

National Research University Higher School of Economics

*As a manuscript*

Ilya Azerkovich

**Evaluating the input of ontological information  
in coreference resolution for Russian language**

Dissertation Summary  
for the purpose of obtaining academic degree  
Doctor of Philosophy in Philology and Linguistics

Academic supervisor  
PhD, Svetlana Toldova

Moscow – 2021

## **Introduction**

Automated coreference resolution serves the task of extracting anaphoric and coreferent expressions in text and linking them together, depending on the entity they refer to. Solving this task is an important step for many fields of natural language processing, including machine translation, question answering systems, and information extraction.

Many researchers, starting from the middle of XX century and up to the present times, have addressed this task (e.g. the works of O. Boguslavskaja and I. Murav'jeva, D. Seleznev, A. Kibrik, T. Winograd, E. Charniak, J. Hobbs, R. Mitkov, M. Poesio, et al.), and at the time of writing this work some impressive results have been achieved in the field. Namely, several key features, allowing to establish the presence or absence of a coreference relation between expressions have been extracted, and they have been employed in a considerable number of automated systems of various architectures, allowing to process texts in a natural language.

Still, for the most part these achievements have been won only for a small number of languages: English, Arabic and Chinese. At the same time, for many other languages, including Russian, only a few of these achievements have been realised. This can be seen, for example, when comparing the variety of systems present for the competitions CoNLL-2012 [10] (English, Arabic and Chinese languages) and Ru-Eval-2014 [22], or mentioned in the overview article [25] (both for Russian language). The presence of such a disparity determines the actuality of this work. As the results of the Ru-Eval-2014 competition have shown, a major part of automated systems oriented at Russian language relies only on morphological and syntactical features to resolve coreference relations. At the same time, the results of a prominent amount of research (i.al. [13; 18–20]) leads to the conclusion that using semantic features can improve the quality of such systems. This raises the question, to what extent precisely does thesaurus information improve the quality of coreference resolution, and in which way it should be used. This work is dedicated to filling the gaps in corpus research and NLP research regarding the role of thesaurus information in building and resolution of coreference chains, and this defines the novelty of this work.

The goal of the research was to develop a system of automated coreference resolution that would use semantic information from open sources to improve the quality of analysis. To achieve the stated goal, the following intermediate tasks had to be completed:

1. To analyse the existing theoretical and practical approaches to the analysis of anaphora and coreference.
2. To study the most widely used algorithms and main systems of automated coreference resolution.
3. To study different methods of extracting and presenting semantic information from sources.

4. To get an understanding of representation of different coreference cases in texts, and most effective uses of different types of semantic information.
5. To research the use of measures of semantic similarity as a way to represent semantic information for automated analysis.
6. To create working prototypes of automated systems of coreference resolution and evaluate the effect of using semantic information on the quality of their performance.

The following statements are presented for the defence:

1. Ontological and semantic information are an important factor for successful realisation of coreference and resolution of coreferential relations.
2. Relations between phrases in a coreferential chain may be established, based upon thesaurus and encyclopaedic information. Such phrases are present in a considerable proportion and can influence the performance of coreference resolution systems.
3. Measures of semantic similarity can be useful for correct automated resolution of coreferential relations.
4. Employing measures of semantic similarity in automated systems of coreference resolution helps improve their performance.

The importance of semantic information for resolution of coreference relations once again establishes coreference and theory of reference in general as not simply morphological or syntactical, but also as discourse phenomena. In addition to this, the classification of coreferential relations, created as a result of this research, and the analysis of their distribution depending on the text genre are of interest from the point of view of pragmatics. This defines the theoretical importance of the work. The practical importance of the work is represented by the improvements in performance, obtained as the result of conducted experiments, and the possibility of further using these results for the task of coreference resolution.

When conducting the research, the method of quantitative analysis was used for studying the distribution of coreference relations in texts, qualitative analysis was used to evaluate the applicability of various measures of semantic similarity to usage in automated systems, as well as programming methods for the creation and evaluation of the performance of automated coreference resolution systems.

The intermediate results of the research were presented at several conferences, such as Dialog–2017, AINL 2018, AIST 2019, GWC 2019.

The contents of the work were published as following papers:

1. Azerkovich I. Employing wikipedia data for coreference resolution in Russian / I. Azerkovich // Artificial Intelligence and Natural Language. AINL 2017. Communications in Computer and Information Science. – Springer, Cham, 2018. – Vol. 789. – P. 107-112.
2. Azerkovich I. Using semantic relatedness measures in coreference

resolution for Russian / I. Azerkovich // NSU Vestnik. Series: Linguistics and Intercultural Communication. – 2019. – Vol. 17. – № 1. – P. 65-77.

