Proceedings of the 13th Workshop on Building and Using Comparable Corpora, pages 29–34
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Mining Semantic Relations from Comparable Corpora through Intersections of Word Embeddings

Špela Vintar[1], Larisa Grčić Simeunović[2], Matej Martinc[3], Senja Pollak[4], Uroš Stepišnik[5]
University of Ljubljana[1,5], University of Zadar[2], Jožef Stefan Institute[3,4]
Aškerčeva 2, SI-1000 Ljubljana, M. Pavlinovića 1, HR-23000 Zadar, Jamova cesta 39, SI-1000 Ljubljana
spela.vintar@ff.uni-lj.si, lgrcic@unizd.hr, matej.martinc@ijs.si, senja.pollak@ijs.si, uros.stepisnik@ff.uni-lj.si

Abstract

We report an experiment aimed at extracting words expressing a specific semantic relation using intersections of word embeddings. In a multilingual frame-based domain model, specific features of a concept are typically described through a set of non-arbitrary semantic relations. In karstology, our domain of choice which we are exploring though a comparable corpus in English and Croatian, karst phenomena such as landforms are usually described through their FORM, LOCATION, CAUSE, FUNCTION and COMPOSITION. We propose an approach to mine words pertaining to each of these relations by using a small number of seed adjectives, for which we retrieve closest words using word embeddings and then use intersections of these neighbourhoods to refine our search. Such cross-language expansion of semantically-rich vocabulary is a valuable aid in improving the coverage of a multilingual knowledge base, but also in exploring differences between languages in their respective conceptualisations of the domain.

Keywords : semantic relations, word embeddings, comparable corpus, karstology, frame-based terminology

1. Introduction

The frame-based approach in terminology (FBT; Faber, 2012; Faber, 2015; Faber & Cabezas-García, 2019) has brought the notion that specialised knowledge can be modelled through conceptual frames which simulate the cognitive patterns in our minds. According to Faber (2012), "[a] frame is thus as an organized package of knowledge that humans retrieve from long-term memory to make sense of the world." Two of the most significant practical contributions of FBT are on the one hand the consolidation between the conceptual and the textual level of domain representation by using specialised corpora for the induction of frames or event templates, and on the other hand the realisation that such frames and templates are not universal but contextually, culturally and linguistically bound.

On a more practical level, the frame-based approach to domain modelling fosters a dynamic and process-oriented view of the concepts, actions, properties and events leading to a deeper understanding of the domain. This is particularly relevant for a domain such as karstology where karst landscapes and landforms are the result of complex and prolonged natural processes occurring in specific environments and under specific sets of conditions.

The broader context for this research is the TermFrame project which employs and extends the frame-based approach to build a visual knowledge base for the domain of karstology in three languages, English, Slovene and Croatian; as well as explores new methods of knowledge extraction from specialized texts (Vintar et al., 2019, Miljkovic et al., 2019, Pollak et al. 2019).

The domain of karstology is conceptualized in terms of events where natural or human agents initiate actions or processes which affect patients in specific ways and thus result in various karst features. In order to explore typical conceptual frames in karstology we devised a domain-specific concept hierarchy of semantic categories, and each category can be described by a set of relations which reveal its typical features. For example, the category of *surface landforms* is typically described by relations that express form, size, location and cause while concepts from the category of *hydrological landforms* are usually defined by the relations cause, location and function.

When building a multilingual knowledge base, identifying such relations is important from the perspective of organising knowledge and ensuring maximum coverage of the domain. For example, COMPOSITION in terms of geological structure plays a crucial role in karstology because karst phenomena can only develop on soluble rocks. It is therefore extremely useful if we can access the entire inventory of expressions denoting COMPOSITION in our corpus, and also compare them between languages as this gives important clues about the domain itself, e.g. the prominence of certain minerals in different geographical regions.

In this research we propose a method to extract expressions pertaining to a specific semantic relation from a comparable English and Croatian corpus by providing a limited number of seed words for each language and relation, then using word embeddings to identify words belonging to same relation class. The seed words in our study are limited to adjectives because of their combinatorial potential within multi-word terms and the observation that semantic relations are frequently expressed through adjectives.

