

Are White Ravens Ever White? – Non-Literal Adjective-Noun Phrases in Polish

Agnieszka Mykowiecka, Małgorzata Marciniak

Institute of Computer Science, Polish Academy of Sciences

Jana Kazimierza 5, Warsaw, Poland

{agn, mm}@ipipan.waw.pl

Abstract

In the paper we describe two resources of Polish data focused on literal and metaphorical meanings of adjective-noun phrases. The first one is FigAN and consists of isolated phrases which are divided into three types: phrases with only literal meaning, with only metaphorical meaning, and phrases which can be interpreted as literal or metaphorical ones depending on a context of use. The second data is the FigSen corpus which consists of 1833 short fragments of texts containing at least one phrase from the FigAN data which may have both meanings. The corpus is annotated in two ways. One approach concerns annotation of all adjective-noun phrases. In the second approach, literal or metaphorical senses are assigned to all adjectives and nouns in the data. The paper addresses statistics of data and compares two types of annotation. The corpora were used in experiments of automatic recognition of Polish non-literal adjective noun phrases.

Keywords: literal/metaphorical meaning, adjective-noun phrases, Polish, annotated corpora

1. Introduction

Natural language utterances are highly ambiguous on various levels. One word can be ambiguous between different inflected forms, syntactic categories and lemmas. One lemma can, in turn, have many different meanings. Usually immersing one word in a larger phrase helps to disambiguate its meaning, e.g. *flying bat* vs. *baseball bat*, but sometimes a specific phrase has a meaning which is very indirectly connected to the meaning of its constituents, or transposes its meaning into new areas, e.g., *to go to bat for* or *white elephant*. There are a lot of different linguistic phenomena which allow for using words in non-literal meanings: idioms, metaphors, hyperbole, metonymy, irony, etc. Their understanding requires deeper language knowledge. Any non-literal usage, easily understood by native language users, constitutes a problem for other speakers and for automatic text understanding systems. Although theoretically end-to-end applications could have dealt with these situations using information of a particular context, in practice, a lot of errors still occur in these cases. For languages for which data resources are not big the problem is even harder. For example, in Polish-English and English-Polish translations, well established metaphors, e.g., *kura domowa* which means ‘housewife’, are translated correctly, but for many other figuratively used phrases we often observe wrong word by word translations, see examples below:

Phrase:	<i>On jest ostatnią nogą z matematyki</i>
Google transl.:	‘He is the last leg in mathematics’
Meaning:	‘He knows nothing in mathematics’
Phrase:	<i>white elephant gifts</i>
Google transl.:	‘prezenty białego słonia’
Meaning:	‘bezużyteczne prezenty’
Phrase:	<i>kura domowa</i>
Google transl.:	‘housewife’
Literal transl.:	‘house hen’
Meaning:	‘housewife’

The above observations make it clear that a priori identifica-

tion of non-literal phrase usage could improve the results of different Natural Language Processing (NLP) applications, like information extraction (IE), question answering (QA) or machine translation (MT). Non-literal usage can theoretically concern any phrase, but to limit manual work needed to prepare the dataset of a size which will be sufficient for NLP purposes, we decided to concentrate on the adjective-noun phrases which very frequently have non-literal meaning. In (Mykowiecka et al., 2018) and (Wawer et al., 2019) we presented the results of various experiments concerning recognition of non-literal Polish adjective-noun (AN) phrases considered in isolation and within a text. In this paper, we describe in detail the datasets which were prepared during these experiments. The corpora which we created might be used for further work on Polish AN phrases and are available from: <http://zil.ipipan.waw.pl/CoDeS>.

2. Isolated Phrases

Although potentially all phrases can be used both literally and metaphorically, in practice we can distinguish three situations:

- for some phrases, e.g. *małe krzesło* ‘small chair’, it is hard to imagine when they get figurative meaning, they are strictly compositional;
- some phrases are used only in figurative meaning, e.g. *żelazny uścisk* ‘iron grip’ which does not have literal meaning in Polish ;
- the third group include phrases which have both literal and non-literal meanings, e.g. *biała flaga* ‘white flag’.

For the first two out of three groups of phrases, it is thus possible to reasonably well judge whether a phrase is used literally or not, without looking into its context. Our corpus was collected in order to test to what degree such automatic distinction is possible in practice for Polish phrases.

