensive Course, which took place the University Language Centre in the second semester of the academic year 2022/23.

The experience was particularly relevant to this PhD project because it provided the opportunity to apply the CroaTPAS resource in a classroom context for the first time to introduce the topic of aspectual verb pairs.

Section 6.1 will provide an overview of the course details, including the course design and information on the participants. Section 6.2 will focus on the chosen language teaching approach and will include some of the game-based teaching activities designed for the course. Section 6.3 will focus on how the CroaTPAS resource (Chapter 4) was used to introduce the topic of verbal aspect (Chapter 2). Sections 6.4 will deal with the design of the test that was administered at the end of the course, while 6.5 will present the test results, which will be analysed following a quantitative approach in section 6.6 and a statistical one in 6.7.

6.1 COURSE OVERVIEW

In the period from February 20th until May 30th, 2023, the author of this Dissertation had the opportunity to teach a 30-hour Croatian Language Intensive Course for beginners at the Language Centre of the University of Pavia[64]. The course was part of a curricular traineeship aimed at developing language teaching skills, which allowed the Language Centre to offer it for free.

As you can see from Figure 1 below, the course took place twice a week – on Mondays and Thursdays – from 5.30 until 7.00 p.m., in order not to interfere with the academic engagements of the students. The course did not grant any ECTS points, but upon attendance of at least 70% of the lessons, a certificate of attendance was awarded and admittance to an A2 exam was granted. For those who passed the exam, an A2 certificate of language proficiency was awarded by the Language Centre.

In order to allow for a fruitful teaching and learning experience, a limit of 15 participants was set. At first, priority was given to PhD students by promoting the course via the PhD mailing list of the University. In a second moment, the call was sent to the rest of the student body, as well as to the University staff.

In Figure 1, you can see the brochure that was used to promote the course. Since the language of the course was Italian, all promotional material was written in Italian.

[64] https://cla.unipv.it/. Website last visited on September 25th, 2023.

Figure 1 – Promotional material for the Croatian Language Intensive Course.

6.1.1 COURSE DESIGN

The Course was designed aiming at an A2 level of language proficiency, which – according to the Illustrative Descriptor Scales of the *Common European Framework of Reference for Languages* (Council of Europe, 2020) – corresponds to the following abilities:

- *Overall oral comprehension*: "Can understand enough to be able to meet needs of a concrete type, provided people articulate clearly and slowly. Can understand phrases and expressions related to areas of most immediate priority (e.g., very basic personal and family information, shopping, local geography, employment), provided people articulate clearly and slowly." (Council of Europe, 2020, p. 48)
- *Overall reading comprehension*: "Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency every-day or job-related language. Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared international vocabulary items." (Council of Europe, 2020, p. 54)
- *Overall oral production*: "Can give a simple description or presentation of people, living or working conditions, daily routines. likes/ dislikes, etc. as a short series of simple phrases and sentences linked into a list." (Council of Europe, 2020, p. 62)
- *Overall written production*: "Can produce a series of simple phrases and sentences linked with simple connectors like "and", "but" and "because." (Council of Europe, 2020, p. 66)
- *Overall oral interaction*: "Can interact with reasonable ease in structured situations and short conversations, provided the other person helps if necessary. Can manage simple, routine exchanges without undue effort; can

> ask and answer questions and exchange ideas and information on familiar topics in predictable everyday situations. Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters to do with work and free time. Can handle very short social exchanges but is rarely able to understand enough to keep conversation going of their own accord." (Council of Europe, 2020, p. 72)

The above-mentioned descriptors were adapted to Croatian by Grgić and Gulešić Machata in the volume *Hrvatski A2: Opisni okvir referentne razine A2* (2017).

While the descriptors helped identifying the course goals, the following textbooks (Figure 2) were used to lay out the lesson plan and pinpoint both the grammatical topics, communicative functions, and lexical areas to be addressed:

- *Hrvatski za početnike 1: Udžbenik i rječnik (A1 i A2)* (Čilaš-Mikulić et al., 2006)
- *Razgovarajte s nama! Udžbenik hrvatskoga jezika za razine A1 i A2* (Čilaš Mikulić et al., 2021)
- *Corso di lingua croata: Livelli A1-B1 del Quadro Comune Europeo di Riferimento per le Lingue* (Gott & Morpurgo, 2020)

Figure 2 – The three textbooks used during the course.

The communicative functions (and lexical areas) that were included in the course syllabus were:

- Introducing oneself (greeting forms, countries, and nationalities)
- Counting entities (numbers)
- Talking about family (family, relations, jobs)
- Describing entities (objects in the room, animals, colours)
- Describing oneself (human qualities and characteristics)

- Ordering at the restaurant (food and drinks)
- Talking about hobbies and everyday life (days of the week, time of the day)

On the other hand, in line with (Udier, 2017), the grammatical topics covered by the course were:

- The alphabet
- Present, past, and future forms of the verb *biti* 'to be' and of the verb *imati* 'to have'
- Nominative forms of the three main inflectional noun classes (see section §1.3.1)
- Irregular nominative plural forms and corresponding sound changes (see section §1.3.1)
- Accusative forms of the three main inflectional noun classes (see section §1.3.2)
- Numbers from 1 to 100 and genitive forms (see section §1.3.3)
- Locative forms of the three main inflectional noun classes (see section §1.3.5)
- Irregular locative forms and corresponding sound changes (see section §1.3.5)
- The locative/accusative alternation to express static locations or goals of motion (see sections §1.3.2 and §1.3.5)
- The alternation between the prepositions *na* 'on' and *u* 'in' (see sections §1.3.2 and §1.3.5)
- Possessive adjectives
- Demonstrative adjectives and pronouns
- Nominative and accusative forms of the personal pronouns
- Present, past, and future forms of the following verb groups: *-ati* > *-am* (e.g., *slušati* 'to listen', *parkirati* 'to park'), *-iti, -jeti* > *-im* (e.g., *kupiti* 'to buy', *voljeti* 'to love'), *-ovati, -evati, -ivati* > *-ujem* (e.g., *stanovati* 'to live'), modal verbs (*morati* 'to must', *trebati* 'to need', *moći* 'to can', *htjeti* 'to want', *smjeti* 'to be able to'), *ići* 'to go', *jesti* 'to eat', and *piti* 'to drink'.
- Past and future forms of *reći* 'to say', *moći* 'to can' and similar verbs.

6.1.2 THE PARTICIPANTS

Of the initial 12 participants, 9 attended the lessons until the end, and 8 were admitted to the final exam. This should be regarded as a high attendance rate given that the Croatian Intensive Course was not a curricular activity and did not provide any ECTS credits.

In terms of education level, 5 students were attending a PhD course, 4 an MA programme, and 3 a BA. In the course of the first few weeks, one BA and 2 MA students dropped out, whereas all 5 PhD students stayed until the end (although one decided not to take the final exam), which proves they are indeed the right audience for such a course.

As for their fields of study, 4 students studied Political Sciences, 3 Geology, 3 Linguistics, 1 Economics, and 1 Engineering. The ones that dropped out studied Political Sciences, Geology and Engineering, which suggests a higher interest in the initiative from the Humanities area.

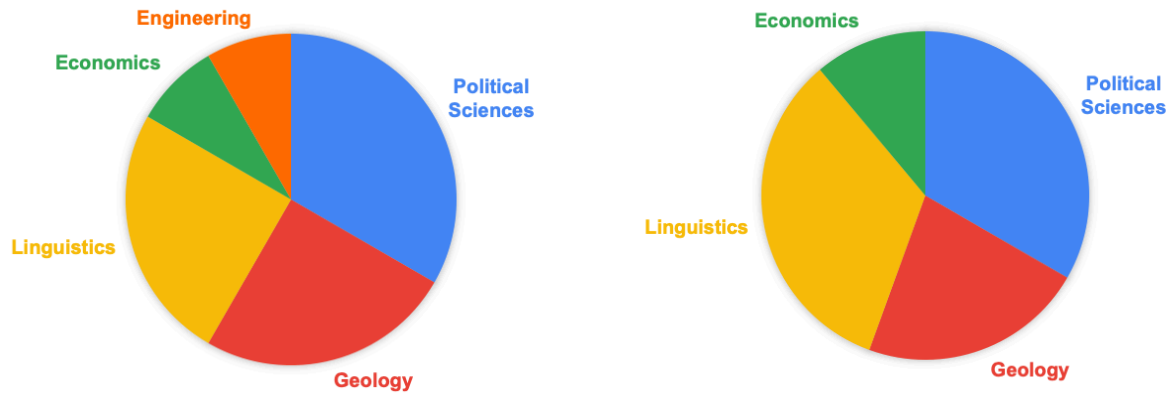


Figure 3 – Academic fields of study of the initial 12 vs. final 9 students of the Croatian Language Intensive Course

All the students attending the course were monolingual Italian native speakers, with the exception of an Italian-Bosnian bilingual student, who grew up in Italy from Bosnian parents and could therefore already speak Croatian, despite having no grammar knowledge.

All students but one had previous knowledge of at least two foreign languages next to Italian, usually English and either French, Spanish, Russian or German. Apart from the bilingual one, only one student had prior knowledge of B/C/S, since they were attending an MA degree in Cooperation and Development and had already lived in Sarajevo.

In terms of gender, 7 students were male and 5 female.

6.2 Language Teaching Principles

The Course was designed following a *Communicative Language Teaching* (CLT) approach. CLT is an overarching approach to language teaching, whose assumptions are based, unlike previous methods, on a cluster of Second Language Acquisition (SLA) principles and foundation stones. CLT extends beyond the grammatical elements of communication and recognizes the importance of social, cultural, and pragmatic aspects. It encourages authentic and meaningful communication, as well as the development of all four components of what Canale and Swain (1980) called *Communicative Competence* (CC):

- *grammatical component:* the ability to use the forms of the language.
- *discourse component:* the ability to understand and produce coherent flows of sentences, in speaking and in writing.
- *sociolinguistic component:* the ability to apply socio-cultural information to communication.
- *strategic component:* the ability to use both verbal and non-verbal strategies to accomplish communicative goals.

Strictly tied to CC is the concept of *communicative proficiency*, which is seen as both fluency and accuracy, and needs to be developed in all four language skills (i.e., writing, reading, speaking, and listening).

As effectively summarized by Brown and Lee (2015, p. 31), Communicative Language Teaching believes in the interdependence of form and function, thus engaging learners in the functional use of language for meaningful purpose, but without discarding the importance of forms.

Importance is also given to real-world contexts. Therefore, language learners are exposed to meaningful input embedded in naturally occurring language, which translates, e.g., into reading and listening activities revolving around original newspaper articles and songs.

In terms of teacher and student roles, CLT teachers take on the role of facilitators and become empathetic coaches in the learning process of their students, who are conceived as active participants in their own learning.

6.2.1 Game-based Language Teaching

The Croatian Language Intensive Course was aimed at beginners who had – for the most part – no prior knowledge of Croatian nor of another Slavic language that they could use to make successful associations between existing knowledge and the new concepts to be learned (*Principle of transfer*).

Therefore, since learners had "very few background schemata to rely on, that is, few existing cognitive pegs on which to hang new words and structures" (Brown & Lee, 2015, p. 139) and needed an adequate dose of repetition, the first half an hour of each lesson was devoted to revising homework and carrying out *game-based language teaching activities* to let the topics sink in.

According to Balboni (2020, p. 29):

> "[The aim] of game-based language teaching [...] is playing and, if possible, winning: it is a self-explanatory goal, there is no getting a good grade, there is no fear of getting a bad one, there is no contraposition between teacher and learner, there is no linguistic judge, but only a referee, which can also be the teacher: Krashen's *rule of forgetting*, according to which acquisition is at its best when one forgets they are learning a language, has its strongest application when it comes to game-based language teaching.[65]

Therefore, we came up with several game-based teaching activities. Hereafter, we will focus only on the following four:

- *Find the odd one out*
- *Scrambled sentences*
- *Divide into groups*
- *Open the box* (available on the *Wordwall* platform[66])

The *Find the odd one out* activity was adapted to several different scenarios, the following one is a lexical revision. Only regular feminine nouns were included in the task, therefore "the odd one out" is odd due to semantic reasons and not morphological ones and it is the word *cipela* 'shoe', which is an item of clothing and not a member of the family, like the rest of the words.

