



NATIONAL RESEARCH UNIVERSITY  
HIGHER SCHOOL OF ECONOMICS

*Svetlana Toldova, Ilya Azerkovich, Yulia Grishina, Alina  
Ladygina, Olga Lyashevskaya, Anna Roytberg, Galina Sim,  
Maria Vasilieva*

**PRE-EXPERIMENTS ON  
ANNOTATION  
OF RUSSIAN COREFERENCE  
CORPUS**

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: LINGUISTICS  
WP BRP 35/LNG/2015

*Svetlana Toldova<sup>1</sup>, Ilya Azerkovich<sup>2</sup>, Yulia Grishina<sup>3</sup>, Alina Ladygina<sup>4</sup>, Olga Lyashevskaya<sup>5</sup>, Anna Roytberg<sup>6</sup>, Galina Sim<sup>7</sup>, Maria Vasileva<sup>8</sup>*

## **PRE-EXPERIMENTS ON ANNOTATION OF RUSSIAN COREFERENCE CORPUS<sup>9\*</sup>**

Building benchmark corpora in the domain of coreference

---

<sup>1</sup> National Research University Higher School of Economics, Faculty of Humanities, School of Linguistics; email: [toldova@yandex.ru](mailto:toldova@yandex.ru).

<sup>2</sup> National Research University Higher School of Economics, Faculty of Humanities, School of Linguistics; email: [iazerkovich@gmail.com](mailto:iazerkovich@gmail.com).

<sup>3</sup> University of Potsdam; e-mail: [jul\\_gr@mail.ru](mailto:jul_gr@mail.ru).

<sup>4</sup> University of Tübingen; e-mail: [aladygina@yahoo.com](mailto:aladygina@yahoo.com).

<sup>5</sup> National Research University Higher School of Economics, Faculty of Humanities, School of Linguistics. Email: [olesar@yandex.ru](mailto:olesar@yandex.ru).

<sup>6</sup> National Research University Higher School of Economics, Faculty of Humanities, School of Linguistics; [cvi@yandex.ru](mailto:cvi@yandex.ru).

<sup>7</sup> Lomonosov Moscow State University, Faculty of Philology; [sim.ge@yandex.ru](mailto:sim.ge@yandex.ru).

<sup>8</sup> Lomonosov Moscow State University, Faculty of Philology, e-mail: [linellea@yandex.ru](mailto:linellea@yandex.ru).

<sup>9\*</sup> This paper is based on our research supported by the Russian Foundation for Basic Research (grant No. 15-07-09306) and made within the framework of the Academic Fund Co-funding Research Program at the National Research University Higher School of Economics (HSE) in 2015 (grant No. 15-09-0288, “Development of a Gold Standard Anaphora and Coreference Corpus”).

and anaphora resolution is an important task for developing and evaluating NLP systems and models. Our study is aimed at assessing the feasibility of enhancing corpora with information about coreference relations. The annotation procedure includes identification of text segments that are subject to annotation (markables), marking their syntactic heads and identifying coreferential links. Markables are classified according to their morphological, syntactic and reference structure. The annotation is performed manually, providing gold standard data for high-level NLP tasks such as anaphora and coreference resolution. The paper reports on inconsistencies in selecting NPs of various types as markables and their borders, and in ways of constructing anaphoric pairs. We consider the types of NPs missed by annotators, and the discourse and semantic factors that may have affected the annotators' judgements.

JEL classification code: Z

Keywords: anaphora, coreference, coreference corpus, Russian language, corpus annotation, inter-annotator agreement.

# 1. Introduction

Building annotated corpora in the domain of coreference and anaphora resolution is an important task for developing and evaluating the corresponding NLP systems and models. As compared to most worldwide languages (English, Chinese, German, etc.), the Russian language still lacks an open-access annotated corpus for high-level NLP tasks such as information extraction or anaphora and coreference resolution. Our paper deals with the first ever experience of building an annotated coreference corpus for Russian (<http://rus-ant.maimbava.net>).

