

ALICJA WÓJCICKA<sup>1,A</sup> & MATEUSZ KOPEĆ<sup>2,B</sup>

<sup>1</sup>University of Warsaw, Poland

<sup>2</sup>Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

<sup>A</sup>alicja.wojcicka@uw.edu.pl ; <sup>B</sup>m.kopec@ipipan.waw.pl

# MENTION DETECTION FOR COREFERENCE RESOLUTION IN POLISH. DEVELOPMENT OF THE FORMAL GRAMMAR

## Abstract

This paper presents the results of an improvement and extension of the Shallow Grammar of Polish, designed for the needs of the *Computer-based Methods for Coreference Resolution in Polish Texts (CORE)* project. The role of the Grammar was to detect nominal groups (i.e. multi-level nested phrases) that could be considered as mentions in coreference resolution tasks. In this article, the reorganization and changes to the Grammar are described, as well as the results of an evaluation of the Polish Coreference Corpus with manual annotations of mentions and coreferential expressions. A comparison of the second version of the Grammar with an evaluation of the first version reveals an improvement to the recall and F1 measures.

**Keywords:** coreference resolution; mention detection; shallow parsing; Formal Grammar for Polish

## 1 Introduction

The research described in this paper is a continuation of works that have been carried out within the *Computer-based Methods for Coreference Resolution in Polish Texts (CORE)* project, funded by the Polish National Science Center<sup>1</sup>; and within the CLARIN-PL Research Infrastructure Project. The aim of those investigations was to create a computer system that could identify coreferential linguistic expressions in a text longer than one sentence. Here, a coreference is understood as a relation between expressions that refer to the same entity (an object, space, time, or situation) in the world of a discourse. In most cases, they are nominal phrases of various types. A detailed report of the research has been published in Ogrodniczuk, Głowińska, Kopeć, Savary and Zawisławska (2015).

### 1.1 Annotation of mentions

One of the stages of the automatic coreference resolution process is ‘mention detection’. A mention is a nominal construct that can be considered as a candidate for membership in a cluster of coreferential expressions. In the CORE project, 4 major types of nominal expressions were regarded as mentions: single-segment nouns and pronouns; nominal groups (nested and not nested); zero

---

<sup>1</sup>Contract number 6505/B/T02/2011/40.

subjects (in Polish, a subject can be omitted in the shallow structure of a sentence); and named entities (Ogrodniczuk et al., 2015, p. 169). In order to identify nominal groups, the shallow parser Spejd, with a Polish grammar, was used. For the needs of the project, the NKJP Grammar (the grammar prepared for NKJP: Narodowy Korpus Języka Polskiego by Katarzyna Głowińska, see Głowińska, 2012) was adopted. A description of the improvements and changes made to the Grammar, as well as the results of the evaluation of the tool, can be found in Ogrodniczuk et al. (2015, pp. 172–176) and Ogrodniczuk, Wójcicka, Głowińska and Kopeć (2014).

As the results obtained during the project were not satisfactory (with only a slight improvement of recall, but a decrease in precision when comparing the NKJP Grammar and the new Grammar<sup>2</sup>), and the methods used in the process of mention detection were state-of-the-art for Polish, it was decided to make a further attempt to expand and improve the Grammar once more.

## 1.2 Tools used in the nominal groups detection process

As mentioned before, the shallow parser Spejd (Przepiórkowski & Buczyński, 2007) was used in the task of the detection of nominal groups. A Spejd grammar consists of a cascade of rules (regular expressions) where, apart from a definition of a match string of words or groups, a left and right context can (but does not have to) be specified. The rules are executed one by one, with the latter rule using the results of the application of the former. With the aid of the *word* and *group* operations, both syntactic words (such as, for example, multi-word forms of verbs) and syntactic groups can be created. In the grammar, the *unify* operation is broadly applied. It checks the agreement of the listed tags of the referenced segments, and if the tags are compatible, it deletes every interpretation that does not agree. The Spejd documentation, as well as the program itself, can be found on the webpage <http://zil.ipipan.waw.pl/Spejd>. The parser operates on texts pre-processed by the morphological analyser Morfeusz (Woliński, 2006) and the Pantera tagger (Acedański, 2010).