2.1. Existing Corpora

For English, there exist two datasets containing literal and non-literal AN phrases. Weak metaphors and phrases which can have both meanings are excluded from these sets. Tsvetkov et al. (2014) published data containing 884 phrases for each literal and metaphorical meaning which create a training set. Additional 100 phrases of each type create a test set. The training set contains 654 adjectives, while the test set contains 167 adjectives from the train set and 33 new ones. The data was used in the several experiments of recognition literal and metaphorical meanings of isolated AN phrases (Shutova et al., 2016; Rei et al., 2017; Bizzoni et al., 2017). Another data was published together with the paper of Gutierrez et al. (2016) and consists of many examples (3991 literal and 4601 metaphorical) of AN phrases containing only 23 adjectives. The goal of the experiment was to learn separately literal and metaphorical matrices which can be used to transform a noun vector into a vector representing the AN phrase consisting of the adjective and the noun.

In contrast to the above approaches, in the FigAN set we included isolated adjective-noun phrases of three categories:

- phrases which are interpreted literally (L);
- phrases which have only non-literal meaning (M);
- phrases which occur in both interpretations (B).

We also wanted to find equilibrium between number of different adjectives and number of different phrases including the same adjective.

2.2. FigAN Corpus

The initial goal of the research was to gather AN phrases which can have non-literal and both literal and non-literal meaning. In the second stage, to make our set more representative, we added also phrases which can possibly have only literal meaning. The collected phrases were additionally classified according to adjective and noun types.

- The adjectives were manually assigned to 55 classes (typology designed for this experiment) which represent such notions as: emotions, quantity, dimension, shape, color, etc.
- Among the nouns we distinguished only two classes: abstract and concrete.

The process of collecting of data was preformed in several steps. First, we prepared (on the basis of native language speakers knowledge) a list of 440 metaphorical phrases functioning in everyday communication. Then, we searched NKJP (National Corpus of Polish, (Przepiórkowski et al., 2012) for frequent adjective-noun phrases with adjectives which occurred in the initial set of 440 phrases. We collected randomly phrases containing the same adjectives to have a set of relatively many phrases for each adjective and to acquire also phrases which have only literal meaning. Then, the phrases were classified by a linguist into one of the three groups described above (B, M and L). The most numerous group, 79 phrases,

was collected for the adjective *czarny* 'black', it consists of 45 literal, 27 metaphorical phrases and only 7 phrases of both types (phrases of the B type are rarer than literal and metaphorical ones). In order to improve the participation of B phrases in our data we looked for them in dictionaries and added these which occurred at least dozen times in our texts. Moreover we added literal and metaphorical phrases for adjectives from the recently collected B phrases. The labels for the final list of phrases was evaluated by two annotators and inconsistencies were discussed in a larger group of project participants.

astronomia (2), cywilizacja (5), czas (1), czynność (1), czystość/brud (3), dobro/zło (2), dźwięk (2), emocje (10), estetyka (2), geografia (1), historia (1), ilość miejsca (1), jednostka miar (1), kolejność (2), kolor (9), kształt (1), majątność (1), materiał (16), mleczność (1), moc/słabość (2), nagość (1), orientacja w przestrzeni (5), otwarty/zamknięty (2), pełny/pusty (2), prawda/fałsz (1), prosty/krzywy (2), przedmiot (1), przeszkoda (1), sen/przebudzenie (1), sfera cielesna/duchowa (3), społeczeństwo (9), stabilność terenu (2), stan ciała/umysłu (5), stan fizyczny (4), stan skupienia (2), substancja (2), światło/ciemność (2), świeży/zepsuty (2), temperatura (2), trudność/łatwość (1), ukształtowanie terenu (3), własność (2), właściwość fizyczna (16), wolność/niewola (1), wymiar (1), zjawisko nadprzyrodzone (1), zjawisko pogodowe (5), zmysły (4), zwierzęta (22), życie/śmierć (2), żywioły (1), żywność (4)	astronomy, civilization, time, action, purity/soil, good/evil, sound, emotions, aesthetics, geography, history, amount of space, unit of measure, order, colour, shape, wealth, material, milkiness, strength/weakness, nudity, orientation in space, open/closed, full/empty, truth/false, straight/crooked, subject, obstacle, sleep/wake, bodily/spiritual sphere, society, terrain stability, state of body/mind, physical condition, physical state, substance, light/dark, fresh/broken, temperature, difficulty/easiness, landform, property, physical property, freedom/slavery, dimension, supernatural phenomenon, weather phenomenon, senses, animals, life/death, the elements, food
--	--

Table 1: Number of adjectives belonging to specific classes.