2. Related work

One of the aims of this study is to leverage word embeddings and a set of seed adjectives expressing semantic relations in order to extract additional adjectives that express the same semantic relation/attribute. This is in essence a set expansion task and previous research on a related subject was conducted by Diaz et al. (2016), who showed that embeddings can be employed for query expansion on domain specific texts. The research concludes that due to strong language use variation in specialized corpora, domain specific embeddings (trained locally on a small specialized corpora) outperform non-topic specific general embeddings trained on a much larger general corpus. A very similar approach for set expansion in the domain of karstology was employed by Pollak et al. (2019) for the purposes of extending terminology.

Previous authors (Duran Muñoz, 2019, Bhat, 1994, Wierzbicka, 1986, Fellbaum et al., 1993, L'Homme, 2002) have already examined the role of adjectives in specialised languages and confirmed their importance in expressing key properties of specialized concepts as well as appearing as parts of multi-word terms. A particularly relevant analysis of semantic relations in complex nominals was performed by Cabezas-García and León-Araúz (2018), who use knowledge patterns and verb paraphrases to construct a frame-based model of semantic categories and the semantic relations occurring between them. They show that a particular combinatorial pattern established for a set of nouns can be extrapolated to the entire semantic category and potentially used for relation induction.

We are also aware of several studies describing the semantic representation of adjectives in ontologies for other domains, e.g. legal (Bertoldi and Chisman, 2007), environment (Campos Alonso and Castells Torner, 2010), plant morphology (Pitkanen-Heikkila, 2015) and waste management (Altmanova et al., 2018).

3. Karstology and the TermFrame Corpus

Karstology is the study of karst, a type of landscape developing on soluble rocks such as limestone, marble or gypsum. Its most prominent features include caves, various types of relief depressions, conical hills, springs, ponors and similar. It is an interdisciplinary domain partly overlapping with surface and subsurface geomorphology, geology, hydrology and other fields.

For the purposes of our research, we used the English and Croatian parts of the TermFrame corpus, which otherwise also contains Slovene as the third language. The comparable corpus contains relevant contemporary works on karstology and is representative in terms of the domain and text types included. It comprises scientific papers, books, articles, doctoral and master's theses, glossaries and textbooks. Table 1 gives basic information about the corpus.

	English	Croatian
Tokens	2,721,042	1,229,368
Words	2,195,982	969,735
Sentences	97,187	53,017
Documents	57	43

Table 1: Corpus information

4. Methods

4.1 Framing karstology

The TermFrame project models the karstology domain using a hierarchy of semantic categories and a set of relations which allow us to describe and model karst events (Vintar et al., 2019). According to the geomorphologic analytical approach (Pavlopoulos et al., 2009), the relations describe different aspects of concepts, such as spatial distribution (HAS_LOCATION; HAS_POSITION), morphography (HAS_FORM; CONTAINS), morphometry (HAS_SIZE), morphostructure (COMPOSED_OF), morphogenesis (HAS_CAUSE), morphodynamics (AFFECTS; HAS_RESULT; HAS_FUNCTION), and morphochronology (OCCURS_IN_TIME). Additional relations were applied for general properties (HAS_ATTRIBUTE; DEFINED_AS), and for research methods (STUDIES; MEASURES).

The research described here focuses on the 5 relations which occur most frequently in the definitions of karst landforms and processes, and they also govern the formation of multi-word terms as illustrated by examples below.

underground cave ⇒ LOCATION (cave) = underground

fluvial sediment ⇒ CAUSE (sediment)=fluvial

enclosed depression ⇒ FORM (depression)= enclosed

gypsum karst ⇒ COMPOSITION (karst)=gypsum

soluble rock ⇒ FUNCTION (rock)=soluble

We thus examined the contexts expressing the selected relations in the TermFrame corpus of annotated definitions (Vintar et al., 2019). From these contexts we obtained lists of seed adjectives for each relation and both languages, which were validated by a domain expert:

LOCATION
English*: coastal, littoral, sublittoral, submarine, oceanic, subsurface, subterranean, subterraneous, subaerial, underground, aquatic, subaqueous, internal, subglacial, epigenic, phreatic, vadose, epiphreatic*

Croatian*: obalni, litoralan, priobalni, podmorski, oceanski, podzeman, freatski, vadozan, podvodan, dolinski, špiljski, epifreatski*

CAUSE
English: fluvial, allogenic, *tectonic, erosional, alluvial, volcanic, lacustrine, solutional, aeolian, periglacial, anthropogenic*

