

Metaphor Annotation in the Corpus of Polish

Magdalena Zawisławska

Institute of Polish Language
Warsaw University

Marta Falkowska

Institute of Polish Language
Warsaw University

Maciej Ogrodniczuk

Institute of Computer Science
Polish Academy of Sciences

Abstract

As a phenomenon involving both conceptual system and language, metaphor has been the subject of interest to researchers representing various disciplines, e.g. psychology, neurology, literary studies, linguistics, natural language processing, etc. Consequently, there is now a substantial body of scientific literature devoted to this topic, yet still predominated by theoretical approach. When we study texts, metaphor turns out to be a more multidimensional and complex issue than the existing theories assume. In order to study properly such complex phenomenon as metaphor, it is essential to start from a restricted research area, and to focus on analyzing authentic texts instead of fabricated ones. This is the main goal of our project — corpus of synesthetic metaphors in Polish. Metaphorical expressions in such a corpus need to be annotated both semantically and grammatically. The paper outlines the analytical procedure employed during the corpus compilation and annotation: it depicts the types of texts included in the corpus and exemplifies lexical items evoking various perceptual frames. Finally, some problematic issues connected with metaphors annotation observable in the material are discussed.

1 Introduction

The main aims of the project include: 1) compiling a corpus of synesthetic metaphors, based on authentic texts from blogs; 2) elaborating a new, efficient method of identifying and analyzing metaphors in discourse; 3) designing tools dedicated to annotating metaphors, analyzing and presenting results; 4) examining grammatical and semantic proprieties of synesthetic metaphor; 5) constructing a model of synesthesia for the Polish language. The result would be the first corpus of synesthetic metaphors in Polish. The corpus, comprising authentic, non-fabricated examples of textual metaphors, both semantically and grammatically annotated, would constitute an extremely significant resource for linguistic research. Moreover, the corpus would be useful for Natural Language Processing, since the very detailed, multilevel analysis

of discursive examples would allow to specify the productivity of certain metaphorical schemata, as well as to identify selectional restrictions of frames, frame elements, and lexemes. The project is in the initial phase. The dedicated tool has been built, and annotators have started to annotate the texts extracted from internet blogs.

2 The Sources of the Corpus

We assume that metaphor basically works in a context, which is why we opt for analyzing the whole texts, and not excerpts. Since we are only interested in synesthetic metaphors, we focus on texts excerpted from blogs devoted to perfume (SMELL), wine, beer, cigars, Yerba Mate, tea, or coffee (TASTE, SMELL, VISION), as well as culinary blogs (TASTE, VISION), music blogs (HEARING), art blogs (VISION), massage and wellness blogs (TOUCH). The reasons for studying synesthetic metaphors are twofold. First of all, their common use in language guarantees that the collected material is rich and varied enough (which should later help to broaden the analysis to include other types of metaphor). Secondly, restricting the research field to the domain of sensory perception ensures a precise indication of the research scope. The synesthetic metaphor category embraces various subtypes. Judycka [1963] cites the taxonomy proposed by E.R. Jaensch, who distinguished synesthesia in the narrow sense (expressions denoting only the perceptual sphere) and synesthesia in the broad sense (emotional synesthesia, such as e.g. *czarna rozpacz* ‘dark despair,’ conceptual symbols, and the so-called complex synesthesia). Judycka herself discriminates between word synesthesia (i.e. etymological, entrenched), e.g. *lęk* ‘fear’ (Proto-Slavic **lęk-ǫ* ‘I bend’), and verbal synesthesia. She also makes a distinction between simple and complex synesthesia, the latter category consists in merging the sensations coming from different perceptual domains, e.g. *chudy* ‘thin/lean’, *thusty* ‘fat/greasy’ [Judycka, 1963, pp. 59–60]. According to Werning *et al.* [2006] synesthetic metaphor embraces all the expressions containing any word primarily denoting a perceptual domain, i.e. either phrases such as *jasny dźwięk* (‘bright sound’), *słodki zapach* (‘sweet smell’), or constructions *jasny umysł* (‘clear mind’), *ciemna sprawa* (‘shady business’). In order to collect the most varied material possible, the project employs the broad sense of the term synesthetic metaphor. Hence, the following types of synesthetic metaphors shall be included, cf.: 1) simple

strong synesthesia, e.g. *jasny dźwięk* ‘clear tone’, *ostry zapach* ‘sharp smell’; 2) simple weak synesthesia, e.g. *czarna melancholia* ‘dark melancholy’, *śłodkie życie* ‘sweet life’; 3) complex strong synesthesia, e.g. *ciężki zapach* ‘heavy odor’, *lekki smak* ‘light taste’; 4) complex weak synesthesia, e.g. *śluzisty dowcip* ‘dirty joke’, *chudy intelekt* ‘meager wit.’ However, the word synesthesia (as Judycka calls it) will be excluded, since expressions of this type require etymological analysis.