The final version of the dataset consists of 2379 Polish adjective-noun phrases containing 259 different adjectives, (Marciniak et al., 2018). The heterogeneous procedure of acquiring the data allowed us to collect both phrases which are relatively frequent and those which occur rather rarely. In Table 2 the number of different adjectives and nouns in all types of phrases are given.

	phrases	different	
		nouns	adjectives
<i>M</i>	1034	753	75
<i>L</i>	1017	762	82
<i>B</i>	328	231	73
<i>all</i>	2379	1526	259

Table 2: Statistics of noun and adjective types for the FigAN dataset

2.3. Non-literal Phrase Usage and Adjective Domains

On the basis of the collected data we check the distribution of three phrase types for different adjective domains. Figure 1 depicts information about numbers of phrases of each type for those adjective domains for which more than 30 different phrases were collected. It can be easily seen that phrases which can have both literal and figurative meaning (B) are usually the least frequent kind of phrase types. It confirms our observation that their acquisition, by just randomly selecting frequent phrases, may not give satisfactory results. There are three domains which have much more examples than the others. In the most numerous group (i.e., the physical feature domain), adjectives more typically form figurative phrases than literal ones. For the color domain the tendency is quite opposite, while the adjectives from the dimension domain form nearly equal numbers of literal and non-literal phrases. Relatively the largest number of type B phrases were formed by adjectives from the society and full/empty group. The groups with the most relative number of non-literal phrases are formed by the adjectives from the sense, tidiness, animal and body part domains.

2.4. Non-literally Used Noun Types

We also analyze the proportion of concrete and abstract nouns in all types of phrases (see Figure 2). For most adjectives, a pair (a noun type and a phrase type) exists which is evidently more common than the other ones. For example, for *deep* the vast majority of examples includes an abstract noun and are non-literal, while for *red* nearly all examples are literal and contain concrete nouns. The most evident exception to this schema is *good* for which the differences in number of phrase types in each group are less evident. Only one pair, i.e., concrete nouns in non-literal phrases (M-concrete), is much smaller than the other ones. The following adjectives form the greatest number of metaphorical phrases: *deep*, *full*, *sharp* and *clear*. The greatest groups of the type B phrases include adjectives *good*, *white*, *bright*, *black* and *dark*.

3. Phrases in Sentence Context

3.1. Existing Corpora

The most widely known corpus containing manual annotation of metaphors is the VU Amsterdam Metaphor Corpus (VU AMC) (Krennmayr and Steen, 2017). It consists of 115 excerpts of the BNC-Baby corpus (Burnard, 2008). They were randomly selected from four types of texts: academic texts, conversations, fiction, and news, and consists

of almost 190,000 lexical units. A procedure for recognition of metaphorical usage of individual words was developed by Steen et al. (2010b). It is based on ideas of Lakoff and Johnson (2008). All words in the corpus were examined according to the procedure if they were related to metaphors. A detailed annotation analysis is published in (Steen et al., 2010a), which shows that 86.3% of words were not related to metaphors and this number varies depending on the type of text (from 81.5% for academic texts to 92.3% in conversations). 12.7% were clear metaphor related words, and the rest were doubtful cases and words that function as metaphor flags, e.g., *as*, *like*.

The VU AMC annotation procedure is used to develop corpora annotated with metaphorical meanings for other languages. Pasma (2011) describes Dutch corpus consisting of news (~51,500 words) and conversations (~50,500 words). The author also reports that news texts contain more metaphor related words (22.1%) than conversations (11.9%). Moreover, adjectives are metaphors almost twice as often as nouns (for news it is 22% to 12%). The procedure is also successfully applied to metaphor identification in Mandarin Chinese (Lu and Wang, 2017). The corpus consists of excerpts from the Lancaster Corpus of Mandarin Chinese (Mcenery and Xiao, 2004) and contains 30,000 words. The metaphor related words are recognized in 3370 words which is only 11.2% of all tokens.

For Polish, there exists a corpus of synesthetic metaphors – Synamet (Zawisławska, 2018). It consists of 1,414 blogs devoted to 11 fields: perfume, wine, beer, yerba mate, coffee, cuisine, music, culture, massage, cosmetics, and wellness. The corpus contains ~685,600 tokens, 9,217 of which are grammatically and semantically annotated metaphorical units. In this research, the procedure of metaphor annotation differs from VU AMC one, as the author use *frames* instead of *domains* and recognises source and target frames.