Nowadays, many research communities in Russia show increasing interest in the development of various NLP tasks. However, many of them are just small start-ups having very few resources. Each community tries to realize its own methodology. There is lack of information exchange among the teams. Moreover, there is lack of open resources that could serve as a Gold Standard and provide the general basis for the evaluation in the field.

In order to overcome this problem we are developing an open corpus of the Russian language annotated with

coreference links. The corpus was primarily designed for the independent evaluation of NLP systems for Russian. The first release of the corpus was completed for evaluating the results of the coreference and anaphora resolution shared task for Russian (Toldova 2014).

Since our goal is to create a corpus that can be used not only for research purposes but also for evaluating anaphora and coreference resolvers for Russian, we have to take into account the capacities of existing NLP systems and access annotation (dis)agreement between humans and machines in the highly subjective domain of discourse, semantic, and NP structure features. Moreover, we measure inter-annotator agreement between two manual annotations on broader material, in order to ensure reproducibility when selecting relevant markables and constructing coreference chains.

The paper is organized as follows. Section 2 outlines related work and motivation for the Russian coreference corpus project. Section 3 details the first experimental task run by both NLP systems and annotators on ten texts, with no guidelines provided. Section 4 reports on the inter-annotator agreement in the second experiment run on 88

texts after annotators have read and discussed the guidelines. We also investigate the typology of NPs most frequently missed by annotators. In Section 5 we discuss pragmatic and semantic factors underlying the annotators' choices.

## **2. Background**

The first annotation scheme for coreference was developed during the first evaluation campaign on automatic coreference resolution, held within the Message Understanding Conference (MUC-7) (Hirschman 1997). The main criteria for the task definition were good inter-annotator agreement, the simplicity and speed of text annotation and the creation of a corpus for independent coreference research. For these reasons, only noun phrases between which an identity relation could be identified were marked as referring expressions. This scheme did not cover more complicated cases like bridging or near-identity.

Both the GNOME corpus (Poesio 2004) for English and the VENEX coreference corpus for Italian (Poesio et al. 2004) include anaphoric annotations according to the MATE annotation scheme developed by Poesio (2004). This

scheme differs from the MUC scheme, the best-known at that time, in that it relies on the discourse model assumption, which implies that discourse entities as part of this model may or may not refer to specific objects of the real world. The main goal of creating this annotation scheme was to create a meta-scheme that could be used in different projects later on. The basic implementation of the scheme includes only the annotation of noun phrases with identity relations; the extended scheme also covers bridging relations and indirect coreference.

Another project in which the MATE scheme was applied is the ARRAU corpus. ARRAU is a corpus of English texts of different genres annotated for anaphoric relations. It includes annotations of not only specific, but also abstract entities, such as events, actions and plans. An additional goal of this annotation scheme was to introduce guidelines to mark ambiguous anaphoric expressions. The annotations included additional information about agreement and bridging (Poesio and Artstein 2008).

The Potsdam Coreference Scheme (Krasavina and Chiarcos 2007) consists of two sub-schemes: the Core scheme and the Extended scheme. The Core scheme was

designed to be basic and includes only nominal coreference, while the Extended scheme also includes bridging relations. This annotation scheme introduces two types of referring expressions, which are called primary and secondary markables. The main difference between them is that the former are always annotated, while the latter are annotated only if they serve as antecedents for some of the primary markables. For each markable, there is a number of attributes to be defined. These are, for example, referential status (discourse-new vs. discourse-old), phrase type, ambiguity, and grammatical role.

The Prague Dependency Treebank (Nedoluzhko et al. 2009) is a Czech corpus of newspaper articles that are annotated at different linguistic levels. Coreference annotations are stored at the textogrammatical level. In this corpus, two types of coreference are annotated: grammatical and textual coreference. Special attention in this corpus is paid to the annotation of bridging relations in this corpus.

