

# Annotation of metaphorical expressions in the *Basic Corpus of Polish Metaphors*

Elżbieta Hajnicz

Institute of Computer Science, Polish Academy of Sciences

ul. Jana Kazimierza 5, 01-248 Warszawa, Poland

hajnicz@ipipan.waw.pl

## Abstract

This paper presents a corpus of Polish texts annotated with metaphorical expressions. It is composed of two parts of comparable size, selected from two subcorpora of the *Polish National Corpus*: the subcorpus manually annotated on morphosyntactic level, named entities level etc., and the *Polish Coreference Corpus*, with manually annotated mentions and the coreference relations between them, but automatically annotated on the morphosyntactic level (only the second part is actually annotated).

In the paper we briefly outline the method for identifying metaphorical expressions in a text, based on the MIPVU procedure. The main difference is the stress put on novel metaphors and considering neologistic derivatives that have metaphorical properties.

The annotation procedure is based on two notions: *vehicle* – a part of an expression used metaphorically, representing a source domain and its *topic* – a part referring to reality, representing a target domain. Next, we propose several features (text form, conceptual structure, conventionality and contextuality) to classify metaphorical expressions identified in texts. Finally, we analyse and evaluate the results of the annotation.

Additionally, some metaphorical expressions are identified as concerning personal identity matters and classified w.r.t. their properties.

**Keywords:** corpus of metaphors, classifications of metaphorical expressions, Polish

## 1. Introduction

Since Aristotle, the traditional theory of metaphor has limited the appearance of metaphorical expressions to poetic language and assumed them to be absent in ordinary everyday language. Metaphor was understood as the property of language not thought.

This scientific view changed in the second half of the twentieth century. One of the most prominent work on the subject is (Lakoff and Johnson, 1980). It considers that metaphor is a universal, conceptual mechanism of human communication. It emerges from a deviation from the central meaning to marginal. Lakoff and Johnson (1980) introduced two basic concepts: a *source domain*, usually more concrete, which is a source of the features which the speaker wants to assign to a concept from a *target domain* adequate in the context, usually more abstract, cf. the classic metaphorical expression LOVE is a JOURNEY.

In this paper we describe the *Basic Corpus of Polish Metaphor* (BCPM). First, we present the procedure for annotation of metaphorical expressions in the corpus, based on lexico-semantic annotation of the corpus (Hajnicz, 2019). Next, we analyse the results of annotation. This task is part of the project *Cognitive and sociocultural analysis of metaphorical expressions in Polish texts* (CORMETAN), aimed at analysis of the distribution of various types of metaphorical expressions in texts of various genres. The way metaphor is used to express people's identities and social relationships and how it influences the speakers' perception of the world will be looked at as well. An additional goal is the automatic detection of metaphorical expressions in Polish texts.

In what follows, we look at other corpora annotated w.r.t. metaphor (cf. section 2.). Next, we describe our corpus (section 3.). The main part of the article deals with the procedure for deciding, when a word or a phrase should be considered metaphorical (section 4.1.) and presenting vari-

ous features used to classify metaphorical expressions (section 4.3.). Finally, we analyse the results of annotation (section 5.).

## 2. Related work

Using computer methods for extensive study of the phenomenon of metaphor is a relatively new approach. Only few publicly available corpora annotated with regard to metaphor in fact exist.

The *VU Amsterdam Metaphor Corpus* (Krennmayr and Steen, 2017) is the most famous and has had the most impact. It is composed of 115 fragments containing about 190,000 lexical units randomly selected from four registers of the BNC-Baby (Burnard, 2008) corpus. It is annotated accordingly to the dedicated annotation procedure, i.e. *Metaphor Identification Procedure Vrije Universiteit Amsterdam* – MIPVU, cf. (Steen et al., 2010). The procedure consists in identifying all lexical units in a text, finding their contextual meaning and their basic meaning. A lexical unit is metaphorical if its contextual and basic meaning differ but the former can be interpreted in comparison with the latter.

The MIPVU procedure is used to annotate other corpora of metaphor, e.g. Badryzlova and Lyashevskaya (2017) for Russian, Reijnierse (2010) for French, Woll (2017) for German, Lu and Wang (2017) for Mandarin Chinese, but it is often modified due to language specificity (Marhula and Rosiński, 2017).

