Proceedings of the 16th Joint ACL-ISO Workshop Interoperable Semantic Annotation (ISA-16), pages 49–58
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Annotating Croatian Semantic Type Coercions in CROATPAS

Costanza Marini, Elisabetta Ježek
University of Pavia Department of Humanities
Corso Strada Nuova, 65, 27100 Pavia PV
costanza.marini93@gmail.com, jezek@unipv.it

Abstract

This short research paper presents the results of a corpus-based metonymy annotation exercise on a sample of 101 Croatian verb entries – corresponding to 457 patters and over 20,000 corpus lines – taken from CROATPAS (Marini & Ježek, 2019), a digital repository of verb argument structures manually annotated with Semantic Type labels on their argument slots following a methodology inspired by Corpus Pattern Analysis (Hanks, 2004 & 2013; Hanks & Pustejovsky, 2005). CROATPAS will be made available online in 2020. Semantic Type labelling is not only well-suited to annotate verbal polysemy, but also metonymic shifts in verb argument combinations, which in Generative Lexicon (Pustejovsky, 1995 & 1998; Pustejovsky & Ježek, 2008) are called Semantic Type coercions. From a sub lexical point of view, Semantic Type coercions can be considered as exploitations of one of the *qualia* roles of those Semantic Types which do not satisfy a verb's selectional requirements, but do not trigger a different verb sense. Overall, we were able to identify 62 different Semantic Type coercions linked to 1,052 metonymic corpus lines. In the future, we plan to compare our results with those from an equivalent study on Italian verbs (Romani, 2020) for a crosslinguistic analysis of metonymic shifts.

Keywords: Semantic Type coercion, Croatian, metonymy

1. Introduction

If we look at the lexicon in its whole, it is possible to identify systematic alternations of meaning that apply not only to single lexical instances but entire classes of words, i.e. patterns of so-called *regular polysemy* (Apresjan, 1973). Some common alternations are *author/work*; *product/producer*; *event/food* or *container/content*.

When dealing with these alternations, however, it is necessary to distinguish between *metonymic* and *inherent polysemy*. In *metonymic shifts*, meaning is extended by conceptual contiguity and a change of referent is required, since one entity is used to denote another which is conceptually associated with it (Ježek, 2016: 59). This is the case, for instance, of the alternation *container/content*, exemplified by sentences such as "I would have eaten the whole fridge", where *fridge* actually stands for the food it contains.

In the case of *inherent polysemy*, on the other hand, there is no sense extension nor change of referent, but only one ontologically complex entity. This is the case, for instance, of alternations such as *information source/ artifact* as in "The book I am reading weighs one kilo" (Pustejovsky & Ježek, 2008: 185), where the lexical item *book* can be understood at the same time as the information it contains and a heavy object. The possibility for more than one of the senses of a complex entity to be activated simultaneously is called *co-predication* and is a prerogative of inherently polysemous words.

In this paper, we are going to present the first results of a metonymy annotation exercise on a sample of Croatian verbs taken from the Croatian Typed Predicate Argument Structures resource (CROATPAS, Marini & Ježek, 2019) (see section 2.1). Since the resource rests on Generative Lexicon Theory (Pustejovsky, 1995 & 1998; Pustejovsky & Ježek, 2008), metonymies are annotated and analysed as Semantic Type Coercions (see section 2.2). The set of semantic labels used for the annotation and the sample choice are covered in section 2.3 and 2.4, respectively.

2. Methodology

2.1 The CROATPAS resource

CROATPAS (Marini & Ježek, 2019) – short for Croatian Typed Predicate Argument Structure resource – is a digital dictionary of Croatian verbs focusing on verbal polysemy, which is currently being developed at the University of Pavia[1] next to its Italian sister project TPAS (Ježek et al., 2014). CROATPAS consists in a repository of verb valency structures whose argument slots have been manually annotated with a set of semantic labels called Semantic Types (henceforth *SemTypes*), following a corpus-based lexicographic methodology inspired by Corpus Pattern Analysis (CPA, Hanks, 2004 & 2013; Hanks & Pustejovsky, 2005).

From a theoretical point of view, CPA rests on the Theory of Norms and Exploitations (TNE, Hanks 2004 & 2013), which differentiates between two types of word uses: conventional ones – the *norms* – and deviations from such norms – the *exploitations*. When applying CPA, lexicographers traditionally focus on identifying normal word usage by mapping standard meanings onto their syntagmatic patterns of use.