## 2 Design of the second version of the Grammar

Comparing the second version to the first version of the Grammar, the syntactic words layer has not been changed, nor have the rules responsible for detecting un-nested nominal groups (for example, [ten bardzo przystojny chłopiec] ‘that very handsome boy’). On the other hand, the group of rules for creating phrases with other phrases nested (for example [siostra [tego bardzo przystojnego chłopca czytającego [książkę o [piratach]]]]) ‘a sister of that very handsome boy who is reading a book about pirates’) has been reorganized and expanded.

The structure of the new version of the Grammar is shown in Figure 1.

### 2.1 Detection of nominal groups

**Valency** The most important modification in relation to the first version of the Grammar consists of the addition of a new grammatical category: valency. The value (or values) of valency determines the syntactical requirements of nouns, participles and adjectives, such as their case or type of prepositional phrase requirement. For example, all the comparative degree forms of adjectives have a valency value **od** (‘than’), because they can occur in constructions such as [drzewo wyższe od [domu]], ‘a tree higher than a house’; whereas a noun *odporność* (‘resistance’) has a valency value **na\_acc** (preposition *na* + accusative case form), because it requires a complement consisting of a preposition *na* and an accusative form of a noun (the preposition *na* can also govern the locative case), as in the group [odporność na [warunki atmosferyczne]], ‘weather resistance’.

<sup>2</sup>The test subset consisted of 530 randomly selected texts from The Polish Coreference Corpus, which were manually annotated for mentions and coreferential expressions (Ogrodniczuk et al., 2014, p. 276).