There are some corpora with a different annotation scheme. For instance, Gordon et al. (2015) describe a corpus of metaphor, in which the basic schema *source domain vs. target domain* of (Lakoff and Johnson, 1980) is enriched with a predefined set of ontological categories represented as a set of conceptually related scenarios.

The only extensively annotated, publicly available Polish metaphor corpus we are aware of is the corpus of synesthetic

metaphors in Polish (Zawisławska, 2016; Zawisławska, 2019). It consists of 1,414 blogs that contain 685,600 tokens, 9,217 of which are grammatically and semantically annotated metaphorical units. Synesthetic metaphors occur when one name is used for various types of sensory perception, e.g. olfactory perception is described by lexemes that primarily activate another sense (e.g. taste), cf. *sweet aroma*.

There also exist a corpus of 1833 short pieces of text sampled from NKJP, composed of 45,000 tokens. Each sample contains at least one adjective-noun phrase which could text literal or metaphorical meaning depending on the context (Wawer et al., 2019).

### 3. The corpus

BCPM is composed of two parts:

- 700 samples of the Polish Coreference Corpus (PCC), randomly selected in a way that balances various registers of texts accordingly to NKJP assumptions, cf. ch. 3 of (Przepiórkowski et al., 2012),
- 2000 samples of a fragment of NKJP 1M considered in the *Składnica* treebank, selected in way that maximises its size and the number of sentences that have parses in *Składnica*, but preserving the balance of registers.

NKJP 1M is a subcorpus of the Polish National Corpus (Polish acronym NKJP) manually annotated on the morphosyntactic level, cf. ch. 6 of (Przepiórkowski et al., 2012). The PCC (Ogrodniczuk et al., 2015), in turn, is randomly selected from the whole NKJP corpus. Therefore, BCPM as a whole is part of NKJP. The PCC part is composed of 200031 tokens (286 tokens per sample on average), whereas the NKJP part includes 144087 tokens (68 tokens per sample on average). Thus, PCC samples are on average 4.2 times longer than NKJP 1M samples, and the PCC part of BCPM is the larger one. The structure of BCPM is motivated by the fact that we plan to examine the interconnections between metaphor and coreference. On the other hand, we want to test, how information about the syntactic structure of an utterance, its predicate-argument structure, selectional preferences etc. influences automatic metaphor detection. Named entities are annotated in NKJP 1M as well, and this information can help in annotation of metaphors, as named entities often play the role of a metaphor's topic.

## 4. The procedure for annotation

The annotation is performed independently by two linguists, and conflicts are resolved by the third. In what follows we present the structure of the whole procedure. The whole procedure, together with the lexico-semantic annotation step is performed by means of the *WebAnno* tool (de Castilho et al., 2016) by means of a web browser.

### 4.1. Identification of a metaphorical expression

A key step for annotating a metaphorical expression is to state whether a particular word or phrase is metaphorical. Our approach is based on the *Metaphor Identification Procedure (MIP)* proposed by Pragglejaz Group (2007) and its modified version *Metaphor Identification Procedure Vrije Universiteit (MIPVU)* described in (Steen et al., 2010).

The procedure is composed of the following steps:

1. Reading the whole text (sample) in order to establish its general meaning and subject.
2. Establishing, which lexical units were used, including multi-word expressions etc. (contextual meaning).
3. Determining, whether another, more basic (more concrete, more precise, not necessarily the most frequent), contemporarily used meaning of each phrase exists, adequate in different contexts (e.g. *HEAD – of a department, state etc. vs. body part*).
4. Stating their common and distinct properties and checking, whether the new meaning can be interpreted through the prism of the old one, distinctly connected to it.

The positive result of the above examination means that we deal with a metaphorical expression. Furthermore:

5. If the meaning adequate in context is not distinguished, but the corresponding “basic” meaning is used in a way that goes far beyond its normal usage, we treat it as metaphorical. For instance, there is no separate meaning for *DRIVE INTO* (in particular, *drive on a tank*) in (1) or for *RAM*, but *capitalism* is not a living being or an object that can drive or ram anything.

