

# Applying MIPVU Metaphor Identification Procedure on Czech

Dalibor Pavlas, Ondřej Vrabel, Jiří Kozmér

Palacký University Olomouc

Křížkovského 511/8, 771 47 Olomouc

dalibor.pavlas@gmail.com, ondra.vrabel@seznam.cz, jiri.kozmer@gmail.com

## Abstract

This paper represents the current state of the research project aimed at modifying the MIPVU protocol for metaphor annotation for usage on Czech-language texts. Three annotators were trained to use metaphor identification procedure MIPVU and annotated 2 short text excerpts of about 600 tokens length, then the reliability of annotation was measured using Fleiss' kappa. The resultant inter-annotator agreement of 0.70 was below kappa values reported by annotators of VU Amsterdam Metaphor Corpus (Steen et al., 2010) and very similar to the agreement that researchers (Badryzlova et al., 2013) got in their first reliability test with unmodified MIPVU procedure applied on Russian texts. Some modifications of the annotation procedure are proposed in order for it to be more suitable for Czech language. The modifications are based on the observations made by annotators in error analysis and by authors of similar projects aimed to transfer MIPVU procedure to Slavic/inflected languages. The functionality of the annotation procedure refinements now have to be tested in the second reliability test.

**Keywords:** Metaphor, MIPVU, MIP, annotation, Metaphor Identification Procedure, inter-annotator agreement, Fleiss' kappa, Czech language

## 1. Introduction

This paper represents the current state of the research project aimed at modifying the MIPVU protocol for metaphor annotation for usage on Czech-language texts. It is the initial stage of creation of Czech metaphor corpus which could be a very valuable resource for several fields of linguistic research (such as computational, cognitive and corpus linguistics).

This initial stage includes:

- 1) Modification of the MIPVU protocol for reliable linguistic metaphor identification in Czech
- 2) Introducing an alternative tag (located in parallel to original MIPVU tags) which, if needed, will allow us to filter out the highly conventionalized cases of metaphors.

The process of modifying the MIPVU procedure is described in the following parts of this work. The addition of the alternative tag for highly conventionalized metaphors is motivated by the desire to use the resulting corpus for training of systems for automatic identification of metaphor.

Lexicalized cases of metaphors can be successfully interpreted using standard word sense disambiguation techniques (Shutova, 2015), which means that if they are labelled metaphorical in training data it may be causing metaphor identification system to be less effective.

Our goal is to keep the data for metaphor usage statistics, so it can be directly comparable with the same statistics available for English, and, at the same time, make the resulting corpus more suitable for computational approaches to metaphor.

## 2. Related work

### 2.1 MIP and MIPVU

Since early ninety-eighties, when conceptual metaphor theory (CMT; Lakoff and Johnson, 1980) was introduced, there has been a great interest in metaphor research. At the same time metaphor, even if we take into account only its manifestation in language, is a very complex phenomenon. It varies from novel and very creative expressions to extremely lexicalized ones, whose

metaphoricity is almost unnoticeable. This caused need for clearly defined guidelines for metaphor identification in text but due to complexity of the task it was not until 2007 before such a procedure was established. It was done by a group of researchers which called themselves Pragglejaz group.

Their method called MIP (Metaphor Identification Procedure; Pragglejaz group (2007)) was then refined in several ways and applied on data from The British National Corpus. The upgraded procedure is called MIPVU and the resulting annotated source is VU Amsterdam Metaphor Corpus (VUAMC; Steen et al., 2010). It consists of approximately 200,000 words taken from the BNC's Baby Corpus and it is divided into four genres: academic, news, fiction, and conversation.

In MIPVU, lexical units (words) whose contextual meanings are opposed to their basic meanings are considered metaphor-related words (MRWs). Annotators establish the basic and the contextual meaning for each word in the corpus using dictionary.

If basic meaning of a word is:

- a) more concrete; what it evokes is easier to imagine, see, hear, feel, smell and taste;
  - b) related to bodily action;
  - c) more precise (as opposed to vague);
- the word is marked as MRW.

The history of a lexical unit is usually not taken into account, which is one of the differences between MIP and MIPVU.

