ADAM KACZMAREK[1,A] & **MICHAŁ MARCIŃCZUK**[2,B]

[1] Computational Intelligence Research Group, Institute of Computer Science, University of Wrocław, Wrocław, Poland

[2] G4.19 Research Group: Computational Linguistics and Language Technology, Department of Computational Intelligence, Wrocław University of Science and Technology, Wrocław, Poland

[A] akaczmarek@cs.uni.wroc.pl ; [B] michal.marcinczuk@pwr.edu.pl

A PRELIMINARY STUDY IN ZERO ANAPHORA COREFERENCE RESOLUTION FOR POLISH

Abstract

Zero anaphora is an element of the coreference resolution task that has not yet been directly addressed in Polish and, in most studies, it has been left as the most challenging aspect for further investigation. This article presents an initial study of this problem. The preparation of a machine learning approach, alongside engineering features based on linguistic study of the KPWr corpus, is discussed. This study utilizes existing tools for Polish coreference resolution as sources of partial coreferential clusters containing pronoun, noun and named entity mentions. They are also used as baseline zero coreference resolution systems for comparison with our system. The evaluation process is focused not only on clustering correctness, without taking into account types of mentions, using standard CoNLL-2012 measures, but also on the informativeness of the resulting relations. According to the annotation approach used for coreference to the KPWr corpus, only named entities are treated as mentions that are informative enough to constitute a link to real world objects. Consequently, we provide an evaluation of informativeness based on found links between zero anaphoras and named entities. For the same reason, we restrict coreference resolution in this study to mention clusters built around named entities.

Keywords: coreference; zero subject; zero anaphora coreference in Polish

1 Introduction

Noun phrase coreference resolution is defined as the problem of determining which noun phrases occurring in a text refer to the same entity or concept in the real world (Rahman & Ng, 2011). While coreference can basically be viewed as a relation of identity of reference, there have also been studies discussing *near-identity*, or a continuum of such a relation, ranging from full identity to non-identity. See, for example, Recasens, Hovy, and Antònia Martí (2011). We did not determine, however, how such an approach could affect the final value of coreference resolution results.

Coreference resolution is often described as an important part of high-level applications e.g. text classification, text summarization, and question answering. Nevertheless, most of the approaches, as well as the evaluation metrics, treat coreference resolution as a generic clustering problem. This study focuses on the information that can be obtained from each coreferential relation, treating mentions with respect to the amount of information they carry about the described object. The study follows the coreference resolution definition from the KPWr corpus, which states that each coreferential cluster is constituted by at least one named entity. As this paper focuses on zero subject coreference, and operates under the assumption that only named entities carry enough information on their own to describe a distinct entity, the coreference resolution system is designed to most accurately assign zero anaphoras to corresponding named entities. End-to-end coreference is usually split into two tasks: identifying *mentions* — coreferential noun phrases — and finding relations between mentions, which from a computational perspective can be described as a clustering problem. In Polish, one can distinguish a subset of coreferential relations connecting zero subjects to other noun phrases — *zero anaphora*. Zero anaphora occurs when an independent clause lacks an explicit subject, like in the following example:

> **Maria** wróciła już z Francji. **Ø-Spędziła** tam miesiąc.

> **Maria** came back from France. **Ø-Spent**$_{singular:feminine:third}$ a month there.

This paper represents an initial step towards developing a machine learning-based system for zero anaphora coreference resolution in Polish, which utilizes a set of language-specialized features.

2   Related work

The automatic detection of zero subjects and zero anaphora coreferences have been the subject of several other studies for languages other than Polish. Russo, Loáiciga, and Gulati (2012) presented a study on improving translation of zero subjects from Italian and Spanish to French, using both rule-based and statistical machine translation methods. Rello, Baeza-Yates, and Mitkov (2012) and Rello, Ferraro, and Gayo (2012) presented an approach to zero subject detection in Spanish and Portuguese using machine learning techniques for distinguishing between explicit subjects, zero subjects and impersonals. Mihăilă, Ilisei, and Inkpen (2010) conducted a study on the distribution, identification, and coreference resolution of zero pronouns in Romanian using a supervised machine learning approach to the latter. Most recently, novel methods for zero coreference resolution in Chinese have been developed. Chen and Ng (2015) have proposed an unsupervised probabilistic model for the joint identification and resolution of zero pronouns by exploiting information about discourse salience, whereas Yin, Zhang, Zhang, and Liu (2016) have proposed a deep neural network model that eliminates the need for hand-crafted features, by generating an abstract representation. Finally, Iida, Torisawa, Hashimoto, Oh, and Kloetzer (2015) have presented an approach to improving zero anaphora resolution in Japanese by using a novel method of subject sharing recognition.