In CROATPAS, our CPA-inspired methodology consists in the following four steps: 1) sampling 250 random concordances from a representative corpus of Standard Croatian for each verb entry, namely the Croatian Web as Corpus (Ljubešić & Klubička, 2014); 2) manually disambiguating its different senses and 3) associating the right SemTypes to the argument slots found in each sense-bound valency structure. The fourth and last step is only possible thanks to our editing environment SKEMA, which is connected to the Croatian Web as Corpus through the *Sketch Engine* corpus management platform (Kilgarriff et al., 2014) and enables annotators to create *patterns* for each retrieved verb sense, such as the ones in Figure 1.

[1] Its first release will contain approximately 200 Croatian verb entries and will be accessible by 2020 on the website of University of Pavia: https://cla.unipv.it/?page_id=53723.

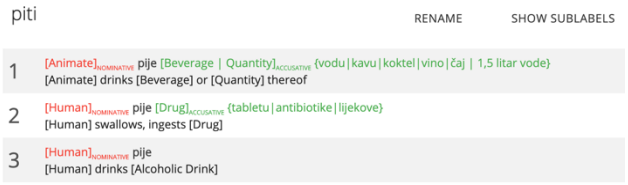

Figure 1: The first 3 patterns from the Croatian verb *piti* (English, *to drink*)

As you can see from the patterns above, the first sense of the Croatian verb *piti* (English, *to drink*) is the most obvious one, namely that of an [Animate] *drinking* a [Beverage]. However, if a [Human] is told to be drinking a [Drug] – such as a pill or antibiotics (Croatian, *tabletu* and *antibiotike*) – then he or she is simply *ingesting* or *swallowing* them. Finally, if we talk of a [Human] drinking (without specifying any direct object), he or she is by default *ingesting an alcoholic drink*.

2.2 Annotating Semantic Type Coercions

In addition to verbal polysemy, CROATPAS also allows lexicographers to annotate metonymic arguments by adding specific sub patterns to existing verb senses (see Figure 2).

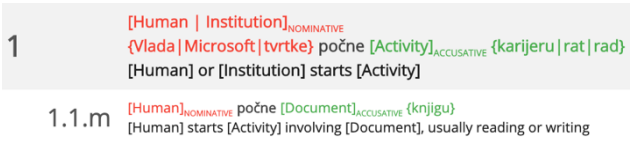

Figure 2: Pattern 1 and its metonymic sub pattern 1.1.m from the Croatian verb *početi* (English, *to begin*)

Despite involving the same verb sense as pattern 1, the metonymic sub pattern 1.1.m is linked only to those concordance lines where there is a mismatch in the SemType of the direct object: namely [Document] instead of [Activity]. This mismatch signals that a metonymic shift is taking place, which in Generative Lexicon Theory takes the name of Semantic Type coercion (Pustejovsky & Ježek, 2008; Ježek & Quochi, 2010). In order to explain this concept, let us look at a couple of sentences provided by Pustejovsky (1995: 115-6) starring a good translational equivalent of the Croatian verb *početi*, namely:

(1) *John began reading a book.*
(2) *John began a book.*

In sentence (1), the verb's second argument – i.e. *reading a book* – denotes an [Activity], whereas in sentences (2) it denotes a [Document] – *a book*. We call Semantic Type Coercion the compositional mechanism which enables us to reconstruct the semantics of the second direct object by forcing – i.e. *coercing* – [Document] into an [Activity] denotation. As pointed out by Ježek & Quochi (2010: 1465), coercion always involves an attested Source Type (e.g. [Document]) which is coerced into a Target Type to fit the verb's selectional requirements (e.g. [Activity]). The shift can involve any argument slot and is graphically represented as follows: [Document] → [Activity].

2.2.1 *Qualia* Exploitation

This being said, if we look at Semantic Type Coercions from a sub lexical point of view, they can be considered exploitations[2] of one of the available *qualia* roles associated with the Source Type not satisfying the verb's selectional requirements (Pustejovsky & Ježek, 2008: 195).

Qualia structure is one of the four levels of representation involved in the computational apparatus of Generative Lexicon (Pustejovsky, 1995 & 1998) and it consists of the four most important semantic properties of any lexical item: its Formal, Constitutive, Telic and Agentive *qualia*. The term *qualia* comes from Latin and is the plural of the word *quale*, which means "what kind?".