- (1) Do Polski kapitalizm wjechał  
To Poland.GEN capitalism.NOM drive.PAST into  
czołgiem i kompletnie nas staranował.  
tank.INST and completely we.ACC ram.PAST  
'Capitalism drove into Poland on a tank and smashed us completely.'

In contrast to the corpus of synaesthetic metaphors (Zawisławska, 2016; Zawisławska, 2019), we decided not to distinguish neologisms which are derivated from the standard meaning of a word as metaphorical. In the case of general language it is hard to determine a boundary between standard and metaphorical derivatives.

### 4.2. Scope of annotation of a metaphorical expression

Often not all parts of an utterance are involved in a metaphor, and sometimes it goes beyond a single utterance. Our annotation is not restricted to single words. We mark its *vehicle* – a part used metaphorically, representing a source domain and its *topic* – a part that refers to reality, that represents a target domain, cf. (Zawisławska, 2019). Note that the figurativeness of an expression emerges from the confrontation of its vehicle and its topic. Both *CAPITALISM* and *POLAND* in (1) refer to reality. However, it is *CAPITALISM* that makes the whole utterance metaphorical, the sentence can be understood literally after deleting it.

Additionally, we select a head of each vehicle – a word that introduces figurativeness to the whole expression. For instance, *PASZTET*, lit. 'pâté' in (2) is used to describe an ugly woman, whereas a phrase *przeteterminowany pasztet*, lit. 'expired pâté' emphasises her old age.

A vehicle of a metaphor is always present in text. However, a topic can be represented indirectly, by the broad context.

We decided to annotate all expressions indicating it. For instance, in (2), the fact that »pâté« posses a palace makes its usage metaphorical. Otherwise, we would have trouble interpreting *zakochać się w przeterminowanym pasztecie* literally as ‘fall in love with expired pâté’. Nevertheless, *ZAKOCHAĆ SIĘ* lies outside the boundaries of the metaphor and cannot be referred to as its topic. Such a shortened sentence could serve as an example of a metaphor with a topic represented indirectly.

- (2) *Może opłacałoby mu się zakochać*  
 perhaps pay.COND he.DAT refl. mark fall in love.INF  
*w przeterminowanym pasztecie z własnym*  
 in expired.LOC pâté.LOC with OWN.INST  
*pałacem.*  
 palace.INST  
 ‘Perhaps it would be profitable for him to fall in love with an old, ugly woman with her own palace’

### 4.3. Classification of metaphorical expressions

All metaphorical expressions identified in a text are classified w.r.t. their various features.

1. *Text\_form* – a form the vehicle of a metaphor takes in a text:

- *word* – the vehicle of a metaphor is composed of a single word, e.g. *STARANOWAĆ* ‘ram’ in (1)
- *phrase* – the vehicle of a metaphor is a phrase, e.g. *wjechać czołgiem* ‘drive on a tank’ in (1), *przeterminowany pasztec* ‘expired pâté’ (2);
- *text* – if a metaphor has a narrative form (e.g. allegory).

2. *Structure* – a conceptual structure of a metaphor:

- *simple* – involves a single source domain X and a single target Y which enable explication in a form *X is Y* and cannot be decomposed on the text level, e.g. *wjechał czołgiem* ‘went on a tank’ and *staranował* ‘rammed’ in (1) or *słodka* (vehicle) *zemsta* (topic) ‘sweet revenge’. Moreover, it has only single topic.
- *relational* – differs from *simple* metaphors in that its vehicle relates two or more topics, e.g. the vehicle *zbudowane* ‘built’ in 3 relates two topics – ‘organisms’ and ‘proteins’.
- *elaborated* – contains additional terms from a source domain emphasising and expanding the metaphorical expression, e.g. *przeterminowany pasztec*, lit. ‘expired pâté’, an old, ugly woman in (2).
- *mixed* – a target domain is described by means of several source domains. An example of a mixed metaphor is (4); the structure of its simplified version (5) is presented on Figure 1.
- *layered* – there are two source domains that are put one onto the other, and each has its own topic from different domains, cf. (6).
- *unknown* – a lexicographer cannot establish the structure of a metaphorical expression.

3. *Characteristics* – specification of a typical source domain for the *X is Y* model:

- *personification* – describing abstracts, objects and animals as people;
- *animisation* – describing abstracts, objects and sometimes people as animals;
- *reification* – describing abstracts, animals and sometimes people as objects;
- *depersonalisation* – describing people as objects or animals in a way depersonalising them.