The OntoNotes corpus includes not only general anaphoric annotations, but also event coreference. It considers both referring expressions expressed through noun phrases and events described through verbs. The coreference

annotation layer is connected to other linguistic layers, such as syntax, propositional structure and word senses. In the corpus, two different types of coreference relations are distinguished: identical coreference and appositive coreference (Pradhan et al. 2007). The first type marks the coreference between specific referents (pronominal, nominal phrases and named entities), excluding generic, underspecified, or abstract entities. Appositives are tagged separately from identity coreference because they express the attributed property of an entity.

In our work, we relied on the recent experience of corpus creation and annotation, trying to rethink the existing methodology and apply it to Russian. Our corpus was designed taking into consideration the existing large-scale linguistic corpora for other European languages that contain coreference annotations: OntoNotes (Hovy et al. 2006), (Pradhan et al. 2007), ARRAU Corpus (Poesio and Artstein 2008) for English, Potsdam Commentary Corpus (Stede 2004) for German, Prague Dependency Treebank (Bejček et al. 2011) for Czech, and VENEX Corpus (Poesio et al. 2004) for Italian, among others. In particular, we studied the annotation principles and annotation schemes used for

coreference corpora presented in Nedoluzhko et al. (2009), Nedoluzhko and Mírovský (2013), and Khudyakova et al. (2011). However, these corpora are based on in-depth annotation schemes: the first one requires syntactic analysis and deals with implicit anaphoric relations such as bridging. The detailed multifactor annotation scheme is implemented in the latter corpus.

One of our main goals was to generate a dataset for the RU-EVAL 2014 Evaluation campaign that could serve as a gold standard to assess the quality of automatic coreference resolution of Russian NLP systems (Toldova et al. 2014). For this reason, a number of decisions and annotation principles were influenced by the evaluation task and similar evaluation campaigns – including the selection of noun phrases amenable to annotation and their attributes, and the mark-up of referring expressions of maximal size (based on the annotation scheme of Krasavina and Chiarcos (2007)). Following the principles of MUC (Hirschman 1997), we determined the types of referring expressions and relations to be annotated, excluding other types such as indirect and event anaphora, bridging relations, and so on. However, following Recasens et al. (2010), we made an

attempt to annotate the so-called “near-identity” cases, where coreferential objects cannot be viewed as completely identical.

In contrast to SemEval 2010 (Recasens et al. 2010), we did not take into account expressions that occurred only once or morphological and syntactical tags, despite the fact that we included them in the output to make manual evaluation easier.

### **3. Experiment one: comparing automatic and manual annotation, no guidelines provided**

Before elaborating annotation principles, we conducted an experiment in order to compare automatic and manual annotation. We gave ten texts to Russian students of linguistics (native speakers of Russian) and to two experts in coreference resolution. Their task was to annotate the referring expressions that occurred more than once and link them into coreference chains. The types of expressions or relations to be annotated were not restricted. Then we tested four coreference resolvers on the same texts and analyzed the discrepancies.

In general, proper names and personal and possessive pronouns related to specific NPs were considered as referring expressions by both systems and annotators. Some types of coreference relations like near-identity (cf. bridging relations between *zapisnaya knizhka* ‘the notebook’ and *oblozhka* ‘the cover of the notebook’) and predicative (event) anaphora (e. g. *On prishel rano. Eto horosho.* ‘He came early. It is fine’) were consistently ignored. Neither systems nor experts annotated zero pronouns as referring expressions.

