The same or just much the same? Problems with coreference from the reader's perspective

Magdalena Zawisławska, University of Warsaw

Maciej Ogrodniczuk, IPI PAN

Abstract

Keywords: coreference, reference, identity, near-identity

The paper presents problems related to coreference annotation in the Polish Coreference corpus. There are three main causes of annotator errors: grammatical (e.g. the lack of an article system in Polish), semantic (the so-called *co-extension*, involving lexical relations between words) and cognitive (the annotators' insufficient real-world knowledge about certain relationships). Apart from provided examples of different kinds of annotation problems, the paper analyzes how coreference relates to identity in extralinguistic reality and in discourse. We also discuss the distinction between coreference and anaphora, as well as dependence of coreference on specific properties of Polish grammar. We also question M. Recasens' theory of near-identity and the need for its detailed classification.

1. Introduction

The main point of the article is to present and analyze issues which appeared in the process of annotation of the Polish Coreference Corpus – the first substantial Polish corpus of that type, created within the *Computer-based methods for coreference resolution in Polish texts (CORE)* project financed by the Polish National Science Centre (contract number 6505/B/T02/2011/40).

The aim of the project, ending in April 2014, is the creation of innovative methods of automatic *coreference resolution* – a task which is usually defined as determining which NPs in a text co-refer, i.e. refer to the same real-world entity. It is usually implemented as a two-step process:

1. identification of such NPs, i.e. *mentions* (in the current task: a group of adjacent words having a nominal head, e.g. pronouns, proper nouns, nominal groups etc.),
2. grouping mentions having an identical referent, i.e. building *coreference clusters*.

For instancje, for the following excerpt:

> *Wisława Szymborska otrzymała Nagrodę Nobla w 1996 r. Komitet Noblowski uhonorował ją za „poezję o ironicznej precyzji”. Część nagrody poetka przekazała na rzecz Fundacji „SERCE”.*
>
> *Wisława Szymborska received the Nobel Prize in 1996. The Nobel Committee awarded her “poetry of ironic precision”. The poet handed over part of the prize to the “SERCE” Foundation.*

the first task should identify the following mentions:

- Wisława Szymborska
- Nagrodę Nobla – the Nobel Prize
- Komitet Noblowski – the Nobel Committee
- ją – her
- *poezję o ironicznej precyzji – poetry of ironic precision*
- *ironicznej precyzji – ironic precision*
- *część nagrody – part of the prize*
- *nagrody – the prize*
- *poetka – the poet*
- *Fundacji „SERCE” – “SERCE” Foundation*

while the second one – the following coreference clusters:

- Wisława Szymborska, ją, poetkę – Wisława Szymborska, her, the poet
- Nagrodę Nobla, nagrody – the Nobel Prize, the prize.

All the remaining mentions are singletons (or clusters containing only one mention):

- Komitet Noblowski – the Nobel Committee
- *poezję o ironicznej precyzji – poetry of ironic precision*
- *ironicznej precyzji – ironic precision*
- *część nagrody – part of the prize*
- *Fundacji „SERCE” – “SERCE” Foundation*

A consistent identification of mentions requires stable and precise annotation guidelines which were created and updated throughout the process of corpus preparation. First of all, nested mentions with different semantic heads are identified (cf. “nagroda” – “the prize” and "część nagrody" – "part of the prize"). For each mention, the most descriptive sequence of words is stored (i.e. not just "poetry", but "poetry of ironic precision"), which includes an extended set of elements within mention contents, i.e., not only adjectives or subordinate nouns in the genitive, but also appositions, subordinate prepositional-nominal phrases or relative clauses.

The coreference resolution task can require even more difficult decisions: taking the most straightforward rule of linking the new mention to the last recent occurrence of another NP results in obviouserrorts , e.g. the pronoun "ją" – "her" in "*Komitet Noblowski uhonorował ją*" – "*The Nobel Committee awarded her"* would be incorrectly recognized as referring to the Nobel Prize and not to the poet (note that this problem occurs only in the Polish example – due to grammatical gender, which modern English is not considered to have). Another important restriction is clear separation of identity-of-reference relations from other potential types of relations, such as the relation between the Nobel Prize and the Nobel Committee.