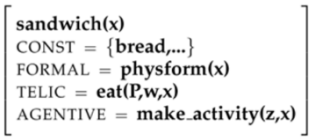

Figure 3: The *qualia* structure of the noun *sandwich* (Pustejovsky & Ježek, 2008: 185)

As we can see in Figure 3, the Constitutive *quale* consists of all the parts that make up the entity we are dealing with – in this case, the sandwich's ingredients. The Formal *quale* answers to the question "What sort of thing is this?" – in this case, a [Physical Entity]. The Telic *quale* – from the Greek word télos, i.e. *end* – expresses the function of the entity denoted by our lexical item – which, for a sandwich, is being eaten. Last but not least, the Agentive *quale* specifies the entity's origin.

If we look at the metonymic sub pattern 1.1.m from Figure 2 under this new light, the Semantic Type Coercion [Document] → [Activity] can be interpreted as an exploitation of either the Telic *quale* "reading" or the Agentive quale "writing", both associated with the qualia structure of any document, since we write so that others can read. It will be the broader context to assign the correct interpretation.

2.3 The System of Semantic Type labels

The list of SemTypes used in CROATPAS is taken from the Italian TPAS resource (Ježek et al., 2014) and belongs to the TPAS ontology (Ježek, 2019), a hierarchically organised set of labels originating from the Brandeis Shallow Ontology (Pustejovsky et al., 2004) currently containing 180 bracketed labels, such as [Human], [Document], and so forth.

Despite looking like ontological categories, SemTypes are semantic classes obtained by "manual clustering and generalization over sets of lexical items found in the argument positions" in valency structures taken from large corpora (Ježek et al, 2014: 891). They are thus able to mirror the way humans talk about entities, states and events through language.

According to Generative Lexicon, SemTypes can be divided into three groups depending on their internal structure:

[2] Be aware that term *exploitation* in this paper may refer to two different frameworks: in section 2.1 it falls within Hank's Theory of Norms and Exploitations, while in section 2.2.1 and 2.3 we generally use it in the expression "*qualia* exploitation", which pertains to Generative Lexicon terminology.

1) *Natural Types* referring to natural concepts characterised only by a Formal and a Constitutive *quale*, e.g. [Animal] or [Natural Landscape Feature];
2) *Artifactual* or *Tensor Types* denoting man-made entities usually possessing also a Telic and an Agentive *quale* to express their purpose and origin, e.g. [Beverage];
3) *Complex Types* characterised by multiple Semantic Types clustered together and normally used to denote inherently polysemous lexical items, e.g. [Institution].

If Tensor Types are characterised by an asymmetrical structure linking their head SemType to a component of its *qualia* structure, as in [Beverage ⊗ Telic Activity (*drinking*)], Complex Types are generally internally symmetrical, as in the case of [Institution = Human Group • Abstract Entity]. Since a dot is used to link together their components, Complex Types are also called Dot Objects.

Artifactual Types are those usually instantiating metonymic shifts via Qualia Exploitation whereas, Complex Types can either allow for *co-predication* or, when only one of their senses is used, for Dot Exploitation. Since differentiating between Qualia Exploitation and Dot Exploitation is not always clear-cut, the TPAS ontology (Ježek, 2019) keeps track of all acknowledged Complex Types by treating them as cases of *multiple inheritance*, i.e. by anchoring them to multiple positions within the SemType hierarchical system as in Figure 4, where [Institution] inherits from both [Abstract Entity] and [Human Group].

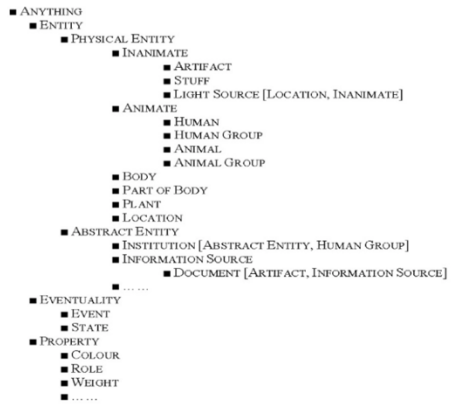

Figure 4: The top-level of the TPAS system (Ježek 2019)

2.4 Verb choice

The verb sample[3] we concentrated on for this metonymy annotation study consists of 44 Croatian aspectual verb pairs[4] and 13 biaspectual verbs taken from the CROATPAS resource (Marini & Ježek, 2019), for a total of 101 verb entries linked to 457 different patterns.