For Polish, there have been several approaches to coreference resolution. Ogrodniczuk and Kopeć (2011), Kopeć and Ogrodniczuk (2012) have proposed a rule-based and a machine learning-based approach for coreference resolution between all potential pairs of mentions, whereas a study conducted by Broda, Burdka, and Maziarz (2012) has provided a machine learning system for coreference resolution for clusters of mentions built around named entities. However, none of these studies have addressed the problem of zero anaphora coreference explicitly, leaving it for further study as a non-trivial task. Recently, there have appeared some studies on mention detection for zero subjects. Kopeć (2014) has presented a machine learning approach and has shown that the performance of dependency parsing itself for this problem is even lower than the majority baseline. Another heuristic approach, referred to as Minos, has been formulated by Kaczmarek and Marcińczuk (2015b) and has achieved state-of-the-art results. This study also provides an in-depth

analysis of verb classification in Polish from the perspective of zero subject detection. Although there have also been studies addressing some theoretical aspects of zero anaphora (e.g. Dunin-Keplicz, 1983), a dedicated approach to the automatic resolution of zero anaphora coreference for Polish has not been developed.

3 Coreference in KPWr corpus

The present study was conducted using a subcorpus of the KPWr corpus (Broda, Marcińczuk, et al., 2012), annotated with coreferential relations. It contains 1035 documents with a total of 45k links[1], of which 22k are connected to zero subjects. The KPWr corpus was annotated by two annotators who worked on separate sets of documents. The annotators followed precise guidelines and were verified by a supervisory linguist. Additionally, the corpus was annotated with all zero subject verbs in the aforementioned documents for the purposes of the development and evaluation of Minos. Coreference in the KPWr is only annotated for mentions that are coreferential with named entities. This approach addresses coreference as an information extraction problem, in which named entities are considered as the only mentions which are informative enough to provide valuable input from coreference resolution. Thus, for each mention in a coreferential cluster there is only one relation annotated — from the mention to the first named entity in the text belonging to the same cluster. The coreferential relations in the KPWr corpus are divided into four categories, based on the type of the mention pointing back to the named entity that is considered to be the head of the coreferential cluster:

- NE-NE — coreference between two named entities,

  *Chcę być tylko bliżej **Loty**, oto cały sekret. (. . .) moje i **Loty** serca rozumieją się tak doskonale.* (KPWr/101515)

  I only wish to be closer to **Charlotte**, that is the secret. (. . .) my heart and **Charlotte's** understand each other so perfectly.

- NE-AgP — coreference between a noun phrase and a named entity,

  *I tylko dzięki **Firefoxowi** nie weszłam na stronę, która usiłowała się otworzyć. (. . .) **Przeglądarka** zidentyfikowała stronę jako niebezpieczną.* (KPWr/101785)

  And only thanks to **Firefox** i did not enter a website that was trying to open. (. . .) **The browser** identified the website as insecure.

- NE-Pron — coreference between a personal pronoun and a named entity,

  *Najczęściej opisywanym (. . .) przypadkiem jest Król Rocka — **Elvis Presley**. Wielu akademików dowodzi, że stał się **on** dzisiaj zjawiskiem religijnym.* (KPWr/101815)

  The most frequently described (. . .) case is the King of Rock — **Elvis Presley**. Many academics prove, that **he** became today a religious phenomenon.

- NE-Zero — coreference between a zero subject and a named entity.

  ***Jakub** z KOL rozwinął temat dwóch z postulatów KOL. **Ø-Powiedział**, że musi być sprawozdanie o wydatkach na mieszkalnictwo.* (KPWr/102027)

  Jakub from KOL explained two demands presented by KOL. **He** said that there must be a survey of household spending.

---

[1]Link is an undirected connection between each pair of coreferential mentions in the cluster.

This paper focuses solely on the last category of relations, which is yet to be addressed directly for Polish. Individual counts of the relation subtypes are shown in Table 1. According to the table, zero subject coreference is the second most numerous group of relations, right after coreference between named entities. It shows how important the problem of zero anaphora coreference is in the context of the whole coreference resolution task for Polish, particularly as the number of zero subject relations is only slightly fewer than the sum of relations for noun phrases and personal pronouns combined.[2]

Table 1: Counts of specific coreference relations in KPWr.

Relation type	NE-NE	NE-AgP	NE-Pron	NE-Zero
Count	4335	1535	835	2053

4   The zero anaphora resolution model

The problem of zero anaphora coreference resolution in Polish has always been seen as a difficult task (Ogrodniczuk & Kopeć, 2011; Kopeć & Ogrodniczuk, 2012; Broda, Burdka, & Maziarz, 2012), especially compared to other categories of coreferential relations. This was the main motivation for conducting a study on zero subject coreference resolution utilizing coreferential links recognized for other types of mentions (i.e., noun phrases, pronouns and proper names). As the existing coreference resolution systems for Polish can proceed independently from zero anaphora, resolving all other types of coreferential relations, the most difficult part of coreference resolution was left to the end of the resolution process. This approach has the advantage of providing the greatest possible amount of information needed to resolve coreferential relations for zero subjects, and was justified by Stoyanov and Eisner (2012).