BAKA	**CIPELA**	DJEVOJKA	MAJKA	KĆERKA
'grandma'	**'shoe'**	'girl'	'mother'	'daughter'

The *Scrambled sentences* activity is useful to let students think about word order and how it changes, for instance, when we are dealing with interrogative forms, as in sentence (1.a), or which position is preferred by temporal adverbs in a non-marked scenario, as in sentence (1.b).

[65] "[Lo scopo] della glottodidattica ludica […] è giocare e, se possibile, vincere: lo scopo è autotelico, non vuole far prendere un bel voto, non incute la paura di un brutto voto, non contrappone docente/studente, non ha un giudice (linguistico) ma un arbitro del gioco, che può anche essere l'insegnante: la *rule of forgetting* di Krashen, secondo cui si acquisisce meglio quando ci si dimentica che si sta acquisendo una lingua, si applica la massimo nella didattica ludica."

[66] https://wordwall.net/it. Website last visited on September 26th, 2023.

(1) a. *tvoj pas Je dalmatiner li ?*
your dog is Dalmatian [interrogative particle]
Je li tvoj pas dalmatiner?
'Is your dog a Dalmatian?'

b. *nisam Danas dobro baš .*
I am not today good quite .
Danas nisam baš dobro.
'I am not doing great today.'

The *Divide into groups* activity was particularly engaging when carried out as a group activity where students were asked to identify two groups of words in the lists presented to them and provide a reason for the separation.

As you can see, the reasons could be semantic, as in the case of group (2.a), where the words highlighted in bold are all positive adverbs, while the rest are negative ones, or inflectional, as in the case of group (2.b), where all the words are nouns ending in *-o*, but the highlighted ones are masculine, while the rest are neuter nouns.

(2)	a.	**ODLIČNO** **'great'**	**JAKO DOBRO** **'very well'**	LOŠE 'badly'	**DOBRO** **'well'**
		NISAM BAŠ DOBRO 'I am not well'	**IZVRSNO** **'excellent'**	TAKO-TAKO 'so and so'	STRAŠNO 'terribly'
	b	SELO 'village'	**AUTO** **'car'**	BRDO 'hill'	PIVO 'beer'
		AVOKADO **'avocado'**	**TANGO** **'tango'**	JEZERO 'lake'	STABLO 'tree'

Finally, we also took advantage of online platforms such as *Wordwall*[67], where game-based language teaching activities can be freely created and shared between teachers. The activity that you can see in Figure 4 was adapted from an already existing template based on the idea of letting students open unknown boxes by selecting a number. They then have to answer the question inside.

[67] https://wordwall.net/it. Website last visited on September 26th, 2023.

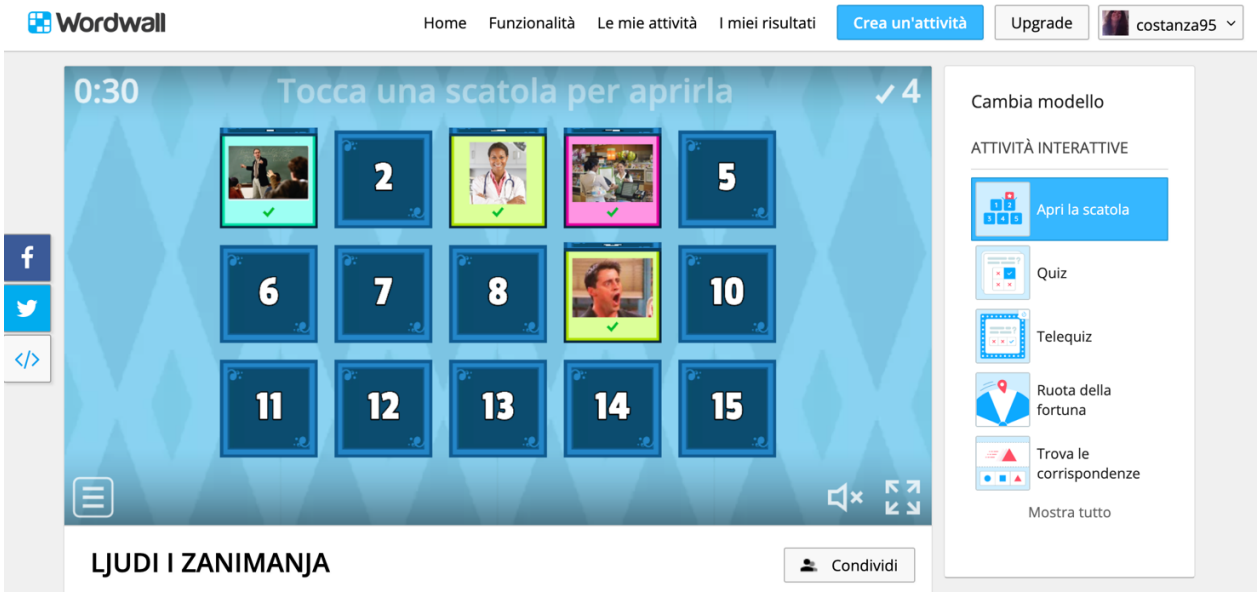


Figure 4 – An *Open-the-box* activity from the *Wordwall* platform

The focus of this specific activity was revising the lexicon of the topic *Ljudi i zanimanja* 'People and professions'. Figure 5 shows the content of Box 9, which features a photo of Joey Tribbiani, a famous character from the American TV series *Friends*, played by actor Matt LeBlanc. The choice of words was between *doktor* 'doctor' (which in Croatian is a more colloquial word for *liječnik* 'doctor') and *glumac* 'actor'.

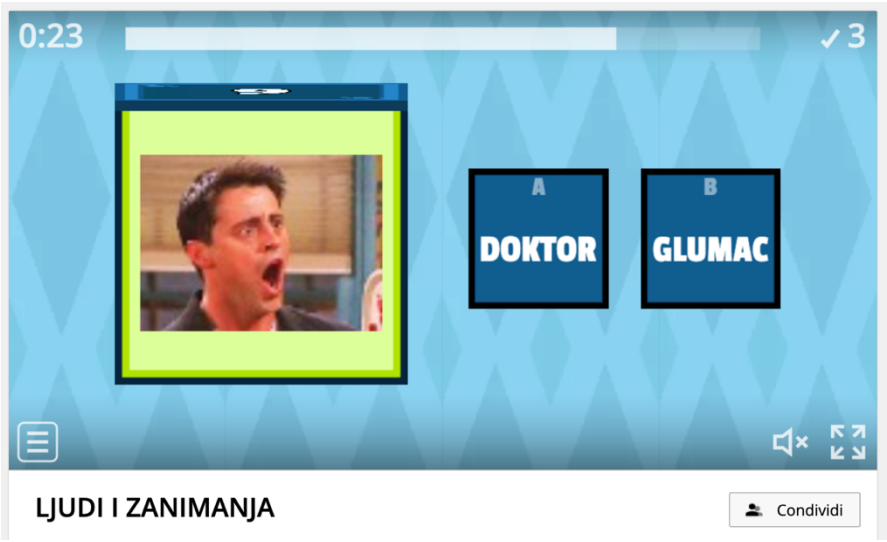


Figure 5 – Box 9 of the *Ljudi i zanimanja* 'People and professions' Wordwall activity.

6.3 TEACHING ASPECT

Last but not least, the 30-hour A2 Croatian Language Intensive Course was also an opportunity to apply the CroaTPAS resource (Chapter 4) in a language teaching environment for the first time.

It was specifically used to introduce the topic of aspectual verb pairs (see Chapter 2) and provide insightful data for classroom discussion.

As a matter of fact, according to Udier (2017, p. 140), the A2 goals in terms of verb aspect knowledge are only basic ones:

> "An A2 speaker [of Croatian] knows that there is a distinction between verbs that express a process (imperfective verbs) and those that express a result (perfective verbs) and that verbs come in aspectual pairs. He knows the most common aspectual pairs (e.g., *jesti/pojesti* 'to eat', *piti/popiti* 'to drink'). An A2 speaker does not yet know the basics of the relationship between aspect and tenses, nor all the aspectual counterparts of all the verbs he knows."[68]

In order to introduce verb aspect for the first time, three common aspectual pairs that are available in the CroaTPAS resource were selected, namely: *jesti/pojesti* 'to eat', *kupovati/kupiti* 'to buy' and *odgovarati/odgovoriti* 'to answer'.

For each aspectual verb pair, the students' attention was brought on the different meaning nuances between the corpus examples featuring the imperfective and perfective variants of the same pairs. For instance, if you look at the corpus examples of the perfective variant *pojesti* 'to eat' (Figure 6), it becomes quite clear that the emphasis is indeed on the completed event.

(3) a. *Što će se dogoditi ako ne **pojedem** špinat do kraja?*
'What will happen if I don't finish eating all the spinach?'
'Cosa succede se non finisco di mangiare tutti gli spinaci?'

b. *Za ručak **sam pojeo** pizzu i popio pola litre Cedevite.*
'For lunch, I ate a [whole] pizza I drank half a litre of Cedevita.'
'A pranzo, mi sono mangiato una pizza e mi sono bevuto mezzo litro di Cedevita.'

c. *Kupio sam previše kruha, a ne mogu ga **pojesti**.*
'I bought too much bread, and I cannot eat it all.'
'Ho comprato troppo pane e non riesco a mangiarlo tutto.'

[68] "Govornik na razini A2 […] zna da postoji distinkcija između glagola kojima se izražava proces (nesvršenim glagoli) i onih kojima se izražava rezultat (svršeni glagoli) te da glagoli imaju svoje vidske parnjake. Poznaje vidske parnjake najčešćih glagola (npr. *jesti / pojesti, piti / popiti*). Osnove odnosa glagolskoga vida i vremena govornik na razini A2 još ne poznaje, kao ni vidske parnjake svih njemu poznatih glagola."

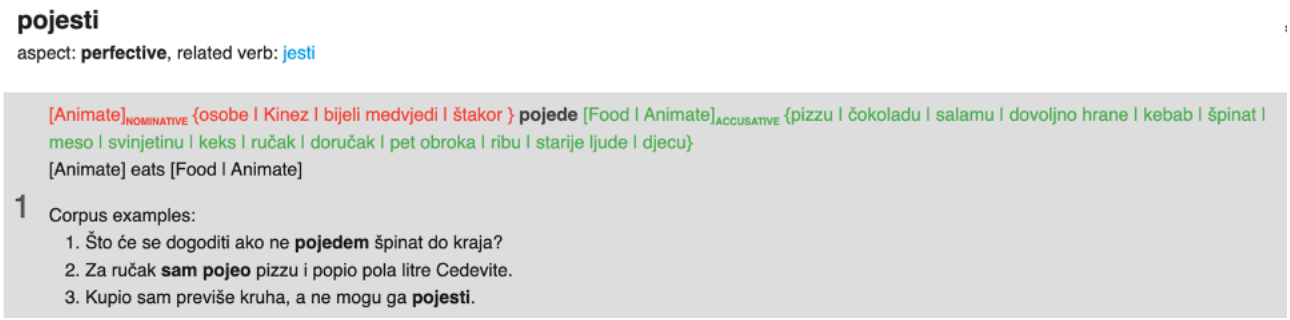


Figure 6 – CroaTPAS pattern 1 and GDEX corpus examples of the perfective verb *pojesti* 'to eat'.

Students were shown the sentences, which were translated collectively (both in Italian and English) so as to make them aware not only of Croatian aspectual markers such as the adverb *do kraja* 'until the end', but also of Italian and in English perfective translation strategies, such as the periphrasis *finire di mangiare* 'to finish eating' in (3.a); the use of the Italian benefactive *mi* 'to me' in (3.b), which conveys the idea that whoever ate the pizza, at it all to their own enjoyment; and the use of *mangiarlo tutto* 'to eat it all' in (3.c).

On the other hand, the sentences exemplifying pattern 1 of the imperfective verb *odgovarati* 'to answer' show the event as an on-going process, both in the present and in the past.

(4) a. *Ne želim ti ništa sugerirati, samo ti **odgovaram** na pitanje.*
'I do not want to suggest anything, I am just answering the question.'
'Non ti voglio suggerire niente, sto solo rispondendo alla domanda.'

b. *Kad **odgovaramo** profesoru, ne smijemo koristiti iste izraze koje koristimo s prijateljima.*
'When answering a professor, we cannot use the same expressions we would use with friends.'
'Quando stiamo rispondendo a un professore, non possiamo usare le stesse espressioni che utilizziamo con gli amici.'

c. *Dječak **nije odgovarao** na pozive roditelja.*
'The boy was not answering to his parents' calls.'
'Il ragazzo non rispondeva alle chiamate dei genitori.'