Half of the sample is made up of the Croatian translational equivalents of a sample of Italian verbs known to trigger Semantic Type Coercions, the so-called *coercive verbs* analysed by Ježek & Quochi (2010); while the other half are the Croatian translational equivalents of a selection of Italian verbs belonging to the language's fundamental vocabulary (FO), i.e. a group of 2,000 words with the highest frequency counts covering about 90% of all Italian written and spoken text (Chiari & De Mauro, 2014: 113). All Croatian translational equivalents were selected consulting the Zanichelli Italian/Croatian bilingual dictionary *Croato compatto* (Aleksandra Špikić, 2017).

3. Results

As a result of our metonymy annotation exercise, we were able to enrich the 457 patterns stored in CROATPAS adding 106 metonymic sub patterns. The metonymic corpus lines justifying these sub patterns are 1,052, a number which is already included in the over 22,000 annotated corpus lines currently linked to the resource.

Patterns	*Sub patterns*	*Tagged corpus lines*	*Metonymic corpus lines*
457	106	22,052	1,052

Table 1: Patterns, sub patterns and corpus lines

This being said, the Reader should keep in mind that the number of metonymic sub patterns does not equal the number of identified Semantic Type coercions (see Appendix 2 for the full inventory). Since different metonymic shifts can occur in the same pattern and even in the same argument slot, we decided to encode them – when possible – within the same sub pattern, as in Figure 5.

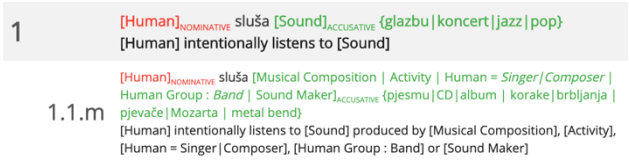

Figure 5: Pattern 1 from the verb *slušati* (English, *to listen*) and its metonymic sub pattern 1.1.m

As we can see above, 5 different Semantic Type coercions are nested within the same sub pattern, namely [Musical Composition] → [Sound], [Activity] → [Sound], [Human = Singer | Composer] → [Sound], [Human Group: Band] → [Sound] and [Sound Maker] → [Sound]. Each of them counts as an instance of the Semantic Type Coercion they stand for, which might have other instances in other sub patterns. All of the coercion instances above occur on the direct object slot of pattern 1 of the verb *slušati* (English, *to listen to*) and are justified by corpus examples such as

[3] See Appendix 1 for a complete list of all the CROATPAS verbs in our sample, together with their TPAS counterparts and English equivalents. In the Italian list, the verbs *sentire* and *guidare* appear twice because 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. On the other hand, one of the verbs from the original list of Ježek & Quochi (2010) has not been taken into account because its Croatian translational equivalent was deemed too polysemous, namely *ići* (Italian, *recarsi*; English *to go*).

[4] Since Croatian is a Slavic language, we usually deal with verb pairs made up of a perfective and imperfective variant, for instance *piti/popiti* (imperfective/perfective - English, *to drink*). All variants are treated and annotated as independent verb entries, in order to collect corpus-based evidence to evaluate to what extent verb meaning depends on aspectual differences.

the ones in Figure 6. We are going to focus on the three highlighted ones.

Figure 6: Corpus lines linked to sub pattern 1.1.m of the CROATPAS verb entry *slušati* (English, *to listen to*)

In the sentence *"U rodilištima bebe slušaju Mozarta i Vivaldija"* (English, *"In maternity wards, babies listen to Mozart and Vivaldi"*), we have two examples of the classic metonymy *author/work*, which in our framework translates to the Semantic Type coercion [Human = Composer] → [Sound]. The same applies to *"Slušam neki metal bend"* (English, *"I am listening to a certain metal band"*), where it is not the group but the music they play that is being listened to, thus giving rise to the coercion [Human Group: Band] → [Sound]. Finally, in *"Slušam korake izgubljenih ljubavnika"* (English, *"I listen to the footsteps of lost lovers"*), the direct object we should be "listening to" is *footsteps*, a lexical item that according to our ontology can be labelled as an [Activity]. However, it is only the [Sound] of said activity which can be heard, thus justifying the coercion [Activity] → [Sound].