4.1   Overview

We based our approach on the cluster-ranking model presented by Rahman and Ng (2011), where for each mention the preceding coreferential clusters are ranked in order to determine the coreferential relations for the mention. In our solution, subsequent zero subjects are assigned to the partial coreferential clusters based on the results of a classifier. The clusters are then updated according to the assignment. The assignment is based on the prediction for pairs consisting of a mention under consideration and a cluster. As we follow the zero subjects in the order of their occurrence in the text, the cluster does not contain zero subjects following the current mention. This order of proceeding zero subjects, as well as updating the coreferential clusters, is based on the observed nature of zero anaphora coreference in Polish. It is very common for numerous subsequent zero subjects to create coreferential chains in consecutive sentences, while having no other coreferential mentions in between. This is especially true for certain categories of documents such as Wikipedia articles. In such cases, there exists a very strong grammatical link indicating a coreference between each adjacent pair of zero subjects, while for mentions of other types the coreferential indicators tend to decrease with growing distance in text. Despite the fact that there might be a coreferential link between a zero subject and a following mention, we consider coreference between two zero subjects to be directed from the latter to the former. We do not introduce such constraints for other mentions, as Dunin-Keplicz (1983), among others, has shown that in Polish there exist some intra-sentential dependencies between zero subject and noun phrases (including personal

---

[2]Note that the counts in the table represent numbers of single links between given mention and a head of the cluster, while total number of links include also all the links induced by transitivity and symmetry of coreference.

pronouns) following them in a text, that can indicate the admissibility or non-admissibility of a coreferential relation between them. Also, it would be not possible to simply adapt the approach presented in Broda, Burdka, and Maziarz (2012), in which personal pronouns are considered to be either directly connected to a named entity or indirectly connected, by being coreferent with an intermediate noun phrase that is coreferent with a named entity. In the case of zero subjects, an indefinite number of intermediate relations would be required, which makes such an approach unviable.

4.2   Features

In our problem, each instance represents a pair consisting of a zero subject and a cluster. We represent such an instance using three sets of features: **mention features** describing only the mention under consideration, **cluster features** describing solely the cluster, and **pair features** describing the relationship between the mention and the cluster. For zero subject mentions, apart from the standard grammatical features we also add features that take into consideration the existence and grammatical case of the relative pronoun *który*, the precedence of other verbs in the sentence, and information about coordinated and subordinated conjunctions to reflect the crucial information concerning the structure of the sentence. For the cluster by itself, we only take into account information about all the mentions which present some insight into their distribution across the document, and about the type of name entity constituting the coreferential cluster. For the mention and the cluster together, we focus primarily on relations between the zero subject mention and the two closest mentions in the cluster — the closest preceding mention and the closest following mention — both denoted as *closest mention*. These features include grammatical agreement, syntactic features (including being object or subject), and features connected to specific sentence constructs that can either indicate a coreferential relation or non-coreferential relation. A detailed list of features is presented in Table 2. Features are extracted from text pre-processed with the WCRFT tagger (Radziszewski, 2013) and MaltParser (Nivre, Hall, & Nilsson, 2006) trained on the Polish Dependency Bank (Wróblewska, 2012). After performing additional pre-processing steps on this feature set, i.e. one-hot encoding of categorical features and unfolding entity recency features for different numbers of sentences under consideration ($m \in \{1..5\}$), we ended up with ~ 150 final features used by the classifier.

4.3   The zero coreference classifier

Although the cluster-ranking model seems intuitively to be a good choice, it was not possible to directly adapt it to our problem. One of the issues was the fact that in our setting training a linear ranking SVM model would result in the disappearance of all the mention-specific features. This is caused by the fact that the training instances are created as differences of feature vectors for positive instances and feature vectors for negative instances within the same mention. Therefore, a decision was made to use a model that is non-linear by default — a random forest model provided by WEKA (Hall et al., 2009) — and we trained the classifier with a probability distribution over classified instances that could be used as a confidence measure of classification results. The confidence of prediction made by the classifier is utilized further in the process of resolution to help disambiguate cases of more than one cluster being classified as coreferential to a single mention. Additionally, by representing the problem as a classification task we gain the advantage of simplicity in representing cases of non-coreferential zero subjects by only creating the negative mention-cluster pair instances, instead of designing empty cluster representations or features dedicated for detecting non-anaphoricity, which in both cases would require great care.

Table 2: Cluster-Mention pair features.