3. Azerkovich I. Using Semantic Information for Coreference Resolution with Neural Networks in Russian / I. Azerkovich // Analysis of Images, Social Networks and Texts. AIST 2019. Communications in Computer and Information Science / eds. W.M.P. van der Aalst [et al.]. – Cham: Springer International Publishing, 2020. – Vol. 1086. – P. 85-93.

## **1. Developing the classification of extralinguistic information**

The first problem that needed to be solved when starting the research, was the theoretical problem of classifying extralinguistic information. Information that may be required to successfully resolve a coreferential relation may be of various nature and obtained from various sources. Therefore, it was necessary to determine, what types of information can be employed for creating a new nomination, and whether it is possible to concentrate on only a few of them.

To answer this question, the classification of coreferential relations was developed, based on the type of additional information present in the element of a coreferential chain compared to its head. This allowed to limit the practical tasks of the research only to studying the information types that were the most represented in the texts.

The classification was created based on the materials of the Russian coreference corpus RuCor, collected from the materials for the competition of automated coreference resolution systems, conducted for the forum RuEval-2014 [22]. The corpus includes 180 texts of different genres: apart from news texts it contains fragments of fiction, internet blogs and articles of Russian Wikipedia, which ensures the genre variety. In total the corpus texts contain 3638 coreferential chains and 16557 coreferential expressions. The final classification is based on differentiating between pragmatic and semantic information. Semantic information, in its turn, is divided into contextual and ontological, and ontological is further split according to types of ontological relations:

- I. Pragmatic information
  1. Metatextual information
  2. Situational information
- II. Semantic information
  1. Contextual information
  2. Ontological information
    - i. Hypo- and hypernymy
    - ii. Synonymy
    - iii. Metonymy
    - iv. Co-hyponymy
    - v. Near-identity and bridging relations

Next, to evaluate the degree to which every type of information was represented, the texts of the corpus were annotated, based on the classification described above. Of all coreferential chains in the corpus, those containing expressions with additional information (it was decided to refer to them as *heteronominative*, after Arutyunova [2]), constituted 18%. The amount of noun phrases, annotated in such way, constituted 10% of the total number of phrases in coreferential chains. It should also be noted, that of the chains consisting of two elements, in 30% of cases the second one was heteronominative.

Of all the types of information, presented in the classification, the most widespread ones were synonymy and hyponymy. This allowed to concentrate on these relations in the consequent research.

## **2. Experiments on integration of ontological information**

For the next step of the research, the following research questions were formulated:

- a) To what extent can the data of Russian sources be used for extraction of semantic information that would later be integrated in automated coreference resolution systems.
- b) In what way can the semantic data of the chosen sources should be represented for future use in such systems.

To answer these questions several experiments were conducted, detailed below.

### **2.1. Evaluation of Wikipedia as a source of ontological information**

The first preparational experiment had as its goal to evaluate the adequacy of the Russian segment of the Wikipedia, online encyclopaedia, from the point of ontological information, present there. The experiment was structured as follows: at first a corpus of news texts dedicated to the set list of named entities was collected, then ontological information from the corresponding articles of Russian Wikipedia was extracted, to later be used to improve the quality of coreference resolution. Next, the architecture of the automated coreference resolution system was developed, based on the pairwise coreference resolution algorithm, and the performance of two of its iterations was compared: with features derived from ontological information integrated in the system, and not integrated.

The set of features, used for analysis, was chosen relying on the seminal work [24], as well the article [8], dedicated to the problem of choosing optimal set of features for machine learning algorithms. In the system three classes of features were used: text-based, morphological, and distance-based. Text-based features consisted of comparison of heads and modifiers, distance-based – of number of tokens between pair members, and morphological – of checking for correspondence of main morphological features of pair members. In the second iteration of the model a binary feature, based upon co-

occurrence of pair members in the text of the same Wikipedia article, was added to the set.

As the result of the experiment, the variation of the system including the feature based on the information from Wikipedia articles demonstrated better quality of coreference resolution than the system without such features (detailed results are presented in Table 1. At the same time, the conclusion was reached that the representation of ontological information had to be better optimized. The experiments devoted to this are described in the next section.