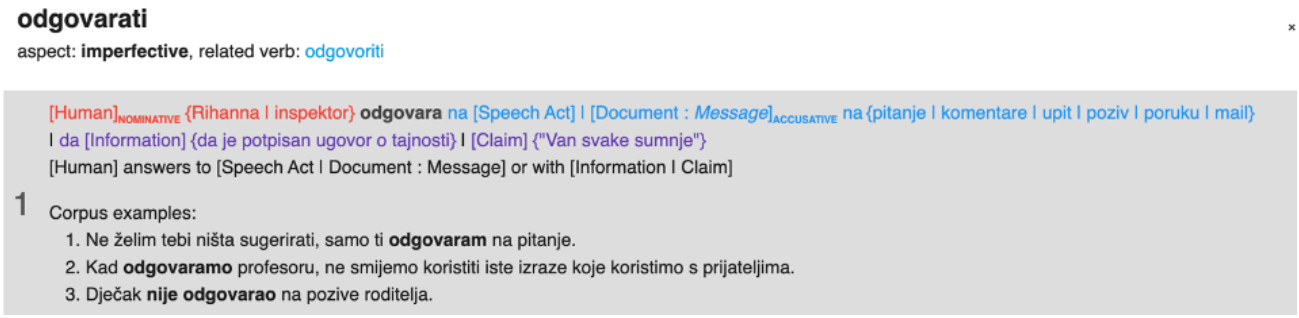


Figure 7 – CroaTPAS pattern 1 and GDEX corpus examples of the verb pair *jesti/pojesti* 'to eat'.

In (4.a) and (4.b), the focus of the students was brought to the fact that the same meaning, to be fully expressed in English or Italian, requires the use of a progressive periphrasis, such as the *-ing* form 'answering' in English or the *stare* 'to stay' + gerund construction in Italian. The last sentence (4.c) provides a similar example, but in the past, thus showing that all verb tenses can actually be adapted to an imperfective view.

Of course, these are a particularly well-selected set of examples that do not however exhaust the spectrum of functions of perfective and imperfective verbs. However, this was deemed quite enough for a first introduction to the topic.

As for aspectual markers that trigger the use of one or the other aspectual variant, students were provided with the essential list from Matovac (2022, p. 229) in their second lesson on the topic.

6.4 TEST DESIGN

In accordance with the testing practices of the University Language Centre of the University of Pavia, as well as those of Centre for Croatian as a Second and Foreign Language of the University of Zagreb (*Croaticum*), the final test was divided into five parts, one for each of the following skills:

- listening comprehension (*razumijevanje slušanog teksta*) (see § 6.4.1).
- reading comprehension and lexical knowledge (*razumijevanje čitanog teksta i poznavanje leksika*) (see § 6.4.2).
- knowledge of grammatical structures (*poznavanje gramatičkih struktura*) (see § 6.4.3).
- written production (*pisana proizvodnja*) (see § 6.4.4).
- oral production (*usmena proizvodnja*) (see section § 6.4.5).

The listening and reading comprehensions were assigned 20 points each, grammar knowledge was assigned 30, while writing and speaking were assigned 15 points each. This choice was made because, in an intensive course of only 30 hours, it was not possible to focus on the development of all five skills equally, especially the productive ones. Therefore, grading the exercises testing the written and oral production in the same way as the rest would have been unfair. Moreover, since a lot of grammar is needed to reach an A2 level in an inflectional language such as Croatian, the grammar section of the exam was assigned more points (30).

6.4.1 LISTENING COMPREHENSION

The listening comprehension was articulated in two listening tasks, both taken from the *Razgovarajte s nama* A1-A2 Udžbenik[69] (Čilaš Mikulić et al., 2021). Both activities were taken from Unit 8 and revolve around the use of Croatian modal verbs, which were one of the topics covered by the course and not included in the grammar section of the exam. Students were able to listen to both tracks three times while taking notes before writing down their answers.

As you can see from the exam excerpt below, task (1.a) was a responsive listening task with open-ended responses, which was slightly adapted from Čilaš Mikulić et al. (2021, p. 152) by adding the context *Alvaro ide u knjižnicu* 'Alvaro is going to the library' to help students in what Richard & Burns (2012) call the *first stage* of listening, i.e., the schemata activating process that helps them prepare for the actual listening part by gaining a hint of what the topic is. The caption was added in case students did not know the word *časopis* 'magazine', which was in the title of the exercise.

[69] Audio tracks are freely accessible at the following address: https://www.hsn.hr/razgovarajte-s-nama-A2/.

Activity (1.b) was also a responsive listening task, but this time with multiple choices, which was adapted from Čilaš Mikulić et al. (2021, p. 158) by adding two extra questions, (h) and (i), in order for both activities to have 5 items each and to test students on their understanding of numbers and time – see the three possible answers to item (h): 15 minutes, 20 minutes and half an hour.

1) Ascolta 3 volte la traccia e rispondi alle domande.

1.a) Smijem li skenirati časopis?

Alvaro ide u knjižnicu.

a) Za što Alvaro treba knjigu? _____
b) Smije li Alvaro posuditi knjige? _____
c) Smije li Alvaro posuditi časopis? _____
d) Smije li Alvaro skenirati časopis?_____
e) Smije li Alvaro skenirati knjigu?_____

1.b) Hoćeš li piti čaj?

f) Alvaro sad: ☐ uči ☐ pije čaj ☐ pije kavu
g) Alvaro treba: ☐ čaj ☐ večeru ☐ odmor
h) Danilo i Erwan će Alvara čekati: ☐ 15 minuta ☐ 20 minuta ☐ pola sata
i) Erwan ide na automat: ☐ po kavu ☐ po večeru ☐ po čaj
j) Alvaro danas: ☐ više ne mora učiti ☐ mora još malo učiti ☐ mora puno učiti

___ / 20

6.4.2 READING COMPREHENSION AND LEXICAL KNOWLEDGE

The reading comprehension was also taken from the textbook *Razgovarajte s nama!* A1-A2, namely from Unit 9 (Čilaš Mikulić et al., 2021, p. 172). As you can see from the exam section below, this task was conceived as an intensive reading activity followed by 5 open-ended responses.

According to Brown & Lee (2015, p. 409), intensive reading activities call students' attention to grammatical forms and other surface structure details for the purpose of understanding literal meaning and implications, which is why this passage was chosen, since it contains several new words

(for instance, *prizemlje* ‘ground floor’ and *vikendica* ‘holiday house’), but in a grammatically and syntactically familiar context, with no subordinate clauses and several locative complements.

2) Leggi il testo e rispondi alle domande

Stanovanje

Ljudi u Hrvatskoj žive u kućama ili stanovima, u zgradama ili neboderima. Kuće mogu imati samo prizemlje ili mogu imati prizemlje i katove. Kuće cesto imaju podrume i tavane. Zgrade imaju katove i široke su. Neboderi su visoki. Neki ljudi imaju i kuću za odmor – vikendicu – na selu ili na moru. U selima na sjeveru Hrvatske možete vidjeti stare drvene kuće. Na obali, i u selima i u gradovima, možete vidjeti tradicionalne kamene kuće.

k) Gdje stanuju ljudi u Hrvatskoj?_____
l) Što kuće mogu imati? _____
m) Kakve su zgrade, a kakvi su neboderi? _____
n) Što je vikendica? _____
o) Gdje možemo vidjeti tradicionalne kamene kuće? _____

___ / 20

6.4.3 KNOWLEDGE OF GRAMMATICAL STRUCTURES

As you can see from the excerpt below, the grammar section of the exam consisted of eight fill-in-the-blank exercises focused on the most important topics covered by the course, namely:

- exercise 3.1 covers both the regular (a, b, c) and irregular forms (d, e) of the *accusative* case, namely, masculine animate nouns that require an *-a* ending such as *prijatelj* ‘friend’ or *pas* ‘dog’ (see section § 1.3.2).
- exercise 3.2 covers both the regular (c, e) and irregular forms (a, b, d) of the *nominative* plural, such as monosyllabic masculine names like *miš* ‘mouse’ or nouns requiring the sibilarisation of *k* into *c* due to the presence of an *-i* in the plural ending, as in the case of *otok* ‘island’ (see section § 1.3.1).
- exercises 3.4 and 3.3 cover, respectively, the alternation between the locative and accusative case for expressing static locations or goals, and the alternation between the prepositions *na* ‘on’ and *u* ‘in’ in both cases (see section § 1.3.2 and 1.3.5). In 3.4, irregular locative forms

requiring the sibilarisation of *k* into *c* (see section § 1.2.6.2) like *Lika* 'Lika' and *otok* 'island' due to the presence of an *-i* in the locative ending were included, namely (c) and (d).

- exercise 3.5, 3.6, and 3.7 cover the three main Croatian verb tenses that an A2 speaker is expected to master, namely *prezent* 'present', *perfekt* 'past', and futur prvi 'simple future', (Udier, 2017, p. 135). Different verb conjugations were included: several verbs in *-ati* > *-am* (e.g., *slušati* 'to listen', *kuhati* 'to cook', *parkirati* 'to park'), *-iti, -jeti* > *-im* (e.g., *kupiti* 'to buy', *tražiti* 'to look for', *voljeti* 'to love', *štedjeti* 'to save up'), the movement verb *ići* 'to go', and the verb *jesti* 'to eat'.
- Finally, the last exercise (4) is devoted to verbal aspect. The exercise was not graded, since A2 speakers are only expected to have a basic understanding of aspect and to have mastered only the most frequent aspectual pairs, like *jesti/pojesti* 'to eat' or *piti/popiti* 'to drink' (Udier, 2017, p. 140). However, it was nevertheless included in the exam to collect data in order to evaluate whether the way the topic was introduced to the class was promising or not.

  The exercise consisted in 10 multiple-choice items where the right aspectual variant of a pair had to be selected depending on the context of the given sentence. Students were supposed to have learnt a list of aspectual markers (e.g., *čim* 'as soon as' for the perfective, *danima* 'for days' for the imperfective) that trigger the use of one or the other variant (Matovac, 2022, p. 229) and were familiar with the pairs featured in the exercise.

3) Grammatica

3.1 Completa con l'accusativo **delle espressioni tra parentesi**

a) Ana voli _____ . (more)

b) Ona lijepo pjeva. Pjeva španjolske _____ . (pjesma *plur.*)

c) Ja kuham svaki dan. Rado kuham _____. (tjestenina)

d) Irina čeka _____ . (prijatelj)

e) Mara ima _____ . (pas)

3.2 Competa le frasi mettendo al plurale **la parola evidenziata.**

a) Micky je **miš**. Mickey i Minnie su _____ .

b) Pag je **otok**. Pag i Krk su _____ .

c) Ivana je **djevojka**. Ivana i Ana su _____ .

d) Ivan je **dijete**. Ivan i Sanja su _____ .

e) To je **brdo**. To su _____ .

3.3 Completa le frasi con la preposizione corretta: *u* **o** *na***.**

a) Marko ide _____ svoju sobu.

b) Vi ne idete _____ film.

c) Nikad nisam išla _____ Francusku.

d) Idemo li zajedno _____ ručak?

3.4 Completa le frasi con il caso corretto: locativo **o** accusativo

a) Nikad nisam išla u _____ . (knjižnica)

b) Djeca se igraju u _____ . (park)

c) Djeca su sada kod bake u _____. (Lika)

d) Idemo na _____ . (otok *plur*.)