Most of the discrepancies were associated with the length of referring expressions. We did not give any instructions about the length of referring expressions so there were two ways to annotate a nominal phrase containing different modifiers (adjectives, participial clauses, possessive or demonstrative pronouns, among others): to include all of the modifiers into a referring expression (the first strategy) or consider only such information that could restrict the set of possible referents (the second strategy). These strategies are shown in (1) below, where the first strategy is represented by the whole group (marked by braces) while the second one is expressed

by a group where only the relative adjective has been added into a markable:

- (1) {*Znamenityj [rossijskij parusnik “Sedov”]*<sub>2</sub>,  
*sovershajuschij krugosvetnoe plavanie*}<sub>1</sub>  
{‘The famous [Russian sailer “Sedov”]<sub>2</sub> going round the world’}<sub>1</sub>}

It was found that most of the native speakers tended to include all of the modifiers of an entity’s head. Furthermore, genitive groups were also considered as a part of referring expressions. On the contrary, some native speakers and systems followed the second strategy.

Other problems were related to distinguishing between generic and specific references. Both annotators and systems marked correctly specific referring expressions, but generic or abstract referents caused a number of discrepancies.

- (2) *Zhiznedejatel’nost’ bobrov mozhet prepyatstvovat’ dorozhnomu dvizheniju. K primeru, v 2011 v Borovske bobry postroili iz gliny i vetok plotinu, iz-za kotoroj na doroge obrazovalos’ iskusstvennoje ozero.*

‘The vital functions of beavers can block road traffic. For example, in 2011, some beavers built in Borovsk a dam of clay and branches, which caused an artificial lake’.

In (2), there are two mentions of beavers, but they have different referents. The first one has a generic referential status and denotes all the beavers in the world, while the second designates some beavers living in Borovsk, which is a subset of the first entity. In some cases, it is not obvious whether the two mentions refer to the same entity or not, and it is a particular problem for languages without articles like Russian.

Abbreviations and their interpretation were annotated as either one or two referring expressions (Trudovoy kodeks Rossiyskoy Federatsii, TK RF ‘Labor Code of the Russian Federation’, ‘LC RF’). Substantivized adjectives (roditeli ego lyubeznoy ‘parents of his beloved [one]’) and NPs with a numeral (47 roninov ‘47 ronins’) were also marked up inconsistently. In cases like ‘The club was called Newton’, ‘the team called X’, the proper name was sometimes not annotated.

The results of the first experiment influenced a number of decisions regarding the annotation principles and guidelines used in the second experiment.

#### **4. Experiment two: comparing manual annotations, under guidelines**

To evaluate inter-annotator consistency, we picked up 88 texts (news, essays, scientific prose, fiction, and wiki-texts, ca. 90,000 tokens). The principles and guidelines for annotation are detailed in Toldova et al. (forthcoming). In brief, the taxonomy used in the guidelines is as follows.

**A. NP reference type.** The specific NPs and pronouns are subject to annotation; the non-specific (abstract and generic) NPs are annotated only if they serve as antecedents for anaphoric pronouns.

**B. Primary vs secondary markables** (Krasavina and Chiarcos 2007). The primary markables (nominal phrases; proper names and titles; 3<sup>rd</sup> person pronouns; 1st person pronouns; demonstrative pronouns; reflexive pronouns; possessive pronouns; relative pronouns) are always annotated and can form coreference chains (marked as a coreference relation). The secondary markables (indefinite NPs; appositions, as in *Pavel Stepanovich Nakhimov*,

*vydajuscshijsya rossijskij admiral* ‘Pavel Stepanovich Nakhimov, the great Russian Admiral’); predicative NPs, cf. *Kometa nazyvalas’ Gallej* ‘The comet was called Gallej’) are only annotated if they are antecedents of an anaphoric pronoun (marked as an anaphoric relation).

**C. NPs of maximal vs of minimal size** (Krasavina and Chiarcos 2007). NPs of maximal size are NPs with all their dependencies (excluding prepositions, relative clauses and postpositional participial constructions). NPs of minimal size include only their syntactic heads. Both NPs of maximal size and NPs of minimal are annotated.

**D. Exceptions to annotation:** 1<sup>st</sup> person pronouns (if they do not refer to the narrator of the text); 2<sup>nd</sup> person pronouns; split antecedents (*Ona skazala emu, chto zhaleet ob ih nedorazumeniyah* ‘She said to him that she regrets their misunderstandings’); discontinuous expressions (*Lap u koshki chetyre* ‘Legs has a cat four’); zero pronouns (*Vasya prishel i Ø zasnul* ‘Vasya came and Ø fell asleep’).