Table 1: Evaluation of system performance

System variant	Precision (P)	Recall (R)	F-measure
Base feature set	0.763	0.739	0.727
Base feature set + Wikipedia	0.777	0.765	0.758

To improve the drawbacks of the system and extend the methods employed for the experiment to other classes of referents beyond named entities, the following steps were taken: a) to use quantifiable features instead of binary ones to represent ontological information; b) to use features that allowed to take into consideration several levels of semantic relations; c) to employ alternative information sources besides Wikipedia.

## 2.2. Evaluating the degree of correlation between semantic similarity and coreference

To realise the suggested steps the following stages of experiments were suggested and conducted. The results of experiments were expected to solve the following tasks:

1. Evaluating the possibility of using measures of semantic similarity for coreference resolution.
2. Building a system of automated coreference resolution, using these metrics.

For solving the first task several measures of semantic similarity were chosen, based on different calculation methods: the length of the ontology path, mutual information of the two notions and degrees of similarity between corresponding Wikipedia articles. The following measures were selected:

1. Based on the length of the ontology path:
  - a. Measure [11] (*rada*)
  - b. Measure [27] (*wp*)
  - c. Measure [15] (*lc*)
2. Based on mutual information:
  - a. Modification of measure [21], described in the article [23] (*res*)
3. Based on textual overlaps:
  - a. Modification of measure [16], described in [7] (*lesk*).

As an additional source of ontological information besides Wikipedia the thesaurus of Russian language RuThes, described in the book [17] was chosen.

Next for all the chosen measures the correlation coefficients between the similarity measure for a pair of mentions and a coreference relation between its members were calculated. The Jaccard measure was chosen as the baseline for comparison.

Table 2: Correlation coefficients of semantic similarity measures

	<i>jaccard</i>	<i>rada</i>	<i>wp</i>	<i>lc</i>	<i>res</i>	<i>lesk</i>
RuThes	0.34	0.56	0.51	0.59	0.30	n/a
Wikipedia	0.34	0.05	0.58	0.35	0.23	0.03
Wikipedia (named entities)	0.6	0.7	0.08	0.6	0.2	0.2

From the correlation data of different measures with coreference annotations presented above it can be seen that results calculated from the RuThes data were in general more representative than those based on Wikipedia data. At the same time, the latter proved representative for named entities. The measures, based on the path between entities, such as *rada*, or its normalized modifications, turned out the most representative of all evaluated measures. The measure of informational content showed to be the least representative for data from both sources.

### 2.3. Creating a prototype system based on statistical machine learning algorithms

As the next step, the quality of different coreference resolution algorithms using the selected measures was evaluated. The first of the systems considered was created based on the model of coreference resolution, described in [14] and [26], which was developed using the decision trees algorithms.

The architecture of the system was built upon a pairwise classification algorithm. For training the features, based on distance between the members of the pair and their morphological characteristics, were used, and syntactical and simple semantic features, such as checks for appositives and proper nouns, were also applied as well. The final feature set corresponds to the main 11 features, mentioned in [26]. In addition to those, the morphological and distance-based features were implemented. During the second stage of the experiment, the measures of semantic similarity for pairs of anaphor and potential antecedents were added to the set.

The evaluation of system performance was conducted with the use of MUC [1] and B<sup>3</sup> [6] metrics, evaluation results are presented in Table 3 below.

Table 3: Quality metrics for variations of system I

	MUC			B <sup>3</sup>		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Variant 1	72.76	59.49	65.46	71.01	44.50	54.71
Variant 2 (Wikipedia)	70.28	59.71	64.56	66.50	44.63	53.41
Variant 2 (RuThes)	72.72	59.43	65.41	71.15	44.44	54.71
Variant 2	73.57	60.01	<b>66.10</b>	71.77	44.93	<b>55.26</b>

These results show that, while using features derived from Wikipedia information helps to improve the recall, using features based on RuThes data helps to improve the precision of coreference resolution. Thus, combining features from the two sources allows to maximize the increase in performance of the system. In this way the usefulness of Wikipedia data despite its generally low correlation factor can be proven.

#### 2.4. Creating a prototype system based on a neural network

During the last stage of the experiments, a system employing a neural network-based model for coreference resolution was tested. Using neural networks has become increasingly popular for many tasks in natural language processing, with neural networks being used for machine translation, speech recognition and generation, etc. Attempts to use these methods for coreference resolution have also been made, e.g. in [9; 12].