3.5 Completa le frasi al presente

a) Njezina sestra izvrsno _____ . (kuhati)

b) Svaki dan _____ u slastičarnicu na kolače. (ići, ja)

c) Brat i sestra _____ novac za novi bicikl. (štedjeti)

d) Alvaro _____ odijelo za svadbu. (tražiti)

3.6 Completa le frasi al perfekt

a) Ona ____ _____ glazbu. (slušati)

b) _____ li _____ psa i mačku? (kupiti, ti)

c) Jučer ____ _____ kremšnite. (jesti, mi)

d) Mama ____ _____ crvene ruže. (voljeti)

3.7 Completa le frasi al futuro

a) Mi _____ _____ kišobran. (nositi)

b) _____ _____ sarmu za ručak. (kuhati, vi)

c) _____ li _____ auto ovdje? (parkirati, ti)

d) _____ _____ vani. (ići, ja)

___ / 30

4) Completa le seguenti frasi con la variante aspettuale **adeguata**

a) Cijeli dan sam _____ filmove. (gledala/pogledala)

b) Čim je Ante _____ na e-mail, bilo je sve u redu. (odgovarao/odgovorio)

c) Upravo sad _____ na poruku. (odgovaram/odgovorim)

d) Katarina je _____ ručak dva sata. (kuhala/skuhala)

e) Carmen je _____ hrvatski jezik za tri godine. (učila/naučila)

f) Čim _____ kavu, moram na WC. (pijem/popijem)

g) _____ sam zadaću za pola sata. (Napisao/Pisao)

h) Odmah ću _____ tvoj novi roman. (kupiti/kupovati)

i) Ti stalno _____ na hranu. (misliš/pomisliš)

j) Irina je _____ cijelu pizzu za pet minuta. (jela/pojela)

6.4.4 WRITTEN PRODUCTION

In the writing section of the exam, in line with genre-based writing pedagogy (Brown & Lee, 2015, p. 432), students were asked to write a journal entry of at least 5 sentences answering the question:

> *Što si radio/radila ovaj tjedan? Koga si vidio/vidjela? Gdje si bio/bila?*
> 'What have you done this week? Who have you seen? Where have you been?'

The topic was chosen to let them express their voice using the past tense, which had been thoroughly dealt with during the course. Student writing was assessed based on the following checklist: consistency with the questions provided in the guidelines (5 points), grammar and syntax (5 points), and vocabulary (5 points) (Brown & Lee, 2015, p. 456). Case inflection, word order, and lexical mistakes were assigned a penalty of 0.5 points each.

6.4.5 ORAL PRODUCTION

Given the limited attention that the intensive course could devote to the development of spoken fluency, the oral part of the exam was assigned 15 points. The task that students were presented with was an interactive task, namely they were individually asked to answer questions in a conversation scenario (Brown & Lee, 2015, p. 385). Each student was asked the first question and an additional two or three questions from the list below, which they had already encountered during the course:

- *Predstavi se, tko si? Odakle si? Koliko imaš godina?*
  'Introduce yourself, who are you? Where are you from? How old are you?'
- *Što radiš u slobodno vrijeme? Imaš li hobije?*
  'What do you do in your free time? Do you have any hobbies?'
- *Kakva je tvoja obitelj? Koliko ima ljudi?*
  'What does your family look like? How many people are in it?'
- *Kako izgleda tvoj tjedan?*
  'What does your week look like?'
- *Kamo ćeš ići na godišnji odmor ove godine? Imaš li planove za ovo ljeto?*
  'Where are you going on holiday this year? Do you have plans for the summer?'
- *Kamo si išao na godišnji odmor prošle godine?*
  'Where did you go on holiday last year?'

Their performance was graded based on an adaptation of the *ISE Foundation Speaking and listening rating scale*[70] developed by the Trinity College of London, which corresponds to an A2 level of the CEFR (Council of Europe, 2020). The original scale goes from 0 to 4 (as you can see from Figure 8), but was adapted, for the needs of the exam, to a 15-point scale.

[70] https://www.trinitycollege.com/resource/?id=7455. Website last visited on September 24th, 2023.

Score	Communicative effectiveness • Task fulfilment • Appropriacy of contributions /turn-taking • Repair strategies	Interactive listening • Comprehension and relevant response • Level of understanding • Speech rate of examiner interventions • Speed and accuracy of response	Language control • Range • Accuracy/precision • Effects of inaccuracies	Delivery • Intelligibility • Lexical stress/intonation • Fluency • Effects on the listener
4	• Fulfils the task very well • Maintains simple exchanges • Says or signals in basic ways that he/she did not follow (eg '*Can you repeat?*')	• Understands short and simple interventions with little repetition • Identifies factual information • Follows conversational speech, sometimes slowed • Sometimes responds slowly	• Uses a sufficient range of basic grammatical structures/lexis and memorised phrases to deal with topics at this level • Shows a sufficient level of grammatical accuracy and lexical precision to deal with simple exchanges • Makes basic mistakes, but most errors do not impede communication	• Mostly intelligible despite noticeable use of non-standard phonemes • Uses basic lexical stress and intonation appropriately • Speaks slowly with frequent pausing and hesitation • Requires some careful listening
3	• Fulfils the task appropriately • Maintains simple exchanges, despite some difficulty, some examiner support is necessary • Attempts to say or signal in basic ways that he/she did not follow (eg '*Repeat?*', '*Mm?*' with a facial expression)	• Understands short and simple interventions but may need repetition • Identifies factual information, sometimes incomplete • Follows slow conversational speech • May make slow responses due to the need to make sense of the input	• Uses a range of basic grammatical structures/ lexis and memorised phrases to deal with topics at this level • Shows a basic level of grammatical accuracy and lexical precision to deal with simple exchanges • Makes basic mistakes – major errors occasionally impede communication	• Mostly intelligible despite noticeable use of non-standard phonemes • Uses basic lexical stress and intonation enough to follow • Speaks slowly with frequent and extended pausing and hesitation • Requires careful listening
2	• Fulfils the task acceptably with support • Examiner support is necessary to keep the interaction going • Attempts to signal in basic ways that he/she did not follow	• Does not always understand interventions, quite often needs repetition • Identifies factual information just enough to respond • Follows slow conversational speech in places • Makes slow responses due to difficulty in making sense of the input	• Uses an acceptable range of basic grammatical structures/lexis and memorised phrases to deal with topics at this level • Shows a basic level of grammatical accuracy and lexical precision – just enough to follow • Makes basic mistakes – major errors sometimes impede communication	• Sometimes unintelligible • Use of non-standard phonemes is very evident • Uses some basic lexical stress and intonation • Speaks slowly with frequent and noticeable pausing and hesitation • Requires careful listening, sometimes difficult to follow
1	• Does not fulfil the task even with support • Difficult to keep the interaction going even with examiner support • May attempt to signal in basic ways that he/she did not follow	• Understands few or no examiner interventions • Does not identify factual information • Has difficulty in following even slow conversational speech • Makes slow responses due to failure to understand input	• Uses some basic grammatical structures/lexis, but does not manage to deal with topics at this level • Does not show an adequate level of grammatical accuracy and lexical precision for simple exchanges • Makes basic mistakes, and major errors often impede communication	• Sometimes or often unintelligible • Does not use basic lexical stress or intonation • Speaks very slowly with frequent and noticeable pausing and hesitation • Requires careful listening, often difficult to follow
0	No performance to assess (candidate does not speak, or does not speak in English).			

Figure 8 - *ISE Foundation Speaking and listening rating scale* by the Trinity College of London.

6.5 TEST RESULTS

As you can see from Table 1, the course participants scored very positive results on the final test, with a mean of 87.5%, a median of 86.5% and a coefficient of variation (CV) of 0.1007956684.

ID	Exam score
S1	84.5%
S2	100%
S3	85.5%
S4	79.5%
S5	97%
S6	73.5%
S7	92.5%
S8	87.5%

Table 1 – Overall final test scores

Both the mean and the median were calculated to understand the central tendency of the set of scores. While the mean has the disadvantage of being affected by a single value being too high or too low compared to the rest, the median is sometimes taken as a better measure of a mid-point. In our case, they are pretty similar. We also calculated the coefficient of variation (CV), which is the ratio of the standard deviation to the mean and is used to provide a measure of the variability of the population in relation to the mean. The lower the CV, the smaller the dispersion, which in our case is very low, showing that the dataset is compact. This means that overall, the test went very well.

The next two Tables provide a more in-depth overview of the test results. Table 2 shows the non-normalised results of the eight participants (S1 to S8) exercise by exercise, while Table 3 shows the same results but in a normalised and thus comparable way. The following abbreviations were used: L for Listening, R for Reading, G for Grammar, VID for verbal aspect, W for Writing and S for Speaking. The scores of the exercise on verbal aspect are between round brackets because they were not included in the final test score, but they are going to be included in our analysis.

ID	L1	L2	R	G1	G2	G3	G4	G5	G6	G7	VID	W	S	Total
S1	8	8	20	4	3	4	4	3	2	4	(7)	11.5	13	84.5
S2	10	10	20	5	5	4	4	4	4	4	(9)	15	15	100
S3	10	6	15	5	4.5	4	3	4	4	3	(5)	13	14	85.5
S4	7	10	20	5	4	2	2	2	2.5	2	(3)	12	11	79.5
S5	10	10	20	5	4.5	4	4	4	4	3	(9)	13.5	15	97
S6	6	6	18	5	3	3	3	2	2	4	(4)	11.5	10	73.5
S7	10	10	19	3	3.5	4	2	4	4	4	(9)	14	15	92.5
S8	10	10	19.5	5	4	4	4	2	3	2	(6)	11	13	87.5

Table 2 – Non-normalised exam results exercise by exercise: Listening (L), Reading (R), Grammar (G), Verb aspect (VID), Writing (W) and Speaking (S)

ID	L1	L2	R	G1	G2	G3	G4	G5	G6	G7	VID	W	S	Total
S1	0.8	0.8	1	0.8	0.6	1	1	0.75	0.5	1	(0.7)	0.77	0.86	0.845
S2	1	1	1	1	1	1	1	1	1	1	(0.9)	1	1	1
S3	1	0.6	0.75	1	0.9	1	0.75	1	1	0.75	(0.5)	0.87	0.93	0.855
S4	0.7	1	1	1	0.8	0.5	0.5	0.5	0.625	0.5	(0.3)	0.8	0.73	0.795
S5	1	1	1	1	0.9	1	1	1	1	0.75	(0.9)	0.9	1	0.97
S6	0.6	0.6	0.9	1	0.6	0.75	0.75	0.5	0.5	1	(0.4)	0.77	0.67	0.735
S7	1	1	0.95	0.6	0.7	1	0.5	1	1	1	(0.9)	0.93	1	0.925
S8	1	1	0.975	1	0.8	1	1	0.5	0.75	0.5	(0.6)	0.73	0.87	0.875

Table 3 – Normalised exam results exercise by exercise: Listening (L), Reading (R), Grammar (G), Verb aspect (VID), Writing (W) and Speaking (S)

6.6. QUANTITATIVE ANALYSIS

In this section, we will go through the results of the test and analyse them following a quantitative approach, in order to collect useful insights on the topics and abilities where students gave their best and worst performances.

No statistically sound generalisation will be attempted in this part of the analysis because of data scarcity, whereas some considerations will be possible in the next section, given that we will not go into the same detail but adopt a more coarse-grained perspective.

The following paragraphs will each focus on one of the five sections of the exam (see section 6.4), starting with the Listening comprehension.

6.6.1 LISTENING COMPREHENSION

As far as listening goes, both activities went well, with students scoring a mean of 8.875 and a median of 10 out of 10 in activity L1, and a mean of 8.75 and a median of 10 out of 10 in activity L2, bringing the overall mean and median at 17.625 and 18.5 out of 20, respectively (Table 4).

ID	**L1**	**L2**	**L1 + L2**
S1	8	8	16
S2	10	10	20
S3	10	6	16
S4	7	10	17
S5	10	10	20
S6	6	6	12
S7	10	10	20
S8	10	10	20
Mean	8.875	8.75	17.625
Median	10	10	18.5

Table 4 – Listening comprehension results.

If we look in detail at the answers they gave, we can notice that two students out of eight had trouble understanding the time that was mentioned in activity L2 (*15 minuta*, '15 minutes'), which might be an indicator of the fact that more time is needed to learn to identify and decode numbers in a listening activity.

6.6.2 READING COMPREHENSION AND LEXICAL KNOWLEDGE

As for the Reading comprehension, this was the activity were students performed at their best, with a mean score of 18.85714286 and a median score of 19.5 out of 20, respectively (Table 5).

ID	R
S1	20
S2	20
S3	15
S4	20
S5	20
S6	18
S7	19
S8	19.5
Mean	18.85714286
Median	19.5

Table 5 – Reading comprehension results.

These results might either suggest that students have mastered this ability, which is not unrealistic giving that they are a group of highly educated individuals that are accustomed to extracting information from written text, or that the chosen reading comprehension and questions were too easy.

6.6.3 KNOWLEDGE OF GRAMMATICAL STRUCTURES

As for the grammatical topics that were included in the test, Table 6 provides an overview of the raw scores for each exercise, including the one on verbal aspect (for the normalised scores see Table 3).