For instance, in (1) we deal with *personification* (a **political system** is an **aggressor**), in (4) we deal with *reification* (a **world** is a **building**) and *animisation* (a **building** is an **animal**).

We decided to distinguish *depersonalisation* which specifies a target domain, not a source one. It is applied for metaphors depreciating the humanity of a person in spite of whether they refer to a person as an object (*pasztec*, lit. ‘pâté’, *szmata* lit. ‘rag’) or an animal (*suka* ‘bitch’). Thus, it describes people as objects or animals in a way depersonalising them. This is especially important from the point of view of assessing the identity issues.

4. *Contextuality* – showing whether and to what extent the figurativeness of an utterance depends on its context:

- *contextual* – interpreting an utterance as metaphorical, with pointing its source and target domains, depends on the context; usually it can be interpreted literally and a topic of metaphor is located outside the utterance;
- *self\_contained* – an utterance can be completely, metaphorically interpreted regardless of the context.

5. *Conventionality*

The conventionality of a metaphorical expression means that it is established in culture and language, and the metaphorical meaning is distinguished and represented in dictionaries. Since our corpus is annotated with PLWORDNET LUs, this annotation serves as a primary source of this classification. Nevertheless, annotators can check whether the meaning is distinguished in other dictionaries<sup>1</sup> Therefore, we annotated conventionality of a metaphorical expression as *standard*, *external* and *novel* – used spontaneously as a result of the free play of associations. Since sometimes dictionaries illustrate a single word meaning with both standard and metaphorical examples, we have added yet another subcategory – *included*.

- (3) *Wszystkie organizmy [...] zbudowane są z białek.*  
 All.NOM organisms.NOM [...] built.NOM.PL are of proteins.GEN  
 ‘All organisms are made of proteins.’

<sup>1</sup>Two dictionaries serve as such secondary sources: WSJP (<https://www.wsjp.pl>) and USJP (Dubisz, 2006), its electronic version <https://s.jp.pwn.pl.html>.

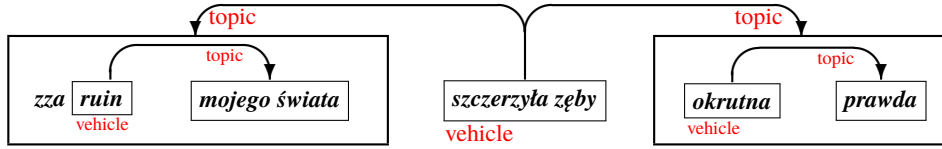


Figure 1: Graphical presentation of a metaphorical expressions with two complex topics and incoherent vehicles

- (4) *Świat* *walił się* *z* *hukiem,*  
 world.NOM crumble.IMPERF.PAST with thud.INST  
 a *zza* *ruin* *szczyrzyła* *zęby*  
 and from behind ruins.GEN bare.IMPERF.PAST teeth  
*najokrutniejsza* *z* *prawd*.  
 cruel.NOM of truth.NOM  
 ‘The world was collapsing with a bang, and from  
 behind its ruins the cruelest truth was baring its teeth.’

- (5) *Zza* *ruin* *mojego* *świata*  
 from behind ruins.GEN my.GEN world.GEN  
*szczyrzyła* *zęby* *okrutna* *prawda*.  
 bare.IMPERF.PAST teeth cruel.NOM of truth.NOM  
 ‘From behind the ruins of my world the cruel truth  
 was baring its teeth.’

The idea of layered metaphor is illustrated by (6). The vehicle *napakowane* ‘packed’ with the topic *wina* on the primary level refer to a source domain meaning ‘full’, ‘stuffed’. However, referring to ‘muscular’ involves another, itself metaphorical interpretation of ‘packed’ – *to pack at a gym*. This additional interpretation forms another layer of metaphoricality of the expression.

- (6) *Napakowane* *owocem,* *muskularne* *niczym*  
 packed.NOM fruit.INST muscular.NOM like  
*Sylvester* *Stallone* *wina*.  
 Sylvester.NOM Stallone.NOM wines.NOM  
 ‘Wines packed with fruit, muscular like Sylvester  
 Stallone.’