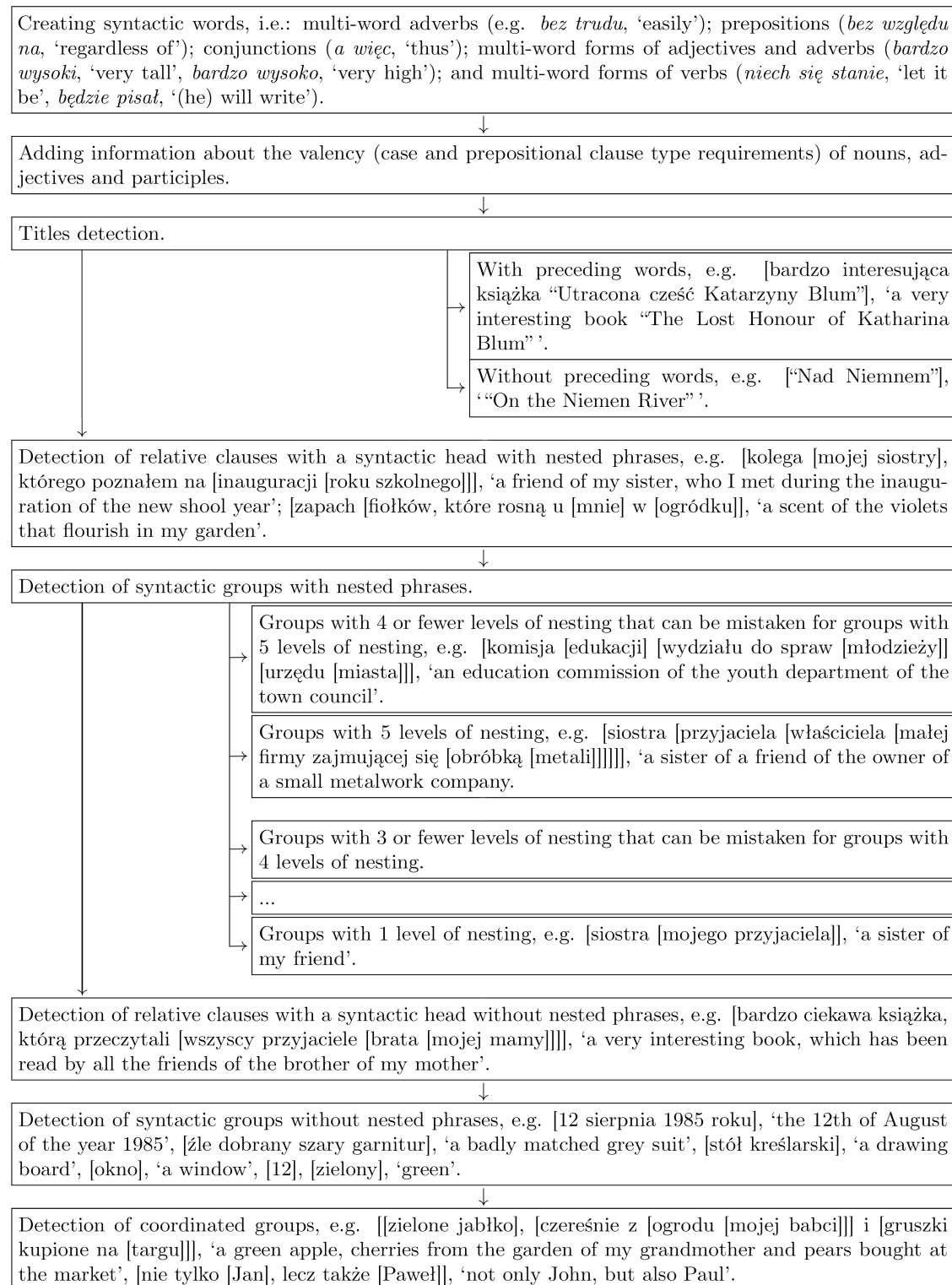


Figure 1: Processing chain in the second version of the Grammar

An adjective *nieznany* ('unknown') requires a complement in the dative case, e.g. *nieznany Janowi człowiek*, 'a man unknown to John'; therefore, it has a valency value **cel** (a Polish abbreviation of *celownik*, meaning 'dative case').

This modification allows prepositional groups and groups with other case requirements than the genitive to be included in the Grammar, and at the same time avoids the creation of incorrect groups. For example, the group [podróż z [Krakowa] do [Warszawy]], 'a journey from Cracow to Warsaw', is recognized as consisting of the head noun *podróż* with two complements: *Krakowa* and *Warszawy*, and not as consisting of the head noun *podróż* with one complementary clause with nesting: [podróż z [Krakowa do [Warszawy]]] because the noun *podróż* has the values **z\_gen** and **do** in the valency category, and the noun *Kraków* has no valency requirements. Of course, there are still some types of groups with an ambiguous structure. In such cases, the interpretation with deeper nesting is chosen because the rules for creating groups with more nesting levels are placed higher in the Grammar. For example, the group [kwiaty stojące w [wazonie stojącym na [stole]]], 'flowers standing in the vase standing on the table', will not have the structure: [kwiaty stojące w [wazonie stojącym] na [stole]]. The second interpretation is odd, but is, theoretically, syntactically correct. The participle *stojący* has both valency requirements: **w\_loc** (preposition *w* + locative case form) and **na\_loc** (preposition *na* + locative case form).

As a source of information about the valency values of nouns, participles and adjectives, the Polish Valency Dictionary, Walenty (Przepiórkowski et al., 2014) and the Polish Coreference Corpus (Ogrodniczuk et al., 2015, pp. 127–148) were used.

**Titles** Titles should be annotated as a single nominal phrase (without nested groups), even if they are syntactically complex (even a sentence can be a title), e.g. ["Utracona cześć Katarzyny Blum"] "The Lost Honour of Katharina Blum"; or ["Nie przyszedłem pana nawracać"] "I Did Not Come to Convert You". For this reason, before the section of the Grammar where the rules for detecting nested phrases are placed, there is a rule for creating groups without nesting from a string of words if it is in quotation marks and begins with an upper case letter. In order to avoid creating a group without nested phrases out of a quotation spoken by someone, the length of the string has been limited to 5 words.

