

Evaluation of Uryupina's coreference resolution features for Polish

1

BARTŁOMIEJ NITOŃ
IPI PAN

The study was cofounded by the European Union from resources of the European Social Fund. Project PO KL “Information technologies: Research and their interdisciplinary applications”, Agreement UDA-POKL.04.01.01-00-051/10-00 and the Computer-based methods for coreference resolution in Polish texts project financed by the Polish National Science Centre (contract number 6505/B/T02/2011/40).

Uryupina's features

2

- Over 350 linguistic features which can be used to recognize coreference
- Language independency

Research goals:

- Verify language independency statement by checking the impact of a certain subset of features on coreference resolution for Polish
- Build coreference resolution tool for Polish language based on acquired results

Uryupina's features classification

3

- Surface similarity (implemented 88)
- Syntactic knowledge (implemented 9 core features)
- Semantic compatibility (omitted for the time being)
- Discourse structure and salience (implemented about 46, omitted in this presentation)
- Anaphoricity and antecedenthood (implemented 4 new features)

Surface features configurations (1)

4

- **all:** all 88 implemented surface features
- **baseline1:** exact match for full names only
- **baseline2:** baseline1 features and head exact matching
- **MED+head:** baseline1 and all MED based features
- **MED-head:** baseline1 features and MED based features without substring selection
- **MED_w-head:** baseline1 and MED measured in words features
- **MED_s-head:** baseline1 and MED measured in symbols features
- **MED_bare-head:** baseline1 and MED based features without length normalizations and substring selection

Surface features configurations (2)

5

- **MED_ante-head:** baseline1 and all MED features with normalization by antecedent length and without substring selection
- **MED_anaph-head:** baseline1 and all MED features with normalization by anaphor length and without substring selection
- **Last:** baseline1 and exact match for last word in mentions
- **First:** baseline1 and exact match for first word in mentions
- **Rarest:** baseline1 and rarest word-based features, each rarest feature is implemented for base forms of words and text forms
- **No_MED:** all implemented features without approximate match features

Surface features configurations (3)

6

- **No_abbrev:** all implemented features without abbrev1 and abbrev2-based features
- **No_rarest:** all features without rarest word-based ones
- **No_rarest_parser:** all features without the rarest word-based ones and features using parsing (i.e., all types of matching except for abbreviation and head matching algorithms)

Normalization functions

7

- *no_case* – ignore case in strings
- *no_punctuation* – strips off all punctuation marks and other auxiliary characters
- *no_determiners* – strips off determiners from text (omitted for the time being)

Syntactic knowledge features

8

- **Post-modification:** checks whether the head is not the last word in mention string
- **Number:** checks the grammatical number of the anaphor or the antecedent
- **Person:** checks the grammatical person of the anaphor or the antecedent
- **Same number:** checks if the anaphor and the antecedent share the same number
- **Same person:** checks if the anaphor and the antecedent share the same person
- **Syntactic agreement:** checks if the anaphor and the antecedent share the same number and person

Anaphoricity and antecedenthood features

9

- Surface
- Syntactic
- Semantic
- Saliency
- Karttunen-motivated features
- *Same_head* features:
 - *Same_head_exist*: checks if there is a mention with same head as given in the preceding text
 - *Same_head_distance*: measure distance between given markable and one with the same head in the preceding text

Experiments conditions

10

- Measure – following CONLL-2011 (Pradhan et al.) average score of MUC, B³ and CEAFE
- Coreference decisions tests were performed using J48, WEKA's implementation of the C4.5 decision tree learning algorithm and weka classifier
- Texts - 390 files sample from Polish Coreference Corpus (<http://zil.ipipan.waw.pl/PolishCoreferenceCorpus>)
- Scores were measured using 10 fold cross-validation
- Test environment – BART (<http://www.bart-coref.org/>)