Feature name	Feature possible values
Mention features	
Number	{Singular, Plural}
Gender	{Masculine, Feminine, Neutral}
Person	{First, Second, Third}
Preceding/Following relative pronoun	{True, False}
Preceding/Following relative pronoun case	{Nom, Acc, Gen, Dat, Inst, Loc, Voc, None}
Preceding/Following coordinated conjunction	{True, False}
Preceding/Following subordinated conjunction	{True, False}
Mention is the first verb in sentence	{True, False}
Cluster features	
Named Entity type	One of 56 named entity categories
Mention recency / Mention count in last m sentences	$\mathbb{Z}^*$, i.e. $\{0, 1, 2, ...\}$
Mention Frequency in terms of words	$\{0, 1\}$
Mention Frequency in terms of sentences	$\{0, 1\}$
Mention-Cluster Pair features	
Percentage of agreement in Number/Person/Gender	$\{0, 1\}$
Closest mention in the same sentence	{True, False}
Closest mention Number, Person, Gender agreement	{True, False}
Closest mention is subject	{True, False}
Closest mention is object	{True, False}
Closest mention grammatical case	{Nom, Acc, Gen, Dat, Inst, Loc, Voc, None}
Closest mention distance in tokens	$\mathbb{Z}^*$
Closest mention distance in sentences	$\mathbb{Z}^*$
Closest mention distance in mentions	$\mathbb{Z}^*$
Closest mention distance in verbs	$\mathbb{Z}^*$
Closest mention preceded by coordinated conjunction	{True, False}
Closest mention preceded by subordinated conjunction	{True, False}
Closest mention is possessive	{True, False}
Closest mention is reflexive	{True, False}
Closest mention is followed by relative pronoun	{True, False}
Case of relative pronoun following closest mention	{Nom, Acc, Gen, Dat, Inst, Loc, Voc, None}

4.4   Training the classifier

Training instances were prepared for each document by generating for each zero subject a positive instance — a pair consisting of a zero mention and the coreferential cluster it belongs to — and a few negative instances consisting of the current zero mention and a cluster it does not belong to. For non-coreferential zero mentions, only negative instances were created, based on coreferential clusters associated with a named entity in the document. Clusters in the training instances were prepared in such a way that they consisted of all non-zero mentions (named entities, agreed noun phrases, personal pronouns) originally present in the cluster, and only zero subjects that occurred in the document before the relevant mention, according to the sequential nature of our approach.

By default. the generated instances are strongly imbalanced in favour of negative ones because we are generating at most one positive instance but several negative instances for a single mention, as there are usually more than two different coreferential clusters in a document. We obtained the best results by limiting the negative instances similarly to a maximum of one negative instance for a single mention.

4.5 The resolution process

The coreference resolution process for a single document can be described as a sequential assignment of zero subjects to clusters to determine their non-anaphoricity. For each zero subject, in the order of their occurrence in the text, and each coreferential cluster with removed zero subjects occurring after current mention, we generate a cluster-mention pair instance. We then classify these pairs and assign the zero subject to the cluster that constitutes a positively classified pair. In cases where more than one pair is classified as coreferential, we break the tie using the positive class probability value from the random forest classifier and. if still necessary, by minimal distance from the current mention to the closest preceding mention in the cluster. We then update the cluster to include the zero subject according to the results and we proceed to the following mention. In a case when all of the instances are classified as non-coreferential, we do not conduct any operations on clusters before proceeding to the next mention.

5 Evaluation

We have evaluated our approach (henceforth called *Crete*) alongside two other tools for coreference resolution for Polish: *Bartek* (Kopeć & Ogrodniczuk, 2012) and *Ikar* (Broda, Burdka, & Maziarz, 2012) with a zero anaphora coreference baseline as described by Kaczmarek and Marcińczuk (2015a).

5.1 Metrics

We used two different approaches to evaluate the coreference resolution systems. In the first approach, we followed the CoNLL 2012 scoring scheme (Pradhan et al., 2014) using three metrics for scoring coreference clusters without taking into account mention types: MUC (Vilain, Burger, Aberdeen, Connolly, & Hirschman, 1995), B^3 (Bagga & Baldwin, 1998) and $CEAF_E$ (Luo, 2005). The final score is computed as a harmonic mean of F_1 scores for these three metrics. There are two main drawbacks of using this scoring scheme. The first is the lack of sensitivity for various levels of informativeness for different relations in coreferential clusters. Thus, in the second approach we employed the Parent metric (Kaczmarek & Marcińczuk, 2015a) to evaluate zero anaphora relations in isolation. This metric allows us to score solely the zero anaphora relations according to the value of information they provide about discourse entities. In our settings, we only take into account relations between zero subjects and named entities, due to the fact that only they are considered to describe real world entities to a degree allowing us to extract valuable information, as mentioned in Section 3.