The Polish Coreference Corpus currently contains over 1750 texts (see Table 1 for other statistics) randomly sampled from the National Corpus of Polish (pol. Narodowy Korpus Języka Polskiego, henceforth NKJP, see http://www.nkjp.pl; Przepiórkowski 2012) and balanced according to NKJP statistics of text genres for Polish. The texts have been manually annotated and super-annotated with mentions and coreference clusters.

Table 1. Statistics of the current version of the Polish Coreference Corpus

Texts	1773
Text size	250-350 words each
Words	503,985
Identity-of-reference relations	168,000
Near-identity relations	4,464
Non-singleton clusters	17,394

2. What is coreference?

Determining coreferential expressions in a text is not an easy task in the first place, even for human annotators. An annotator has to decide whether given phrases indeed refer to the same object, based on various syntactic, semantic and pragmatic indicators. Anna Wierzbicka (2010: 61) recognized the identity as a universal and elementary semantic unit, expressed by the phrase *the same*. In the mental spaces theory of G. Facounnier and M. Turner (2002: 95-96) the identity is also defined as one of the basic, vital relations in conceptual blending. The authors emphasize, however, that recognition of identity, sameness and equivalence is, in fact, a product of a very complex mental process. For example, we have to connect mental spaces of a baby, a child, a teenager and an adult through the relation of personal identity, despite fundamental physical differences involved.

Sometimes the identity relation can be interpreted differently depending on the knowledge of the world. For example, for an average speaker, there is no identity between a caterpillar, a chrysalis and a butterfly, although the relation between those beings is identical to that between a baby, a teenager and an adult.

2.1. Specificity of Polish grammar and coreference

Coreference – as a textual phenomenon – is not universal and it is strictly determined by properties of a particular language. The differences between coreference clusters in Polish and in English are seen in the example from M. Bulgakov "Master and Margarita", cf.:

> Procuratorowi skurcz wykrzywił policzek. Powiedział cicho:
> – Wprowadźcie **oskarżonego**.
> Natychmiast dwóch legionistów wprowadziło między kolumny z ogrodowego placyku dwudziestosiedmioletniego **człowieka** i przywiodło **go** przed tron procuratora. **Człowiek ów** odziany był w stary, rozdarty, błękitny chiton. Na głowie **miał** biały zawój przewiązany wokół rzemykiem, ręce związano **mu** z tyłu. Pod jego lewym okiem widniał wielki siniak, w kąciku ust miał zdartą skórę i zaschłą krew. **Patrzył** na procuratora z lękliwą ciekawością.
> *Mistrz i Małgorzata* (The Master and Margarita) by Michał Bułhakow (Mikhail Bulgakov), translated from the Polish by Irena Lewandowska, Witold Dąbrowski, published by Czytelnik (publishing house), Warsaw 1988.
>
> The Procurator's cheek twitched and he said quietly:
> `Bring in **the accused**.'
> At once two legionaries escorted **a man** of about twenty-seven from the courtyard, under the arcade and up to the balcony, where they placed **him** before the Procurator's chair. **The man** was dressed in a shabby, torn blue chiton. His head was covered with a white bandage fastened round his forehead, his hands tied behind his back. There

> was a large bruise under the man's left eye and a scab of dried blood in one corner of his mouth. **The prisoner** stared at the Procurator with anxious curiosity.
>
> *The Master and Margarita* by Mikhail Bulgakov, translated from the Russian by Michael Glenny, printed in Great Britain by Collins Clear-Type Press London and Glasgow © 1967

The differences between coreference clusters in the same text in Polish and English result from such grammatical properties of the Polish language as:

1) free word order in the sentence,
2) rich inflection,
3) zero-subject clauses,
4) no system of articles.