3 Methodology of Description

Many Polish studies on metaphor use the interaction theory of metaphor. The problematic thing about this theory though is that it lacks a clear definition of *interaction* — Richards [1976] employs the terms *interanimation* and *cooperation*, while Black [1979] speaks of *isomorphism*. In Dobrzyńska’s [1984] interactive conception of metaphor, metaphor comes down to predication. The majority of recent corpus-based studies utilize either the conceptual metaphor theory, formulated by Lakoff and Johnson [1988], or the conceptual integration theory [Fauconnier and Turner, 2002]. Lakoff and Johnson view metaphor as a primarily conceptual phenomenon consisting in mapping across domains (from the source domain onto the target domain). However, it has not been precisely stated what the term *conceptual domain* stands for, and how the domain’s structure is supposed to be reconstructed. Moreover, the metaphors are presented as very general schemata *X is Y*, which results in the metaphors’ grammatical properties exhibited in the text being neglected. On the other hand, the conceptual integration theory formulated by Fauconnier and Turner is highly intricate, and it is quite difficult to put it into practice. The definition of mental space is as vague as that of the conceptual domain. Moreover, the analysis of metaphor as a linguistic phenomenon is beyond doubt insufficient. Recent works on metaphor have employed the Fillmorean frame semantics framework in order to account better for metaphor’s both cognitive and linguistic properties [Dancygier and Sweetser, 2014; Dodge *et al.*, 2015; Sullivan, 2013]. Also in the project *MetaNet: A Multilingual Metaphor Repository* two different methodologies are employed: the conceptual metaphor theory, formulated by Lakoff and Johnson, as well as the FrameNet ontology¹. Frames are defined generally as mental structures organizing human experiences. The term *frame* is understood in different ways by various researchers; they also call it: *schema*, *script*, *semantic frame*, *cultural model*, *cognitive model*, *domain*, *gestalt* [Tannen, 1979, p. 15]. In linguistics, the most prominent theory is frame semantics by Fillmore [1982]. Fillmore posits that the meaning of lexical units, phrases, grammatical and syntactic constructions resides in schematic phenomena, such as our beliefs, experiences or typical actions. However, the essential thing is that a frame is seen as an ordered structure within which there are categories (*slots*), and their values (*fillers*).

In the project, metaphorization process is seen as frame shifting. i.e. a ‘semantic reanalysis process that reorganizes existing information into a new frame.’ [Coulson, 2001]. It

¹<https://www.icsi.berkeley.edu/icsi/gazette/2012/05/metanet-project>

means that some elements of a frame evoking specific sensations (e.g. smell) as its topic may become reorganized under the influence of a vehicle activating a frame of some other sensory perception (e.g. hearing), cf. Example 1.

- (1) *Dochodzi zapach delikatnych kwiatów, szyprowy oddech mchu, a baza mruczy rozkosznie delikatnym piżmem.*

‘The **smell** of delicate flowers is drawing near, the chypre **breath** of moss, and the base note **purrs** contentedly with delicate musk.’

According to Petersen *et al.* [2007] we assume that frame shifting is not accidental but it depends on the structures of both frames (the source and the target one) — elements and values of the source frame are mapped on the analogous attributes of the target frame. We are aware that frame structure does not exhaust the potential of the metaphorical mapping. We also consider connotations² of analyzed words.

4 Metaphor Identification

The biggest challenge to be faced is the creation of metaphor identification procedure. The metaphoricity of certain expressions may be perceived quite differently by various language users. Hence, metaphor forms rather a continuum ranging from the most typical phenomena, through more peripheral ones, up to the borderline cases residing on the verge of the literal and the figurative meanings. There are several different procedures of recognizing metaphors in discourse. For instance, we have at our disposal the *met*system* [Fass, 1991] or Krishnakumaran and Zhu’s system [2007], which is based on the WordNet ontology, the MIP system [Group, 2007; Semino, 2008], the MIPVU system [Steen *et al.*, 2010], as well as metaphor identification using nouns and verbs clustering [Shutova *et al.*, 2010]. The metaphor recognition procedure draws on the systematic character of conceptual metaphor as it is seen by Lakoff and Johnson [1988]. In this method, the starting point consists in determining the basic set of metaphorical expressions that represent the source domain to target domain mappings. Subsequently, nouns denoting various concepts belonging to the target domain, and verbs denoting situations evoking the source domain are collected. In this way, a verbal lexicon pertaining to the source domain comes into being. The final stage involves browsing the corpus in order to note the expressions from the target domain with verbs from the source domain. However, the procedure cannot be adapted for our purposes, since it boils down to equating every metaphorical expression to a noun-verb combination.