3.1 The most frequent Semantic Type coercions

In our annotation exercise, we managed to identify a total of 179 Semantic Type coercions of 62 different kinds (see Appendix 2 for the full list). Table 2 portrays the 15 most frequent coercions in our inventory. Since we did not extract the number of corpus lines each Semantic Type coercion is exemplified by, the figures in the third column report the *coercion instances*, i.e. the amount of times each coercion appears in a different sub pattern or in a different argument slot within the same sub pattern.

Rank	*Semantic Type Coercion*	*Coercion instances*
1	Area > Human Group	25
2	Area > Institution	21
3	Area > Human Group: Football Team	6
4	Artifact > Activity	6
5	Business Enterprise > Road Vehicle	6
6	Musical Composition > Sound	6
7	Concept > Human Group	5
8	Sound Maker > Sound	5
9	Activity > Sound	4
10	Beverage > Activity	4
11	Building > Activity	4
12	Event > Location	4
13	Food > Activity	4
14	Bomb > Sound	4
15	Document > Activity	3

Table 2: Our 15 most frequent Semantic Type coercions

As we can see from the data, the most frequently annotated Semantic Type coercion in our sample happens to be [Area] → [Human Group], which makes up for 25 out of the 179 attested occurrences of our 62 different Semantic Type Coercions. As for the second and the third most frequent coercions, we can say that they not only share the same Source Type as the most frequent one, but their Target Types are also somewhat hierarchically related, since [Human Group] is one of the constituents of the Complex Type [Institution] and [Football Team] is a hyponym of [Human Group]. The metonymic sub pattern 2.1.m in Figure 7 encoding the Semantic Type coercion [Area] → [Human Group: Football Team] will give us an idea of how this specific coercion works.

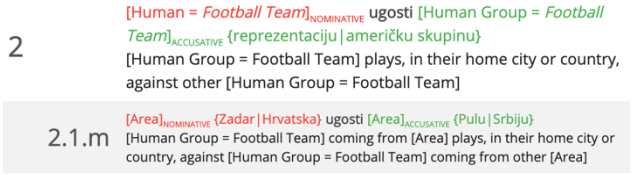

Figure 7: Pattern 2 and its metonymic sub pattern 2.1.m from the Croatian verb *ugostiti* (English, *to host*)

When saying a sentence like "*Hrvatska će ugostiti Srbiju u četvrtfinalu*" (which translates to "*Croatia will host Serbia for the quarter final"*), the SemType [Area] is coerced into a [Football Team], since what the speaker actually means is that the Croatian national team will play against the Serbian one, and not the respective geographical areas.

3.2 The most *coercive* Croatian verbs

The CROATPAS verbs giving rise to the most Semantic Type coercions are the following: *tutnjati* (English, *to rumble*) with 11 coercions to be traced back to only 2 observed patters; *odjekivati* (English, *to* echo) with 10 coercions and only 3 patterns; *okrenuti* (English, *to turn*) with 9 SemType coercions and 16 patterns, followed by both the perfective and imperfective variant of the Croatian equivalent of *to listen* – namely *slušati* and *poslušati* – both with 3 recorded senses and 9 metonymic sub patterns each.

Since after these first five verbs the number of SemType coercions drastically diminishes to 5 or less for the rest of the sample, it is not unreasonable to suggest that *verbs of hearing* are particularly well suited to trigger metonymic shifts within their valency structure. To give an idea of the mechanisms at play in these sound-focused coercions, take a look at Figure 8.

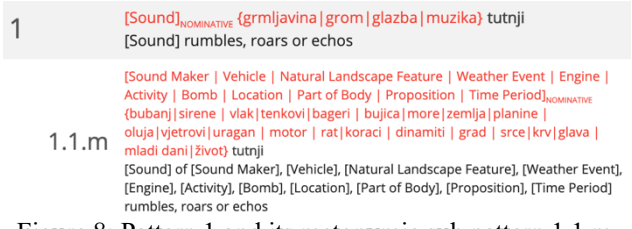

Figure 8: Pattern 1 and its metonymic sub pattern 1.1.m from the Croatian verb *tutnjati* (English, *to rumble*)

As we can see, pattern 1.1.m lists all the SemTypes of the entities whose sound can rumble, roar or echo (e.g. [Vehicle], [Weather Event], [Engine], [Sound Maker]…) and provides also some particularly well-suited examples between square brackets, such as *vlak* (English, *train*), *oluja* (English, *storm*) and motor (English, *engine*). In all of these instances, a *qualia* role of the entity in object

position is exploited and coerced into a [Sound], like in the case of *sirene* (English, *sirens*), whose Telic *quale* is "producing a sound".