3.2. FigSen Corpus

Our FigSen corpus consists of 1833 short fragments of text selected from the NKJP (National Corpus of Polish, (Przepiórkowski et al., 2012)). To limit manual work, we decided not to annotate random sentences, but we selected fragments which included at least one of the phrases we identified building our set of B type adjective-noun phrases (described in the previous section). As some of the phrases do not occur within the NKJP corpus in both senses, finally, we have examples of 165 different phrases from the FigAN set. The corpus is built from over 45,000 tokens including punctuation marks and excerpt delimiters. Each excerpt consists of one to three sentences and the average length is 24.5 tokens. The part-of-speech annotation is done with the help of the Concraft2 tagger (Waszczuk, 2012).

The set undergone two annotation procedures which differ both in annotation level (phrase vs. word) and in detailed scope of the elements annotated as M. In the first one, all grammatically correct occurrences of all adjective-noun phrases were annotated at the phrase level either as literal (L) or figurative (M). The M label was assigned not only to strict metaphorical phrases but also to other phrase occurrences which are non-literal but cannot be treated as a

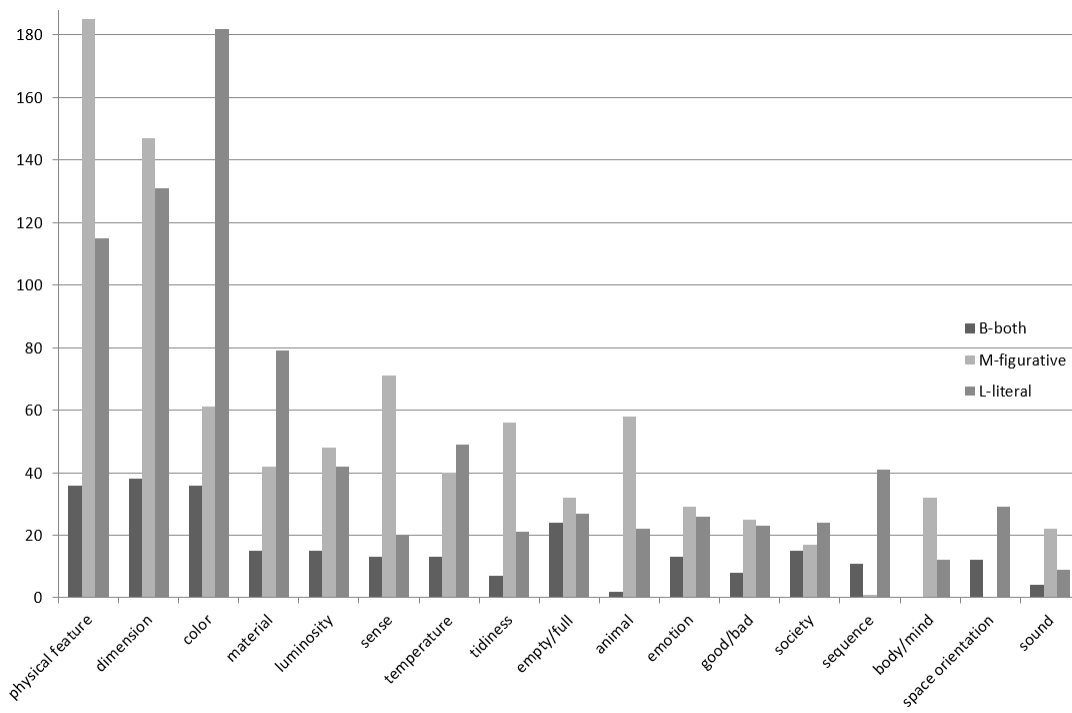


Figure 1: The number of the phrase types inside groups with adjectives belonging to different semantic domains. Only groups of the cardinality above 30 are shown.