The most widely known metaphor recognition procedure is MIPVU [Steen *et al.*, 2010], which constitutes a modified and elaborated version of the MIP (*Metaphor Identification*

²The term *connotation* we use as defined by Apresjan. Apresjan ‘understands connotation of a lexical item as the insignificant but steady attributes of the expressed thereby notion that are the embodiment of the assessment accepted in this linguistic community of the existing object or fact of the reality considering that connotations usually characterize the basic or the original meanings of words, and that they materialize through the figurative meanings.’ [Davletbaeva *et al.*, 2015, p. 160].

Procedure) as it has been proposed by the Pragglejaz group [Group, 2007; Semino, 2008]. The MIP comprises the following stages: 1) reading the whole text in order to establish its general meaning; 2) determining the lexical units that have been used in the text; 3) determining the meanings of the lexical units. The procedure requires that every lexeme's meaning was determined in the given context, i.e. how the word 'applies to an entity, relation, or attribute in the situation evoked by the text' [Group, 2007, p. 3]. The next step consists in determining whether each of the words has a different, more basic sense, activated in other contexts (e.g. *głowa kapusty* 'a head of cabbage' and *głowa* 'head' denoting a body part). According to the authors, more basic meanings of a lexeme include, cf.

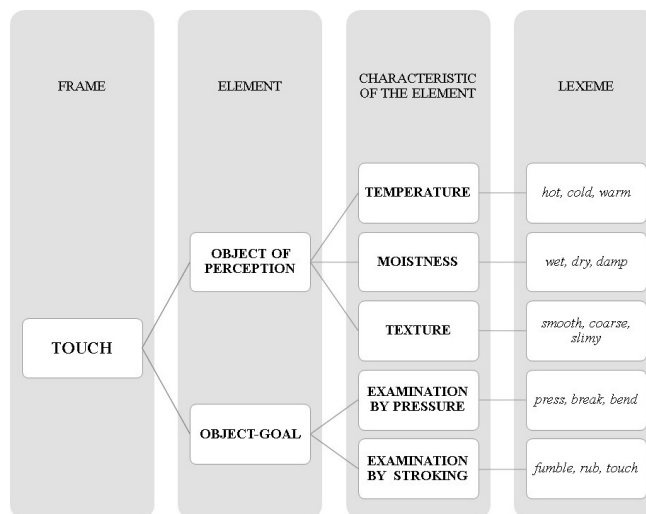
- concrete, physical meanings (what the word evokes is possible for the speaker to imagine; it can be seen, heard, felt, smelled, or tasted)
- meanings evoking bodily action
- more precise, clear meanings
- historically older meanings (however, it does not imply that a more basic meaning is also the most frequent one).

The MIPVU procedure [Steen *et al.*, 2010] has been enriched by adding the following principles: word class boundaries may not be crossed for lexical units represented by the same shape (i.e. meaning of a verb cannot be compared to a meaning of a noun), word etymology is hardly ever considered. Additionally, metaphors have been divided into subtypes: *direct metaphors*, *implicit metaphors*, *personifications*, and there is a separate class grouping problematic expressions (*borderline cases of metaphor*). Metaphor signals appearing in the text are taken into account. However, there are also doubts concerning this procedure, since too much attention is paid to data excerpted from dictionaries. It also takes great pains to establish 'the more basic' meanings. During the first stage of the project, the investigators are required to analyze texts on their own in order to indicate the occurrence of synesthetic metaphors. The MIPVU, as the most frequently applied procedure, is now tested during the annotation process. Yet, if we find that the procedure does not fit well the Polish material, a new procedure of metaphor identification shall be elaborated.

5 Metaphor Annotation Procedure

On the basis of the preliminary analysis, a set of basic frames for the five perceptual domains — VISION, HEARING, TOUCH, TASTE, SMELL was prepared, together with a lexicon containing vocabulary that evokes each of the frames (cf. Figure 1). Separate frames were elaborated for the so-called *complex synesthesia* as well as most for basic non-perceptual domains (e.g. BODY PARTS, MACHINE, TIME, PLACE, etc.). However, since it is conceivable that an important element of a frame could get overlooked, the annotators are granted the right to add the necessary elements during the corpus annotation.