3.3 Semantic Type coercions and clause roles

If we look at the clause roles where Semantic Type coercions take place (see Table 3), we can see that approximately half of the observed metonymic shifts take place in the subject slot, nearly 40% involves the verb's direct object and 14% indirect complements.

Argument slots	*Coercion instances*	*Coercion %*
Subject	85	47.5 %
Object	69	38.5 %
Indirect complements	25	14 %
Total	179	100 %

Table 3: Semantic Type Coercions by clause roles

Even though subjects, objects and indirect complements are not equally distributed across the verb sample, the percentages in Table 2 still demonstrate that all argument slots can be good candidates for metonymies to take place.

3.4 Source Types and Target Types

As previously mentioned, Semantic Type Coercions can also be analysed in terms of Source Type and Target Type (Ježek & Quochi: 2010). As we could have already guessed from the most coercive verbs mentioned in section 3.2, the most frequent Target Type is [Sound], which appears in 39 Semantic Type coercions instances out of 179. The second most frequent Target Type is [Human Group] (30 instances), followed by [Activity] (29) and [Institution] (20), which – if considered as a hyponym of [Human Group] – would actually cause the latter to become the most frequent Target Type overall.

As for Source Types, as it was to be expected from the data in Table 2, the most frequent Type is [Area], appearing in 53 coercion instances, followed by [Human] (16 instances) and both [Activity] and [Business Enterprise] at 11. Since [Event] – hypernym of [Activity] – is used as Source Type in 7 more Semantic Type coercions, it might be worth looking at an example. We are talking, for instance, of alternations like [Activity] → [Sound], which are triggered by words such as *korake* (English, *steps*) when used as direct objects of verbs such as *slušati* (English, *to listen*).

4. Conclusions

In this paper, we have presented the first results of a metonymy annotation exercise on a sample of 101 Croatian verb entries taken from the semantic resource CROATPAS (Marini & Ježek, 2019), a digital repository of verb argument structures manually annotated with Semantic Type labels on their argumental structure. At present, the resource contains 457 patterns and 106 metonymic sub patterns. The overall number of annotated corpus lines is 22,052, of which 1,052 are linked to the 106 metonymic sub patterns they provide evidence for. We explained the mechanism underlying how metonymy works in our chosen framework and provided an overview of the set of semantic labels we used, together with a clarification of our verb choice. Our results show that [Area] → [Human Group] proves to be our most frequent Semantic Type Coercion, appearing 25 out of 179 times. Sound verbs such as *tutnjati* (English, *to rumble*), *odjekivati* (English, *to echo*) and *slušati*/*poslušati* (English, *to listen*) position themselves amongst the most coercive verbs in the sample: a result supported also by the fact that the most frequent Target Type, appearing in 39 coercion instances out of 179, is [Sound]. On the other hand, the most frequent Source Type is [Area], a finding which agrees with the data on the most frequent Semantic Type coercions overall. From a tentative analysis of clause role predisposition to Semantic Type Coercion, all argument slots seem to be able to enable the shift. In order to give a stronger claim to our results and evaluate the CROATPAS resource, we plan on involving other annotators and devise a task to measure the degree of Inter Annotator Agreement. Once evaluated, we believe that our inventory of manually annotated metonymic corpus lines could be used as training data to develop an automatic metonymy recognition method. Current on-going work is focussed on comparing our results with an equivalent annotation performed in the TPAS resource on the set of Italian verbs which corresponds to the first half of our Croatian sample (Romani, 2020). We expect this comparison to provide crosslinguistic insights on the linguistic and cognitive basis of metonymic shifts.

5. References

Apresjan, Y.D. (1973). Regular polysemy. In: *Linguistics*, 142, pp. 5-32.

