

The Impact of Morphology Processing Quality on Automated Anaphora Resolution for Russian

Maxim Ionov
m.ionov@corp.mail.ru

Andrey Kutuzov
andrey.kutuzov@corp.mail.ru

Mail.ru Group,
HSE, MSU

Dialogue 2014
Bekasovo, 07.06.2014





Каждый фермер, у которого есть осел, бьет его.

Антецедент анафора Относительное местоимение Антецедент анафора Местоимение

Table of Contents

1 System overview

- Rule-based approach
- ML approach

2 Experiments and evaluation

- System evaluation
- Hybrid approach
- Influence of morphological processing quality

3 Conclusion

System overview

- Open source anaphora resolution system for Russian: **An@phora**
- Created specially for RU-EVAL

System overview

- Open source anaphora resolution system for Russian: **An@phora**
- Created specially for RU-EVAL ⇒
 - Pronominal anaphora only
 - Pronouns: personal and possessive pronouns, reflexives and relatives
 - No complex and ambiguous cases: ellipsis, verbal, clausal, split antecedents, etc.

System overview

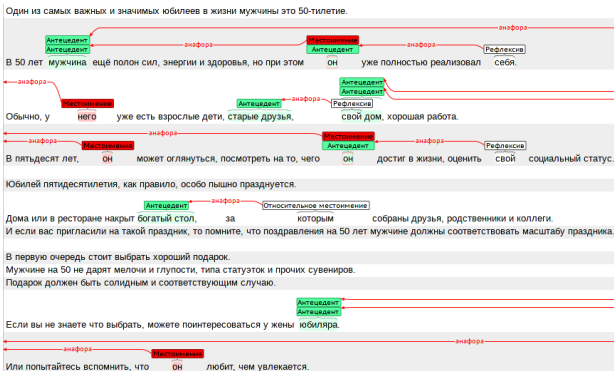
- Open source anaphora resolution system for Russian: **An@phora**
- Created specially for RU-EVAL ⇒
 - Pronominal anaphora only
 - Pronouns: personal and possessive pronouns, reflexives and relatives
 - No complex and ambiguous cases: ellipsis, verbal, clausal, split antecedents, etc.
- Freeling as an underlying morphological analyzer (however, it is possible to use any other)
- Available on <http://ling.go.mail.ru/anaphora>

Table of Contents

- 1 System overview
 - Rule-based approach
 - ML approach
- 2 Experiments and evaluation
 - System evaluation
 - Hybrid approach
 - Influence of morphological processing quality
- 3 Conclusion

Rule-based approach: overview

- Sliding analysis window of N words
- Searching to the left from anaphoric pronoun within the window



Rule-based approach

Each type of pronouns has different set of constraints:

- agreement in gender
- agreement in number
- comma between the antecedent and the anaphor
- ...

Table of Contents

- 1 System overview
 - Rule-based approach
 - ML approach
- 2 Experiments and evaluation
 - System evaluation
 - Hybrid approach
 - Influence of morphological processing quality
- 3 Conclusion

ML approach: overview

- Anaphora resolution is a binary classification task
- For a list of NP candidates classifiers decides if it is a true antecedent
- Random forest classifier trained on RU-EVAL training set¹
- Probability threshold: ignoring predictions with low probability

¹92 texts, 70k words

ML approach: overview

- Anaphora resolution is a binary classification task
- For a list of NP candidates classifiers decides if it is a true antecedent
- Random forest classifier trained on RU-EVAL training set¹
- Probability threshold: ignoring predictions with low probability

- Preprocessing:
 - POS tagging
 - NP chunking (O_{NP})

¹92 texts, 70k words

ML approach: features #1

- Length of the candidate group in characters
- Length of the candidate group in words

ML approach: features #1

- Length of the candidate group in characters
- Length of the candidate group in words
- Distance between pronoun and the candidate in characters
- Distance between pronoun and the candidate in words
- Distance between pronoun and the candidate in groups

ML approach: features #1

- Length of the candidate group in characters
- Length of the candidate group in words

- Distance between pronoun and the candidate in characters
- Distance between pronoun and the candidate in words
- Distance between pronoun and the candidate in groups

- Grammatical number of the candidate
- Grammatical number of the pronoun
- Do numbers of the candidate and the pronoun agree?

ML approach: features #2

- Grammatical case of the candidate
- Grammatical case of the pronoun
- Do cases of the candidate and the pronoun agree?

ML approach: features #2

- Grammatical case of the candidate
- Grammatical case of the pronoun
- Do cases of the candidate and the pronoun agree?

- Is the candidate a proper name?
- Number of the occurrences of candidate in the text
- Pronoun type
- Pronoun itself

ML approach: feature selection

- Features with feature importances ≥ 0.05 (most important first)
 - Distance in characters (*0.193*)
 - Distance in words (*0.121*)

ML approach: feature selection

- Features with feature importances ≥ 0.05 (most important first)
 - Distance in characters (*0.193*)
 - Distance in words (*0.121*)
 - Distance in groups
 - Length of candidate in characters
 - Pronoun
 - Number of occurrences of candidate in text
 - Case of the candidate
 - Type of the pronoun

Table of Contents

- 1 System overview
 - Rule-based approach
 - ML approach
- 2 Experiments and evaluation
 - System evaluation
 - Hybrid approach
 - Influence of morphological processing quality
- 3 Conclusion

Table of Contents

- 1 System overview
 - Rule-based approach
 - ML approach
- 2 Experiments and evaluation
 - System evaluation
 - Hybrid approach
 - Influence of morphological processing quality
- 3 Conclusion

System evaluation

- Rule-based system was evaluated on the whole training corpus
- ML system was evaluated on two subcorpora excluded from the training corpus