Experiment 1: surface similarity

11

Configuration	no	no_case	no_punctuation	full	all
all	0.72	0.72	0.72	0.72	0.72
baseline1	0.69	0.70	0.69	0.70	0.70
baseline2	0.69	0.69	0.69	0.69	0.69
MED+head	0.70	0.70	0.70	0.70	0.70
MED-head	0.71	0.71	0.71	0.71	0.71
MED_w-head	0.69	0.70	0.69	0.70	0.69
MED_s-head	0.72	0.72	0.72	0.72	0.72
MED_bare-head	0.70	0.70	0.70	0.70	0.71
MED_ante-head	0.72	0.72	0.72	0.72	0.72
MED_anaph-head	0.72	0.72	0.71	0.72	0.72
last	0.69	0.70	0.69	0.70	0.70
first	0.69	0.70	0.69	0.70	0.70
rarest	0.72	0.72	0.72	0.72	0.72
no_MED	0.71	0.71	0.71	0.71	0.70
no_abbrev	0.72	0.72	0.72	0.71	0.71
no_rarest	0.70	0.70	0.70	0.70	0.70
no_rarest_parser	0.70	0.70	0.70	0.70	0.70

Experiment 1: conclusion

12

- Using normalizations for Polish coreference resolution can result in slight, but not very noticeable increase, best for *no_case* normalization
- Worst normalization is *no_punctuation*, even than without using normalization
- Best surface configurations: all, MED_s-head, MED_ante-head, MED_anaph-head, rarest and no_abbrev
- Best approaches: rarest words, MED algorithm (specially based on signs)
- Configurations using head words obtain slightly lower scores than those not using it

Experiment 2: adding complex features

13

Configuration (F-score)	CEAFM	CEAFE	MUC	B3	average
syntactic	0.71	0.77	0.00	0.83	0.53
all + synt	0.75	0.80	0.53+	0.84+	0.72
MED_s-head + synt	0.76-	0.80-	0.48-	0.84-	0.71-
MED_ante-head + synt	0.77	0.82	0.49	0.85-	0.72
MED_anaph-head + synt	0.77-	0.81-	0.49	0.85-	0.72
rarest + synt	0.77+	0.82	0.51+	0.85+	0.73+
no_abbrev + synt	0.74-	0.79-	0.52	0.83	0.72
same_head	0.71	0.77	0.00	0.83	0.53
all + same_head	0.61-	0.66-	0.45-	0.72-	0.61-
MED_s-head + same_head	0.71-	0.77-	0.44-	0.81-	0.67-
MED_ante-head + same_head	0.71-	0.76-	0.44-	0.81-	0.67-
MED_anaph-head + same_head	0.73-	0.78-	0.45-	0.82-	0.68-
rarest + same_head	0.76	0.82	0.50	0.84	0.72
no_abbrev + same_head	0.61-	0.66-	0.45-	0.72-	0.61-
synt + same_head	0.72	0.78	0.07	0.83	0.56
all + synt + same_head	0.57-	0.62-	0.45-	0.68-	0.58-
MED_s-head + synt + same_head	0.68-	0.74-	0.44-	0.78-	0.65-
MED_ante-head + synt + same_head	0.70-	0.76-	0.45-	0.80-	0.67-
MED_anaph-head + synt + same_head	0.69-	0.75-	0.44-	0.79-	0.66-
rarest + synt + same_head	0.74-	0.80-	0.49-	0.83-	0.71-
no_abbrev + synt + same_head	0.57-	0.61-	0.45-	0.68-	0.58-

Experiment 2: conclusion

14

- Best configuration: rarest configuration + syntactic features
- Rarest based features are very good predictors of coreference in the Polish language
- Syntactic features do not provide any advantage on other surface configurations than rarest based ones
- *Same_head* features affect coreference resolution in a very negative way
- Using only syntactic information and/or *same_head* features does not produce satisfying results
- Surface similarity features are indispensable in coreference resolution for Polish and no sufficient score is likely to be obtained with higher-level features only

Thank you...