Croatian: fluvijalni, alogeni, tektonski, erozijski, aluvijalan, vulkanski, lakustrijski, eolski, periglacijalni, antropogeni

FORM
English: *polygonal, vertical, dendritic, shallow, enclosed, elongated, flat, steep, cavernicolous, detrital*

Croatian: vertikalan, ravnocrtan, strm, kavernozan, horizontalan, mrežast, longitudinalan, kružan, razgranat, ulegnut, uravnjen

COMPOSITION
English: *carbonate, limestone, dolomitic, sedimentary, sulfate, calcareous, carboniferous, silicate, sulfuric, diagenetic, siliceous, clay, volcanoclastic*

Croatian: karbonatni, vapnenački, dolomitski, sedimentan, sulfatni, kalcitan, karbonski, sulfatni, glinovit, sedreni, stijenski,klastičan,sedreni

FUNCTION
English: impermeable, permeable, solutional, hydrothermal, speleological, geological, soluble, porous, depositional, regressive, undersaturated

Croatian: nepropustan, propustan, speleološki, geološki, topiv, porozan, taložan, urušan

4.2 Word embeddings

Our initial assumption was that the word embeddings of a set of adjectives expressing a specific semantic relation, such as CAUSE, FORM or COMPOSITION, share a certain semantic component which can be used to extract other adjectives expressing the same relation.

To test this assumption, we first train FastText embeddings (Bojanowski et al., 2017) on the English and the Croatian part of the TermFrame corpus respectively (see Section 3). Embeddings were calculated for all the words that appear in the corpus at least three times and we use a skip-gram model with an embedding dimension of 100. For each seed adjective expressing a specific semantic relation, we use embeddings to find a set of 100 closest words according to the cosine distance. In order to find words of similar semantic provenance that express a specific semantic relation, in the next step we calculate all non-empty intersections between these sets of 100 closest words for all possible subsets of a set of adjectives for each relation. These subsets range in size from 10 to 2, since 10 is the largest subset of seed adjectives for a relation, for which a non-empty intersection was returned. All words found in these intersections are retained as candidate words that express a specific relation and are used in manual evaluation (see Section 5). For example, (see examples (1) and (2) below), the intersection of the closest embeddings for a subset of 5 English input words for LOCATION (*coastal, littoral, oceanic, submarine, subterranean*) yields the single word *nonmarine* as intersection, while the intersection for the subset of 3 Croatian input words for FORM (*horizontalan, kružan, vertikalan*) yields 8 words in the intersection:

(1) SIZE: 5
SUBSET: coastal, littoral, oceanic, submarine, subterranean INTERSECTION: nonmarine

(2) SIZE: 3 SUBSET: horizontalan, kružan, vertikalan INTERSECTION: okomito, sjecište, vodoravan, inverzan, okomit, nepravilan, presjecište, konveksan

5. Results and Discussion

Intersections were computed for subsets of input words ranging from maximum 10 to 2 words, whereby most intersections were empty for larger subsets and only started yielding results from size 7 downwards (see Table 2).

Our first observation is that both in English and Croatian a large majority of extracted words are adjectives and other words functioning as premodifiers in multi-word terms, thus illustrating that the embeddings capture also syntactic properties.

Since the overall goal of the experiment is to extract words pertaining to the same semantic relation, we first report the total number of extracted words and the number of correctly predicted ones, i.e. belonging to the same semantic class as the input words (Table 2).

	location		function		form		composition		cause	
	en	cr	en	cr	en	cr	en	cr	en	cr
N	357	228	147	152	164	152	293	244	183	181
C	118	88	68	43	108	97	184	197	88	132
P	0.33	0.39	0.46	0.28	0.66	0.64	0.63	0.80	0.48	0.73

Table 2: Precision per semantic relation and language (N = number of extracted words, C = correct, P = precision (C/N))

A quick glance at Table 2 shows that the numbers of extracted words are slightly lower for Croatian, which is possibly due to the difference in the size of corpora, but the overall lowest and highest precisions are also found for Croatian candidates. Next we observe large differences between individual semantic relations, both in terms of precision of prediction and the yield, but relatively similar performance across both languages. The largest number of correctly extracted candidates is achieved for COMPOSITION, where an input of only 13 words allows us to extract 184 English and 197 Croatian expressions for geological or chemical composition, e.g. *lithoclast, calcitic, azurite, loessic, gneiss, chalky, magmatic, pyrite, framestone, siliclastic* and *kalkarenit, laporovit, škriljac, glinenac, piroksenit, fliški* etc. Many of the extracted expressions are highly specialised and occur in the corpus with a very low frequency, yet their membership in the semantic class could still be correctly predicted.