### 2.2 Applications of MIPVU to different languages

Yulia Badryzlova and her colleagues (2013) modified the MIPVU protocol for Russian-language texts and attempted to extend annotation to the level of conceptual mappings "deep annotation".

They measured the inter-annotator agreement on texts using original MIPVU and their modified version and compared it with the results of the same tests made by Steen and his colleagues (2010) in the process of establishing MIPVU procedure. In the second test their resulting inter-annotator agreement outperformed the agreement reported for VUAMC. The project was then discontinued, but recently Badryzlova and Lyashevskaya

(2017) renewed the pursuit for creation of Russian metaphor corpus. They used an annotation procedure based on MIPVU but modified in several ways. In their project, linguistic metaphor annotation is added as a new layer to SynTagRus, the Russian syntactical dependencies treebank.

Justina Urbonaitė (2015) examined metaphors of law related concepts in English and Lithuanian using MIPVU procedure for annotation. Although unable to report inter-annotator agreement as she was the only annotator, her work offered very useful remarks on applying MIPVU on an inflected language.

For the current stage of our project we are using a model similar to work of (Badryzlova et al., 2013) and are trying to utilize the findings and observations from all the three above mentioned sources.

### 3. Reliability test

We annotated two text excerpts each of about 600 tokens length. First excerpt (598 tokens) belonged in the fiction genre and was taken from short story “Zasraný vánoce” by Michal Viewegh. The second one (611 tokens) was taken from a transcription of proceedings of European Parliament. These transcriptions are available from the parallel corpus InterCorp (Rosen et al., 2017), which is a part of The Czech National Corpus project.

Dictionary of Standard Czech Language (Vácha et al., 1971; abbreviation SsJČ is commonly used) and Dictionary of Standard Czech (Kroupová et al., 2005; SsČ) were used to establish basic meanings.

Two of the 3 annotators were Ph.D. students and the remaining one was a Master's student, all of them in the field of linguistics and with prior experience in conceptual metaphor studies.

The reliability of the annotation was measured using Fleiss' kappa, a statistical measure of inter-annotator agreement which corrects for chance agreement between analysts (Artstein and Poesio, 2008).

In this first reliability test, the annotators were trained in MIPVU protocol and instructed to follow it. The annotation was performed in the manner similar to reliability tests in the process of making VUAMC, which means the annotators worked only with plain text and marked each lexical unit with either 1 (MRW) or 0 (non-MRW). The Fleiss' kappa calculation as well as determination of the cases of disagreement was carried out by a Python program designed specifically for this task.

The results can be seen in Tab. 1.

Text	Tokens	Percentage unanimous			Fleiss' κ
		Not MRW	MRW	Total	
Viewegh	598	87.46	4.85	92.31	0.65
Europarl	611	76.76	10.97	87.73	0.72
Total Fleiss' κ					0.70

Table 1: Resultant inter-annotator agreement

The minimum thresholds accepted for Fleiss' kappa are commonly stated to be 0.67, 0.7 or 0.8 (Artstein and Poesio, 2008; Badryzlova et al., 2013), more important is the comparison of the resultant inter-annotator agreement with the agreement observed on VUAMC and with the work (Badryzlova et al., 2013). See the comparison in Tab. 2.

Applying MIPVU on Czech; 3 annotators, 1209 tokens	Russian corpus of conceptual metaphor; 3 annotators, approx. 2000 tokens  (Badryzlova et al. 2013)	Russian corpus of conceptual metaphor; 3 annotators, approx. 2000 tokens  (Badryzlova et al. 2013)	VU Amsterdam Metaphor Corpus; 4 annotators, 1921 tokens  (Steen et al. 2010)
Reliability test 1	Reliability test 1	Reliability test 2	Reliability test 6
0.70	0.68	0.90	0.85

Table 2: Comparison of inter-annotator agreement in other MIPVU projects

It shows that our kappa is yet below the desired numbers and very similar to the agreement that Badryzlova and her colleagues got in their first reliability test with unmodified MIPVU procedure.

## 4. Error analysis and proposed modifications

### 4.1 Cases of disagreement

The table 3 shows disagreement count for both annotated texts in total and in respect of different parts of speech. Part of speech which in both annotated excerpts manifested most of the disagreement were verbs, followed by prepositions in case of the fiction text by Michal Viewegh, and by nouns in the case of European Parliament proceedings.