ID	G1	G2	G3	G4	G5	G6	G7	VID
S1	4	3	4	4	3	2	4	7
S2	5	5	4	4	4	4	4	9
S3	5	4.5	4	3	4	4	3	5
S4	5	4	2	2	2	2.5	2	3
S5	5	4.5	4	4	4	4	3	9
S6	5	3	3	3	2	2	4	4
S7	3	3.5	4	2	4	4	4	9
S8	5	4	4	4	2	3	2	6
Mean	4.625	3.8	3.625	3.25	3.125	3.28	3.25	6.5
Median	5	4	4	3.5	3.5	4	3.5	6.5

Table 6 – Grammar results

Exercise G1 was focused on accusative forms and is the exercise where students obtained on average the highest score, with a mean of 4.625 and a median of 5 out of 5. Two of the three mistakes that were made concerned the inflection of the irregular masculine noun *pas* 'dog', which not only requires an *-a* ending to mark animacy, but is also an example of *nepostojano a* 'mobile a', thus becoming *psa*. The only other mistake was made by the Italian-Bosnian bilingual student, who had trouble grasping the concept of case and applying it in exercises, since case inflection is already an automatic process for them.

Exercise G2 was on the creation of nominative plural forms and included five sentences. The mean score of the students on this exercise was 3.8 and the median one 4 out of 5. The three answers that they usually got wrong are: the plural form of *otok* 'island', where they failed sibilarise *k* into *c*; the spelling of the plural form of *dijete* 'child', *djeca* 'children'; and, surprisingly, the plural of the regular neuter noun *brdo* 'hill', which two students inflected as a monosyllabic masculine noun creating the agrammatical form **brdovi*. No mistakes were made in creating the plural form of the monosyllabic masculine noun *miš* 'mouse', *miševi* (which required also a vowel change due to the presence of the palatal sound *š* before the plural ending), and of the regular feminine noun *djevojka* 'girl', *djevojke*.

Exercise G3 focused on the alternation between the prepositions *na* 'on' and *u* 'in', which takes place in complements expressing static locations or goals of motion and depends on the semantics of the noun introduced by the prepositions themselves: *na* 'on' for surfaces, open spaces and activities and *u* 'in' for closed spaces (see section § 1.3.2 and 1.3.5). The exercise included four sentences and, given that the mean score was 3.625 and the median one was 4 out of 4, it is one of the exercises where students performed at their best. Only three mistakes were made on three different sentences. The one that nobody got wrong is *Vi ne idete na film* 'You are not going to the movies', where the goal of motion is the metonymic activity of "watching a movie". Given that there are no repeated mistakes and that two out of three were made by the same student, the reason for the mistakes could be attributed to the individual.

Exercise G4 focuses on the alternation between locative and accusative to express static locations or goals of motion (see section § 1.3.2 and 1.3.5). The mean score was 3.25 and the median 3.5 out of 4. Four mistakes out of six involved the accusative plural of *otok* 'island', which became **otoce* instead of *otoke*, probably due to the interference with the nominative and locative plural forms *otoci* and *otocima*, where the presence of the *-i* sound requires a sibilarisation from *k* to *c*, which is however not required in the accusative plural in *-e*. This is a clear case of overgeneralization of an exception.

Exercise G5 was on the present. The mean score obtained by students was 3.125 and the median 3.5 out of 4. Mistakes were only made in the conjugation of verbs of the *-iti, -jeti* > *-im* group, namely *tražiti* 'to look for' and *štedjeti* 'to save up', while no mistakes were made conjugating the *-ati* > *-am* verb *kuhati* 'to cook', nor the irregular verb *ići* 'to go'. This could be due to the fact that the *-ati* group was the first presented to the class.

Exercise G6 was on the past. The mean score obtained by students was 3.28 and the median 4 out of 4. Also in this case, most mistakes involved verbs of the *-iti, -jeti* > *-im* group, namely *kupiti* 'to buy', and *voljeti* 'to love', but also the irregular verb *ići* 'to go'.

Exercise G7 was on the future, its mean score was 3.25 and its median score 3.5 out of 4. The few mistakes that were made involved one of two scenarios: (1) the interrogative form in (c) *Hoćeš li parkirati auto ovdje?* 'Are you going to park here?', which requires the long form of the auxiliary *htjeti* 'to want' in first position; or (2) the inversion of the lexical verb and of the auxiliary required when no subject pronoun is used and the verb phrase is at the beginning of the sentence, as in (b) *Kuhat ćete sarmu za ručak* 'You will cook sarma for lunch' and (d) *Ići ću vani* 'I will go out'. The affirmative form with the short auxiliary, as in sentence (a) *Mi ćemo nositi kišobran* 'We will take the umbrella', on the other hand, was always correctly produced.

Finally, for what concerns the exercise on verbal aspect, as you can see from Table 6, this is where the students' performances were a bit lower, scoring both a mean and a median of 6.5 out of 10. Indeed, although students knew the meaning of all the aspectual verb pairs featured in the exercise and were familiar with the list of aspectual markers (Matovac, 2022, p. 229) used in the sentences, this is clearly a more demanding exercise than the previous ones.

Table 7 provides a more detailed overview of the answers to this exercise. In Column 1, you can find the aspectual marker that was used in the sentence, which is also highlighted in bold in the actual exercise sentences reported in Column 2. Column 3 lists the aspectual values of the correct answers that ought to have been selected. Columns 4 to 11 list the answers of each student (S1-S8): "X" signals the correct answers, while "no" the wrong ones. Finally, Column 12 lists the number of mistakes that were made for each sentence in Column 2, which were ranked from the one with the least mistakes to the one with most mistakes.

Marker	**Sentences from the VID exercise**	**Correct variant**	**S5**	**S3**	**S7**	**S1**	**S2**	**S4**	**S6**	**S8**	**N° of mistakes**
dva sata 'for two hours'	d) *Katarina je kuhala/*skuhala ručak* ***dva sata****.* 'Katarina has been cooking lunch for two hours.'	imperfective	X	X	X	X	X	X	X	no	1
cijeli dan 'all day'	a) ***Cijeli dan*** *sam gledala/*pogledala filmove.* 'I have been watching movies all day.'	imperfective	X	X	X	X	X	no	X	no	2
upravo sad 'right now'	c) ***Upravo sad*** *odgovaram/*odgovorim na poruku*. 'I'm answering the message right now'	imperfective	no	X	X	no	X	X	X	X	2
čim 'as soon as'	f) ***Čim*** **pijem/popijem kavu, moram na WC.* 'As soon as I drink coffee, I have to go to the toilet'	perfective	X	X	X	no	X	no	X	X	2
za tri godine 'in three years'	e) *Carmen je *učila/naučila hrvatski jezik* ***za tri godine.*** 'Carmen has learnt Croatian in three years.'	perfective	X	no	no	X	X	X	no	X	3
stalno 'constantly'	i) *Ti* ***stalno*** *misliš/*pomisliš na hranu*. 'You are constantly thinking about food.'	imperfective	X	no	X	X	X	no	no	X	3
za pet minuta 'in five minutes'	j) *Irina je *jela/pojela cijelu pizzu* ***za pet minuta***. 'Irina has eaten a whole pizza in five minutes.'	perfective	X	X	X	X	X	no	no	no	3
čim 'as soon as'	b) ***Čim*** *je Ante *odgovarao/odgovorio na e-mail, bilo je sve u redu.* 'As soon as Ante answered the e-mail, everything was OK.'	perfective	X	no	X	X	X	no	no	no	4
za pola sata 'in half an hour'	g) *Napisao/*Pisao sam zadaću* ***za pola sata****.* 'I have done my homework in half an hour.'	perfective	X	no	X	no	X	no	no	X	4
odmah 'immediately'	h) ***Odmah*** *ću kupiti/*kupovati tvoj novi roman.* 'I will buy your new novel immediately.'	perfective	X	no	X	X	no	no	no	X	4

Table 7 – Students' answers to the exercise on verbal aspect

By looking at the results, we can say that the sentences where fewer mistakes were made were those containing imperfective aspectual markers, namely (d) *Katarina je kuhala/*skuhala ručak* ***dva sata*** 'Katarina has been cooking lunch for two hours' (only one student got it wrong), (a) ***Cijeli dan*** *sam gledala/*pogledala filmove* 'I have been watching movies all day' (only two students out of eight got it wrong), and (c) ***Upravo sad*** *odgovaram/*odgovorim na poruku* 'I'm answering the message right now' (only two students got it wrong). On the other hand, 50% of the students (4 out of 8) did not identify the right aspectual variant in sentences (b), (g) and (h), featuring the perfective markers *čim* 'as soon as', *za pola sata* 'in half an hour' and *odmah* 'immediately'. This might point to the fact that imperfective scenarios are more easily understood and recognized by beginners.

Moreover, we believe it is important to stress that one of the three students that got 9 out of 10 answers right had prior knowledge of another Slavic language (Russian) that they could transfer to Croatian, while the other is the Italian-Bosnian bilingual student who could already use aspectual variants despite not being aware of their underlying mechanisms. The third student is an Italian native speaker, who was however extremely motivated and achieved the best test results in the whole class.

6.6.4 WRITTEN PRODUCTION

In the writing section of the exam, students were asked to write a journal entry of at least 5 sentences to answer the question: *Što si radio/radila ovaj tjedan? Koga si vidio/vidjela? Gdje si bio/bila?* 'What have you done this week? Who have you seen? Where have you been?'

ID	**W**
S1	11.5
S2	15
S3	13
S4	12
S5	13.5
S6	11.5
S7	14
S8	11
Mean	13
Median	13

Table 8 – Writing results

As you can see from Table 8, the students' performances went quite well, with a mean and median score of 13 out of 15. They all understood the assignment and described their week, thus starting from

a base of 5 points. A variety of spelling and grammar mistakes were made, which did not however prevent the reader from understanding the different texts.

6.6.5 ORAL PRODUCTION

Finally, Table 9 presents the results of the speaking part of the exam, which went well considering that the mean score was 13.25 and the median 13.5 out of 15.

ID	**S**
S1	13
S2	15
S3	14
S4	11
S5	15
S6	10
S7	15
S8	13
Mean	13,25
Median	13,5

Table 9 – Speaking results

The students' performances were graded based on an adaptation of the *ISE Foundation Speaking and listening rating scale* developed by the Trinity College of London (see Figure 8), which corresponds to an A2 level of the CEFR (Council of Europe, 2020). For instance, the students who were given more than 12.5 out of 15 demonstrated good *communicative effectiveness* (they were able to maintain simple exchanges and to signal in basic ways when they were not following), *interactive listening* (they could understand short interventions, identify factual information, and follow the conversation even if sometimes slowed), *language control* (they used a sufficient range of basic grammatical structures and lexis making only basic mistakes that did not impede communication) and *delivery* (intelligible despite the use of non-standard phonemes and frequent pausing and hesitation). The two students (S4 and S6) who got lower marks were still able to fulfill the task and maintain simple exchanges, but more examiner support was needed ad they needed a greater amount of repetition. They answered more slowly because they needed more time to make sense of input and were not always able to identify all the factual information conveyed in the questions. They were able to use a range of basic grammatical structures, words, and memorized phrases, making some basic mistakes and sometimes major ones thus impeding communication. The delivery was nonetheless intelligible despite the use of non-standard phonemes and frequent and extended pausing and hesitation.

Considering the length of the course and the few speaking opportunities they had, this was a very successful result.

6.7 STATISTICAL ANALYSIS

Based on the normalised exam results of Table 3 and assuming that said results can be used as an indicator of the subjects' level of acquisition of the topics covered in the course, we are going to carry out a statistical analysis aimed at answering to the following five questions:

I. Is there a difference in the students' level of acquisition of the accusative case and of the nominative plural? (see section § 6.7.1)
II. Is there a difference in the students' level of acquisition of the accusative case and of the locative? (see section § 6.7.2)
III. Is there a difference in the students' level of acquisition of the present tense, the past and the future? (see section § 6.7.3)
IV. If we were to compare groups of topics that are related to each other, is there a difference in the students' level of acquisition of (a) nominative plural and accusative forms; (b) the locative/accusative alternation and the use of the prepositions *na* 'on' and *u* 'in'; (c) the three verb tenses; (d) verbal aspect? (see section § 6.7.4)
V. Is there a difference in the students' performances depending on the skill that was tested, i.e., speaking, reading, grammar, writing and speaking? (see section § 6.7.5)

In order to translate these questions into statistical hypotheses, we had first to consider the fact that we are working with a small dataset, which made us lean towards non-parametric statistical tests that can be used also with non-normally distributed data, such as the Wilcoxon Paired Test (Gibbons & Chakraborti, 2011; Hollander & Wolfe, 1999) and Friedman Test (Hogg & Wolfe, 1987; Hollander & Wolfe, 1999).