The first stage of the annotation is distinguishing those excerpts from the analyzed text where forms representing the perceptual lexicon enter into syntactic relations. Annotators



Rysunek 1: A demonstration fragment of the TOUCH frame

check whether the heads or the dependents of these phrases evoke the same perceptual frame (a literal expression, e.g. *śmierdziało lawendą* 'it stank of lavender'), a different perceptual frame (a strong synesthetic metaphorical expression, e.g. *pachniał donośnie* 'it was smelling loudly'), or else a completely different conceptual frame (a weak synesthetic metaphorical expression, e.g. *miał czarne myśli* 'he's been having dark thoughts'). The most important element of the annotation is to identify the topic of the metaphor. We understand the topic according to Wróblewski [1998] as the element in the text that is construed within the realistic convention, that is the referent of the metaphoric predication. The same topic (referent) can be connected with several metaphorical expression forming a cluster, cf. Example 2.

- (2) *Encens Mythique d'Orient dość szybko zmierza do ambrowego akordu bazy. A ta jest cudownie ciepła, głęboka, słodkawa i jednocześnie piżmowa, przy tym bardzo zmysłowa. Pachnie zaskakująco donośnie i bardzo długo – otaczając noszącego prawdziwie luksusową aurą.*

'Encens Mysthique d'Orient makes its way towards the amber accord of the base quite quickly. And the latter is wonderfully warm, deep, sweetish and musk at the same time. What's more, it's really sensual. It smells surprisingly loudly and very long, surrounding its user with an air of true luxury.'

According to the above-mentioned assumptions, a substantial fragment of the text: *ambrowego akordu bazy. A ta [baza] jest cudownie ciepła, głęboka, słodkawa i jednocześnie piżmowa, przy tym bardzo zmysłowa. Pachnie zaskakująco donośnie i bardzo długo – otaczając noszącego prawdziwie luksusową aurą* constitutes a single metaphor cluster, since its referent is smell, evoked by the lexical unit *baza* (*zapachu*) '(fragrance) base.' As it can be seen, a metaphor cluster does not necessarily coincide with the utterance's borders. Therefore, in such cluster the annotator has to isolate smaller units. Metaphorical unit (MU) is defined as word forms or phrases

which are used metaphorically in the given context, i.e. they combine lexemes primarily belonging to different perceptual frames or to other types of frames. For instance, in the MC presented above, the following MUs may be distinguished:

- *baza (zapachu)* ‘(fragrance) base’ [VISION and SMELL]
- *ambrowy akord* ‘amber accord’ [HEARING and SMELL]
- *[baza] jest cudownie ciepła* ‘[the base] is wonderfully warm’ [TOUCH and SMELL]
- *[baza] jest cudownie głęboka* ‘[the base] is wonderfully deep’ [SMELL and COMPLEX SYNAESTHESIA]
- *[baza] jest cudownie słodkawa* ‘[the base] is wonderfully sweetish’ [SMELL and TASTE]
- *[baza] pachnie zaskakująco donośnie* ‘[the base] smells surprisingly loudly’ [SMELL and HEARING]

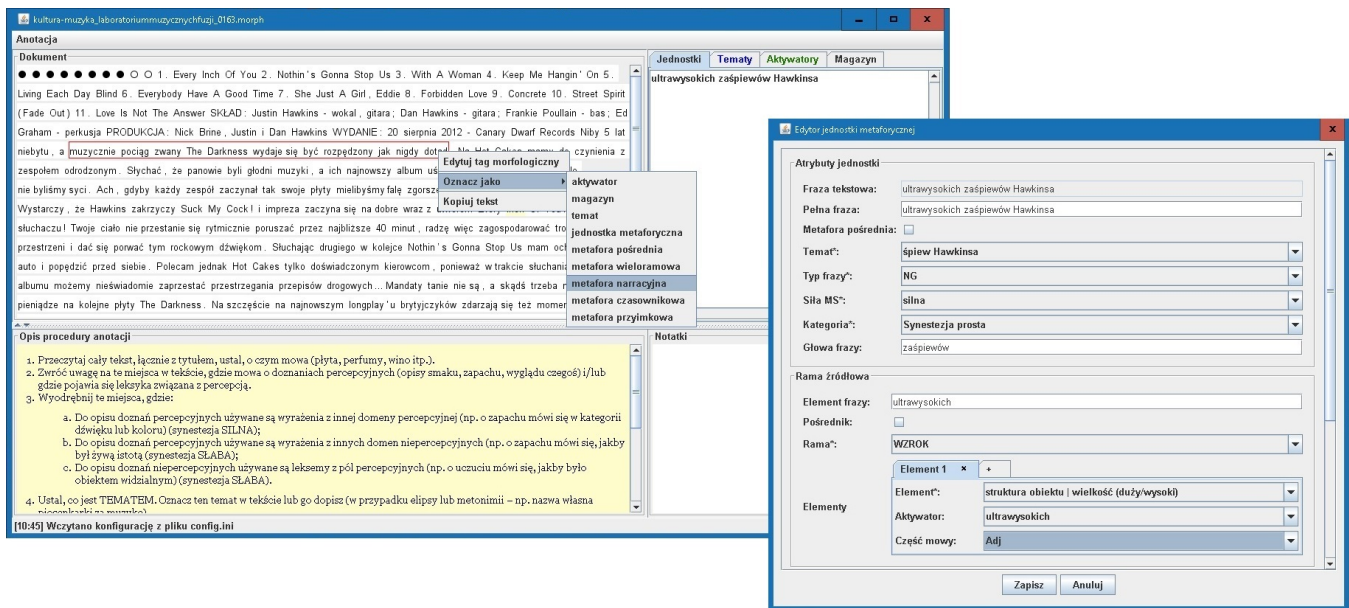
First of all, the annotator will have to complete the elliptical MUs with the missing components, indicate the topic and the modifier, and describe their grammatical properties. Afterwards, perceptual frames (or other types of frames) and their specific elements connected with the metaphorical expression will be chosen. For instance, the expression *baza (zapachu)* ‘(fragrance) base’ is elliptical in that its topic is missing. Its primary meaning is ‘a foundation, a bottom part.’ In the text, the word is used as a perfumery term — *baza zapachu* — i.e. the latest fragrance component to become perceptible. Hence, in that metaphor, *zapach* ‘smell’ (topic) is conceived of in terms of the visual perception. In contrast, the expression *ambrowy akord* ‘amber accord’ links the SMELL frame with the HEARING frame. The form *ambrowy* serves as the topic (which activates the SMELL frame, and within the frame — the element OLFACTORY SENSATION), and the word *akord* evokes the HEARING frame. Afterwards, the annotator will classify the MU under the appropriate semantic category of metaphors. The categories’ set shall be formed during the preliminary analysis of the samples. The categories that emerge at this stage include e.g. strong/weak synesthetic metaphor, simple/complex synesthesia. A separate *magazyn* (‘storehouse’) category dedicated to expressions whose metaphoricality is doubtful to the annotator.