5.2 Evaluation settings

We conducted evaluation on the KPWr subcorpus mentioned in Section 3 using a 10-fold cross validation, with the folds balanced in terms of documents and relation counts, and also prepared to reflect the document type distribution across the corpus. We evaluated the zero anaphora coreference resolution for different settings of mentions and clusters. For mentions, we used two settings: the gold standard zero mentions and mentions obtained using Minos (*Minos Mentions*). For clusters, we used the following settings: gold standard clusters (without zero-subject verbs), clusters obtained by *Ikar*, and clusters obtained by *Bartek*. To provide an insight into both the

Table 3: CoNLL: Results of coreference resolution for zero anaphora in different settings. Comparison performed for both settings of clusters: gold standard clusters without zero subjects and system result clusters for Ikar and Bartek without zero subjects. Similarly we performed comparison for gold standard zero mentions and zero mentions provided by Minos.

System	MUC			B³			CEAF$_E$			CoNLL
	P[%]	R[%]	F_1[%]	P[%]	R[%]	F_1[%]	P[%]	R[%]	F_1[%]	F_1[%]
Gold Standard Zero Mentions										
Baseline	100	76.44	86.65	100	88.84	94.09	87.76	99.06	93.07	**91.27**
Ikar	66.24	92.41	77.17	79.26	93.96	85.99	93.17	73.06	81.90	**81.69**
Crete	94.19	95.39	94.78	96.32	97.46	96.89	97.38	96.7	97.04	**96.24**
Minos Mentions										
Ikar	73.36	87.41	79.77	85.95	92.12	88.93	81.47	91.00	85.97	**84.89**
Crete	96.76	88.54	92.47	98.00	93.81	95.86	93.69	98.03	95.81	**94.71**

(a) Gold Standard Clusters

System	MUC			B³			CEAF$_E$			CoNLL
	P[%]	R[%]	F_1[%]	P[%]	R[%]	F_1[%]	P[%]	R[%]	F_1[%]	F_1[%]
Gold Standard Zero Mentions										
Baseline	79.64	52.43	63.23	94.30	78.56	85.71	76.18	90.40	82.68	**77.21**
Ikar	53.35	68.91	60.14	74.46	82.99	78.50	77.49	65.14	70.78	**69.81**
Crete	79.04	69.42	73.92	91.16	84.74	87.84	82.93	88.44	85.59	**82.45**
Minos Mentions										
Ikar	58.87	64.91	61.74	80.67	81.70	81.18	77.62	73.27	75.38	**72.77**
Crete	79.71	63.75	70.84	92.48	82.19	87.03	80.76	89.59	84.95	**80.94**

(b) System Result Clusters for Ikar

System	MUC			B³			CEAF$_E$			CoNLL
	P[%]	R[%]	F_1[%]	P[%]	R[%]	F_1[%]	P[%]	R[%]	F_1[%]	F_1[%]
Gold Standard Zero Mentions										
Baseline	75.19	44.06	55.56	94.74	74.79	83.59	74.16	89.90	81.28	**73.48**
Bartek	53.89	70.76	61.18	74.70	82.45	78.39	77.62	64.37	70.37	**69.98**
Crete	76.84	75.26	76.04	88.94	85.35	87.11	85.36	86.31	85.83	**82.99**
Minos Mentions										
Bartek	59.62	63.24	61.38	82.25	80.43	81.33	76.68	74.14	75.39	**72.70**
Crete	78.23	68.68	73.14	91.02	82.51	86.55	82.49	87.89	85.15	**81.61**

(c) System Result Clusters for Bartek

impact of this study on solely the zero anaphora resolution, and the potential impact on the end-to-end coreference resolution, we tested each combination of mentions and clusters settings, which produced 6 configurations.

5.3   Results

Table 3 presents the results of the evaluation using the CoNLL scheme. For gold standard clusters, we evaluated *Crete* (the system presented in this paper) alongside the naive zero anaphora baseline for Polish presented by Kaczmarek and Marcińczuk (2015a), denoted by *Ikar*. Additionally, we provided a baseline score for gold standard clusters without zero anaphora. The evaluation shows that by using *Crete*, we achieved a significant improvement in the overall result when working on gold standard clusters in both settings, compared to both *Ikar* and the baseline without co-referential relations for zero subjects. In the case of *Ikar*, however, we can observe that while the recall improves, precision decreases, causing the overall score to be lower than for the baseline. For the scores on system result clusters, we provide two baselines — one for *Ikar* and one for *Bartek*. Analogically to the baseline for gold standard clusters, their scores reflect system performance without zero anaphora coreference. The evaluation provides a comparison of *Crete*, working on both baseline results, with the results obtained from *Ikar* and *Bartek*, with zero anaphora coreference results included for gold standard and Minos zero mentions. For *Crete*, in both settings we can observe similar results. When compared to the baselines we can see up to an 8% improvement, and even up to 14% compared to *Bartek* and *Ikar* with zero anaphora coreference included. For *Bartek* and *Ikar*, we can again observe that with an increasing number of coreferential relations found (reflected in recall), we have decreasing precision, meaning that we also find many incorrect zero coreference relations. This is reflected in a final score lower than the corresponding baselines.