Chiari, I. & De Mauro, T. (2014). The New Basic Vocabulary of Italian as a Linguistic Resource. In: *Proceedings of the 1st Italian Conference on Computational Linguistics (CLiC-it)*. Pisa, Italy.

Hanks, P. (2004). Corpus Pattern Analysis. In: *Proceedings of the 11th EURALEX International Congress*. Lorient, France.

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. Cambridge: The MIT Press.

Hanks, P. & Pustejovsky, J. (2005). A Pattern Dictionary for Natural Language Processing. In: *Revue française de linguistique appliquée*, 10 (2), pp. 63-82.

Ježek, E. (2016). *The lexicon: An introduction*. Oxford: Oxford University Press.

Ježek, E. (2019). Sweetening Ontologies Cont'd: Aligning bottom-up with top-down ontologies. In: *Proceedings of CREOL 2019*. Graz, Austria.

Ježek, E. & Quochi, V. (2010). Capturing Coercions in Texts: A First Annotation Exercise. In: *Proceedings of the 7th conference on International Language Resources and Evaluation (LREC)*. Valletta, Malta.

Ježek, E., Magnini, B., Feltracco, A., Bianchini, A., Popescu, O. (2014). T-PAS: A resource of Typed Predicate Argument Structures for linguistic analysis and semantic processing. In: *Proceedings of the 9th conference on International Language Resources and Evaluation (LREC)*. Reykjavik, Iceland.

N. Ljubešić & F. Klubička (2014). {bs, hr, sr} WaC – web corpora of Bosnian, Croatian and Serbian. In: *Proceedings of the 9th Web as Corpus Workshop*, pp. 29–35.

Kilgarriff, O., Baisa, V., Bušta, J., Jakubíček, M., Kovár, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: ten years on. In: *Lexicography 1(1)*, pp. 7-36.

Marini, C. & Ježek, E. (2019) CROATPAS: A Resource of Corpus-derived Typed Predicate Argument Structures

for Croatian. In: *Proceedings of the 6th Italian Conference on Computational Linguistics (CLiC-it).* Bari, Italy.
Pustejovsky, J. (1995). *The Generative Lexicon.* Cambridge: The MIT Press.
Pustejovsky, J. (1998). The semantics of lexical underspecification. In: *Folia Linguistica* 32.
Pustejovsky, J. & Ježek, E. (2008). Semantic Coercion in Language: Beyond Distributional Analysis. In: *Italian Journal of Linguistics*, vol. 20, pp. 181-214.
Romani, E. (2020). *Searching for Metonymies in Natural Language Text. A corpus-based study on a resource for Italian for Verbs.* BA Thesis, University of Pavia.
Špikić, A. (2017). *Croato compatto: dizionario croato/italiano e italiano/croato*, Zanichelli: Bologna.

Appendix 1:

The Croatian verb entries from CROATPAS used for our Semantic Type Coercion exercise[5]

	CROATPAS	TPAS	English translations
1	bacati/baciti	lanciare	to throw
2	čitati/pročitati	leggere	to read
3	čuti	sentire*	to hear
4	čuvati/očuvati	conservare	to preserve
5	dirati/dirnuti	toccare	to touch
6	djelovati	agire	to act
7	dočekivati/dočekati	accogliere	to welcome
8	dolaziti/doći	arrivare	to arrive
9	dovršavati/dovršiti	completare	to complete
10	gostiti/ugostiti	ospitare	to accommodate
11	informirati	informare	to inform
12	isključivati/isključiti	escludere	to exclude
13	jesti/pojesti	mangiare	to eat
14	kontaktirati	contattare	to contact
15	kriti/sakriti	nascondere	to hide
16	liječiti/izliječiti	curare	to heal
17	napredovati	avanzare	to advance
18	obavještavati/obavijestiti	avvisare	to apprise
19	objašnjavati/objasniti	precisare	to specify
20	objavljivati/objaviti	annunciare	to announce
21	odjekivati/odjeknuti	echeggiare	to echo
22	okretati/okrenuti	girare	to turn
23	organizirati	organizzare	to organise
24	osjećati/osjetiti	sentire*	to feel
25	osnovati/osnivati	fondare	to found
26	padati/pasti	cadere	to fall
27	parkirati	parcheggiare	to park
28	piti/popiti	bere	to drink
29	početi/započeti	cominciare	to commence
30	podvrgnuti	sottoporre	to submit

[5] Verbs marked by an asterisk (*) appear twice.