Such cases will be investigated at a later stage by the whole research team. The tool implemented for metaphor annotation is shown in Figure 2. On the left side of the figure we can see a window with the text being analyzed. A fragment of the text can be tagged as: 1) *aktywator* (‘activator’), i.e. the phrase that activates a frame; 2) *magazyn* (‘storehouse’); 3) *temat* (‘topic’); 4) *jednostka metaforyczna* (‘metaphoric unit’). The annotator can also add the topic in the case of ellipsis. The annotation of MU takes place in a separate window. The annotator has to choose a topic for the MU, the type of the phrase (e.g. verbal phrase, adjective phrase, nominal phrase ect.), and describe the source frame (*rama źródłowa*), and the target frame (*rama docelowa*). The annotator also has to choose an activator of each frame and provide its grammatical description. The outcome of the MU analysis comprises: 1) a detailed grammatical categorization of the MU; 2) a general

metaphorical scheme of the MU, e.g. HEARING → SMELL; 3) lexical items activating the MU’s frame together with their grammatical description, i.e., for instance a combination of a lexeme primarily activating the HEARING frame with lexemes from the SMELL frame. It will allow to establish the productivity of each of the lexemes; 4) a detailed metaphor scheme of the MU, e.g. OLFACTORY SENSATION IS A SET OF SOUNDS; 5) the MU semantic categorization, e.g. strong synesthetic metaphor, simple synesthesia.

6 Problems with Metaphor Annotation

Corpus-based research of metaphor provides insight into the textual functions of metaphors. Most theoretical works on metaphor tend to regard it as a syntactic scheme *X is Y*. Soskice [1985, p. 19] argues that a lot of disputes over metaphor is rooted in the fact that researchers rarely specify whether they talk of the syntactic form of a metaphor or of its logical structure, and — on top of all that — they often confuse the two. Soskice believes that metaphor does not manifest itself in just one, specific form, since it is identified not only on the basis of syntactic criteria, but also semantic and pragmatic ones. The multiformity of metaphors in texts is also noticed by Goatly and Cameron. It follows clearly from Goatly’s [1997] analysis of examples from literature and from the *Bank of English* that metaphors are expressed not only by nouns, but by other parts of speech as well. Cameron [2003, pp. 88–89] in her corpus-based study has distinguished metaphors of different grammatical forms (verb and preposition metaphors), and comparison metaphors involving two incongruous domains arranged to form an explicit comparison. Cameron has compared the number of metaphors expressed by different linguistic forms, and it enabled her to determine that almost 50% of metaphors come in the form of verbs. Metaphorical expressions involving nouns (of the type *A=B*, e.g. *Man is a wolf*) make up no more than 5% of the corpus. Deignan [2005, p. 178] focused on the English word *blossom*, which may either be a collective noun denoting flowers, or a verb meaning ‘to produce flower(s).’ As she reports, the noun had 167 occurrences in the literal meaning and was used only twice in the figurative sense, whereas the verb occurred 5 times in its literal meaning, and as many as 55 occurrences of the verb represented the figurative meaning. As Wróblewski [1998, p. 31] rightly points out, in the description of metaphorical expressions the conceptions that prove to be the most useful are those that account for the lexicalized elements of metaphors. However, if we were to consider any hypothetical elements of the metaphor (e.g. as it happens in the case of an obligatory reduction of all metaphorical expressions to the *X is Y* schema), the study’s accuracy would diminish, since it would in fact consist in a subjective and arbitrary interpretation of metaphor. Therefore, in the project we try to be as close to the actual text form of a metaphoric expression as possible. Unfortunately, synesthetic metaphors in analyzed discourse are highly complex, coalesced; they form long chains of clusters evoking different kinds of perceptions at the same time, cf. Example 3.