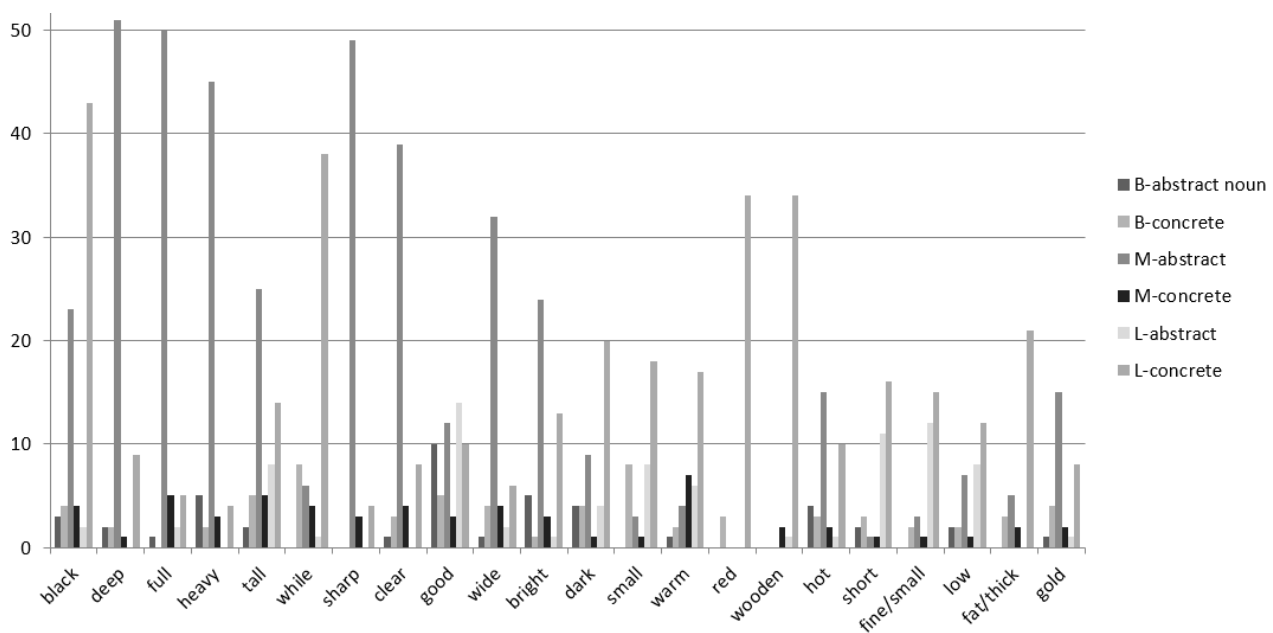


Figure 2: The distribution of the noun types (abstract vs. concrete) within each type of phrases for the phrases with the most frequent adjectives (at least 30 phrases).

pure metaphor in the sense defined in (Lakoff and Johnson, 2008). As the result of the annotation process, apart from the 1833 examples which were searched for intentionally, another 2279 occurrences of AN phrases within the selected sentences were also annotated. In total, in the entire corpus 2315 phrase types were identified (4112 occurrences), but only 269 of them occurred more than once. This subset is called FigSen-1 (Polak et al., 2019). In the second annotation schema (FigSen-2, (Marhula et al., 2019)) we went down to the word level. We were

still interested in noun and adjective phrases but we annotated all adjective and noun occurrences in the corpus. The annotation was done by two researchers specialising in metaphors in Polish. The annotators adapted the procedure for recognition of metaphorical usage of individual words developed for the VU Amsterdam Metaphor Corpus. A slightly modified procedure is described in (Marhula and Rosiński, 2018), while (Marhula and Rosiński, 2017) presents in more details the difficulties that arise when the method is applied to Polish. However, the general ideas of

	adj		ppas		adj+ppas	subst		ger		subst+ger	total	
	nb	%	nb	%	nb	nb	%	nb	%	nb	nb	%
<i>M</i>	1184	19.1	106	19.0	1290	1306	11.0	81	21.2	1397	2687	14.2
<i>L</i>	5004	80.9	453	81.0	5457	10520	89.0	301	78.8	10821	16278	85.8

Table 3: Statistics of *M/L* annotations in the FigSen corpus for each grammatical category occurring within AN phrases.

the annotation schema, in particular referring to the Lakoff and Johnson (2008) concept of metaphors, are preserved. Examples of excerpts containing the phrase *pierwsze skrzypce* together with the annotation on the word level are given below. It can be interpreted literally e.g ‘first violin’ in a context of an orchestra or a chamber, or it might mean ‘a leader of something’ in the case of a metaphor interpretation.

- *Oznacza to_L, że odtąd [pierwsze_M skrzypce_{M]} w państwie_L grać będzie premier_L.*
It means that from now on, the prime minister will take [the lead] in the country.
- *15 listopada_L 1983 został członkiem_L grupy_L [pierwszych_L skrzypiec_L] Filharmoników_L Berlińskich_L*
On November 15, 1983, he became a member of the [first violin] group of the Berlin Philharmonic.
- *... on był na pierwszym_M planie_M, w przeciwieństwie do poprzednich_M lat, gdzie to ona grała [pierwsze_M skrzypce_{M]}*
... he was at the forefront, unlike previous years, she was [the leader]

An inter-annotator agreement was tested on 51 excerpts consisting of 1246 tokens. In this fragment, there are 555 adjectives and nouns which were annotated by two people. 14 words were differently annotated, and the Cohen’s kappa was equal to 0.899, so the annotators obtained very good agreement. As the kappa was high and the procedure for annotation was very time-consuming, we divided the corpus into two parts which were annotated separately by one person. The final annotation was reviewed by removing minor inconsistencies and omissions which was done by one of the annotators. 180 decisions were changed, the label *M* was changed into *L* in 54 cases, and in the opposite way in 126 cases. Table 3 contains information regarding how many adjectives (regular adjectives and past participles fulfilling adjective roles), nouns and gerunds are annotated as having a literal and metaphorical meaning in the final annotation.