**E. Annotation attributes.**

1.type of anaphora (reference to NPs, predicates, appositives, direct speech, others)

2. type of chain (anaphoric or coreference)

3. head of NP (noun; personal, possessive, reflexive, demonstrative, relative pronouns)

Firstly, we measured the percentage of NPs marked by both annotators as having the same length: 95 per cent. Secondly, we evaluated the consistency of annotators when selecting NPs relevant for our corpus (see Table 1). Both annotators marked 5160 NPs out of a total of 7009.

We measured the observed agreement as  $A_o=A/N$  (where A corresponds to the number of markables marked by both annotators and N corresponds to the overall number of all markables selected by both annotators) and the expected agreement using Cohen's Kappa coefficient (Cohen 1960):

$$A_e^k=(A_y*A_n^*+B_y*B_n)/N^2$$

where N corresponds to the overall number of markables selected by both annotators;  $A_y$  and  $B_y$  correspond to the number of markables marked by the first and second annotator respectively while  $A_n$  and  $B_n$  correspond to the

number of markables not selected by the first and second annotator respectively).

In our case,  $A_o = 0.74$  and  $A_e^k = 0.22$ , so the inter-annotator agreement viewed as  $(A_o - A_e^k) / (1 - A_e^k) = 0.66$ . We also used Mitkov's metric  $\mu = 2 * A / (A_1 + A_2)$  (Mitkov et al. 2000), where  $A$  is the number of markables marked only by both annotators, and  $A_1$  and  $A_2$  correspond to the number of markables marked by the first and second annotator, respectively, and got  $\mu = 0.85$ .

**Table 1. Inter-annotator consistency when selecting relevant markables.**

<b>Markables</b>	<b>Marked by the second annotator</b>	<b>Not marked by the second annotator</b>	<b>Total</b>
Marked by the first annotator	5160	1430	6590
Not marked by the first annotator	419	0	419
Total	5579	1430	7009

Next, we assessed the way our annotators constructed anaphoric pairs. We measured both the observed agreement  $A_o=0.69$  and Mitkov's metric  $\mu=0.79$ .

As for consistency when constructing coreference chains,  $A_o=0.64$  and  $\mu=0.69$ . Since it was not obligatory to mark certain types of NPs according to our instructions to annotators (such as appositions, predicatives, pronouns in direct speech and some other dubious cases), this may have considerably affected our consistency. Thus, we evaluated our inter-annotator agreement, excluding all such markables from analysis, and obtained  $A_o=0.65$  and  $\mu=0.69$ .

Finally, we considered the type of NPs missed by one annotator when working with the corpus (see Table 2).

**Table 2. Types of missed NPs.**

Type of group	%
Reflexive pronouns	3.76
Relative pronouns	6.20
Anaphoric pronouns	12.47
Possessive pronouns	6.48
Noun groups	71.08
Total	2.35

As most were full NPs, we decided to analyze them more closely. It turned out that the head nouns of 38.04 per cent of such NPs were proper names (for a closer analysis see Table 3).

**Table 3. Types of missed proper names.**

Type	%
Locations	35,84
Persons	27,27
Organizations	15,84
Other	21,04
Total	100

## **5. Discussion**

The annotators' mistakes and inconsistencies can be considered as evidence of interesting phenomena underlying the process of coreference resolution. Specific features of referential ambiguity in languages without articles had an influence on the inter-annotator agreement. In particular, the following factors may affect annotators' performance: (a) the length of referring expression; (b) the distinction between the new referent vs. previously mentioned.; (c) syntactic position of referring expression; (d) discourse

factors, (e) semantic shifts. Below we discuss the latter two factors in detail.