The results of the evaluation using the Parent metric are presented in Table 4. For gold clusters, we compared *Crete* with the naive zero baseline implemented in *Ikar*. In this setting we achieved a very significant improvement (at least 30%) over the baseline in both mention settings, which shows that we can extract much more accurate information from the text. Using system results clusters, we can observe that while for gold standard mentions we obtain even higher results than in the setting with gold standard clusters and Minos mentions, for the end-to-end setting with Minos mentions and system result clusters we obtain a much lower score, indicating that only $\sim 45\%$ of relations between zero subjects and named entities in the result are correct. Additionally, we can observe that working on clusters obtained using *Ikar* we get better results than on clusters obtained using *Bartek*. This may be caused by the fact that *Ikar* was developed using an entity mention model, while *Bartek* uses a mention-pair approach. For *Bartek* we obtained slightly better results than *Ikar*, and both these tools performed significantly worse than *Crete*.

5.4   Feature importance

We also evaluated the feature set used in our system, using relative feature importances for the random forest classifier. The plot of these importances (see Figure 1) shows that about 25% of the features are quite useful, while the rest give us very little value in terms of classifying mention-cluster pairs. Table 5 presents the features with the highest importance, together with the features with the lowest importance. The list of the most important features mainly contains features which address the frequency of occurrence of mentions from clusters, measured either in a number of words or in the number of sentences containing its mentions. As expected following the observations from Section 4.1, there is also a feature denoting that the closest mention preceding a zero subject is also a zero subject. Additionally, it appears that the most important agreement feature between a cluster and a zero mention is the agreement between grammatical genders. Features with the lowest importance contain mainly rare or unlikely phenomena, such as relative pronouns in the vocative case or the closest following mention being a zero subject. The latter is impossible due to the system design. The most unexpected feature on the list is that of the closest

Table 4: PARENT: Results of zero coreference resolution from information extraction perspective. This results provide scores related only to assignment of zero subjects to correct named entities in text, what reflects the utility of information provided by systems.

System	Gold Standard Zero Mentions			Minos Zero Mentions		
	P[%]	R[%]	F_1[%]	P[%]	R[%]	F_1[%]
Gold Standard Clusters (without zero verbs)						
Ikar	16.64	45.09	24.30	16.64	45.09	24.30
Crete	73.76	77.73	**75.69**	64.61	48.69	**55.53**
System Result Clusters (without zero verbs)						
Ikar	15.39	45.43	23.00	12.22	32.79	17.81
Bartek	23.00	34.57	27.26	18.03	23.08	20.25
Crete (Ikar clusters)	58.18	63.95	**60.93**	53.47	39.36	**45.44**
Crete (Bartek clusters)	51.88	62.90	56.86	42.10	32.74	36.83

preceding mention being the subject in its sentence. As subjects in Polish are often omitted in following sentences, thus inducing the occurrence of zero subjects, this seemed at first to be an important feature. There are two possible causes for this phenomenon. Firstly, other features, such as cases of the closest preceding mention, may be covering the same situations in a more clear or distinctive way. Secondly, the performance of the dependency parser may not be good enough to provide valuable input into the coreference resolution system.
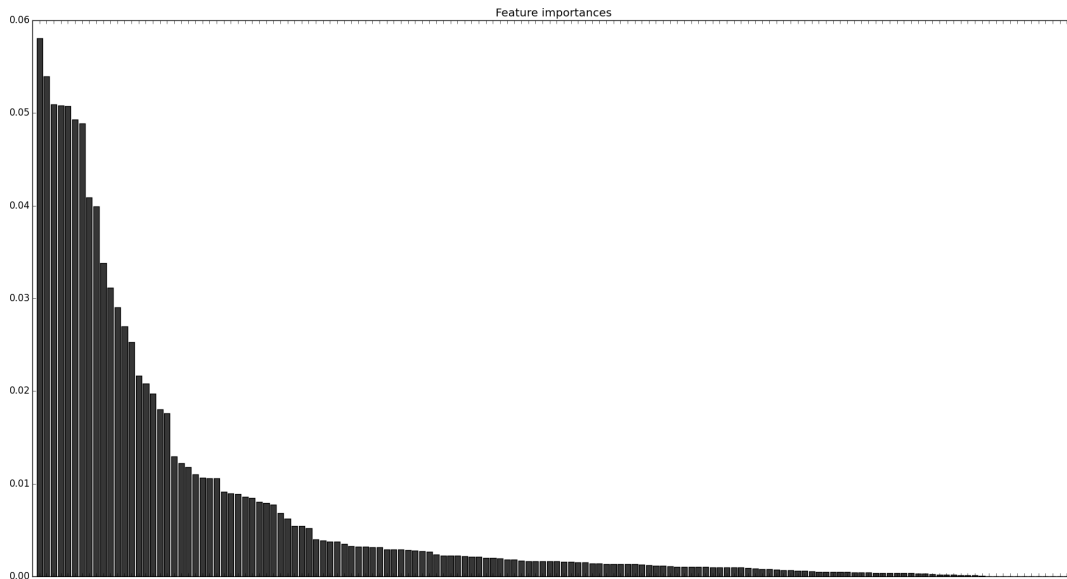

Figure 1: Feature importances chart for our random forest classifier.