The model used in the research was based on a feed-forward network that consisted of two consecutive parts: the mention encoder and the ranker model. The mention encoding implies transformation of the input pair of an anaphor and its potential antecedent into a matrix representation, which is transferred to the second module. The ranker model is an additional layer of the neural network that activates after the mention encoding step and accepts as input the matrix representations, generated by previous layers.

As the input vector features morphological and textual features of mentions and of the pair as a whole were used, as well as embeddings of the two mentions. Same as in the previous experiments, variations of the model including and not including semantic features were compared to each other.

The comparison of performance of the model variations is presented in Table 4 below. On the one hand, the performance in general is considerably better than during the previous experiment, but the same tendencies can be observed. Features derived from thesaurus information allow to increase the precision of the results, while lowering the recall, and features derived from Wikipedia data produce the opposite result. Joint usage of information from both sources achieves the highest metrics.



Table 4: Quality metrics for variations of system II

	MUC			B <sup>3</sup>		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Model 1	68.3	60.7	64.3	56.8	64.4	60.4
Model 2 (RuThes)	69.3	72.9	71.0	57.1	62.4	59.7
Model 2 (Wikipedia)	64.1	67.9	66.0	56.6	65.9	60.9
Model 2	69.3	73.0	<b>71.1</b>	56.8	68.2	<b>62.0</b>

Conduction of the series of experiments described above produced the following results:

- Metrics of semantic similarity correlate with presence of a coreference relation in the pair of mentions and, consequently, can be used as additional features for automated coreference resolution.
- Measures of semantic similarity, calculated based on data from different sources, can differ for noun phrases of different classes, in particular, common names and proper names.
- Using features, derived from semantic information, increases the quality of automated coreference resolution independent from the strategy of resolution employed or the exact algorithm realised within the system.

### 3. Conclusion

As the result of conducting the research presented here, the following results were obtained:

- A classification of coreferential relations was created, based on the character of information added to derivative nominations: pragmatical or semantical, and, in the case of semantical, contextual or ontological. On the one hand, it offers an insight into coreference resolution from the semantical and pragmatical point of view, and, on the other hand, offers theoretical grounds for using semantically oriented features for automated coreference resolution.
- Based on the developed classification, distributive analysis of coreference relations in a corpus of Russian was conducted on the texts of RuCor, and conclusions about connection existing between text genre and presence of certain coreference relations in it were drawn.
- A possibility of using semantic similarity measures between notions as features for coreference resolution was shown, as their values correlate to presence or absence of coreference relations between mentions.

- Several prototype systems of automated coreference resolution were developed, using features based on semantic information. The prototypes differed in representations of semantic information used, as well as system architectures and algorithms they were built around. Despite this, all researched systems demonstrated better performance than their analogues that did not employ semantic features.

As far as further research based on the results of this work is concerned, further development and improvement of prototypes of coreference resolution systems presented here should be considered. On the one hand, it is expected to be possible to optimize the semantic features used in the analysis – both by testing other semantic similarity measures and using embeddings-based methods. What is more, alternative feature sets and algorithms and paradigms of analysis should be taken into consideration, such as cluster-ranking models or graph-based methods. Seeing active developments and improvements these methods have brought into other areas of natural language processing, further exploring the potential of neural networks for coreference resolution is also of interest.

## References

1. A Model-Theoretic Coreference Scoring Scheme / M. Vilain [et al.] // Proceedings of the 6th Message Understanding Conference (MUC-6). – 1995. – P. 45-52.
2. Arutyunova N.D. Nominatsija i tekst / N.D. Arutyunova // Jazikovaja nominatsija. Vidi naimenovanij / eds. B.A. Serebrennikov, A.A. Ufimtseva. – 1977. – P. 304-355.
3. Azerkovich I. Employing wikipedia data for coreference resolution in Russian / I. Azerkovich // Artificial Intelligence and Natural Language. AINL 2017. Communications in Computer and Information Science. – Springer, Cham, 2018. – Vol. 789. – P. 107-112.
4. Azerkovich I. Ispol'zovanije mer semanticheskoj blizosti dl'a raspoznavanija koreferentsii v russkom jazike / I. Azerkovich // NSU Vestnik. Series: Linguistics and Intercultural Communication. – 2019. – Vol. 17. – № 1. – P. 65-77.
5. Azerkovich I. Using Semantic Information for Coreference Resolution with Neural Networks in Russian / I. Azerkovich // Analysis of Images, Social Networks and Texts. AIST 2019. Communications in Computer and Information Science / eds. W.M.P. van der Aalst [et al.]. – Cham: Springer International Publishing, 2020. – Vol. 1086. – P. 85-93.
6. Baldwin B. Algorithms for Scoring Coreference Chains / B. Baldwin, A. Bagga // The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference. – 1998. – P. 563-566.
7. Banerjee S. Extended Gloss Overlaps as a Measure of Semantic Relatedness / S. Banerjee, T. Pedersen // Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence. – 2003. – P. 805--810.