POS	Viewegh	Europarl	Sum of disagreement
Nouns	6	18	24
Verbs	18	30	48
Adjectives	6	6	12
Adverbs	5	4	9
Prepositions	11	16	27
Conjunctions	0	1	1
All POS	46	75	121

Table 3: Disagreement count

It is noteworthy that while the annotated excerpt of European Parliament proceedings shows more disagreements in annotation it nevertheless shows higher inter-annotator agreement (as seen in Tab. 1). This is caused by the fact that more than twice as many metaphors are present in the text compared to the other excerpt. This corresponds with the findings of Steen and his colleagues (2010) that from the four registers, (academic, news, fiction, and conversation) only conversation had lower frequency of MRWs than fiction texts.

Part of the disagreement in verb annotation seems to be caused more by a bias of individual annotators than a systematic pattern in the annotation protocol. In case of the European Parliament proceedings one of the annotators did not marked several metaphorically used lexical units as MRWs. The reason was that in case of some verbs the annotator overlooked personifying connection between the verb and its subject if the latter was highly abstract (e.g. luck, possibility, right or

freedom), the annotator have realized this omission immediately after the annotation course was finished.

The approach we have chosen for dealing with disagreements in preposition annotation is showed in chapter 4.2.

## 4.2 Prepositions

In English and presumably in many languages, prepositions are the most metaphor-rich part of speech as they are reported to account for 38.5-46.9% of metaphor-related words in VUAMC (Steen et al., 2010). Czech prepositions are more homonymous than prepositions in English and there was a substantial disagreement between the annotators.

Just like Badryzlova and her colleagues (2013) did, we made a list of major prepositions' basic meanings. We followed the Czech linguistic tradition where prepositions' meanings are distinguished by grammatical case (Veselková, 1986; Štícha et al., 2013). This helped to filter out homonymy and made it possible to choose just one basic meaning.

Take for example these expressions containing preposition “za”. While it is clear that in sentences 3) and 4) “za” is a MRW, in the case of 1) and 2) both meanings are clearly distinct but equally concrete and bodily related.

- 1) Petr stojí za mnou; Petr stands behind me
- 2) Chytil jsem ho za nohu; I caught him by the leg
- 3) Za 2 roky to bude hotové; It will be done in 2 years
- 4) Vyměnil jsem kolo za auto; I traded the bike for the car

If we distinguish between “za” in instrumental (expression 1)) and in accusative 2), we can have basic meaning for each one, moreover “accusative za” standing for basic meaning of this preposition in sentences 3) and 4) which both are MRWs.

## 4.3 Reflexive pronouns “se/si” and auxiliary verbs

Reflexive pronouns “se/si” are used either when the subject and object of the sentence are identical 5) or as an integral part of a reflexive verb whose lexical meanings they often determine. The presence of a reflexive pronoun “se/si” can result in a complete shift of meaning as illustrated in 6).

- 5) umyji se; I will wash myself
- 6) rozvést / rozvést se; to develop (an idea) / to divorce

Expectably, the original MIPVU procedure does not account for this phenomenon. The table 4 shows its effect on an actual annotated sentence.

Annotated sentence	Když	se	před	třemi	lety	rozvedl [...]
Original MIPVU	0	0	1	0	0	1
Modified MIPVU	0	0	1	0	0	0

Table 4: Annotation of a sentence where reflexive pronoun causes a shift of meaning

The highlighted tokens, when treated as separate lexical units, will render the basic meaning of the word “rozvedl” to be “he developed/expanded (something)” and the contextual meaning which is “he got divorced” should therefore be a MRW. On the other hand, the expression “se” + “rozvedl”, when counted as one lexical unit which is distinct from “rozvedl”, has an equal basic meaning to

the contextual one, so it is annotated as not-MRW, which matches better with the general sense of the sentence.<sup>1</sup>

Similarly, Czech auxiliary verbs such as “bych” are considered integral parts of the full verb’s conjugation forms.

Therefore for reflexive pronouns “se/si” and auxiliary verbs we applied the same policy as annotators of VUAMC used for phrasal verbs in English, which means that they count as one lexical unit altogether with the full verb.

On the other hand, meanings commonly expressed by phrasal verbs in English tend to be expressed by prefixes in Czech which are already parts of the word as seen in 7).