3.3. Corpus Analysis

Quite often, only one element of a metaphorical AN phrase has a metaphorical meaning. For example, in the phrase *gorzka prawda* ‘bitter truth’, which always has a metaphorical sense, the noun *truth* usually has a literal sense and only *gorzka* ‘bitter’ has a metaphorical sense. Whereas in the phrase *rodzina to nie tylko bezpieczna przystań* ‘family is not only a safe haven’, the AN phrase *bezpieczna przystań* ‘safe haven’ has a metaphorical sense because of

the noun *haven*, the adjective *safe* has the literal meaning. While *haven* is a safe place for boats after potentially dangerous voyages. Table 4 gives the numbers of AN phrases depending on the *M/L* meaning of their components.

type	nb	example
LL	2761	(planeta z) <i>gęstą atmosferą</i> ‘(a planet with) dense atmosphere’
LM	251	<i>kluczowe ogniwo</i> (twórczości Pestalozziego) key link of Pestalozzi’s work’
ML	489	<i>barwny opis</i> ‘colorful description’
MM	61	<i>gęsta atmosfera</i> (tej prozy) ‘the dense atmosphere (of this prose)’

Table 4: Number of occurrences of AN phrases in which non-literal meaning is expressed by an adjective (ML), by a noun (LM), by both phrase elements (MM), or phrases with literal meaning (LL) in the FigSen-2 corpus.

Numbers shown in Table 4 confirm that quite often only one phrase element is used non-literally while the other retains its standard meaning. What is more, such a metaphorical meaning could be used in several different phrases, e.g., *czysta abstrakcja*, *czysta ciekawość*, *czyste brzmienie*, *czyste konto* ‘clean abstraction/curiosity/sound/account’ and sometimes they are noted in lexicons. For all the examples from the previous phrases we can assign one of the 29 senses of the adjective *clean* in the Polish Wordnet (plWordNet, (Piasecki et al., 2009)). Thus the task of recognizing metaphors is intertwined with the word sense disambiguation task. The identification of metaphorical word/phrase usage may be easier than precise sense disambiguation and can improve the results of NLP applications. To illustrate more clearly this statement, for adjectives and nouns, we analyse the number of senses represented in plWordNet, see Table 5. In this statistics, we don’t take into account past participles and gerunds as they are not numerous in plWordNet, and their base forms are infinitives, so it is difficult to analyse them automatically. Relatively large number of nouns are not represented in the plWordNet due to many proper names in the FigSent corpus. Moreover, data was automatically processed what caused a certain number of errors, especially in the choice of base forms. For example, the tagger assigned to the noun *konwulsje* ‘convulsions’ a plural form as the base form, while in the plWordNet it is represented by a singular form *konwulsja*. Adjectives are slightly less represented in plWordNet, so there is no meaning for *skompromitowany* ‘discredited’, *przegniły* ‘rotten’ and *niepretensjonalny* ‘unpretentious’. The vast majority of words with one sense has literal meaning – 95.4%, *wyboisty* ‘bumpy’ is an example of adjective with one meaning in plWordNet, but metaphor-

type	nb.	lack		1 sens		2–4 senses		5–9 senses		senses 10 \geq		Max
all adj	1390	42	3.0%	306	22.0%	609	48.1%	319	22.9%	55	3.9%	29
adjM	181	5	2.7%	10	5.5%	71	39.2%	64	35.4%	31	17.1%	29
adjL	1319	38	2.9%	298	22.6%	629	47.7%	301	22.8%	54	4.1%	26
all subst	3938	463	11.6%	1310	33.3%	1654	42.0%	448	11.4%	63	1.6%	25
substM	442	11	2.5%	65	14.7%	217	49.1%	122	27.6%	27	6.1%	25
substL	3742	455	12.2%	1275	34.1%	1551	41.4%	403	10.8%	58	1.5%	23

Table 5: Number of senses of different words in the FigSen corpus.

ical sense in our data. The phrase *wyboista droga* ‘bumpy road’ is used in the context of someone’s life in the corpus. The noun *gangrena* ‘gangrene’ is very common used metaphorically as ‘destruction’, but in plWordNet it has one literal meaning — a type of tissue death. The statistics shows that words with more senses are more likely interpreted metaphorically.