Some of those purely linguistic facts cause problems during annotation, especially the fact that there are no articles in Polish. Therefore, the most difficult task for the annotators was to distinguish between definite and indefinite referents in the text. In the following example, the annotator incorrectly created one cluster in which he placed all forms of the word *asystent (assistant)*, although they were not coreferential, e.g.:

> *Każdy szanujący się poseł ma* ***asystenta****.* ***Asystentami*** *są z reguły ludzie młodzi, ale nie brakuje również szczerze zaangażowanych emerytów. Poglądy polityczne* ***asystenta*** *powinny być zbieżne z linią szefa. Pracują jako wolontariusze tak jak Marek Hajbos,* ***asystent*** *Zyty Gilowskiej. Poseł Adam Bielan (rzecznik PiS) na przykład płaci* ***asystentom*** *za wysyłanie korespondencji. Obecny minister sprawiedliwości Grzegorz Kurczuk zaczynał partyjną działalność jako* ***asystent*** *Izabelli Sierakowskiej. W ministry poszedł też były* ***asystent*** *Józefa Oleksego Lech Nikolski. Posłowie nie poprzestają na jednym* ***asystencie****.*
>
> *Every decent Member of Parliament has* ***an assistant****.* ***Assistants*** *are usually young people, but there are also genuinely involved senior citizens. Political views of* ***an assistant*** *should coincide with that of their boss. They work as volunteers like Marek Hajbos,* ***the assistant*** *of Zyta Gilowska. The Member of Parliament Adam Bielan (spokesman of PIS), for example, pays his* ***assistants*** *for sending his mail. Present Minister of Justice Grzegorz Kurczuk started his party activity as* ***an assistant*** *of Izabela Sierakowska. Lech Nikolski, the former* ***assistant*** *of Józef Oleksy, was also appointed as a minister. Members of Parliament are usually not content with having just one* ***assistant****.*

2.2. Coreference and reference

It seems quite obvious that coreference must strictly bind up with reference. In all definitions, coreference is described as a phenomenon involving two or more expressions in a text signifying the same object. This means that some phrases cannot be coreferential since they do not have any referents, like for example indefinite pronouns. The analysis of the texts from the corpus shows that there are contexts where indefinite pronouns have reference and form coreferential clusters with other expression in the text, e.g.:

> *Jeśli* ***coś*** *przestanie być potrzebne, można* **to** *usunąć z dysku, zwalniając miejsce na inne zasoby.*
>
> *If* ***something*** *is no needed any more,* ***it*** *can be removed from the disc, to free space for other resources.*
>
> *Moja Wdowa, ona zawsze coś trafnego zacytuje. I właśnie zacytowała niedawno* ***coś śmiesznego****,* ***coś śmiesznie bolesnego****.* ***Coś****, co sobie ledwie uzmysławiałem jako jedną z przyczyn mego obrzydzenia się ludzką skórą,* **tego** *odpowiednio sam nie potrafiłem skrótowo nazwać.*
>
> *My widow, she always quotes something appropriate. And recently she quoted* ***something funny****,* ***something pitifully funny****.* ***Something*** *that I barely was aware of as a reason for my disgust with human skin, but I couldn't have appropriately named* ***it*** *in such a brief manner myself.*

The examples above show that coreference is not just a simple continuation of nominal group reference, but it is mostly determined by the semantic and pragmatic context and emerges only in the text.

2.3. Coreference and anaphora

On the other hand, coreference is not just a textual phenomenon, it also involves various paralinguistic elements. Thus coreference is not identical with anaphora. Though coreference and anaphora are strictly correlated in the text, it is essential to differentiate between these two phenomena. Anaphora is a purely linguistic tool which makes the text coherent. Usually, anaphora determines coreference, but there are cases when they do not coincide.