A metaphor can be composed of a chain of metaphorical expressions, that usually refer to the same topic. Since we annotate each metaphorical expression on its own, we add information that a metaphor continues.

Finally, for the sake of analysis of issues connected with people’s identity, we consider an additional category identity that takes values: gender, sexuality, family, friend/foe, ethnicity, individual/group, class, age, liking, views, other. Contrary to other categories, several of them can be adequate for a single metaphorical expression.

## 5. Results of the annotation

The process of annotation has turned to be hard and slow. Therefore, we decided to annotate only the PCC part of the corpus. Below we discuss the result of annotation of 343 samples containing 98,336 tokens performed by two annotators for each sample.

We use Scott’s (Scott, 1955)  $\pi$  statistics to calculate the interannotator agreement:

$$(7) \quad \pi = \frac{P(A) - P(E)}{1 - P(E)},$$

$$(8) \quad P(A) = \frac{\mathbf{n}_b}{\mathbf{t}},$$

$$(9) \quad P(E) = \sum_{k \in K} \left( \frac{\mathbf{n}_k}{2\mathbf{t}} \right)^2,$$

where  $T$  is a set of tokens,  $K$  is the set of classes for annotation, and  $\mathbf{t}, \mathbf{k}$  are cardinalities of  $s T, K$ , respectively. Next,  $\mathbf{n}_b$  is a number of all common choices and  $\mathbf{n}_k$  are total choices of a class  $k \in K$ .

The main step of the annotation is to choose whether a particular word or phrase is a vehicle for a metaphoric expressions. The total number of metaphorical expressions in the corpus is 8547, their average number in a sample is 16. It means the only 5,5% of tokens are considered metaphoric<sup>2</sup>. If we include tokens annotated by neither of annotators in  $\mathbf{n}_b$ , then we obtain  $P(A) = 0.94$ ,  $P(B) = 0.89$  and hence  $\pi = 0.41$ . However, if we limit ourselves to tokens annotated by at least one annotator, we obtain  $P(A)_/ = 0.28$ . However,  $\mathbf{n}_{k/} = 2\mathbf{t}$ , and we cannot calculate  $\pi$ .

In what follows, the calculations of the inter-annotators agreement of the various features of metaphorical expressions is calculated only for tokens considered vehicles by both annotators. Hence, from now  $\mathbf{t} = 2410$ .

In Table 1 we present the distribution of annotators’ choices for various features of metaphorical expressions. They are source information for calculating the  $\pi$  statistics (Table 2). The symbol ‘\*’ means that the annotator did not make any choice for the particular feature<sup>3</sup>.

The values of the Scott’s  $\pi$  statistics are low, especially w.r.t. to their  $P(A)$  part. The reason is that the distribution of all features is unbalanced. The most dominating classes for all features are presented in the lower part of Table 2.

Domination of each type (class) for every feature is justified. Domination of ‘\*’ for Identity results from that the personal identity issues are not involved in every metaphor. The most balanced feature is Characteristics. However, the domination of reification is surprising – we have supposed that this feature is not adequate for most MEs (‘\*’ value). Observe, that this feature has the lowest  $P(A)$  statistics.

Such domination causes that  $P(E)$  is high, due to the way its is calculated. Therefore, Scott’s (Scott, 1955)  $\pi$  statistics is not appropriate for such unbalanced data (Cohen’s (Cohen, 1960)  $\kappa$  does not differ from it with this respect). Both underestimate the quality of the annotation.

5651 <sup>2</sup>The whole number of tokens includes punctuation marks etc.  
<sup>3</sup>This is the *WebAnno* convention.