3.4. Annotation Schemata Comparison

To compare two different approaches to non-literal phrase usage definitions, we transformed VU-like word level annotation to phrase level labels. The label *L* is assigned to an AN phrase if both elements are annotated as literal, while *M* is assigned if any of two elements (or both) has a metaphorical sense. Table 6 shows phrases with the most numerous examples together with the statistics of their labels in two annotation schema. For some phrases, the label distributions are exactly the same, but sometimes they differ quite significantly. The most prominent reason of these discrepancies are different choices of the most basic meaning of particular words. For example, the phrase *kosmiczna katastrofa* ‘cosmic/huge disaster’ is annotated in the opposite ways because in the VU-like annotation, ‘cosmic’ means ‘huge’ (after plWordNet) while in the second schema the more common understating of this word as related to the universe is assumed. Similarly, 23 literal interpretations of *false note* can be explained by the assumption that the basic meaning of the *note* is related directly to sound, not to graphical symbol. Some differences are due to contradictory assumptions made by annotators in cases when the context is very limited.

4. Corpus Usage and Future Plans

The FigAN and FigSen corpora were already used in the experiments for automatic identification of Polish metaphorical AN phrases. Experiments performed on the first set (Mykowiecka et al., 2018) confirmed that it is relatively easy to differentiate L and M phrases even if they are isolated. Using Word2vec 300 dimensional word representation (Mykowiecka et al., 2017) and simple two layer dense network we are able to classify such phrases with the F1 value equal to 0.89. Phrases of type B are much harder to classify; when all three classes were taken into account, the overall F1 value decreased to 0.79, and for the B class it was equal only 0.49. Nonetheless, as B phrases are relatively not very common, the initial classification into just two classes could also help eliminating some text interpretation errors. Proper classification of the B phrases is possible only in context, thus the need of building corpora in

which every phrase occurrence is annotated either as L or M. The FigSen-2 corpus was used to train the first model for automatic recognition of non-literal usage of Polish adjectives and nouns within AN phrases (Wawer et al., 2019). We obtained macro F1 equal to 0.81 with the help of LSTM net with the additional CRF layer and Wikipedia-trained Polyglot (Al-Rfou et al., 2013) word embeddings on the input level. Further experiments, both with the corpus extension and net architecture, changes are planned. The selection of excerpts for the FigSen corpus was focused on phrases which might have both metaphorical or literal senses. Additional phrases which occurred in these sentences are usually represented by single examples. In the future, to make our data set more robust, we want to supplement the corpus with the excerpts including more examples for phrases having only one meaning: literal or metaphorical one. We would also like to check how much the results of automatic identification of metaphorical phrases can improve results of a specific NLP application, e.g., machine translation.

5. Acknowledgements

This work was supported by the Polish National Science Centre project 2014/15/B/ST6/05186 *Compositional distributional semantic models for identification, discrimination and disambiguation of senses in Polish texts*.

6. Bibliographical References

- Bizzoni, Y., Chatzikiyiakidis, S., and Ghanimifard, M. (2017). “deep” learning : Detecting metaphoricity in adjective-noun pairs. In *Proceedings of the Workshop on Stylistic Variation*, pages 43–52. Association for Computational Linguistics.
- Burnard, L. (2008). *Reference Guide to BNC Baby (second edition)*.
- Gutierrez, D., Shutova, E., Marghetis, T., and Bergen, B. (2016). Literal and metaphorical senses in compositional distributional semantic models. In *Proceedings of ACL 2016 (short papers)*.
- Krennmayr, T. and Steen, G. (2017). VU Amsterdam Metaphor Corpus. In Nancy Ide et al., editors, *Handbook of Linguistic Annotation*, pages 1053–1071. Springer International Publishing.
- Lakoff, G. and Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.
- Lu, X. and Wang, B. P.-Y. (2017). Towards a metaphor-annotated corpus of mandarin chinese. *Language Resources and Evaluation*, 51(3):663–694, Sep.