## 5.1. Discourse factors

According to Table 3, proper names were missed, even if they designated specific referents. Two mentions of a proper name tended not to be marked as co-referring if they were separated by a considerable amount of text. Inanimate objects like companies, planets and locations were also usually missed in the annotation.

This phenomenon may have a cognitive background because these referents do not play a crucial role in the discourse structure. If the mentions of a proper name occur after a considerable amount of text or the entity is located at the right end of the Silverstein animacy hierarchy (Silverstein 1976) then they have a low activation cost and are not salient in the discourse. This is also the case with other inanimate objects expressed by common nouns even if they have definite referential status, like *starom derevenskon dome* ('the old house') and *dom* ('the house') in example (3).

(3) *Nepriyatnosti nachalis', kogda v starom derevenskom dome poyavilas' taksa Funtik. ... Inogda po vecheram lyagushka prikhodila v dom.*

'The troubles started when the taksa Funtik appeared in the old country house... Sometimes in the evenings a frog came to the house.'

## **5.2. Semantic shifts**

While annotating the corpus, we encountered several difficult cases of establishing identity relation. Most of them were qualified as examples of near-identity.

First, one of the problems we found was related to the entities that changed their denotation within a text. An obvious example of this phenomenon is sport teams whose membership changes every year. For example, the German national football team 1974 and the German national football team 2014 could be resolved as two different entities or as mentions of one abstract entity - the German national football team. A similar case is named entities whose name or status have been altered. For instance, in our corpus there are articles about the history of some countries, and over time their names have been changed (Samoa -

German Samoa - Western Samoa). Such entities required additional decisions. As they differ from pragmatical and extra-linguistic information, they were treated as independent referents by supervisors.

There was a difficulty in establishing the identity relation in particular object classes, such as mental products (books, works of art, albums, inventions, etc.). In some cases, their reference scope is not clear since they can represent either a mental product itself or its releases. Thus, some annotators linked all of the mentions in one chain, while others distinguished between mentions of an abstract entity and its material representation, cf. (4).

(4) [*Ipad*]<sup>1</sup> *sovershil perevorot na rynke planshetov. Uroven' prodazh [ipad-ov]<sup>2</sup> uvelichivaetsya s kazhdym godom*

‘The Ipad has set the world's tablet market alight. The sales of Ipads increase each year.’

The mentions of Ipad are included in different coreference chains: the first one is resolved as the name of the invention, i.e. a mental product, while the second one is considered as a release of this invention.

In addition, standard metonymy caused lots of discrepancies, as can be observed in example (4). The most common type of this relation is shown in example (5), where the object *brilliantovoj broshju* ('a diamond brooch') is described using the material that is made from *eti samye brilliandy* ('these diamonds'). Similarly, in example (6), the capital's name *Vashington* ('Washington') is used to represent the government of the country *oficialnyj Vashington* ('official Washington').

(5) *Tut k Svete podoshla vysokaya dama c brilliantovoj broshju na grudi. U Svety sp'orlo dykhanie i ona protyanula ladon', chtoby pogladit' eti samye brilliandy.*

'Then a tall madame on whose breast there was a diamond brooch came to Sveta. It took Sveta's breath away and she held out her hand to caress these diamonds.'

(6) *Oficialnyj Vashington obratilsya k pravitel'stvu Rossii s zayavleniem...*

'Washington made a formal statement to the Russian government...'

The above-mentioned cases caused discrepancies in the annotation and highlight open questions that need to be

addressed in the future. These cases should be taken into consideration in the annotation guidelines and as well as incorporated into the analysis of near-identity.