31	pokušavati/pokušati	tentare	to try
32	posjećivati/posjetiti	visitare	to visit
33	posuđivati/posuditi	prestare	to lend
34	preferirati	preferire	to prefer
35	prekidati/prekinuti	interrompere	to interrupt
36	preporučivati/preporučiti	consigliare	to advise
37	približavati/približiti	avvicinare	to approach
38	pripadati/pripasti	appartenere	to belong
39	raditi/uraditi	funzionare	to work
40	rezervirati	riservare	to book
41	slijetati/sletjeti	atterrare	to land
42	slušati/poslušati	ascoltare	to listen
43	snimati/snimiti	riprendere	to shoot
44	spasavati/spasiti	salvare	to save
45	stizati/stići	raggiungere	to reach
46	tutnjati	rimbombare	to rumble
47	tužiti/optužiti	accusare	to accuse
48	ubijati/ubiti	uccidere	to kill
49	ujedinjavati/ujediniti	unire	to unite
50	upravljati	dirigere	to manage
51	uzlaziti/uzaći	salire	to rise
52	voditi/provoditi	guidare*	to lead
53	voziti	guidare*	to drive
54	zaključivati/zaključiti	concludere	to conclude
55	završavati/završiti	finire	to finish
56	žderati/požderati	divorare	to devour
57	zvati/pozvati	chiamare	to call

Appendix 2:

The complete list of the Semantic Type Coercions resulting from our annotation exercise[6]

Rank	Semantic Type Coercion	Raw frequency
1	Area > Human Group	25
2	Area > Institution	21
3	Area > Human Group: Football Team	6
4	Artifact > Activity	6
5	Business Enterprise > Road Vehicle	6
6	Musical Composition > Sound	6
7	Concept > Human Group	5
8	Sound Maker > Sound	5
9	Activity > Sound	4
10	Beverage > Activity	4
11	Building > Activity	4
12	Event > Location	4
13	Food > Activity	4
14	Bomb > Sound	3
15	Document > Activity	3
16	Document > Narrative	3
17	Event > Sound	3
18	Activity > Food	2
19	Activity > Information	2
20	Activity > Location	2
21	Artwork > Activity	2
22	Business Enterprise > Flying Vehicle	2
23	Business Enterprise > Location	2
24	Container > Beverage	2
25	Engine > Sound	2
26	Flying Vehicle > Human	2
27	Food > Flavour	2
28	Human > Document	2

[6] The Coercions ranked 58 (*srce* > Sound) and 59 (*suze* | *smijeh* | *smiješak* > Emotion) do not have a proper Source Types but only source lexical items due to the fact that they belong to idiomatic patterns. In the first case, *srce* (English, *heart*) can be coerced into a sound since hearts usually have a heartbeat. As for the second case, although the words *suze* (English, *tears*), *smijeh* (English, *laughter*) and *smiješak* (English, *smile*) are all coerced into the emotions they typically represent, they cannot be grouped into a shared SemType since some of them are [Physical Entities] (e.g. *suze*), while others are [Activities] (e.g. *smijeh* and *smiješak*).

29	Human > Flying Vehicle	2
30	Human > Information	2
31	Human > Information: Advice	2
32	Human > Road Vehicle	2
33	Human > Sound	2
34	Human > Speech Act	2
35	Human Group > Sound	2
36	Part of Language > Sound	2
37	Physical Entity > Activity	2
38	Proposition > Sound	2
39	Route > Activity	2
40	Activity > Asset: Victory	1
41	Area > Activity: Car Race	1
42	Asset > Money Value	1
43	Business Enterprise > Food	1
44	Container > Food	1
45	Deity > Information: Advice	1
46	Device > Asset	1
47	Human > Musical Composition	1
48	Human > Part of Language	1
49	Institution > Money Value	1
50	Location > Activity	1
51	Location > Sound	1
52	Metal > Asset: Award	1
53	Musical Instrument > Sound	1
54	Natural Landscape Feature > Sound	1
55	Part of Body > Sound	1
56	Part of Language > Activity	1
57	Physical Entity > Smell	1
58	*srce* > Sound	1
59	*suze* | *smijeh* | *smiješak* > Emotion	1
60	Time Period > Sound	1
61	Vehicle > Sound	1
62	Weather Event > Sound	1