Syntactic groups such as [ta bardzo interesująca książka "Utracona cześć Katarzyny Blum"], 'this very interesting book "The Lost Honour of Katharina Blum"' are considered as groups without nested phrases. For this reason, a list of the words which often precede a title has been defined (names of written items such as a *książka* 'book', *opowiadanie* 'short story', *wiersz* 'poem', *raport* 'report', *artykuł* 'paper', etc.) and used in the rules for creating such groups.

**Relative clauses** Relative clauses found together with a superior nominal group are annotated as mentions. Although relative clauses are very hard to detect automatically, there is a group of rules responsible for that task in the Grammar. Several problems related to relative clauses can be observed.

The main problem is how to determine the boundary of a relative clause. A punctuation mark can, but does not have to, indicate the end of a clause; or the clause can consist of a complex sentence with more than one verb phrase. In the Grammar, a relative clause is defined as follows: It is a string of syntactic words and groups in which: there is a nominal group (nested or not), with a comma, then the relative pronoun, such as *który* or *jaki* ('which' / 'that'), or with a prepositional group with such a pronoun, e.g. *na którego* ('on which'), where there is an agreement of gender and number between the superior nominal group and the pronoun, and then there is a string of syntactic words and groups with exactly one finite verb form or with a number of verb forms connected with a conjunction; there is an agreement of valency (case or type of prepositional group requirement) between the verb form and the relative pronoun (or the prepositional group with a relative pronoun); on the end of the string there is a punctuation mark (a comma, full stop, exclamation mark, or question mark).

The rules governing the creation of relative clauses are divided into two sub-groups, which are placed at two different stages of processing: a superior nominal phrase at the beginning of a relative clause (the head of the whole group) can contain nested phrases, or not. If there are nested phrases in the superior phrase, the rule for detecting such a relative clause has to be placed before the other rules for creating groups with nesting. For example, a rule that detects the clause [siostra [chłopca, którego znam]], ‘a sister of a boy who I know’, must be placed before the rule that would create the smaller group [siostra [chłopca]]. In this example, the form *chłopca* has the same value for gender as the form *którego*, and the noun *siostra* has another value for gender, so there is no doubt that the structure of the group has to be as it is mentioned above, not [siostra [chłopca], którego znam]. On the other hand, if there is a sentence [siostra [chłopca], którą znam], ‘a sister of a boy that I know’ (I know a sister, not a boy), the forms *siostra* and *którą* have the same value for gender (and number). Relative clauses with an ambiguous structure can also be identified if in the head of a clause there is more than one form with the same gender and number values, e.g. the sentence *siostra dziewczynki, którą znam*, ‘a sister of a girl that I know’ can have both structures assigned: [siostra [dziewczynki, którą znam]] and [siostra [dziewczynki], którą znam]. In such cases, the first interpretation is chosen, despite the fact that the second can also be correct in some contexts.

Relative clauses with a syntactic head without nesting are placed after the rules responsible for creating nested groups, but before the rules that detect groups without nested phrases. For example, the group [bardzo ciekawa książka, którą czytałem], ‘a very interesting book that I have read’, has to be created before the detection of the non-relative group [bardzo ciekawa książka], ‘a very interested book’.

For subordinate relative clauses, some nested nominal groups can occur, and they should also be detected. Thus, before relative clauses can be created, the Grammar should detect all such groups. For this reason, the context-dependent rules responsible for creating both nested and un-nested nominal phrases inside a subordinate relative sentence are placed before the rules for detecting whole relative clauses.