6.7.1 NOMINATIVE AND ACCUSATIVE

The first question we want to answer to is whether there is a statistical difference in the performance of subjects in exercise G1 (accusative) and exercise G2 (nominative plural forms) (see Table 3). In order to do that, for each subject (S1, S2, etc.), the normalised G2 score is subtracted from their G1 score and the median of these subtractions is calculated (see Table 4). If the median value is zero, then the subjects' performances are averagely the same, which will be our H_0.

ID	G1-G2
S1	0.2
S2	0
S3	0.1
S4	0.2
S5	0.1
S6	0.4
S7	-0.1
S8	0.2
Median	**0.1 (± 0.1505940617)**

Table 10 – G1-G2 scores

As you can see above in Table 10, the median value of G1-G2 is 0.1 (± 0.1505940617). The Wilcoxon Paired test we ran returned a *p*-value of 0.0625, which is higher than the significance level alpha = 0.05. Therefore, we can accept the H_0 and say that there are no statistical differences between the scores.

Provided that test scores can tell us something about acquisition, this might mean that both the nominative and accusative case have been successfully acquired by the students.

6.7.2 ACCUSATIVE AND LOCATIVE

The second question we want to answer to is whether a statistical difference can be observed in the performances concerning exercise G1 (accusative) and G4 (alternation accusative/locative) (see Table 3). Also in this case, for each subject, their G4 score is subtracted from the G1 score and the median subtraction value is calculated (see Table 11) and, if the median is zero, then the performances can be considered averagely the same, which will be our H_0.

ID	G1-G4
S1	-0,2
S2	0
S3	0,25
S4	0,5
S5	0
S6	0,25
S7	0,1
S8	0
Median	**0.1 (± 0. 2150581317)**

Table 11 – G1-G4 scores

As you can see from Table 11, the median value of G1-G4 is 0.1 (± 0.2150581317). A Wilcoxon Paired Test on the data returned a *p*-value of 0.1875, which is higher than the significance level alpha = 0.05. Therefore, we can accept the H0, affirm that there are no statistical differences between the exercise scores, which could be interpreted as proof that the locative and accusative cases can be successfully disambiguated and have been indeed mastered by learners.

6.7.3 PRESENT, PAST AND FUTURE

The third question we want to answer to is whether there is a statistical difference between the scores of exercises G5 (present tense), G6 (past tense) and G7 (future tense) from Table 3, hypothesizing that there is actually no difference (H_0).

In this case, since the comparison is between three groups, the statistical test needed is Friedman, which returns a *p*-value of 0.69. Since the *p*-value is indeed higher than the significance level alpha = 0.05, we can once again accept the zero hypothesis and say that there is no statistical difference between the scores. This means that all the tenses have been acquired similarly.

6.7.4 VERBAL ASPECT

The fourth question we want to answer to is whether there there is a statistical difference between the scores of the verbal aspect exercise (VID) and the scores of these three clusters of exercises: G1_G2 (accusative and nominative plural); G3_G4 (accusative/locative alternation and *na*/*u* alternation); and G5_G7 (present, past, and future tense). The zero hypothesis (H_0) is that there is no difference.

In order to be able to compare the scores of the three clusters, we first created merged scores for each of them. Let us take the cluster G1_G2 as an example: for each subject, the raw score of G1 (Table 2) was added to the raw score of G2 and the sum was normalised. Table 12 provides an overview of all the *p*-values returned by the tests that we ran on the scores.

	***p*-value**	**H_0**	***p*-value (adjusted)**	**H_0 (adjusted)**	***alpha***	***p*-value rank**
Friedman (all groups)	*0.0436	rejected			0.05	
G5_G7 vs. VID	*0.0234	rejected	0.1404	accepted	0.05	1
G3_G4 vs. VID	*0.0391	rejected	0.1173	accepted	0.05	2
G1_G2 vs. VID	0.0781	accepted	0.1562	accepted	0.05	3
G3_G4 vs. G5_G7	0.4062	accepted	0.6093	accepted	0.05	4
G1_G2 vs. G5_G7	0.4531	accepted	0.54372	accepted	0.05	5
G1_G2 vs. G3_G4	0.7031	accepted	0.7031	accepted	0.05	6

Table 12 – An overview of both the non-adjusted and adjusted *p*-values returned by the Friedman (first raw) and Wilcoxon tests we ran on the scores: *p*-value (adjusted) = *p*-value * 6 (number of tests) / *p*-value rank.

The first test that we ran on all the scores was Friedman, to see if any statistical difference was detected between the subjects' performances on the four different clusters of exercises (considering verbal aspect as a cluster of its own). As you can see from the first row of Table 6, Friedman returned a *p*-value of 0.0436, which is lower than alpha = 0.05, thus allowing us to reject the null hypothesis that no statistical difference can be observed between the groups.

At this point, to try to identify which group of performances is indeed statistically different from the rest, we decided to run paired comparisons between the different groups using Wilcoxon, which did highlight a difference between G5_G7 vs. VID and G3_G4 vs. VID.

However, this statistically significant difference is erased if the *p*-value is adjusted following Benjamini & Hochberg (1995) adjustment, which is advised to guarantee test significance (specifically to maintain the alpha value) when passing from a multiple test like Friedman to post-hoc comparisons with Wilcoxon-Paired Test. Therefore, since all the adjusted *p*-values are higher than the significance level alpha = 0.05, we can accept the zero hypothesis and conclude that no significant differences can be identified between the clusters, after all.

This result may seem counterintuitive, but it is justified by the fact that, when carrying out Wilcoxon paired comparisons, the data at our disposal shrinks and there is actually not enough information to answer the test's question with the desired confidence level.

However, a statistical difference is clearly there (as supported by Friedman), and even if we do not have enough data to say with enough confidence which are the actors that are responsible for it, we can reasonably suppose – by taking account the non-adjusted *p*-values – that such difference may be linked to the scores of the verbal aspect exercise (VID).

This means that the VID scores are plausibly different from the rest of the dataset, which would provide further evidence that this topic is indeed more advanced and is not to be included in an A2 proficiency test.

6.7.5 COMPARING SKILLS

The fifth and last question we would like to answer is whether there might be a statistical difference in the students' performances depending on the skill that was tested, i.e., speaking (S), reading (R), grammar (G), writing (W) and speaking (S)? The zero hypothesis (H_0) is that there is no difference.
As in the previous section, also in this case, to answer this question, it was necessary to group together the scores of the different exercises to create the five different clusters representative of the five identified abilities. Verbal aspect was included in the grammar cluster. Table 13 provides an overview of all the *p*-values returned by the tests that we ran on the scores.

	p-value	H_0	*p*-value (adjusted)	H_0 (adjusted)	*alpha*	*p*-value rank
Friedman (all groups)	*0.0284	rejected			0.05	
G vs. S	*0.0078	rejected	0.078	accepted	0.05	1
R vs. G	*0.0312	rejected	0.156	accepted	0.05	2
R vs. W	0.0781	accepted	0.2603333333	accepted	0.05	3
L vs. G	0.0781	accepted	0.19525	accepted	0.05	4
G vs. W	0.1484	accepted	0.2968	accepted	0.05	5
R vs. S	0.3125	accepted	0.5208333333	accepted	0.05	6
W vs. S	0.3281	accepted	0.4687142857	accepted	0.05	7
L vs. R	0.4062	accepted	0.50775	accepted	0.05	8
L vs. W	0.5	accepted	0.5555555556	accepted	0.05	9
L vs. S	1	accepted	1	accepted	0.05	10

Table 13 – An overview of both the non-adjusted and adjusted *p*-values returned by the Friedman (first raw) and Wilcoxon tests we ran on the scores: *p*-value (adjusted) = *p*-value * 10 (number of tests) / *p*-value rank

Now, as in the previous section the first test that was run on all the scores was Friedman, to see if any statistical difference, which was the case, as you can see from the first row of Table 7. Friedman returned a *p*-value of 0.0284, which is lower than alpha = 0.05, thus allowing us to reject the null hypothesis that no statistical difference can be observed between the groups.

At this point, to identify which cluster is responsible for the detected difference, paired comparisons were run between the different groups using Wilcoxon, which did highlight a difference between Grammar vs. Speaking, as well as between Reading vs. Grammar.

Although this difference is lost as soon as we carry out the adjustment of the *p*-value suggested by Benjamini & Hochberg (1995), the former results of the Wilcoxon tests show us that the statistical difference highlighted by Friedman is to be found in the comparisons Grammar vs. Speaking and Reading vs. Grammar. Table 14 might provide a possible interpretation.

ID	Grammar - Speaking	Reading - Grammar	Reading - Writing
S1	-0.09166666667	0.225	0.2333333333
S2	-0.025	0.025	0
S3	-0.1208333333	-0.0625	-0.1166666667
S4	-0.1708333333	0.4375	0.2
S5	-0.0625	0.0625	0.1
S6	-0.01666666667	0.25	0.1333333333
S7	-0.1625	0.1125	0.01666666667
S8	-0.1166666667	0.225	0.2416666667
Median	-0.1041666667	0.16875	0.1166666667
SD	0.0579921998	0.1572290754	0.1269637415

Table 14 – The most different skill pairs when we look at the scores

As we can see from Table 14, if we look at the skill pairs with the most different scores, the three top-three pairings are:

- Grammar vs. Speaking, with a median difference of | -0.1041666667 |
- Reading vs. Grammar, with a median difference of | 0.16875 |
- Reading vs. Writing, with a median difference of | 0.1166666667 |

This means that, on average, subjects have been better in the Speaking than in the Grammar part by +0.1041666667; have scored on average +0.16875 in the Reading than in the Grammar; and + 0.1166666667 in the Reading than in the Writing.

Since only the first two pairings had initially been highlighted as significant by Wilcoxon (Table 7) and they both include Grammar scores, the statistical difference between skills suggested by Friedman might once again be traced back to the presence of exercises on verbal aspect in the Grammar component, which has lowered the average grades.

CONCLUSIONS

This Dissertation wanted to give a contribution to the study and teaching of Croatian verbal aspect, focusing on its interaction with *Aktionsart* and verb polysemy.

After providing a thorough introduction to the Croatian language (Chapter 1), to the concepts of verbal aspect and *Aktionsart* (Chapter 2), as well as to those of valency and verb polysemy (Chapter 3), the focus shifted to the *Croatian Typed Predicate Argument Structures* (Chapter 4).

The CroaTPAS resource, which currently contains 180 Croatian verbs, is accessible online at the address https://croatpas.baisa.cz/ and has been specifically tailor-made to portray verb polysemy, proved to be a powerful tool both to identify phenomena of interest such as *asymmetrical verb pairs*, (Chapter 5), and to introduce Croatian aspectual verb pairs in a classroom context (Chapter 6).

In order to be applied to these different scenarios, however, CroaTPAS was first evaluated (see section § 4.3) via an online survey based on a multiple-choice sense disambiguation task. The survey answers were compared against a benchmark representing CroaTPAS's sense distinctions and the *Jaccard similarity index* was used as a measure of the agreement between the respondents and the resource's annotation. Since the mean similarity score of the collected survey answers is at 91.36% (± 5.12) and datasets with a Jaccard index above 85% are considered highly stable, we can conclude that the collected answers form a proper cluster, thus corroborating CroaTPAS's manual annotation.

In light of this, CroaTPAS was used to collect data to shed light on the intricate relationship between aspect and *Aktionsart* in Croatian aspectual verb pairs. Special attention was devoted to aspectual pairs lacking one or more meanings that are however available in the sense inventory of their aspectual counterpart (*asymmetrical verb pairs*). The formalism of event structures was used to model both symmetric and asymmetrical meanings and has proven especially adequate to provide a graphical portrayal of the relationship between perfective and imperfective variants, which allows for a view of perfectivity in line with Comrie's *blob* metaphor. According to our results, asymmetries seem to be indeed caused by constraints posed by the *Aktionsart* of the meanings involved and all *Aktionsarten* (states, processes, and transitions) seem to have the potential to pose certain constraints on the aspectual viewing of the events they encode and to prevent verbs from displaying meanings that are available to their aspectual counterparts.

Finally, in Chapter 6, an overview of the 30-hour Croatian Language Intensive Course for beginners (A2 level) that took place at Language Centre of the University of Pavia in the second semester of the academic year 2022/23 was provided.

The Intensive Course was a great opportunity to design a syllabus following a CLT (Communicative Language Teaching) approach and exploit the CroaTPAS resource to introduce verbal aspect to the classroom.