- 7) zesilit; turn up

## 4.4 Set expressions

Dealing with set expressions, we followed remarks on MIPVU made recently by the main author of VUAMC (Steen, 2017), which is to treat each word of set expression as a lexical unit itself. This renders the demarcation line between metaphor and idiom unclear. On the other hand, using dictionaries to determine set expressions as (Badryzlova et al., 2013) did, seemed to be problematic because unlike the dictionaries used in the original MIPVU procedure, dictionaries available for Czech are neither corpus based, nor contemporary.

## 5. Summary

So far, we have applied MIPVU on Czech texts and tested inter-annotator agreement. Direct transferability of the MIPVU procedure to Czech language turned out to be problematic, which we expected, as the same complications were reported by researchers applying the procedure on Russian (Badryzlova, 2013) and Lithuanian (Urbonaitė, 2015).

After the error analysis, we have proposed several minor modifications of the guidelines in order to make them more suitable for Czech and we plan to conduct second reliability test as soon as possible.

The next step after successfully transferring MIPVU to be used on Czech texts would be to annotate the data with an additional tag for highly lexicalized metaphors. It is meant to work not by asking whether the contextual meaning is different from basic one but rather whether there is a literal word in use which can express the given contextual meaning. If there is not, it is probably a highly conventionalized metaphor.

Nevertheless, there are several yet unanswered questions regarding this approach, the most important one being if annotators will agree sufficiently on those cases.

## 6. Acknowledgements

This work was funded by Ministry of Education, Youth and Sport of the Czech Republic as a part of the project “Počáteční fáze tvorby korpusu metafor v češtině” (Grant number IGA\_FF\_2018\_026).

<sup>1</sup> In the first course of annotation the interconnection of words is realized simply by giving the reflexive pronoun (or an auxiliary verb) always the same value of metaphoricality which is given to its corresponding verb. This naive method is justifiable because this stage of the project only serves to refine the annotation manual. It is not suitable for actual corpus generation as it would influence the metaphor usage statistics.

## Bibliographical References

- Arstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4): 554–596.
- Badryzlova Y., Lyashevskaya O. (2017). Metaphor Shifts in Constructions: the Russian Metaphor Corpus. In AAAI Spring Symposium Series, pp. 127-130.
- Badryzlova, Y., Shekhtman, N., Isaeva, Y., Kerimov, R. (2013). Annotating a Russian corpus of conceptual metaphor: a bottom-up approach. In Proceedings of the First Workshop on Metaphor in NLP. Atlanta, GA: Association for Computational Linguistics, pp. 77–86.
- Kroupová, L. et al. (2005). Slovník spisovné češtiny pro školu a veřejnost: s Dodatkem Ministerstva školství, mládeže a tělovýchovy České republiky. Praha: Academia.
- Lakoff, G., Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.
- Pragglejaz Group (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1):1–39.
- Shutova, E. (2015). Design and Evaluation of Metaphor Processing Systems. *Computational Linguistics*, 41(4): 579-623.
- Steen, G. (2017). Identifying metaphors in language. In Semino, E., Demjén, Z. (Eds.) *The Routledge Handbook of Metaphor and Language*. London: Routledge, chapt. 5.
- Steen, G., Aletta, G., Dorst, J., Herrmann, B., Kaal, A. A., Krennmayr, T., Pasma, T. (2010). A method for linguistic metaphor identification: From MIP to MIPVU. Amsterdam, John Benjamins.
- Štícha, F., et al. (2013). *Akademická gramatika spisovné češtiny*. Praha: Academia.
- Urbonaitė, J. (2015). Metaphor identification procedure MIPVU: an attempt to apply it to Lithuanian. *Taikomoji kalbotyra*, (7):1–26.
- Vácha, J., editor, et al. (1971). *Slovník spisovného jazyka českého*. Praha: Academia.
- Veselková, J., et al. (1986). *Mluvnice češtiny. 2, Tvarosloví*. Praha: Academia.

## Language Resource References

- Rosen, A., Vavřín, M., Zasina, A. J. (2017): InterCorp, 10.0, Institute of the Czech National Corpus, Charles University, Prague. Available from: <http://www.korpus.cz>