phrase	All	FigSen-1 phrase annot.		FigSen-2 VU-like annot.	
		L	M	L	M
<i>pełne garście</i> ‘handful’	216	42	174	52	164
<i>gorzki smak</i> ‘bitter taste’	136	72	64	68	68
<i>głęboka rana</i> ‘deep wound’ or ‘deeply wounded’	91	65	26	65	26
<i>cierpki smak</i> ‘sour taste’, ‘sour grapes’	57	35	22	35	22
<i>falszywa nuta</i> ‘false note’, ‘deceitfully’	56	0	56	23	33
<i>czyste ręce</i> ‘clean hands’	33	8	25	10	23
<i>kosmiczna katastrofa</i> ‘cosmic/huge disaster’	29	0	29	29	0
<i>czysta karta</i> ‘clean page’	27	4	23	9	18
<i>miękkie nogi</i> ‘soft legs’, strong emotions	24	1	23	24	0
<i>miękkie lądowanie</i> ‘soft landing’	22	1	21	9	13
<i>słodki owoc</i> ‘sweet fruits’, something good	22	13	9	13	9
<i>twardy sen</i> ‘sound sleep’, ‘fast asleep’	21	0	21	0	21

Table 6: M/L labels distribution for the most frequent AN phrases in the FigSen dataset for two annotation procedures.

- Marhula, J. and Rosiński, M. (2017). Co oferuje MIPVU jako metoda identyfikacji metafory? *Polonica*, XXXVII (37).
- Marhula, J. and Rosiński, M. (2018). Chapter 9 Linguistic metaphor identification in Polish. In *Metaphor identification in multiple languages: MIPVU around the world*. <https://osf.io/phf9q/>.
- Mcenery, A. and Xiao, Z. (2004). The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, May 26-28, 2004, Lisbon, Portugal*. European Language Resources Association.
- Mykowiecka, A., Marciniak, M., and Rychlik, P. (2017). Testing word embeddings for Polish. *Cognitive Studies / Études Cognitives*, 17:1–19.
- Mykowiecka, A., Marciniak, M., and Wawer, A. (2018). Literal, metaphorical or both? detecting metaphoricity in isolated adjective-noun phrases. In *Proceedings of the Workshop on Figurative Language Processing*, pages 27–33, New Orleans, Louisiana, June. Association for Computational Linguistics.
- Pasma, T. (2011). *Metaphor and register variation: The personalization of Dutch news discourse*. Ph.D. thesis, Vrije Universiteit Amsterdam.
- Piasecki, M., Szpakowicz, S., and Broda, B. (2009). *A Wordnet from the Ground Up*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- Adam Przepiórkowski, et al., editors. (2012). *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN, Warsaw.
- Rei, M., Bulat, L., Kiela, D., and Shutova, E. (2017). Grasping the finer point: A supervised similarity network for metaphor detection. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1537–1546. Association for Computational Linguistics.
- Shutova, E., Kiela, D., and Maillard, J. (2016). Black holes and white rabbits: Metaphor identification with visual features. In Kevin Knight, et al., editors, *HLT-NAACL*. The Association for Computational Linguistics.
- Steen, G., Dorst, L., Herrmann, J. B., Kaal, A., and Krennmayr, T. (2010a). Metaphor in usage. *Cognitive Linguistics*, 21, 11.
- Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A., Krennmayr, T., and Pasma, T. (2010b). *A method for linguistic metaphor identification. From MIP to MIPVU*. Number 14 in *Converging Evidence in Language and Communication Research*. John Benjamins.
- Tsvetkov, Y., Boytsov, L., Gershan, A., Nyberg, E., and Dyer, C. (2014). Metaphor detection with cross-lingual model transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 248–258. ACL.
- Waszczuk, J. (2012). Harnessing the CRF complexity with domain-specific constraints. The case of morphosyntactic tagging of a highly inflected language. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 2789–2804.
- Wawer, A., Marciniak, M., and Mykowiecka, A. (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.
- Zawisławska, M. (2018). *Metaphor and Senses: The Synamet Corpus: A Polish Resource for Synesthetic Metaphors*. Peter Lang.

7. Language Resource References

- Marciniak, Małgorzata and Mykowiecka, Agnieszka and Polak, Justyna and Rychlik, Piotr. (2018). *FigAN: Adjective-Noun Polish Metaphorical Phrases*. Institute of Computer Science Polish Academy of Sciences.
- Marhula, Joanna and Rosiński, Maciej and Marciniak, Małgorzata and Rychlik, Piotr and Mykowiecka, Agnieszka. (2019). *FigSen-2: Annotation of Metaphoric Occurrences of Polish Adjective-Noun Phrases in Text*. ICS PAN, Warsaw.
- Polak, Justyna and Marciniak, Małgorzata and Rychlik, Piotr and Mykowiecka, Agnieszka. (2019). *FigSen-1: VU-like Annotation of Metaphoric Occurrences of Polish Adjectives and Nouns in Text*. Institute of Computer Science Polish Academy of Sciences.