Table 5: List of the most and the least important features in random forest classifier.

Most important features	Importance
Cluster term frequency	0.067586
Closest preceding is zero subject	0.058995
Mention-Cluster gender match	0.051871
Cluster sentence frequency	0.046494
Mention-Cluster preceding entity recency	0.046359
Closest preceding predicate distance	0.040065
Least important features	Importance
Closest preceding is subject	0.000025
Closest Following Person is second	0.000002
Mention preceding/following relative case is vocative	0.000000
Closest preceding gender is masculine	0.000000
Mention person is not defined	0.000000
Closest following is zero subject	0.000000

6   Conclusions

In this study, we have presented a preliminary machine learning approach to zero anaphora coreference in Polish. The presented system performs significantly better than the currently existing baseline systems, which do not address the problem of zero anaphora coreference directly. We have developed a dedicated set of features based on cluster-mention dependencies, and we have used a sequential approach which allows the utilization of certain grammatical properties of zero subjects. The results which we obtained from gold standard data show a lot of promise for zero anaphora coreference resolution. However, the evaluation of the end-to-end setting showed that there is still much room for improvement for both CRETE and the underlying pre-processing systems it depends on when processing zero anaphora coreference in more realistic settings. This discrepancy in results allows us to observe how crucial mention detection and coreference resolution of other types of mentions are for zero anaphora coreference. Furthermore, we have observed that, despite fairly positive results for the evaluation with CoNLL metrics that measure generic clustering quality, the value of information that can be extracted from the obtained results is still not satisfactory.

7   Future work

There are several aspects of end-to-end coreference resolution for zero anaphora in Polish which require improvement. One is the improvement of zero subject detection, which we plan to achieve by combining a machine learning approach with knowledge incorporated in the current state-of-the art heuristic approach (Kaczmarek & Marcińczuk, 2015b). Another aspect is the improvement of coreference resolution for other mentions, as we have shown that it has a significant impact on the final results. Finally, we plan to extend the machine learning approach used to resolve the coreferential relations of zero anaphoras, e.g. by combining the global cluster features using Recurrent Neural Networks, in a manner similar to that presented in Wiseman, Rush, and Shieber (2016).

References

Bagga, A., & Baldwin, B. (1998). Algorithms for scoring coreference chains. In *The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference* (pp. 563–566).

Broda, B., Burdka, L., & Maziarz, M. (2012). IKAR: An improved kit for anaphora resolution for Polish. In *Proceedings of COLING 2012: Demonstration papers* (pp. 25–32). COLING 2012, Mumbai, December 2012.

Broda, B., Marcińczuk, M., Maziarz, M., Radziszewski, A., & Wardyński, A. (2012). KPWr: Towards a free corpus of Polish. In N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey: European Language Resources Association (ELRA).

Chen, C., & Ng, V. (2015). Chinese zero pronoun resolution: A joint unsupervised discourse-aware model rivaling state-of-the-art resolvers. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Vol. 2, *Short Papers*, pp. 320–326). Association for Computational Linguistics. `https://doi.org/10.3115/v1/P15-2053`

Dunin-Keplicz, B. (1983). Towards better understanding of anaphora. In A. Zampolli & G. Ferrari (Eds.), *EACL 1983, 1st Conference of the European Chapter of the Association for Computational Linguistics, September 1-2, 1983, Pisa, Italy* (pp. 139–143). Association for Computer Linguistics. `https://doi.org/10.3115/980092.980116`

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18. `https://doi.org/10.1145/1656274.1656278`

Iida, R., Torisawa, K., Hashimoto, C., Oh, J.-H., & Kloetzer, J. (2015). Intra-sentential zero anaphora resolution using subject sharing recognition. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 2179–2189). Lisbon, Portugal: Association for Computational Linguistics. `https://doi.org/10.18653/v1/D15-1260`

Kaczmarek, A., & Marcińczuk, M. (2015a). Evaluation of coreference resolution tools for Polish from the information extraction perspective. In *Proceedings of the Workshop: The 5th Workshop on Balto-Slavic Natural Language Processing, BSNLP 2015*.