In line with the testing practices of the University Language Centre of the University of Pavia, and of the Centre for Croatian as a Second and Foreign Language of the University of Zagreb (*Croaticum*), the final test was divided into five parts: (1) listening comprehension; (2) reading comprehension and lexical knowledge; (3) knowledge of grammatical structures; (4) written production; (5) and oral production.

The course participants scored very positive results in all test components, with an overall mean of 87.5, and a median of 86.5 out of 100. Since we were given this chance, statistical tests were run on the test results' dataset to try to identify possible statistical differences in the students' level of acquisition of the accusative case and of the nominative plural (see section § 6.7.1), of the accusative case and of the locative (see section § 6.7.2), of the present, past and future tenses (see section § 6.7.3), of all the above and verbal aspect (see section § 6.7.4) and of the different skills that were tested (speaking, reading, grammar, writing and speaking (see section § 6.7.5).

Out of all the above, the only statistical difference was found in the scores concerning verbal aspect, a result which provides further evidence that this topic is indeed more advanced than the others and that, although A2 speakers know that there is a distinction between perfective and imperfective verbs and are familiar with the most common aspectual pairs, testing this knowledge in a proficiency test is indeed challenging.

APPENDIX 1

A complete list of the Croatian verb entries in CroaTPAS[71].

	CroaTPAS	T-PAS	English translation
1	bacati/baciti	lanciare	to throw
2	brojiti/izbrojiti	contare	to count
3	čekati (*impf.*)	attendere	to wait
4	čitati/pročitati	leggere	to read
5	čuti (*bias.*)	sentire*	to hear
6	čuvati/očuvati	conservare	to preserve
7	dirati/dirnuti	toccare	to touch
8	djelovati (*impf.*)	agire	to act
9	dočekivati/dočekati	accogliere	to welcome
10	dolaziti/doći	arrivare	to arrive
11	dovršavati/dovršiti	completare	to complete
12	glasati/izglasati	votare	to vote
13	gostiti/ugostiti	ospitare	to accommodate
14	ići (*impf.*)	recarsi	to go
15	informirati (*bias.*)	informare	to inform
16	isključivati/isključiti	escludere	to exclude

[71] Italian verbs marked by an asterisk (*) appear twice (see § 3.3.1), while the Italian *coercive* verbs from Ježek & Quochi (2010) are underlined. When in a pair, Croatian imperfective verbs always come first (e.g., in the pair *bacati/baciti*, *bacati* is the imperfective variant and *baciti* di perfective one), otherwise aspectual values are expressed: *impf.* (imperfective), *pf.* (perfective), *bias.* (biaspectual). In the Italian list, the verbs *sentire* and *guidare* appear twice because, since they both display a highly polysemous behaviour, we decided to create entries for more than one of their Croatian translational equivalents, namely *čuti* 'to hear' and *osjećati/osjetiti* 'to feel' for the first, *voditi/provoditi* 'to lead' and *voziti* 'to drive' for the second.

17	ispravljati/ispraviti	correggere	to correct
18	izazivati/izazvati	provocare	to cause
19	izlagati/izložiti	esporre	to display
20	izroniti (*pf.*)	emergere	to emerge
21	jesti/pojesti	mangiare	to eat
22	kontaktirati (*bias.*)	contattare	to contact
23	kontrolirati (*bias.*)	controllare	to control
24	kriti/sakriti	nascondere	to hide
25	kupovati/kupiti	acquistare	to buy
26	liječiti/izliječiti	curare	to heal
27	napredovati (*bias.*)	avanzare	to advance
28	naređivati/narediti	imporre	to order
29	nastavljati/nastaviti	proseguire	to continue
30	navoditi/navesti	citare	to quote
31	obavještavati/obavijestiti	avvisare	to apprise
32	objašnjavati/objasniti	precisare	to specify
33	objavljivati/objaviti	annunciare	to announce
34	odgovarati/odgovoriti	corrispondere	to correspond
35	odjekivati/odjeknuti	echeggiare	to echo
36	oduzimati/oduzeti	togliere	to subtract
37	okretati/okrenuti	girare	to turn
38	organizirati (*bias.*)	organizzare	to organise
39	osjećati/osjetiti	sentire*	to feel
40	osnovati/osnivati	fondare	to found

41	otkrivati/otkriti	rivelare	to reveal
42	ovlašćivati/ovlastiti	autorizzare	to authorise
43	označávati/označiti	segnare	to mark
44	padati/pasti	cadere	to fall
45	parkirati (*bias.*)	parcheggiare	to park
46	piti/popiti	bere	to drink
47	poboljšavati/poboljšati	migliorare	to improve
48	početi/započeti	cominciare	to commence
49	podnositi/podnijeti	subire	to endure
50	podučavati/podučiti	insegnare	to teach
51	podvrgnuti (*pf.*)	sottoporre	to submit
52	pokušavati/pokušati	tentare	to try
53	pomoći/pomagati	assistere	to assist
54	ponavljati/ponoviti	ripetere	to repeat
55	posjećivati/posjetiti	visitare	to visit
56	posjedovati (*impf.*)	possedere	to possess
57	posuđivati/posuditi	prestare	to lend
58	pozivati/pozvati	invitare	to invite
59	preferirati (*bias.*)	preferire	to prefer
60	prekidati/prekinuti	interrompere	to interrupt
61	preporučivati/preporučiti	consigliare	to advise
62	približavati/približiti	avvicinare	to approach
63	pripadati/pripasti	appartenere	to belong
64	prodavati/prodati	vendere	to sell

65	pronalaziti/pronaći	ritrovare	to find
66	pružati/pružiti	tendere	to tend
67	raditi/uraditi	funzionare	to work
68	rezervirati (*bias.*)	riservare	to book
69	silaziti/sići	scendere	to descend
70	siliti/prisiliti	costringere	to force
71	slati/poslati	mandare	to send
72	slijetati/sletjeti	atterrare	to land
73	slušati/poslušati	ascoltare	to listen
74	snimati/snimiti	riprendere	to shoot
75	spašavati/spasiti	salvare	to save
76	sprečavati/spriječiti	impedire	to hinder
77	stizati/stići	raggiungere	to reach
78	studirati (*bias.*)	studiare	to study
79	širiti/proširiti	diffondere	to spread
80	štovati/poštovati	rispettare	to respect
81	tumačiti/protumačiti	interpretare	to interpret
82	tutnjati (*impf.*)	rimbombare	to rumble
83	tužiti/optužiti	accusare	to accuse
84	ubijati/ubiti	uccidere	to kill
85	učiti/naučiti	imparare	to learn
86	ujedinjavati/ujediniti	unire	to unite
87	upravljati (*impf.*)	dirigere	to manage
88	upućivati/uputiti	avviare	to start

89	uvjeravati/uvjeriti	convincere	to convince
90	uzlaziti/uzaći	salire	to rise
91	voditi/provoditi	guidare*	to lead
92	voziti (*impf.*)	guidare*	to drive
93	vraćati se/vratiti se	rientrare	to return
94	vući/izvući	trarre	to draw
95	zaključivati/zaključiti	concludere	to conclude
96	zamišljati/zamisliti	immaginare	to imagine
97	završavati/završiti	finire	to finish
98	zvati/pozvati	chiamare	to call
99	žderati/požderati	divorare	to devour
100	željeti/poželjeti	desiderare	to wish

APPENDIX 2

A complete list of the abbreviations from Chapter 5.

ACT	activity
e	event
INGR	ingressive
P	process
S	state
T	transition
T_{acc}	accomplishment
T_{ach}	achievement

REFERENCES

Agrell, S. (1908). *Aspektänderung und Aktionsartbildung beim polnischen Zeitworte*. Häkan Ohlssons Buchdruckerei.

Babić, S. (2002). *Tvorba riječi u hrvatskome književnome jeziku*. Hrvatska akademija znanosti i umjetnosti i Nakladni zavod Globus.

Babić, Z. (1995). *Fonetika i fonologija hrvatskog jezika: Udžbenik za 1. Razred gimnazije i tehničkih škola*. Školska knjiga.

Baisa, V., Tiberius, C., Ježek, E., Colman, L., Marini, C., & Romani, E. (2020). Skema: A New Tool for Corpus-driven Lexicography. *Proceedings of the 19th EURALEX International Congress*, 523–528.

Balboni, P. (2020). *Fare educazione linguistica. Insegnare italiano, lingue straniere e lingue classiche*. UTET.

Barić, E., Lončarić, M., Malić, D., Pavešić, S., Peti, M., Zečević, V., & Znika, M. (2005). *Hrvatska Gramatika* (4th ed.). Školska knjiga.

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

Basile, G., Casadei, F., Lorenzetti, L., Schirru, G., & Thornton, A. M. (2010). *Linguistica generale*. Carrocci Editore.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, *57*(1), 289–300.

Bertinetto, P. M. (1986). *Tempo, aspetto e azione nel verbo italiano. Il sistema dell'indicativo*. Accademia della Crusca.

Bertinetto, P. M., & Delfitto, D. (2000). Aspect vs. Actionality: Why they should be kept apart. In *Tense and Aspect in the Languages of Europe*. Mouton de Gruyter.

Bertinetto, P. M., & Squartini, M. (1995). An attempt at defining the class of 'Gradual Completion Verbs'. In *Temporal Reference, Aspect, and Actionality. 1: Semantic and Syntactic Perspectives.* Rosenberg & Sellier.

Binnick, R. I. (1991). *Time and the verb: A guide to tense and aspect*. Oxford University Press.

Brown, H. D., & Lee, H. (2015). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Pearson.

Browne, W. (1993). Serbo-Croat. In *The Slavonic Languages*. Routledge.

Browne, W., & Alt, T. (2004). *A Handbook of Bosnian, Serbian and Croatian*. Slavic and Eurasian Language Resource Center (SEELRC).

Brozović, D. (1991). Fonologija hrvatskoga književnog jezika. In *Povijesni pregled, glasovi i oblici hrvatskoga književnog jezika*. HAZU – Globus.

Brozović, D. (1997). Narječja hrvatskoga jezika. In Hrvatski leksikon 2. Leksikografski zavod Miroslav Krleža.

Brugmann, K. (1904). *Kurze vergleichende Grammatik der indogermanischen Sprachen*. Trübner.

Canale, M., & Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, *1*, 1–47.

Chiari, I., & De Mauro, T. (2014). The New Basic Vocabulary of Italian as a Linguistic Resource. *Proceedings of the 1st Italian Conference on Computational Linguistics*.

Cinkova, S., Holub, M., Rambousek, A., & Smejkalova, L. (2012). A database of semantic clusters of verb usages. *Proceedings of the Eighth International Conference on Language Resources and Evaluation,* 3176–3183.

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge University Press.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.

Cvikić, L., & Jelaska, Z. (2007). Složenost ovladavanja glagolskim vidom u inojezičnome hrvatskome. *Lahor*, *2*(4), 190–216.

Čilaš Mikulić, M. (2012). *Verbal Aspect in Croatian as a Second and Foreign Language* [PhD]. University of Zagreb.

Čilaš Mikulić, M., Gulešić Machata, M., & Udier, S. L. (2021). *Razgovarajte s nama! Udžbenik hrvatskoga jezika za razine A1 i A2* (1st ed.). Hrvatska Sveučilišna Naklada.

Čilaš-Mikulić, M., Gulešić Machata, M., Pasini, D., & Udier, S. L. (2006). *Hrvatski za početnike 1: Udžbenik i rječnik (A1 i A2)*. Hrvatska Sveučilišna Naklada.

Dahl, Ö. (1985). *Tense and Aspect Systems*. Blackwell.

De Mauro, T. (1980). *Guida all'uso delle parole*. Editori Riuniti.

De Mauro, T. (2016). I*l Nuovo Vocabolario di Base della lingua italiana*. Available online at: https://www.dropbox.com/s/mkcyo53m15ktbnp/nuovovocabolariodibase.pdf?dl=0

Dowty, D. R. (1979). *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montagues PTQ.pdf*. D. Reidel Publishing Company.

Dörnyei, Z. (2007). Research Methods in Applied Linguistics: Quantitative, Qualitative and Mixed Methodologies. Oxford University Press.

Eller, W., & Hladczuk, J. (1992). *International Handbook of Reading Education*. Greenwood.

Fife-Schaw, C. (2006). Questionnaire Design. In *Research Methods in Psychology* (3rd ed.). Sage.