The most obvious example of the situation where an anaphor is present in a textalthough one cannot speak of coreference in the sentence, involves a noun phrase which is used predicatively, e.g.:

Chcę być **architektem** *i* **nim** *zostanę.*
I want to be ***an architect****, and I will become* ***one****!*

We cannot speak of coreference either when a noun phrase signifies an activity, an action or a state, e.g.:

Awantura *trwała cały rok, nie dało się* ***jej*** *przerwać.*
The argument *lasted the whole year, it was not possible to stop* ***it****.*

The anaphor can also refer to a clause, while coreferential expressions cannot, e.g.:

Zmierzchało *i bardzo go* **to** *przeraziło.*
It was growing dark *and* ***it*** *horrified him very much.*

It is also possible that coreferential expressions form a cluster, but there is no anaphor in the text. This happens when the noun phrase contributes new semantics content, e.g.:

Jan *wrócił ze Stanów.* ***Młody prawnik*** *był zachwycony wizytą.*
John *has returned from USA.* ***The young lawyer*** *was delighted by the visit.*

Nie mogę znieść **Marii**. **Idiotka** *popsuła mi samochód.*
I can't stand **Mary**. ***The***[1] ***idiot*** *broke my car.*

2.4. Coreference and other lexical relations

The most problematic case for annotators working on the coreference corpus was the so-called *co-extension*. It means that two or more phrases refer to objects which occur in the same conceptual field. The phrases can be related by various semantic relations, e.g. hypo-

[1] The anaphor in Polish does not exist due to the lack of articles.

/hyperonymy, meronymy, antonymy, etc. Such a relation very often makes it difficult to decide whether the phrases are coreferential or not.

In the first example, the annotator created a cluster in which s/he put the phrases *mity (myths)* and *mitologia (mythology),* while a myth is just a meronym of a mythology, and a mythology is a holonym of a myth, cf.:

> *[...]* **mity** *są niezastąpionym narzędziem dla psychologa, usiłującego prześledzić wzorce ludzkich zachowań . Wysiłki archeologów, religioznawców, antropologów doprowadziły z jednej strony do porzucenia eurocentrycznego spojrzenia na* ***mitologię****...*
>
> *[...]* **myths** *are irreplaceable tools for a psychologist, who is trying to follow through standards of human behavior. Efforts of archeologists, specialists in religious studies, anthropologists brought on the one hand giving up the Eurocentric look at the* ***mythology*** *...*

In the second example, the annotator could not decide whether s/he should establish the identity connection between the words *okupacja (occupation)* and *wojna (war).* It is obvious that those two words have something in common (country occupation is usually a result of a war, therefore it would be closest to the type of WordNet relation called *entailment*), but they are not coreferential, cf.

> *Od czasu* **okupacji***... - Ale tu je masz z powrotem, w metryce, i musisz ich używać w urzędowych papierach - powiedział oschle dyrektor i podsunął mi nowy blankiet do wypełnienia . - Kiedy to jest stara metryka, którą mi odtworzono zaraz po* **wojnie***.*
>
> *After the* **occupation** *... "But here you have them back, in your birth certificate and you have to use them in official papers," said the headmaster coldly, and gave me a new form to fill in. "But this is an old birth certificate, which was reconstructed after* ***the war****."*

The kind of problems like those listed above were the reason why Marta Recasens and her team working on the coreference corpus for Catalan and Spanish decided to introduce near-identity relations. Recasens stated that identity is some kind of a continuum, ranging from full identity to non-identity, and it is necessary to introduce additional links between words. Her typology of near-identity (2010: 151) is very broad and detailed, cf.:

A. Name metonymy

a. Role

b. Location

c. Organization

d. Information realization

e. Representation

f. Other

B. Meronymy

a. Part_Whole

b. Stuff_Object

c. Set_Set

C. Class

a. More specific

b. More general

D. Spatio-temporal function

a. Place

b. Time

c. Numerical function

d. Role function

We decided to avoid this detailed classification and limit possible relations between phrases to just identity and near-identity. The analysis of near-identity links in our corpus showed that annotators have had much bigger problems with deciding what type of near-identity to select than with establishing identity links between coreferential expressions. There are some cases where two annotators made different decisions and linked the same phrases in a given text as identical or near-identical. But different classification mostly arose due to 1) too shallow syntactic analysis, 2) confusing word meaning (and various lexical relations) with coreference, 3) the annotator’s lack of specific knowledge.