Rysunek 2: The annotation tool

- (3) [...] *neroli brzmi tak jak zazwyczaj, czyli głównie zielono i kwiatowo, z lekką nutą octanową (to jaśmin), która wydaje mi się nie na miejscu. Składniki oud — to czone grzybem drewno, pewna smolistość — na chwilę ustępują świergotowi neroli, ale potem łagodnie obejmują przewodnictwo, a neroli dziwnie zaokrągla całość.*³

‘[...] neroli⁴ sounds [HEARING] as usual, which is mainly greenly [VISION] and flowery [SMELL] with a light [COMPLEX PERCEPTION] acetate⁵ [SMELL] note [HEARING] (it is a jasmine [SMELL]) which seems to be out of place. Oud⁶ [SMELL] components — wood eaten by fungus [VISION], some of pitch-blackness [VISION] — for a moment give in for a chirping [HEARING] of neroli [SMELL], but later gently take the leadership [PERSONIFICATION] while neroli [SMELL] strangely rounds off [VISION] the whole.’

Another problem that makes metaphor annotation a challenge is the specificity of Polish language. As the rich morphology allows to omit subject, there are sometimes very long sequences of sentences with no overt subject at all, cf. Example 4.

- (4) *Habit Rouge, znany nam dzisiaj, to zapach już po reformulacji, dostosowany do współczesnych nosów. Mimo to jest zaskakująco uważany za klasyka męskich perfum. Otwiera się głośno, przykurzonym aromatem cytrusów,*

³<http://nostrhills.blox.pl/html>

⁴Bitter orange oil.

⁵Ester, a chemical compound derived from an acid.

⁶Dense, dark and fragrant resin derived from the tropical agar tree, which is thought to originate in Assam in India (<http://www.gq-magazine.co.uk/article/best-mens-oud-fragrances-guide>).

znany chociażby z *Eau de Guerlain*. *Przypomina w tym miejscu wodę kolońską, ale nie taką, której używał tata.*⁷

‘**Habit Rouge** [subject] that we know today is a fragrance after reform, well adapted to contemporary noses. Yet, [it] is deservedly considered as a classic of men’s perfumes. [It] opens loudly with a dusty aroma of citrus which we know for instance from Eau de Guerlain. [It] is a reminiscent of cologne, but not like the one your father used to apply.’

Quite often a coreference chain of names in a metaphoric cluster that is connected with one referent differs in gender and sometimes in number, cf. Example 5:

- (5) *Wącham próbkę i już na wstępie mi się odechciewa. Nieśmiertelne, nudne i schematyczne niczym stylizyka Porsche, brzmienie – oparte na esencjonalnym połączeniu kardamonu, cytrusów i drewna sandałowego. [...] Znacicie to wrażenie, że zapach pachnie w sposób tak nieznośnie nośny, a jednocześnie jego agresywne i wszędobylskie brzmienie nie do końca trafia w wasze gusta? W pierwszym odruchu chcecie to paskudztwo z siebie zmyć, byle tylko zamknęto jadalnię i wreszcie zamilkło! [...].*⁸

‘I smell the **tester** (feminine) and at the beginning I already don’t feel like smelling it. The everlasting **sound** (neuter), boring, and schematic as the Porsche style — based on strong combination of cardamom, citrus, and sandal wood. [...] Do you know this feeling when the

⁷<http://www.opinie-perfumy.pl/guerlain-habit-rouge-perfumy-dla-drwala/>

⁸<https://perfumomania.wordpress.com/2015/12/22/fcuk/friction/for/him/czyli-sandalowiec-dla-sandalomasochistow/>

smell (masculine) smells in an unbearably popular way while at the same time its violent, and all-pervasive **sound** (neuter) not quite appeals to your taste. Your first impulse is to wash away this **muck** (neuter) in order to shut its gob, and finally make it to go quiet. [...]’