On the other hand, the LOCATION relation is more difficult to capture because it may refer to the position of an entity within the karst system, its position relative to some other entity or its position relative to the land or sea. The retrieved words include many geographical names, e.g. *Baltic*, *Bahamian;kvarnerski, mosorski*, which we do not count as positives for the simple reason that our annotation scheme uses a different semantic relation (HAS_POSITION) for toponyms.

Next, we measure the precision of the predicted relation for each intersection, and we report average precision for each subset size and each language (see Table 3 and Table 4). We use precision@M denoting the number of true predictions divided by the number of all words in the intersection, and precision@5 where the size of the intersection is fixed to 5 words. In this case, a perfect precision is not possible for intersections containing less than 5 words and intersections containing more than 5 words are truncated. For the example (1) above, precision@M = 1 and precision@5 = 0.2.

As mentioned before, most intersections for larger subsets (English 8-10 input words, Croatian 7-10 input words) were empty, except for COMPOSITION in English. This would indicate that the most suitable subset size ranges from 2 to 6 input words. In English, poorest results were obtained for FUNCTION, where the intersections of subsets 4-6 contained only a single word (*sluggish*), which expresses manner of (water) movement but not function. Results for FORM, COMPOSITION and CAUSE were however promising in that they yielded highly accurate predictions, e.g. *zigzag, honeycomb, steep, curvilinear, elliptical, coalescent, sharp, semicircular, asymmetric, sinusoidal, pinnacled, undulating* for FORM and *compressional, geogenic, preglacial, bioclastic, erosional, disolutional, orogenic, tensional* etc. for CAUSE.

	location		function		form		composition		cause	
subset size	p@M	p@5	p@M	p@5	p@M	p@5	p@M	p@5	p@M	p@5
10							1	0.20		
9							1	0.20		
8							1	0.21		
7	0	0					0.99	0.24	1	0.20
6	0.36	0.07	0	0	1	0.2	0.98	0.28	0.78	0.16
5	0.45	0.13	0	0	1	0.22	0.95	0.35	0.65	0.16
4	0.45	0.17	0.01	0	1	0.31	0.91	0.44	0.60	0.20
3	0.42	0.22	0.10	0.03	0.94	0.47	0.85	0.53	0.60	0.30
2	0.37	0.29	0.26	0.13	0.70	0.55	0.75	0.59	0.56	0.39

Table 3: Precision of English predicted words per subset size

	location		function		form		composition		cause	
subset size	p@M	p@5	p@M	p@5	p@M	p@5	p@M	p@5	p@M	p@5
6	0	0								
5	0	0	0.33	0.20	1	0.20	0	0	0.50	0.10
4	0.10	0.05	0.33	0.28	0.92	0.20	0.69	0.20	0.53	0.15
3	0.28	0.16	0.32	0.30	0.78	0.28	0.79	0.35	0.65	0.27
2	0.33	0.30	0.32	0.20	0.72	0.49	0.79	0.62	0.72	0.55

Table 4: Precision of Croatian predicted words per subset size

FUNCTION also had the lowest yield of meaningful expressions in Croatian, with only one non-empty intersection for subset 5, but on the other hand the entire range of karst-related studies was retrieved by intersecting *geološki* and *speleološki (3)*:

(3) SIZE: 2
SUBSET: geološki, speleološki
INTERSECTION: arheološki, biospeleološki, geomofološki, tipološki, geoekološki, biološki, mitološki, kršološki, ontološki, geoekološka, aerološki, fiziološki, paleokrški, speleomorfološki, drološki, geokronološki, etnološki, paleontološki, filološki