Some cases of “near-identity” arose because of insufficiently deep analysis of sentence structure. For example, in the sentence below, we have hidden predicative usage of nominal phrases:

> *Donald Tusk przybył na spotkanie nie jako premier, ale jako ojciec Kasi.*
>
> *Donald Tusk arrived at the meeting not as a prime minister, but as a father of Kasia.*

In this case there is no “near-identity” between the phrases *prime minister* and *father of Kasia*, although both of them describe two different roles of Donald Tusk. However, both nominal phrases serve as predicates – they have no reference to the object, they just describe two features of Donald Tusk (who is the prime minister of Poland and the father of Kasia).
Another example of near-identity marking was caused by a different syntactic phenomenon:

> *Have you read “Gone with the Wind”? No, but I’ve seen it.*

The dialog is about two different objects: the book “Gone with the Wind” written by Margaret Mitchell and the film “Gone with the Wind” directed by Victor Fleming. Of course there is something common in those two different entities: the film tells a similar (but not identical) story as the book. But it is not a “meta-object” like Recasens wants us believe, but a very simple ellipsis:

> *Have you read “Gone with the Wind”? No, but I’ve seen [THE FILM] based on it [THE BOOK].*

Such ellipsis is very common in language and it may cause various misunderstandings.

Most examples of near-identity marked in the corpus were in fact very typical semantic relations like homonymy, meronymy, metonymy, element of a set or, sometimes, hyperonymy, e.g.:

> ***Cała Warszawa** była właściwie jednym wielkim cmentarzem. W nasz dom uderzyło kilkanaście rozmaitych pocisków. Ginęli ludzie, mnóstwo ludzi! Na podwórku, już tak po 15 sierpnia, praktycznie codziennie był pogrzeb przed kapliczką. **Warszawa** była bardzo pobożna…*
>
> ***The whole Warsaw** was, in fact, a giant graveyard. Our home was hit by a dozen or so various shells. People were dying, plenty of people! After the 15th there were*

> *funerals in the courtyard, in the front of the chapel, almost every day.* ***Warsaw*** *was veryreligious...*

In this example, the word *Warsaw* is used in two different meanings 1. ‘city’, 2. ‘people living in the city’. In this case there is no coreference between those phrases, although they look the same.

The second example, where a big fire in ancient Rome is described, is much more difficult. There are phrases in the text which describe various elements of the occurrence which we called *fire*, but actually they are not coreferential with the phrase *fire*. It is the case of mero-/holonym or entailment lexical relation, cf.:

> *Teraz* ***pożar*** *zwalił się na budy kupieckie na Velia Carinae, pochłonął je jednym łykiem, łapczywie i prędko, po czym uderzył zaraz szeroką ścianą* **ognia** *na skupisko suburskie. Wielopiętrowe domy stawały w* **płomieniach** *jedne za drugimi. Z ogarniętych* **ogniem** *insul ludzie nie mieli czasu uciekać. [...] Krzycząc i nawołując się rozpaczliwie ludzie biegali tam i z powrotem, nie znajdując dla siebie wyjścia z* ***morza ognia****. Wpół oszaleli, w płonących tunikach, pędzili na oślep przed siebie, wpadali w* ***płomienie*** *i ginęli.*

> *The* ***fire*** *came down on the merchant’s ramshackle houses at Velia Carinae, ravaged them with one gulp, ravenously and fast, whereupon it hit the suburban center with a broad wall of* ***fire****. Multi-storey houses burst into* ***flames*** *one by one. People had no time to run from homesteads engulfed by blaze. [...] Screaming and desperately calling one another, people were running back and forth, unable to find an exit from the* ***sea of fire****. Half-crazy, in burning tunics, they were running headlong, rushing into* ***flames*** *and dying.*