The process of detection of two types of relative clauses is described below:

- A. Processing chain of an exemplary clause: [siostra [chłopca], którą [brat [mojego kolegi]] spotkał w [Warszawie]], (‘a sister of a boy who [a sister] was met in Warsaw by a brother of my friend’).
1. Detection of nominal groups with nested phrases inside a subordinate relative clause:
    - a) Detection of nested syntactic group(s).  
Detected group: [mojego kolegi], (‘my friend’)
    - b) Detection of the whole group.  
Detected group: [brat [mojego kolegi]], (‘a brother of my friend’)
  2. Detection of the nominal groups without nested phrases inside a subordinate relative clause.  
Detected clause: [Warszawie], (‘Warsaw’)
  3. Detection of the nested groups inside the nominal syntactic head of a relative clause.  
Detected clause: [chłopca], (‘a boy’)
  4. Detection of the whole relative clause.  
Detected clause: [siostra [chłopca], którą [brat [mojego kolegi]] spotkał w [Warszawie]]
- B. Processing chain of an exemplary clause: [siostra [chłopca, którego [brat [mojego kolegi]] spotkał w [Warszawie]]], (‘a sister of a boy who [a boy] was met in Warsaw by the brother of my friend’).
1. and 2. — As described above.
  3. Detection of a relative clause.  
Detected clause: [chłopca, którego [brat [mojego kolegi]] spotkał w [Warszawie]],

(‘a boy who was met in Warsaw by the brother of my friend’)

4. Detection of a nominal group with a nested relative clause.

Detected clause: [siostra [chłopca, którego [brat [mojego kolegi]] spotkał w [Warszawie]]]

**Syntactic groups with nested phrases** Below the rules governing relative clauses, there is a section of rules for creating nested (non-relative) nominal syntactic groups. Within this section, the rules are ordered from the ones that detect groups with the deepest embedding (with a maximum of 5 nested phrases) to the ones that create groups with the shallowest embedding (with only 1 nested phrase). The rules for detecting syntactic groups with the same number of nested phrases are divided into types and are ordered from the broadest to the narrowest. Table 1 shows some exemplary types of the groups which are detected by the rules from this section.

Table 1: Exemplary types of groups with nested phrases

Number of nested phrases	Example of a type of group	Example of a realisation of the type
4	[sth + preposition [sth + preposition [sth [sth(gen) + preposition [sth]]]]]	[biedniaków przy [ognisku w [rejonie [budynków na [Karolinie]]]]] (‘banqueters by the campfire in the area of buildings in Karolin’)
4	[sth [sth(gen) [sth(gen) [sth(inst) [sth(gen)]]]]]	[środowiska [Systemu [zarządzania [bazami [danych]]]]] (‘environment of the database management system’)
3	[sth [sth(gen)] + preposition [sth] + preposition [sth [[sth] and [sth]]]	[wypalenia [traw] na [nieużytkach rolnych] pomiędzy [miejscowościami [[Racot] i [Nielęgowo]]]] (‘burning out grass on wastelands between the localities of Racot and Nielęgowo’)
3	[sth + preposition + [sth [sth(gen) [sth(gen)]]]]]	[pogląd na [kodyfikację [prawa [Bawarów]]]] (‘an opinion about a codification of the law of the Bavarians’)
3	[sth [sth(gen) [[sth(gen) [sth(gen)], [sth], [sth]]]]]	[ogromne ilości [[dwutlenku [węgla], [metanu], [freonu]]]] (‘enormous amounts of carbon dioxide, methane, freon’)
2	[sth [sth(gen) [sth(gen)]]]	[szereg [aktualnych spraw [świata]]] (‘a number of current matters of the world’)
1	[sth + preposition [sth]]]	[sadzawkę z [fontanną]] (‘a pool with a fountain’)

Every type of syntactic group with nesting is detected according to the same mechanism. Firstly, the deepest nested phrases are detected, so that in the end the whole group is created with phrases on the same level of nesting processed from the right to the left. The processing chain for syntactic groups with nested phrases is described below.

A. Processing chain of an exemplary clause: [środowiska [Systemu [zarządzania [bazami [danych]]]]] (‘environment of the database managing system’).

1. Detection of the deepest nested phrase in the context of a group with 5 nested phrases.

Detected syntactic group: [danych]

The context of the whole group must be specified; otherwise, every single noun would be marked as a nominal group, which would lead to an incorrect annotation. E.g., in the group *danych tej osoby* ('data for this person') the noun *danych* ('data') should not be detected as a mention (i.e. the system has to recognize the structure [danych [tej osoby]]).