Results for English also show a positive linear correlation between the subset size and precision@M (especially for the relations FORM, COMPOSITION AND CAUSE), and a negative linear correlation between the subset size and precision@5. This phenomenon can be explained with the fact that at large subset sizes there are less than five words in the intersection which has a negative impact on precision@5, but as the few extracted examples are likely to be correct, it has a positive impact on precision@M. On the other hand, at small subset sizes the number of words in the intersection will increase, which has a positive effect on precision@5 but also negatively affects precision@M, since the percentage of correctly retrieved words in the intersection decreases. The results for Croatian also show a strong negative linear correlation between the subset size and precision@5, while for precison@M the correlation somewhat varies between relations, ranging from being negative for LOCATION, CAUSE and COMPOSITION, to no correlation for FUNCTION, and to a positive correlation for the FORM relation. This means that for Croatian a larger subset size does not necessarily guarantee that a larger percentage of extracted examples will be correct.

To understand why relations perform differently in such an experimental setting we must consider their conceptual role within the frame-based domain model. It is clear that there can be an almost indefinite number of words used to describe the form of an entity in the karst landscape - think just of the multitude of underground forms found in caves. The embeddings thus successfully capture about one hundred expressions for FORM in each language, yet miss words like *ravničast, ponikvast, kavernozan, terasast, klifast, zaravnjen* etc. On the other hand, not all karst landforms have functions in the karstologic event, and the number of possible causes is also limited. For CAUSE, certain suffixes seem especially productive and allow us to extract relevant expressions – often cognates – on this basis: -genic/-gen, -genijski, -genski (*epigenic, geogenic, cryogenic, orogenic, biogenic, pathogenic, hypogenic, glacigenic, rheogenic / epigenijski, orogenski, egzogen, kemogen, zoogen, biogen, kriogen*); -glacial/-glacijalan (*preglacial, subglacial, fluvioglacial, englacial, proglacial, supraglacial / glacijalan, proglacijalan, interglacijalan, postglacijalan, fluvioglacijalan, periglacijalan*), -luvial/-luvijalan (*alluvial, eluvial, colluvial, pluvial, deluvial / iluvijalan, proluvijalan, delovijalan, diluvijalan, koluvijalan*).

In all experiments reported above we measure precision but not recall. To measure recall we would need to have a list of true positives for each relation, which could only be created manually by inspecting, for instance, all adjectives in the corpus and labelling them with relations, which has not been done as yet.

Finally, during evaluation we noted several ambiguous examples which in some contexts could refer to causes, while in others they denote composition, function or form. For Croatian, some overlap was found between the lists of expressions denoting COMPOSITION and FUNCTION (e.g. *vodopropusan* [permeable]), and for English between COMPOSITION and CAUSE (e.g. *magmatic, sediment, igneous*). Indeed such cases show that some relations are closer than others, and that specialised vocabulary is inherently multidimensional and context-dependent.

6. Conclusions

We explore semantic relations in a comparable English and Croatian corpus of karstology focusing on the adjectives and other premodifiers in multi-word terms. By assuming the frame-based domain model we identify groups of seed adjectives according to the semantic relation they express in the multi-word terms (e.g. FORM, LOCATION, FUNCTION), whereby the conceptual frame provides guidance as to which relations are expected for each concept category.

Against these background assumptions we attempt to extract attributes pertaining to the same relation using word embeddings computed on the two domain-specific corpora. We use subsets of seed adjectives as input and intersect their closest neighbours to extract candidate English and Croatian words.

Results are relatively similar across the two languages, but show high variability in precision between relations, with poor performance for the FUNCTION relation and slightly better for LOCATION. On the other hand, for the other three relations (COMPOSITION, FORM, CAUSE) results seem highly promising in that for both languages the intersections yield relevant candidates with high precision, despite the relatively small size of the domain-specific corpora. Our approach illustrates that word embeddings trained on small specialised corpora can be used to predict the semantic relations in a frame-based setting.

As future work we plan to explore the possibility of modelling karstological processes and events using analogies between semantically related pairs of concepts. It appears that the cognitive dimensions of frame-based knowledge modelling have interesting parallels within the spatial logic of word embeddings.

It is also possible to imagine a scenario where word embeddings and intersections of related words can be used to develop a frame-based model for a new domain, or more specifically to help discern the relations.

Another line of future work will consider cross-lingual query expansion, where we will try to extract adjectives expressing a specific relation in the target language by using only seed terms from the source language. In order to do this we would first need to align embeddings for both languages into a common vector space by using one of the existing methods, e.g., the one proposed in Conneau et. al (2017) that also employs FastText embeddings. Leveraging this procedure we would be able to expand the set of adjectives in a target language with terms that are not clearly associated with the target language seed terms but do however express the same relation.