But the main issue with establishing identity relations between phrases was the lack of the annotators’ competence in some fields. For example, in the first example, the annotator did not know that Johann Mühlegg, a German sportsman, was also a triple champion. Therefore an annotator did not mark a connection between the phrases *trzykrotny złoty medalista (triple gold medal winner)* and *Mühlegg*, cf.:

> *Jedynym moim pożywieniem w ostatnich trzech dniach były węglowodany – powiedział* ***trzykrotny złoty medalista****. Sportowcy niemieccy są ogromnie zaskoczeni wiadomością o pozytywnym wyniku testu antydopingowego* ***Mühlegga****.*

> *"During the last three days my only food was carbohydrates," said* ***the triple gold medal winner****. German sportsmen are very surprised by the news that* ***Mühlegg*** *tested positive for drugs.*

The second example is even more difficult. The recipient had to know that the players in the Silesian football team Ruch Chorzów wear blue undershirts and the players from the Warsaw football team Polonia wear black ones, cf.:

> *W trzecim kwartale 2010 roku* ***Ruch Chorzów*** *zarobił na czysto aż 5,5 mln zł. Wiadomość o zysku* **Niebieskich** *na pewno ucieszy jego kibiców. [...] Większość tych pieniędzy pochodziła ze sprzedaży do* **Polonii Warszawa** *dwóch reprezentantów Polski, grających w Ruchu w poprzednim sezonie. Za transfer Artura Sobiecha* ***"Czarne Koszule"*** *zapłaciły milion euro.*
>
> *In the third quarter of 2010,* ***Ruch Chorzów*** *earned a clear 5.5 million zlotys. The news about the profit of* **the Blues** *will surely please their supporters. [...] Most of this money came from the sale of two Polish team representatives who had played in Ruch to* **Polonia Warsaw** *during the previous season.* ***"Black Shirts"****. paid one million euro for the transfer of Artur Sobiecha.*

As stated above, most examples of near-identity were a result of either mixing different levels: syntax, semantics and reality, or of insufficient knowledge of a specific field. Therefore there is no point introducing not only a detailed typology of near-identity by Recasens, but also near-identity relations in general.

Of course, there are some examples which show that coreference does not necessarily mean identity of the object in the reality, e.g.:

> *W miejscu dawnej jezdni ryją buldożery. Bez problemu można dojechać ul. Bandurskiego, a że nawierzchnia* ***Retkińskiej*** *była znana jako jedna z najbardziej dziurawych w mieście, nikt nawet specjalnie nie skarży się na utrudnienia w ruchu.* ***Nowa Retkińska*** *będzie miała i sygnalizację u zbiegu z ul. Krzemieniecką, i chodnik (spory odcinek ulicy był go całkowicie pozbawiony).*
>
> *In the place of the former roadway, bulldozers are churning up the ground. There is no trouble getting to the Bandurski Street, and since the surface of the* ***Retkińska***

> ***street*** *was full of potholes of anyway, nobody really complains about impediments to traffic.* ***The new Retkińska Street*** *will feature both traffic lights at the junction of Krzemieniecka street and a pavement (which a large section of the street was missing).*

There are phrases in the text above which denote the same object in reality (the Retkińska Street) but they are not coreferential. In the real world, Retkińska is still the same street, although renovated, but the author of the text wants the reader to see this object as two different objects – the old Retkińska street with potholes and the new one with traffic lights and the4 pavement. In this case, reality does not matter – what is more important is the world created in the discourse. The author gives us a very clear and explicit syntactic and semantic signal – using the phrase *new Retkińska Street* – that he wants us to recognize the referent of the phrase as a different one. This example shows that coreference is more dependent on the discourse logic than on the real world logic. Recasens writes that "coreference relations between DEs depend on criteria of identity largely determined by the linguistic and pragmatic context" (Recasens et al. 2011; 1142).