Sometimes the excess of ellipsis in an analyzed texts demands that annotators have to fill in essential elements of metaphorical expressions, cf. Example 6.

- (6) *Mineralność* wynagradza w *nim* niską kwasowość i już po *pierwszym nosie* czuć, że otwierając **2014** dokonaliśmy dzieciobójstwa.⁹

‘Minerality [of the wine] compensates **its** [the wine] low acidity, and just after **the first nose** [after the tester smelled the wine for the first time] we feel that when we open **2014** [the vintage of the wine] we have committed a infanticide.’

Another hard task is to discriminate lexicalized and novel metaphors in discourse. Text analysis has shown that in some contexts a seemingly dead metaphor may become alive. Knudsen [2003] has described it as opening of a closed metaphor. Müller [2009] has suggested to dispose of the division into the two classes. Instead, she has distinguished dead metaphors (where conceiving the target domain in terms of the source domain is entirely inactive in an average language speaker), sleeping metaphors (with low-level activation), and alive metaphors (high activation level). The discussion confirms that it is difficult to determine the borders of metaphor in a text. In the project, we employ the proposal put forward by Cameron and Maslen [2010], who introduced the term *potential metaphor*. Cameron and Maslen [2010, p. 102] suggest that what should be identified is potentially metaphorical words, which does not entail that the words must be regarded as such by all language users. It enables the annotators to take into account both entrenched and novel, creative metaphors. We distinguished a special category for the highly lexicalized metaphors that serve as terms in the professional discourse (e.g. *a note* in perfume industry, *a body* in wine industry). We call that category *premetafora* (‘premetaphor’). We also have separate category *magazyn* (‘storehouse’) which is intended for all questionable examples.

Last but not least, the fundamental role of a maximally broad context in the metaphor analysis must be considered. Quite often, if we consider just a limited context we may even overlook a metaphor, cf. Example 7.

- (7) *Jaka szkoda, że ta przecednej urody aria jest tak lakoniczna, dyskretna, ulotna i delikatna – bo aż by się chciało zatracić i przepaść z kretesem w tym zniewalającym brzmieniu...*¹⁰

‘It’s too bad that this utterly beautiful aria is so laconic, discreet, light, and delicate — because one would like to become completely engrossed and disappear in this captivating sound.’

⁹<http://blogi.magazynwino.pl/prange/>

¹⁰<https://perfumomania.wordpress.com/2015/10/26/acqua-di-parma-colonia-intensa-czyli-kwintesencja-ulotnej-doskonalosci/>

At the first sight, it looks like a literal description of an opera fragment. In reality, the excerpt concerns the perfume *Aqua di Parma*, therefore it is a metaphor.

7 Conclusion

The corpus resulting from the project would be the second metaphor corpus in the world, and the very first corpus of synaesthetic metaphors. We believe that by comprising authentic, non-fabricated examples of textual metaphors, both semantically and grammatically annotated, it would constitute an extremely significant resource for linguistic research. Since synaesthetic metaphor is attested in all natural languages, the results of our analyzes could be applied in comparative studies, e.g. with German [Werning *et al.*, 2006]. Moreover, the corpus would be useful for Natural Language Processing, since the analysis of discursive examples would allow to specify the productivity of certain metaphorical schemata, as well as to identify selectional restrictions of frames, frame elements, and lexemes. We reckon that the results of our analyzes will permit to establish (on a limited basis, for the time being) to what extent the metaphorization processes are systematic, and whether it is possible to construct a software system that would automatically analyze metaphors in discourse. It is also assumed that compiling the corpus shall only be a prelude to further research on metaphor in various discourse types. We believe that the procedures of metaphor identification and analysis, the tools designed for text annotation and presentation of results, as well as the results achieved during our research will constitute a good point of departure for compiling a large corpus of metaphors.

Acknowledgements

The work reported here was carried out within the research project financed by the Polish National Science Centre (contract number 2014/15/B/HS2/00182) and was partially financed as part of the investment in the CLARIN-PL research infrastructure funded by the Polish Ministry of Science and Higher Education.

Literatura

- Max Black. More about metaphor. In A. Ortony, editor, *Metaphor and Thought*, pages 19–41. Cambridge University Press, Cambridge, New York, 1979.
- Lynne Cameron and Robert Maslen. Identifying metaphors in discourse data. In Maslen Robert Cameron, Lynne, editor, *Studies in Applied Linguistics. Metaphor Analysis. Research Practice in Applied Linguistics, Social Sciences and the Humanities*, pages 97–115. Equinox Pub., London, 2010.
- Lynne Cameron. *Metaphor in Educational Discourse*. Advances in Applied Linguistics. Continuum, London, New York, 2003.
- Seana Coulson. *Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction*. Cambridge University Press, Cambridge, New York, 2001.