7. Acknowledgements

This work is supported by the Slovenian Research Agency grant J6-9372 TermFrame: Terminology and knowledge frames across languages, 2018-2021 and by European Union's Horizon 2020 research and innovation programme under grant agreement No. 825153, project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media).

8. Bibliographical References

Altmanova, J., Grimaldi, C., Zollo, S. D. (2018). Le rôle des adjectifs dans la catégorisation des déchets. In F. Neveu, B. Harmegnies, L. Hriba et S. Prévost (Eds.), *SHS Web Conferences* 46, 6ème Congrès Mondial de Linguistique Française. Université de Mons, Belgique, pp. 1-15.

Bhat, D.N.S. (1994). The adjectival category: Criteria for differentiation and identification. Amsterdam: John Benjamins Publishing Co.

Bertoldi, A., Chishman, R.L. (2007). Improving Legal Ontologies through Semantic Representation of Adjectives. ICSC 2007, pp. 767-774.

Bojanowski, P. et al. (2017) Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5, pp: 135-146.

Cabezas-García, M., and León-Araúz, P. (2018). Towards the inference of semantic relations in complex nominals: A pilot study. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), pp. 2511-2518.

Campos, Alonso, A. and Castells, Torner S. (2010) Adjectives and collocations in specialized texts: lexicographical implications. In A. Dykstra, T. Schoonheim (Eds.), Proceedings of the XIV EURALEX International Congress. Leeuwarden/Ljouwert: Fryske Akademy - Afûk. pp. 872-881.

Conneau, A., Lample, G., Ranzato, M. A., Denoyer, L. and Jégou, H. (2017). Word translation without parallel data. *arXiv preprint arXiv:1710.04087*.

Diaz, F., Bhaskar M., Craswell, N. (2016). Query expansion with locally-trained word embeddings. *arXiv preprint arXiv:1605.07891*

Durán-Muñoz, I. (2019/forthcoming). Adjectives and their Keyness. A Corpus-based Analysis in English Tourism. *Corpora*, 14 (3). Edinburgh University Press.

Faber, P. (Ed.), (2012). A Cognitive Linguistics View of Terminology and Specialized Language. Berlin and New York: Mouton De Gruyter.

Faber, P. (2015). Frames as a framework for terminology. In Kockaert, H.J., Steurs, F. (Eds) *Handbook of Terminology*. Vol 1. Amsterdam and Philadelphia: John Benjamins, pp. 14-33.

Faber, P., Cabezas-García, M. (2019) Specialized Knowledge Representation: From Terms to Frames. *Research in Language*, 17(2): 197-211.

Fellbaum, C., Gross, D., Miller, K. (1993). Adjectives in WordNet. In G. Miller et al. (Eds.) Five Papers on WordNet. Tehnical Report 43, Cognitive Science Laboratory, Princeton University, pp. 26-39.

L'Homme, M.C. (2002). What can Verbs and Adjectives Tell us about Terms? Paper presented at TKE 2002, Nancy, France.

Miljković, D., Kralj, J., Stepišnik, U., Pollak, S. (2019). Communities of related terms in a karst terminology co-occurrence network. *Proceedings of eLex19*, Sintra, Portugal. pp. 357-373.

Pavlopoulos, K., Evelpidou, N., Vassilopoulos, A. (2009). *Mapping Geomorphological Environments*. Springer, Berlin Heidelberg.

Pitkänen-Heikkilä, K. (2015) Adjectives as terms. *Terminology*, 21 (1):76-101

Pollak, S., Repar, A., Martinc, M., Podpečan, V. (2019). Karst exploration: extracting terms and definitions from

karst domain corpus. *Proceedings of eLex19*, Sintra, Portugal. pp. 934-956.
Vintar, Š., Saksida, A., Vrtovec, K., Stepišnik, U. (2019). Modelling specialized knowledge with conceptual frames: the TermFrame approach to a structured visual domain representation. *Proceedings of eLex19*, Sintra, Portugal. pp. 305-318.
Wierzbicka A. (1986) What's in a noun? (Or: How do nouns differ in meaning from adjectives?) *Studies in Language* 10:353-389.