Table 1: The distribution of annotators’ choices for various ME features

feature	names of classes and their cardinality
structure	elaborated: 461, layered: 29, mixed: 285, relational: 500, simple: 3535, unknown: 7
conventionality	*: 2, external: 292, included: 214, novel: 552, standard: 3757
characteristics	*: 1449, animisation: 366, depersonification: 58, personification: 635, reification: 2309
contextuality	*: 2, contextual: 333, self_contained: 4482
text_form	phrase: 569, text: 23, word: 4226

Table 2: Statistics for annotators’ choices of features

feature	structure	conventionality	characteristics	contextuality	text_form	identity
both	1591	1807	1231	2206	2068	1539
$P(A)$	0.66	0.75	0.51	0.92	0.86	0.64
$P(E)$	0.56	0.63	0.34	0.87	0.78	0.56
$\pi$	0.23	0.33	0.26	0.35	0.35	0.17
class name	simple	standard	reification	self_contained	word	*
number	3535	3757	2309	4482	4226	3594
part	0.73	0.78	0.48	0.93	0.88	0.75

Table 3: Statistics for annotators’ choices of topics

topics’ number	equals number	equals part	overlaps number	overlaps part
4373	992	0.23	1397	0.32

Identity causes additional problems, since it is a multi-choice feature. Therefore, the real number of evaluated classes  $\mathbb{k}$  satisfies the inequality  $\mathbb{k} < \mathbb{k} < 2^{\mathbb{k}}$ . The evaluation based on the identity of whole classes, not particular elements, has to be underestimated.

The next feature to be estimated for the quality of metaphorical expressions’ annotation in the BCMP is the choice of topics for each ME vehicle. It is hard to use any sophisticated statistics here, since most metaphorical expressions has topics, but each time they are different sets of tokens and their number varies from 0 to 2. What is more, we expect annotators to mark all elements of a phrase being a topic. It is especially important for complex metaphors, cf. example (4) and its visualisation on Figure 1. Nevertheless, some annotators indicate only topics’ heads, some include punctuation marks and other does not etc. Therefore, we decided to consider two cases:

1. the complete equality of topics’ annotations interpreted as sets of tokens;
2. the overlapping of such topics’ annotations.

In both cases the number of topics has to be equal and each topic should have its counterpart. Therefore, even if one topic is identical in both annotations, we cannot accept it as consistent if one of annotators indicated another topic.

The results of the evaluation of the topics assignment is present in Table 3. The actual evaluation values (the ‘part’ columns) are similar to  $P(A)$  element of  $\pi$  and  $\kappa$  statistics. These results show that choosing the proper set of topics for

a metaphorical expression is a hard task.

Finally, we analyse the annotation of the Continuation relation. Usually, metaphorical expressions occur in a text independently, sometimes forming complex expressions as i (4). Connected separate MEs are rare. Some annotators probably completely ignored annotating this relation. Consequently, there is 1501 occurrences of the relation in the whole corpus (31% of MEs), with 59 chosen consistently (4% of relations). Remember, however, that relations link two metaphorical expressions, source and target, and the last one could be an ME chosen by the one annotator only.

## 6. Conclusions

In this paper, we have discussed the scheme of annotation of metaphorical expressions in a corpus of samples of Polish texts, in which metaphorical expressions are annotated. The procedure for deciding whether we deal with a metaphorical expression or not is based on the *Metaphor Identification Procedure Vrije Universiteit* procedure. For each metaphorical expression, its vehicle and topic(s) are identified in text, and the expression is classified w.r.t. its various features.

The procedure for annotation turned to be complicated. Even though the instruction for annotation (Hajnicz et al., 2020) was very detailed, annotators has had problems in determining whether a particular expression is metaphorical or not. The identification of all topics (and their text scope) of a particular ME turned to be hard as well. On the other hand, the task of choosing the features of MEs was performed satisfactorily, perhaps because of the domination of the one class for each feature. Such situation may lead to their overusing, but this is a consequence of the choice of features.

The procedure of superannotation of the corpus is nearing completion. Next we plan experiments with automatic detection of metaphorical expressions in text, but we are afraid that our corpus is now too small to obtain good results in this

task. The annotation of the corpus should be continued.

## 7. Acknowledgements

These research was financed by the Polish National Science Centre, within the project **2018/29/B/HS1/01773** *Cognitive and sociocultural analysis of metaphorical expressions in Polish texts* (CORMETAN).