4. Conclusion

The process of corpus annotation allowed us to test the ability to recognize coreference in the text by humans. Our experience shows that coreference is a very complex and multifaceted phenomenon. It is much more that just anaphora and reference. Coreference combines various different aspects, such as real-world knowledge, pragmatic context of the text, semantics and purely grammatical features of the phrases being analyzed.

We noticed that most errors in coreference clusters made by the annotators were a result of disturbance of communication on three different levels:

1) grammatical: e.g. the syntactic analysis of the relation in a clause was not deep enough or the lack of articles made it difficult to establish whether the phrases really referred to the same object,

2) semantic: lexical relations between words caused putting them in one cluster, although there was, in fact, no coreference,

3) cognitive: the annotator did not have sufficient factual knowledge to link words which were actually coreferential.

During the annotation of the corpus we have also verified M. Recasens' near-identity hypothesis. Although the annotators sometimes had problems with establishing whether given phrases were coreferential, we noticed that there was no recurrence of the near-identity links. In the same text, two annotators linked very different phrases as near-identical or often did not use that link at all. This means that there is no need to introduce such a complicated categorization of near-identity as the one proposed by Recasens. In fact, it is questionable whether there is a need to consider near-identity links in the coreference corpus at all.

We hope that our analysis becomes a valuable source of information for creators of future coreference corpora for other inflectional and free-word-order languages.. We believe that they could particularly benefit from our experience with annotating clauses with zero subjects. The Polish Coreference Corpus will be primarily used for implementation of computer algorithms and tools for effective automatic identification of coreference. Their creation is necessary for further development of research on numerous important issues situated at the crossroads between linguistics and computer science, such as machine translation, information retrieval and extraction or automatic summarization. The latter can, for instance, build upon coreference resolution for decoding expressions impossible to interpret without their antecedents even though they were not informative enough to be extracted for the summary. The Corpus can be also a useful source of linguistic information for research into text cohesion and coherence.

References

Fauconnier, Gilles, Turner, Mark. 2002. *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities*, New York: Basic Books.

Kunz, Kerstin Anna. 2010. *Variation in English and German Nominal Coreference*, Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang.

Libura, Agnieszka. 2010. *Teoria przestrzeni mentalnych i integracji pojęciowej. Struktura modelu i jego funkcjonalność,* Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Ogrodniczuk, Maciej, Zawisławska, Magdalena. 2012. *Semantic Approach to Identity in Coreference Resolution Task*. B. Glimm, A. Krüger (ed.), KI 2012, LNCS 7526, 241–244, Heidelberg: Springer.

Przepiórkowski, Adam, Bańko, Mirosław, Górski, Rafał Ludwik, Lewandowska-Tomaszczyk, Barbara, (ed.), 2012. *Narodowy Korpus Języka Polskiego* [Eng.: National Corpus of Polish], Warszawa: Wydawnictwo Naukowe PWN.

Recasens, Marta, Hovy, Eduard, Marti, M. Antonia. 2010. *A Typology of Near-Identity Relations for Coreference (NIDENT)*. Nicoletta Calzolari and Khalid Choukri and Bente Maegaard and Joseph Mariani and Jan Odijk and Stelios Piperidis and Mike Rosner and

Daniel Tapias (ed.), *7. International Conference on Language Resources and Evaluation, LREC 2010. European Language Resources Association (ELRA), Valletta, Malta*.

Recasens, Marta, Hovy, Eduard, Marti, M. Antonia. 2011. *Identity, non-identity, and near-identity: Addressing the complexity of coreference*. Lingua 121(6), 1138–1152.

Topolińska, Zuzanna. 1984. *Składnia grupy imiennej.* Zuzanna Topolińska (ed.), *Gramatyka współczesnego języka polskiego. Składnia*, Warszawa: PWN.

Wierzbicka, Anna. 2010. *Semantyka. Jednostki elementarne i uniwersalne*, Lublin: Wydawnictwo UMCS.