2. Detection of the next phrases in order from the deepest to the shallowest. (The context of a group with 5 nested phrases is still specified.)

Detected syntactic groups (in the given order):

[bazami [danych]]

[zarządzania [bazami [danych]]]

[Systemu [zarządzania [bazami [danych]]]]

3. Detection of the whole group: [środowiska [Systemu [zarządzania [bazami [danych]]]]].

The most problematic groups in this section are the names of departments in offices, faculties in universities, and other complex names of parts of institutions, e.g. Komisja Bezpieczeństwa Dróg Urzędu Miasta ('The Road Security Commission of the Town Council'); or Instytut Anglistyki Uniwersytetu Warszawskiego ('The Institute of English Studies of the University of Warsaw'). The proper structure of these groups can be presented as follows: [Komisja [Bezpieczeństwa [Dróg]] [Urzędu [Miasta]]] and [Instytut [Anglistyki] [Uniwersytetu Warszawskiego]]. According to the processing chain of the Grammar, these groups should have incorrect structures: [Komisja [Bezpieczeństwa [Dróg [Urzędu [Miasta]]]]] (a group with 4 nested genitive phrases, such as [córka [siostry [ojca [kolegi [mojego szefa]]]]], 'a daughter of a sister of the father of a friend of my boss') and [Wydział [Anglistyki [Uniwersytetu Warszawskiego]]] (the same type of nesting, but with only 2 nested genitive phrases). This error arises because the nesting in the correct structures is not as deep as in the incorrect ones, and the nested phrases are all in a genitive case, which matches the pattern of the correct type of group with the nesting [sth of [sth of [sth...]]].

The rules that create a group with 4 levels of nesting are placed higher in the Grammar than the rules responsible for detecting groups with only 2 levels of nesting. For that reason, a modification of the rule order has been made, so that the rules recognizing the type [sth of [sth of [sth]] of [sth of [sth]]] have been placed before the rules that detect groups with 4 levels of nesting; and the rules for creating groups of the type [sth of [sth] of [sth]] have been placed before the rules for detecting groups with 2 levels of nesting. The same modifications have been made for all analogous types of groups. Additionally, in order to avoid an incorrect structure [córka [siostry [ojca]] [kolegi [mojego szefa]]], certain lexical restrictions have been introduced. For example, the type of group [sth of [sth of [sth]] of [sth of [sth]]] can be recognized only if the first and the fourth nouns, respectively, are names of a part of an institution and of an institution.

Besides relative clauses, complementary sentence clauses for nouns have also been taken into account. However, this part of the Grammar is still in development. The aim of this is to detect groups such as [czas, żeby przygotować [podziemne państwo]], ('the time to prepare the underground state'); and [powszechne przekonanie, że [ujawnienie [sprawy]] zapobiegło [planowanym na [13 kwietnia] kolejnym rytualnym samobójstwom]], ('a widespread opinion that the disclosure of the matter prevented the next ritual suicides, which were scheduled for the 13th of April'). The mechanism for creating such groups is analogous to the processing chain for relative clauses. The only modification involves replacing the relative pronoun with a conjunction, such as *że*, *żeby*, or with the questioning particle *czy*, and in adding the special restriction that in the head of the group only a noun can occur that requires a sentence complementation.

**Syntactic groups without nesting** The section of the rules for creating nominal groups without nesting is divided into several sub-sections. These are ordered as follows: special group types, such as addresses, dates, and hours; numeral groups; substantive groups (consisting of two or more nouns, e.g. appositions, names of persons); nominal groups modified by adjective(s); adjectival groups; and groups with abbreviations. At the end of this part of the Grammar, all single

nouns, numerals and adjectives are also marked as groups. The rules within each of these sub-sections are arranged from the broadest to the narrowest, e.g. phrases such as [bardziej znanych warszawskich agencji towarzyskich] ('better-known Warsaw escort agencies') will be created before clauses such as [najmniejsze podejrzenia] ('the slightest suspicions'); and dates such as [dnia 21 sierpnia 1985 roku] ('the 22<sup>st</sup> day of August of the year 1985') will be created before partial dates such as [21 sierpnia] ('the 22<sup>st</sup> of August') or [sierpień 1985] ('August 1985').