Filip, H. (1999). *Aspect, eventuality types, and nominal reference*. Garland.

Filipović, R. (1977). Some Basic Principles of Languages in Contact reinterpreted. *Studia Romanica et Anglica Zagrabiensia*, *43*, 157–166.

Fillmore, C. J. (1975). An Alternative to Checklist Theories of Meaning. Proceedings of the First *Annual Meeting of the Berkeley Linguistics Society*, 123–131.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift Für Philisophie Und Philosophische Kritik*, *100*, 25–50.

Gaio, S., Borgo, S., Masolo, C., Oltramani, A., & Guarino, N. (2010). Un'introduzione all'ontologia DOLCE. *AIDAinformazioni*, 1-2/2010.

Garey, H. B. (1957). Verbal aspects in French. *Language*, *33*(2), 91–110.

Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric Statistical Inference* (5th ed.). Chapman & Hall/CRC Press, Taylor & Francis Group.

Gojmerac, M. (1980). *Glagolski vid u hrvatskome ili srpskom i njemačkom jeziku.* University of Zagreb.

Gorbov, A. (2020). Aktionsart. *Encyclopedia of Slavic Languages and Linguistics Online*.

Gott, V., & Morpurgo, A. (2020). *Corso di lingua croata: Livelli A1-B1 del Quadro Comune Europeo di Riferimento per le Lingue*. Hoepli.

Graffi, G. (2019). *Due secoli di pensiero linguistico: Dai primi dell'Ottocento a oggi*. Carocci Editore.

Grgić, A., & Gulešić Machata, M. (2017). *Hrvatski A2: Opisni okvir referentne razine A2*. Filozofski fakultet Sveučilišta u Zagrebu, FF press.

Grimshaw, J. (1990). *Argument structure*. The MIT Press.

Ham, S. (2017). *Školska gramatika hrvatskoga jezika*. Školska knjiga.

Hanks, P. (2012). How People use words to make Meanings. Semantic Types meet Valencies. In *Input, Process and Product: Developments in Teaching and Language Corpora*. Masaryk University Press.

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. The MIT Press.

Hanks, P. (2017). Lexis. In *Oxford Handbook of Computational Linguistics* (2nd ed.). Oxford University Press.

Hanks, P. (1996). Contextual Dependency and Lexical Sets. *International Journal of Corpus Linguistics*, 1(1).

Hanks, P. (2004). Corpus Pattern Analysis. *Proceedings of the 11th EURALEX International Congress*.

Hanks, P., & Ježek, E. (2008). Shimmering lexical sets. *Proceedings of the 13th EURALEX International Congress*.

Hanks, P., & Pustejovsky, J. (2005). A Pattern Dictionary for Natural Language Processing. *Revue Française de Linguistique Appliquée,* 10, 63–82.

Hogg, R. V., & Wolfe, D. A. (1987). *Engineering Statistics*. Macmillan.

Hollander, M., & Wolfe, D. A. (1999). *Nonparametric Statistical Methods*. John Wiley & Sons, Inc.

Hui, C. H., & Triandis, H. C. (1989). Effects of Culture and Response Format on Extreme Response Style. J*ournal of Cross-Cultural Psychology*, 20(3).

Isačenko, A. V. (1968). *Die russische Sprache der Gegenwart. Formenlehre*. Halle.

Jelaska, Z. (2004). *Fonološki opisi hrvatskoga jezika: Glasovi, slogovi, naglasci.* Hrvatska Sveučilišna naklada.

Ježek, E. (2016). *The lexicon: An introduction*. Oxford University Press.

Ježek, E. (2018). Partecipanti impliciti nella struttura argomentale dei verbi. In *La Grammatica delle Valenza. Spunti teorici, strumenti e applicazioni*. Franco Cesati.

Ježek, E. (2019). Sweetening Ontologies Cont'd: Aligning bottom-up with top-down ontologies. P*roceedings of the Joint Ontology Workshops 2019*.

Ježek, E., Magnini, B., Feltracco, A., Bianchini, A., & Popescu, O. (2014). T-PAS: A resource of corpus-derived Typed Predicate-Argument Structures for linguistic analysis and semantic processing. *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*.

Ježek, E., Marini, C., & Romani, E. (*forthcoming*). Verb sense and Argument Semantics: From Linguistic Theory to Lexicographic practice. In *Patterns of meaning in lexicography and lexicology*. De Gruyter.

Ježek, E., & Quochi, V. (2010). Capturing Coercions in Texts: A First Annotation Exercise. *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC).*

Johnstone Young, T. (2016). Questionnaires and Surveys. In *Research Methods in Intercultural Communication: A Practical Guide*. Wiley Blackwell.

Kenny, A. (1963). *Action, Emotion And Will*. Routledge & Kegan Paul.

Kilgarriff, A. (1997). I Don't Believe in Word Senses. *Computers and the Humanities*, 31(2).

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36.

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., & Rychlý, P. (2008). GDEX: automatically finding good dictionary examples in a corpus. *Proceedings of the XIII EURALEX International Congress*, 425–432.

Kolaković, Z. (2018). *Dvoaspektni glagoli – razlike između (p)opisa u priručnicima i stanja u korpusu s posebnim osvrtom na uporabu izvornih govornika* [PhD]. University of Zagreb & University of Regensburg.

Kolaković, Z. (2020). Par – nepar – aspektni par*. *Jezikoslovlje*, *21*(2), 103–147.

Kordić, S. (1997). *Serbo-Croatian*. Lincom Europa.

Landau, E., Lončarić, M., Horga, D., & Škarić, I. (1999). Croatian. In *Handbook of the International Phonetic Association: A Guide To The Use Of The International Phonetic Alphabet*. Cambridge University Press.

Levin, B. & Rappaport Hovav M. (2005) *Argument realization*. Cambridge University Press.

Łazorczyk, A. A. (2010). *Decomposing Slavic aspect: The role of aspectual morphology in Polish and other Slavic languages* [PhD]. University of Southern California.

Ljubešić, N., & Klubička, F. (2014). {bs, hr, sr} WaC – web corpora of Bosnian, Croatian and Serbian. *Proceedings of the 9th Web as Corpus Workshop*, 29−35.

Malić, D. (2005). Razvoj hrvatskoga književnog jezika. In *Hrvatska gramatika*. Školska knjiga.

Marini, C. (2020). *CROATPAS: A Resource of Corpus-derived Typed Predicate Argument Structures for Croatian* [Master's]. University of Pavia.

Marini, C. (2022b). CroaTPAS: A Survey-based Evaluation. *Proceedings of The18th Joint ACL - ISO Workshop on Interoperable Semantic Annotation (ISA-18)*.

Marini, C. (2022a). CroaTPAS: uspoređivanje značenja vidskih parnjaka s namjerom istraživanja odnosa između vida, aktionsarta i glagolske polisemije u hrvatskom. *Croatica: Časopis Za Hrvatski Jezik, Književnost i Kulturu*, 46(66), 67–80.

Marini, C., & Ježek, E. (2019). CROATPAS: A Resource of Corpus-derived Typed Predicate Argument Structures for Croatian. *Proceedings of the 6th Italian Conference on Computational Linguistics (CLiC-It)*.

Matovac, D. (2022). *Basic Croatian Grammar: For Croatian Language Learners*. Hrvatska sveučilišna naklada.

Matovac, D. (2014). The semantics of the preposition po in Croatian language. *Croatica et Slavica Iadertina*, *10/1*(10), 3–29.

Mereu, L. (2020). *Semantica della frase*. Carocci Editore.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11), 39–41.

Moens, M., & Steedman, M. (1988). Temporal ontology and temporal reference. *Computational Linguistics*, *14*(2), 15–28.

Naylor, K. E. (1980). Serbo-Croatian. In *The Slavic literary languages—Formation and development*. Yale Concilium on International and Area Studies.

Novak Milić, J. (2010). Što je što u aspektologiji. *Lahor: Časopis Za Hrvatski Kao Materinski, Drugi i Strani Jezik*, *2*(10), 125–143.

Piasevoli, V., & Vranješ, M. (2014). La didattica del croato e del serbo nella Sezione di Studi di Lingue Moderne per Interpreti e Traduttori dell'Università di Trieste. *Quaderni CIRD*, *9*.

Polančec, J. (2015). Što nam Aktionsart glagola može reći o njegovu vidu. In *Višejezičnost kao predmet multidisciplinarnih istraživanja*. Srednja Europa i Hrvatsko društvo za primijenjenu lingvistiku (HDPL).

Polančec, J. (2020). *A Typology of Aspect-Actionality Interactions* [PhD]. University of Zadar.

Polančec, J. (2018). Osamostaljeni izvedeni nesvršeni glagoli u hrvatskom jeziku. *Suvremena Lingvistika*, *44*, 113–138.

Polančec, J. (2022). Vidski neupareni glagoli u hrvatskom jeziku. *Croatica: Časopis za hrvatski jezik, književnost i kulturu*, 46(66), 105-133.

Pustejovsky, J. (1991). The Syntax of Event Structure. *Cognition*, 41, 47-81.

Pustejovsky, J. (1995). *The Generative Lexicon*. The MIT Press.

Pustejovsky, J., Hanks, P., & Rumshisky, A. (2004). Automated Induction of Sense in Context. *Proceedings of the 20th International Conference on Computational Linguistics (COLING)*.

Pustejovsky, J., & Ježek, E. (2012). *Introducing Qualia Structure*. University of Pavia, Brandeis University.

Pustejovsky, J., & Ježek, E. (2008). Semantic Coercion in Language: Beyond Distributional Analysis. *Rivista Italiana di Linguistica*, 20(1), 123–138.

Pustejovsky, J., & Stubbs, A. (2012). N*atural Language Annotation for Machine Learning*. O'Reilly.

Rappaport Hovav, M. (2008). Lexicalized meaning and the internal temporal structure of events. In *Crosslinguistic and Theoretical Approaches to the Semantics of Aspect*. John Benjamins.

Rappaport Hovav, M. & Levin, B. (1998). Building Verb Meanings. In: *The Projection of Arguments: Lexical and Compositional Factors*. CSLI Publications.

Recanati, F. (2012). Compositionality, flexibility, and context-dependence. In: *The Oxford Handbook of Compositionality*. Oxford University Press.

Renau, I., & Nazar, R. (2011). Propuesta metodológica para la creación automática de patrones léxicos usando el Corpus Pattern Analysis. *Actas Del 27° Congreso de La SEPLN.*

Richard, J., & Burns, A. (2012). *Tips for teaching listening: A practical approach*. Pearson Education.

Rosner, B. (1983). Percentage Points for a Generalized ESD Many-Outlier Procedure. *Technometrics*, 25, 165–172.

Rundell, M. (2002). *Macmillan English Dictionary for Advanced Learners*. Macmillan Education.

Ryle, G. (1949). *The Concept of Mind*. Hutchinson's University Library.

Silić, J. (1978). An approach to the study of aspectuality in the Croatian literary language. In *Contrastive analysis of English and Serbo-Croatian* (pp. 42–70).

Silić, J., & Pranjković, I. (2007). *Gramatika hrvatskoga jezika: Za gimnazije i visoka učilišta* (2nd ed.). Školska knjiga.

Simone, R. (2013). *Nuovi fondamenti di linguistica*. McGraw-Hill.

Sinclair, J. (1991). *Corpus, Concordance, Collocation.* Oxford University Press.

Smith, C. (1991). *The parameter of aspect*. Springer.

Sussex, R., & Cubberley, P. (2006). *The Slavic Languages*. Cambridge University Press.

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

Špikić, A. (2017). *Croato compatto: Dizionario croato/italiano e italiano/croato*. Zanichelli.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Klincksieck.

Udier, S. L. (2016). The Croatian language at the beginning of the 21st century. In *Croatia at First Sight. Textbook of Croatian Culture*. FF press.

Udier, S. L. (2017). Gramatička kompetencija. In *Hrvatski A2: Opisni okvir referentne razine A2*. FF press.

Van Valin, R. D., & LaPolla, R. J. (1997). *Syntax: Structure, Meaning, and Function*. Cambridge University Press.

Vendler, Z. (1957). Verbs and Times. *The Philosophical Review*, *66*(2), 143–160.

Verkuyl, H. J. (1972). *On the Compositional Nature of the Aspects*. D. Reidel Publishing Company.

Verkuyl, H. J. (1993). *A Theory of Aspectuality*. Cambridge University Press.

Zumel, N., & Mount, J. (2014). *Practical Data Science with R.* Manning.