Kaczmarek, A., & Marcińczuk, M. (2015b). Heuristic algorithm for zero subject detection in Polish. In P. Král & V. Matoušek (Eds.), *Text, Speech, and Dialogue: 18th International Conference, TSD 2015, Pilsen, Czech Republic, September 14-17, 2015: Proceedings* (pp. 378–386). Cham: Springer International Publishing. (*Lecture Notes in Computer Science*, 9302). `https://doi.org/10.1007/978-3-319-24033-6_43`

Kopeć, M. (2014). Zero subject detection for Polish. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* (Vol. 2, *Short Papers*, pp. 221–225). Gothenburg, Sweden: Association for Computational Linguistics. `https://doi.org/10.3115/v1/E14-4043`

Kopeć, M., & Ogrodniczuk, M. (2012). Creating a coreference resolution system for Polish. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)* (pp. 192–195). Istanbul, Turkey: European Language Resources Association (ELRA).

Luo, X. (2005). On coreference resolution performance metrics. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05 (pp. 25–32). Stroudsburg, PA, USA: Association for Computational Linguistics. `https://doi.org/10.3115/1220575.1220579`

Mihăilă, C., Ilisei, I., & Inkpen, D. (2010). *Zero pronominal anaphora resolution for the Romanian language*.

Nivre, J., Hall, J., & Nilsson, J. (2006). Maltparser: A data-driven parser-generator for dependency parsing. In *Proceedings of LREC-2006* (pp. 2216–2219).

Ogrodniczuk, M., & Kopeć, M. (2011). Rule-based coreference resolution module for Polish. In *Proceedings of the 8ᵗʰ Discourse Anaphora and Anaphor Resolution Colloquium (DAARC 2011)* (pp. 191–200). Faro, Portugal.

Pradhan, S., Luo, X., Recasens, M., Hovy, E., Ng, V., & Strube, M. (2014). Scoring coreference partitions of predicted mentions: A reference implementation. In *Proceedings of the 52nd Annual Meeting of the*

Association for Computational Linguistics (Vol. 2, *Short Papers*, pp. 30–35). Baltimore, MD, USA: Association for Computational Linguistics. `https://doi.org/10.3115/v1/P14-2006`

Radziszewski, A. (2013). A tiered CRF tagger for Polish. In R. Bembenik, Ł. Skonieczny, H. Rybiński, M. Kryszkiewicz, & M. Niezgódka (Eds.), *Intelligent tools for building a scientific information platform: Advanced architectures and solutions* (pp. 215–230). Berlin: Springer. (*Studies in Computational Intelligence*, 467). `https://doi.org/10.1007/978-3-642-35647-6_16`

Rahman, A., & Ng, V. (2011). Narrowing the modeling gap: A cluster-ranking approach to coreference resolution. *Journal of Artificial Intelligence Research*, *40*(1), 469–521. Retrieved 5 May 2017, from `http://dl.acm.org/citation.cfm?id=2016945.2016958`

Recasens, M., Hovy, E., & Antònia Martí, M. (2011). Identity, non-identity, and near-identity: Addressing the complexity of coreference. *Lingua*, *121*(6), 1138–1152. `https://doi.org/10.1016/j.lingua.2011.02.004`

Rello, L., Baeza-Yates, R., & Mitkov, R. (2012). Elliphant: Improved automatic detection of zero subjects and impersonal constructions in Spanish. In *EACL 2012: The 13th European Chapter of the Association for Computational Linguistics*. Avignon, France: Association for Computational Linguistics.

Rello, L., Ferraro, G., & Gayo, I. (2012). A first approach to the automatic detection of zero subjects and impersonal constructions in Portuguese. *Procesamiento de Lenguaje Natural*, (49), 163–169.

Russo, L., Loáiciga, S., & Gulati, A. (2012). Improving machine translation of null subjects in Italian and Spanish. In *EACL '12: Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 81–89). Stroudsburg, PA, USA: Association for Computational Linguistics. `http://dl.acm.org/citation.cfm?id=2380943.2380953`

Stoyanov, V., & Eisner, J. (2012). Easy-first coreference resolution. In *Proceedings of COLING 2012: Technical papers* (pp. 2519–2534). COLING 2012, Mumbai, December 2012. `http://www.aclweb.org/anthology/C12-1154`

Vilain, M., Burger, J., Aberdeen, J., Connolly, D., & Hirschman, L. (1995). A model-theoretic coreference scoring scheme. In *Proceedings of the 6th Conference on Message Understanding*, MUC6 '95 (pp. 45–52). Stroudsburg, PA, USA: Association for Computational Linguistics. `https://doi.org/10.3115/1072399.1072405`

Wiseman, S., Rush, A. M., & Shieber, S. M. (2016). *Learning global features for coreference resolution.* Association for Computational Linguistics. `http://www.aclweb.org/anthology/N/N16/N16-1114.pdf`

Wróblewska, A. (2012). Polish dependency bank. *Linguistic Issues in Language Technology*, *7*(1). Retrieved from `http://journals.linguisticsociety.org/elanguage/lilt/article/view/2684.html`

Yin, Q., Zhang, W., Zhang, Y., & Liu, T. (2016). A deep neural network for Chinese zero pronoun resolution. Retrieved 5 May 2017, from `http://arxiv.org/abs/1604.05800`