Table 1: Subcorpora for evaluation

	Test set, texts	Train set, texts
Subcorpus 1 (S1)	5	106
Subcorpus 2 (S2)	13	95

Baseline algorithm

Table 2: Baseline: choose nearest NP

	Precision	Recall	F-measure
Baseline	37.2%	35.7%	36.4%

RB and ML systems performance

Table 3: Comparison of ML and RB systems performance

	Precision	Recall	F-measure
RB, all	56.50%	51.6%	53.90%
RB, S1	45.02%	46.58%	45.79%
ML, S1	52.54%	43.51%	47.60%
RB, S2	63.57%	56.45%	59.80%
ML, S2	65.11%	45.67%	53.69%

Table of Contents

- 1 System overview
 - Rule-based approach
 - ML approach
- 2 Experiments and evaluation
 - System evaluation
 - Hybrid approach
 - Influence of morphological processing quality
- 3 Conclusion

Hybrid approach

- ML approach shows higher precision than rule-based
- RB shows higher recall

Hybrid approach

- ML approach shows higher precision than rule-based
- RB shows higher recall
 - ⇒ if ML system returns a result we are confident about it
- So we used ML first and RB as a back-off

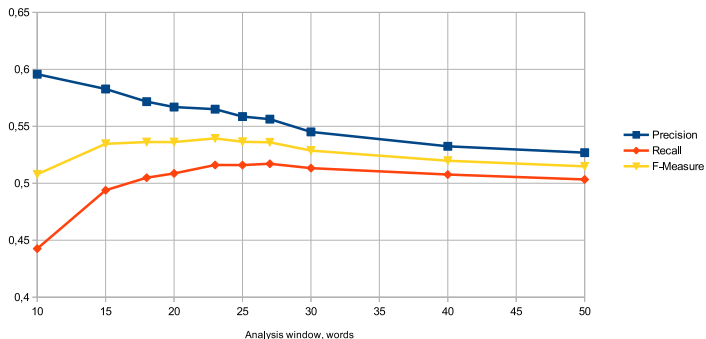
All systems performance

Table 4: Comparison of ML, RB and hybrid systems performance

	Precision	Recall	F-measure
RB, all	56.50%	51.60%	53.90%
RB, S1	45.02%	46.58%	45.79%
ML, S1	52.54%	43.51%	47.60%
HY, S1	49.49%	53.72%	51.52%
RB, S2	63.57%	56.45%	59.80%
ML, S2	65.11%	45.67%	53.69%
HY, S2	62.22%	56.46%	59.20%

Performance and analysis window length

Performance of anaphora resolution



Frequent errors

- Proper names

Frequent errors

- Proper names
- Animate and inanimate nouns

Frequent errors

- Proper names
- Animate and inanimate nouns

У спутницы Олдмэна на лице траурная вуаль, она несет свои глаза на подносе.

Антецедент анафора Местоимение анафора Рефлексив

Frequent errors

- Proper names
- Animate and inanimate nouns

У спутницы Олдмэна на лице траурная вуаль,
она несет свои глаза на подносе.

- Antecedent beyond the limits of analysis window

Frequent errors

- Proper names
- Animate and inanimate nouns

У спутницы Олдмэна на лице траурная вуаль,
она несет свои глаза на подносе.

- Antecedent beyond the limits of analysis window
- Cataphora

Frequent errors

- Proper names
- Animate and inanimate nouns

У спутницы Олдмэна на лице траурная вуаль,
она несет свои глаза на подносе.

- Antecedent beyond the limits of analysis window
- Cataphora
- Direct speech

Table of Contents

- 1 System overview
 - Rule-based approach
 - ML approach
- 2 Experiments and evaluation
 - System evaluation
 - Hybrid approach
 - Influence of morphological processing quality
- 3 Conclusion

Influence of morphology pre-processing

To measure the impact of quality of morphological processing on overall quality we fixed lemmatization errors manually:

Influence of morphology pre-processing

To measure the impact of quality of morphological processing on overall quality we fixed lemmatization errors manually:

- Added proper names to Freeling dictionary

Influence of morphology pre-processing

To measure the impact of quality of morphological processing on overall quality we fixed lemmatization errors manually:

- Added proper names to Freeling dictionary
- Fixed gender annotations

Influence of morphology pre-processing

To measure the impact of quality of morphological processing on overall quality we fixed lemmatization errors manually:

- Added proper names to Freeling dictionary
- Fixed gender annotations
- Tuned number and case probabilities

Influence of morphology pre-processing

To measure the impact of quality of morphological processing on overall quality we fixed lemmatization errors manually:

- Added proper names to Freeling dictionary
- Fixed gender annotations
- Tuned number and case probabilities
- etc. . .

Influence of morphology pre-processing

To measure the impact of quality of morphological processing on overall quality we fixed lemmatization errors manually:

- Added proper names to Freeling dictionary
- Fixed gender annotations
- Tuned number and case probabilities
- etc. . .

⇒ F-measure increased from 45.6% to 53.9%

Table of Contents

- 1 System overview
 - Rule-based approach
 - ML approach
- 2 Experiments and evaluation
 - System evaluation
 - Hybrid approach
 - Influence of morphological processing quality
- 3 Conclusion

To sum it up...

- We built **An@phora**, open source anaphora resolution system for Russian

To sum it up...

- We built **An@phora**, open source anaphora resolution system for Russian
- Quality of underlying morphological analysis greatly influences anaphora resolution quality

To sum it up...

- We built **An@phora**, open source anaphora resolution system for Russian
- Quality of underlying morphological analysis greatly influences anaphora resolution quality
- Our fixes will be contributed to Freeling upstream

Thank you!
Any questions?

Check our live demo and code repository:
<http://ling.go.mail.ru/anaphora>