- Barbara Dancygier and Eve Sweetser. *Figurative language*. Cambridge textbooks in linguistics. Cambridge University Press, New York, 2014.
- Raisa Gubaydullovna Davletbaeva, Salima Aybulatovna Tagirova, Lidiya Fatikhovna Abubakirova, Dinara Damirova Khisamova, Shamilya Aflyakhovna Makhmutova, and Zaliya Damirova Khanova. The Connotative Component as Part of the Systemic Lexical Meaning of a Word (The Case of the Russian and Bashkir Languages). *Mediterranean Journal of Social Sciences*, 6(5 S4):158–166, 2015.
- Alice Deignan. *Metaphor and Corpus Linguistics*, volume 6 of *Converging Evidence in Language and Communication Research*. John Benjamins Publishing, Amsterdam, Philadelphia, 2005.
- Teresa Dobrzyńska and Maria Renata Mayenowa. *Metafora*, volume 4. Zakład Narodowy im. Ossolińskich, Wydawn. Polskiej Akademii Nauk, Wrocław, 1984.
- Ellen Dodge, Jisup Hong, and Elise Stickles. Metanet: Deep semantic automatic metaphor analysis. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 40–49, Denver, Colorado, June 2015. Association for Computational Linguistics.
- Dan Fass. met*: A Method for Discriminating Metonymy and Metaphor by Computer. *Computational Linguistics*, 17(1):49–90, 1991.
- Gilles Fauconnier and Mark Turner. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, New York, 2002.
- Charles J. Fillmore. *Frame semantics*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea, 1982.
- Andrew Goatly. *The language of metaphors*. Routledge, London, New York, 1997.
- Pragglejaz Group. MIP: A method for identifying metaphorically used words in discourse. *Metaphor and symbol*, 22(1):1–39, 2007.
- Irmina Judycka. Synestezja w rozwoju znaczeniowym wyrazów. *Prace Filologiczne*, pages 59–78, 1963.
- Susanne Knudsen. Scientific metaphors going public. *Journal of Pragmatics*, 35(8):1247–1263, 2003.
- Saisuresh Krishnakumaran and Xiaojin Zhu. Hunting Elusive Metaphors Using Lexical Resources. In *Proceedings of the Workshop on Computational Approaches to Figurative Language*, pages 13–20, Rochester, New York, April 2007. Association for Computational Linguistics.
- George Lakoff and Mark Johnson. *Metafora w naszym życiu*. PIW, Warszawa, 1988.
- Cornelia Müller. *Metaphors Dead and Alive, Sleeping and Waking: A Dynamic View*. University of Chicago Press, 2009.
- Wiebke Petersen, Jens Fleischhauer, Hakan Beseoglu, and Peter Bücker, editors. *The Baltic International Yearbook of Cognition, Logic and Communication*, volume 3: A Figure of Speech. 2007.
- Ivor Armstrong Richards. *The Philosophy of Rhetoric*. The Mary Flexner lectures. Oxford University Press, London, 1976.
- Elena Semino. *Metaphor in discourse*. Cambridge University Press, Cambridge, UK, New York, 2008.
- Ekaterina Shutova, Lin Sun, and Anna Korhonen. Metaphor Identification Using Verb and Noun Clustering. In C.-R. Huang and D. Jurafsky, editors, *Proceedings of the 23rd International Conference on Computational Linguistics, COLING 2010*, pages 1002–1010, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
- Janet Martin Soskice. *Metaphor and religious language*. Clarendon Press; Oxford University Press, Oxford, London, New York, 1985.
- Gerard J. Steen, Aletta G. Dorst, J. Berenike Herrmann, Anna Kaal, Tina Krennmayr, and Trijntje Pasma. *A method for linguistic metaphor identification: From MIP to MIPVU*, volume 14. John Benjamins Publishing, Amsterdam, Philadelphia, 2010.
- Karen Sullivan. *Frames and Constructions in Metaphoric Language*. John Benjamins Publishing, Amsterdam/Philadelphia, 2013.
- Deborah Tannen. What's in the Frame? Surface Evidence for Underlying Expectations. In D. Tannen, editor, *Framing in discourse*, pages 14–56. Oxford University Press, New York, 1979.
- Markus Werning, Jens Fleischhauer, and Hakan Beseoglu. The Cognitive Accessibility of Synaesthetic Metaphors. In N. Miyake R. Sun, editor, *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, page 2365. 2006.
- Piotr Wróblewski. *Struktura, typologia i frekwencja polskich metafor*. Wydawnictwo Uniwersytetu w Białymstoku, 1998.