**Coordinated groups** In the Grammar, coordinated groups are also taken into consideration. A coordinated group can be nested in a broader clause, e.g. [znajomość [[świata] i [ludzi]]] ('knowledge of the world and of people'), as well as being the most external group with a nested coordinate, e.g. [[znajomość [świata]], [obyście w [towarzystwie]] oraz [urok osobisty]] ('knowledge of the world, good manners and personal charm'). The first type of coordinated group is detected in the part of the Grammar responsible for creating groups with nested phrases. The second type is created at the very end of the grammar. In the above example, groups with nesting such as [znajomość [świata]] and [obyście w [towarzystwie]] will be created first (by the rules placed in the section responsible for the detection of syntactic groups with nested phrases), and groups without nested mentions, such as [urok osobisty], will be detected in the section with the rules recognizing nominal groups without nesting. At the end of the Grammar, there is a rule that checks if the syntactic heads of all coordinated groups have the same value of grammatical case. If so, a coordinated group is created. Untypical coordinated groups are also detected, e.g. [zielone [gruszka] i [jabłko]] ('a green pear and a (green) apple'); or [[komisja katastralna] i [geodezji]] ('a cadastral and geodesy commission').

### 3 Errors in the automatic detection of nominal groups and their reasons

Not every type of group was automatically detected by the Grammar correctly. Several reasons for these errors can be identified.

1. Long and complicated clauses with untypical nesting patterns are often only partially recognized. For example: [czyścić [podróży], turystyczno-podróżne purgatorium, w którym [[nasze zabiegane organizmy] i [spłoszone dusze]] doznają [oczyszczenia], zanim przekroczą [próg [nieznanego]]], ('a purgatory of a journey, a purgatory of tourism and travel, in which our busy bodies and frightened souls are purified, before they cross the threshold of the unknown'). In this group, there is a problem with the boundary of a relative clause. The word *zanim* is not a coordinating conjunction, so the relative clause recognized by the parser is shorter (and will end with *oczyszczenia*).

2. Discontinuous groups created errors, especially those with parentheses, e.g. [stoiskach (działach) w [domach towarowych (handlowych)] z [artykułami spożywczymi] o [obsadzie [co najmniej 3 osób]]], ('stands (departments) in a department store (shopping centre) with food products and at least 3 persons of staff').

3. Errors in the earlier stages of processing, and especially errors with the Pantera tagger. In some cases, Pantera incorrectly recognized the values for the case, number or gender of nouns, adjectives and participles. In the Grammar, restrictions in the case, number and gender agreement between nouns and adjectives/participles are applied: nominal groups also have to meet certain syntactic requirements of verbs, nouns, adjectives or participles. If the tagger fails, some rules of the Grammar cannot be correctly applied. For example, in the text [górze trzech stronach [maszynopisu]] ('no more than three pages of a typescript') the word *górze* was recognized as a noun (while in fact, it is a particle), so it was not included in the nominal group by the parser. In the text [kamieniami symbolizującymi [pięć kontynentów]] ('stones symbolizing five continents'), a sub-group [pięć kontynentów] should have been recognized as an accusative case value (as the participle *symbolizującymi* governs the accusative), but the tagger marked the word *pięć* as a



nominative form. Thus, the group [kamieniami symbolizującymi [pięć kontynentów]] could not be created.

4. Errors caused by unknown words. Not every word in the corpus has a representation in the dictionary used by the tagger. For this reason, there are some words marked with the tag “ign” (ignored form). Among the unknown forms there are, first of all, proper names (of locations, persons, companies etc.), as well as misspelled words and digits. Only digits were taken into account in the Grammar, and they were included in the rules responsible for detecting dates and numeral expressions. All other unknown words were ignored. This led to dissimilarities between the manually and automatically annotated corpora.