## 8. Bibliographical References

- Badryzlova, Y. and Lyashevskaya, O. (2017). Metaphor shifts in constructions: the Russian Metaphor Corpus. In *The AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding*.
- Burnard, L. (2008). Reference guide to BNC Baby. Internet.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- de Castilho, R. E., Éva Mújdricza-Maydt, Yimam, S. M., Hartmann, S., Gurevych, I., Frank, A., and Biemann, C. (2016). A web-based tool for the integrated annotation of semantic and syntactic structures. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pages 76–84, Osaka, Japan.
- Stanisław Dubisz, editor. (2006). *Uniwersalny słownik języka polskiego PWN*. Wydawnictwo Naukowe PWN, Warsaw, Poland.
- Gordon, J., Hobbs, J. R., May, J., Mohler, M., Morbini, F., Rink, B., Tomlinson, M., and Wertheim, S. (2015). A corpus of rich metaphor annotation. In Ekaterina Shutova, et al., editors, *Proceedings of the 3rd Workshop on Metaphor in NLP (Meta4NLP 2015)*, pages 56–66. Association for Computational Linguistics.
- Hajnicz, E., Zawisławska, M., and Maciejewska, M. (2020). Zasady znakowania wyrażeń metaforycznych w tekstach polskich. Manuscript.
- Hajnicz, E. (2019). Annotation of metaphorical expressions in the basic corpus of polish metaphors. submitted to 12th International Conference on Language Resources and Evaluation (LREC-2020).
- Krennmayr, T. and Steen, G. J. (2017). VU Amsterdam Metaphor Corpus. In Nancy Ide et al., editors, *Handbook of Linguistic Annotation*, pages 1053–1072. Springer-Verlag, Dordrecht, Holland.
- Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press, Chicago, IL.
- Lu, X. and Wang, B. P.-Y. (2017). Towards a metaphor-annotated corpus of Mandarin Chinese. *Language Resources and Evaluation*, 51(3):663–694.
- Marhula, J. and Rosiński, M. (2017). Co oferuje MIPVU jako metoda identyfikacji metafory? *Polonica*, XXXVII:23–36.
- Ogrodniczuk, M., Głowińska, K., Kopeć, M., Savary, A., and Zawisławska, M. (2015). *Coreference in Polish: Annotation, Resolution and Evaluation*. Walter De Gruyter.
- Pragglejaz Group. (2007). Mip: A method for identifying

metaphorically used words in discourse. *Metaphor and Symbol*, 22(1):1–39.

- Adam Przepiórkowski, et al., editors. (2012). *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN, Warsaw, Poland.
- Reijnierse, G. (2010). Making MIP operational for French. practical and theoretical issues concerning the choice of a dictionary. In *presented at 8th International Conference on Researching and Applying Metaphor (RaAM 10)*.
- Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, 19(3):321–325.
- Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., and Pasma, T. (2010). *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. Number 14 in *Converging Evidence in Language and Communication Research*. John Benjamins Publishing Company, Amsterdam.
- Wawer, A., Mykowiecka, A., and Marciniak, M. (2019). Detecting word level metaphors in Polish. In Zygmunt Vetulani et al., editors, *Human Language Technologies as a Challenge for Computer Science and Linguistics – 2019*, pages 87–91. Wydawnictwo Nauka i Innowacje.
- Woll, K. (2017). *The Adjustment of MIPVU (Metaphor Identification Procedure Vrije Universiteit) to German*. Master thesis, Vrije Universiteit Amsterdam, Amsterdam, Holland.
- Zawisławska, M. (2016). SYNAMET—A microcorpus of synesthetic metaphors. preliminary premises of the description of metaphor in discourse. *Cognitive Studies/Études Cognitives*, 16:07–118.
- Zawisławska, M. (2019). *Metaphor and Senses. The Synamet Corpus: A Polish Resource for Synesthetic Metaphors*. Peter Lang, Berlin, Germany.

## 9. Language Resource References

- NKJP Consortium. (2012). *Manually annotated subcorpus of National Corpus of Polish (NKJP 1M)*. NKJP Consortium, <http://clip.ipipan.waw.pl/NationalCorpusOfPolish>, 1.0.
- G4.19 Group at Department of Artificial Intelligence, Wrocław University of Technology. (2015). *Polish wordnet plWordNet*. Department of Artificial Intelligence, Wrocław University of Technology, <http://plwordnet4.clarin-pl.eu/>, 2.1.
- Maciej Ogrodniczuk. (2019). *Polish Coreference Corpus*. Institute of Computer Science, Polish Academy of Sciences, <http://zil.ipipan.waw.pl/PolishCoreferenceCorpus>, 1.5.