8. Bengtson E. Understanding the Value of Features for Coreference Resolution / E. Bengtson, D. Roth // Proceedings of the Conference on Empirical Methods in Natural Language Processing. – Honolulu, Hawaii: Association for Computational Linguistics, 2008. – P. 294-303.
9. Clark K. Improving Coreference Resolution by Learning Entity-Level Distributed Representations / K. Clark, C.D. Manning // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). – 2016. – P. 643-653.
10. CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes / S. Pradhan [et al.] // Joint Conference on EMNLP and CoNLL - Shared Task. – 2012. – P. 1-40.
11. Development and Application of a Metric on Semantic Nets / R. Rada [et al.] // IEEE Transactions on Systems, Man and Cybernetics. – 1989. – Vol. 19. – № 1. – P. 17-30.
12. Fernandes E.R. Latent Structure Perceptron with Feature Induction for Unrestricted Coreference Resolution / E.R. Fernandes, C.N. dos Santos, R.L. Milidiú // Proceedings of the Joint Conference on EMNLP and CoNLL: Shared Task. – 2012. – P. 41-48.
13. Harabagiu S.M. Text and knowledge mining for coreference resolution / S.M. Harabagiu, R.C. Bunescu, S.J. Maiorano // 2nd Meeting of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Conference. – The Association for Computational Linguistics, 2001. – P. 55-62.
14. Kutuzov A.B. The impact of morphology processing quality on automated anaphora resolution for Russian / A.B. Kutuzov, M. Ionov // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue" (2014). – 2014. – Vol. 13. – № 20. – P. 232-240.
15. Leacock C. Combining local context with wordnet similarity for word sense identification / C. Leacock, M. Chodorow // WordNet: an electronic lexical database. – MIT Press, 1998. – P. 265-283.
16. Lesk M. Automatic Sense Disambiguation: How to Tell a Pine Cone from an Ice Scream Cone / M. Lesk // Proceedings of the 1986 SIGDOC Conference. – 24-26: ACM, 1986. – P. 24-26.
17. Loukachevitch N. V. Tezaurusi v zadachah informatsionnogo poiska / N. V. Loukachevitch. – 2011. – 512 p.
18. Markert K. Comparing knowledge sources for nominal anaphora resolution / K. Markert, M. Nissim // Computational Linguistics. – 2005. – Vol. 31. – № 3. – P. 367-401.
19. Ponzetto S.P. Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution / S.P. Ponzetto, M. Strube // Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. – Association for Computational Linguistics, 2006. – Vol. 33. – P. 192-199.
20. Rahman A. Coreference resolution with world knowledge / A. Rahman, V.

Ng // Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. – The Association for Computer Linguistics, 2011. – P. 814-824.

21. Resnik P. Using Information Content to Evaluate Semantic Similarity in a Taxonomy / P. Resnik // IJCAI. – 1995. – P. 448-453.

22. RU-EVAL-2014: Evaluating anaphora and coreference resolution for Russian / S. Toldova [et al.] // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue” (2014). – 2014. – Vol. 13. – P. 681-694.

23. Seco N. An Intrinsic Information Content Metric for Semantic Similarity in WordNet / N. Seco, T. Veale, J. Hayes // ECAI. – IOS Press, 2004. – P. 1089-1090.

24. Soon W.M. A machine learning approach to coreference resolution of noun phrases / W.M. Soon, D.C.Y. Lim, H.T. Ng // Computational Linguistics. – 2001. – Vol. 27. – № 4. – P. 521-544.

25. Sysoev A.A. Coreference resolution in Russian: state-of-the-art approaches application and evolvment / A.A. Sysoev, I.A. Andrianov, A.Y. Khadzhiiskaia // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue” (2017). – 2017. – P. 341-352.

26. Toldova S. Coreference Resolution for Russian: The Impact of Semantic Features / S. Toldova, M. Ionov // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue” (2017). – 2017. – Vol. 16. – P. 339-348.

27. Wu Z. Verb Semantics and Lexical Selection / Z. Wu, M. Palmer // ACL. – 1994. – P. 133-138.