## 4 Evaluation

The evaluations of the first and the second versions of the Grammar were carried out under the same conditions. 530 texts from the Polish Coreference Corpus were randomly selected. The results of the evaluation of the first version of the Grammar, cited from (Ogrodniczuk et al., 2014, p. 276), as well as of the second version are shown in Tables 2 and 3.

Table 2: Evaluation results of the NKJP Grammar and the first version of the PCC Grammar

Mention detection results		NKJP Grammar	PCC Grammar, Ver. 1
	Precision	66.07%	66.04%
	Recall	63.36%	63.99%
	F1	64.69%	65.00%

Table 3: Evaluation results of the second version of the PCC Grammar

Mention detection results		PCC Grammar, Ver. 2
	Precision	64.69%
	Recall	66.76%
	F1	65.71%

The slight improvement to recall and F1 resulted in a decrease in precision. Thus, in order to improve the results, the ignored words have to be taken into consideration, and the errors in the earlier stages of processing (especially errors in the tagger) have to be eliminated. A change in the parser should also be considered, since the Spejd parser does not handle discontinuous expressions.

## References

- Acedański, S. (2010). A morphosyntactic brill tagger for inflectional languages. In H. Loftsson, E. Rögnvaldsson, & S. Helgadóttir (Eds.), *Advances in Natural Language Processing: 7th International Conference on NLP, IceTAL 2010, Reykjavik, Iceland, August 16–18, 2010, Proceedings* (pp. 3–14). Berlin: Springer. <http://dx.doi.org/10.1007/978-3-642-14770-8>
- Głowińska, K. (2012). Anotacja składniowa. In A. Przepiórkowski, M. Bańko, R. L. Górski, & B. Lewandowska-Tomaszczyk (Eds.), *Narodowy Korpus Języka Polskiego* (pp. 107–127). Warszawa: Wydawnictwo Naukowe PWN.
- Ogrodniczuk, M., Głowińska, K., Kopeć, M., Savary, A., & Zawisławska, M. (2015). *Coreference. Annotation, resolution and evaluation in Polish*. Berlin: Walter de Gruyter.
- Ogrodniczuk, M., Wójcicka, A., Głowińska, K., & Kopeć, M. (2014). Detection of nested mentions for coreference resolution in Polish. In A. Przepiórkowski & M. Ogrodniczuk (Eds.), *Advances in*

- Natural Language Processing: 9th International Conference on NLP, PolTAL 2014 Warsaw, Poland, September 17–19, 2014: Proceedings* (pp. 270–277). Cham: Springer. <http://dx.doi.org/10.1007/978-3-319-10888-9>
- Przepiórkowski, A. & Buczyński, A. (2007). Spejd. Shallow Parsing and Disambiguation Engine. In Z. Vetulani (Ed.), *Proceedings of the 3rd Language & Technology Conference* (pp. 340–344). Poznań: Wydawnictwo Poznańskie.
- Przepiórkowski, A., Hajnicz, E., Patejuk, A., Woliński, M., Skwarski, F., & Świdziński, M. (2014). Walenty: Towards a comprehensive valency dictionary of Polish. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014* (pp. 2785–2792). Reykjavík: European Language Resources Association (ELRA).
- Woliński, M. (2006). Morfeusz — a practical tool for the morphological analysis of Polish. In M. A. Kłopotek, S. T. Wierzchoń, & K. Trojanowski (Eds.), *Intelligent Information Processing and Web Mining: Proceedings of the International IIS: IIPWM '06 Conference held in Ustron, Poland, June 19–22, 2006* (pp. 511–520). Berlin: Springer. <http://dx.doi.org/10.1007/3-540-33521-8>

---

## Acknowledgment

This paper is co-funded by the European Union, with financial resources from the European Social Fund, as part of Project PO KL “Information technologies: Research and their interdisciplinary applications”.

The authors declare that they have no competing interests.

The authors' contribution was as follows: concept of the study: first author; concept development: both authors; data collection and analyses: first author; evaluation: second author; the writing: first author.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (<http://creativecommons.org/licenses/by/3.0/pl/>), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.