### References

- Bejček, E., J. Hajič, J. Panevová, J. Mírovský, J. Spoustová, J. Štěpánek, P. Straňák et al. 2011. "Prague Dependency Treebank 2.5."
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20 (1): 37–46.
- Hirschmann L. 1997. MUC-7 coreference task definition. Version 3.0, Proceedings of the 7th Message Understanding Conference (1997).
- Hovy, E., M. Marcus, M. Palmer, L. Ramshaw and R. Weischedel. 2006. "OntoNotes: the 90% solution." In *Proceedings of the human language technology conference of the NAACL, Companion Volume: Short Papers*, 57-60. Association for Computational Linguistics.
- Khudyakova, M. V., G. B. Dobrov, A. A. Kibrik and N. V. Loukachevitch. 2011. "Computational modeling of referential choice: Major and minor referential options." In *Proceedings of the CogSci 2011 Workshop on the Production of Referring Expressions*. Boston (July 2011).
- Krasavina, O. and C. Chiarcos. 2007. "PoCoS: Potsdam coreference scheme." *Proceedings of the Linguistic Annotation Workshop*. Association for Computational Linguistics.
- Mitkov, R., R. Evans, C. Orasan, C. Barbu, L. Jones and V. Sotirova. 2000. "Coreference and anaphora: developing annotating tools, annotated resources and annotation strategies." *Proceedings of the Discourse Anaphora and*

- Anaphora Resolution Colloquium (DAARC'2000)*.  
Lancaster, UK: 49–58.
- Nedoluzhko, A. and J. Mírovský. 2013. “How dependency trees and tectogramatics help annotating coreference and bridging relations in Prague Dependency Treebank.” *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*. Prague, 244-251.
- Nedoluzhko, A., J. Mírovský, R. Ocelák and J. Pergler. 2009. “Extended coreferential relations and bridging anaphora in the prague dependency treebank.” *Proceedings of the 7th Discourse Anaphora and Anaphora Resolution Colloquium (DAARC 2009)*, Goa, India, 1-16.
- Poesio, M. and R. Artstein. 2008. “Anaphoric Annotation in the ARRAU Corpus.” *LREC*.
- Poesio, M., R. Delmonte, A. Bristot, L. Chiran and S. Tonelli. 2004. The VENEX corpus of anaphora and deixis in spoken and written Italian. *University of Essex*.
- Poesio, M. 2004. “Discourse annotation and semantic annotation in the GNOME corpus.” *Proceedings of the 2004 ACL Workshop on Discourse Annotation*. Association for Computational Linguistics.
- Pradhan, S. S., L. Ramshaw, R. Weischedel, J. MacBride and L. Micciulla. 2007. “Unrestricted coreference: Identifying entities and events in OntoNotes.” In *International Conference on Semantic Computing, 2007. ICSC 2007*. IEEE, 446-453.
- Recasens, M., L. Màrquez, E. Sapena, M. A. Martí, M. Taulé, V. Hoste, , and Y. Versley. 2010. Semeval-2010 task 1: Coreference resolution in multiple languages. In *Proceedings of the 5th International Workshop on Semantic Evaluation* (pp. 1-8). Association for Computational Linguistics.
- Silverstein, M. 1976. *Hierarchy of features and ergativity*.

- In *Grammatical categories in Australian languages*, ed. by Robert M. W. Dixon (Hrsg.), 112-171. Canberra: Australian Institute of Aboriginal Studies.
- Stede, M. 2004. "The Potsdam commentary corpus." *Proceedings of the 2004 ACL Workshop on Discourse Annotation*. Association for Computational Linguistics.
- Toldova S., Yu. Grishina, A. Ladygina, M. Vasilyeva, G. Sim, I. Azerkovich (forthcoming). Russian coreference corpus. In *Yearbook of Corpus Linguistics and Pragmatics*, Special Issue. Springer.
- Toldova S., A. Roytberg, A. Ladygina, M. Vasilyeva, I. Azerkovich, M. Kurzukov, G. Sim, D. Gorshkov, A. Ivanova, A. Nedoluzhko and Y. Grishina. 2014. "RU-EVAL-2014: Evaluating anaphora and coreference resolution for Russian."

**Svetlana Toldova**

National Research University Higher School of Economics,  
(Moscow, Russia). Faculty of Humanities, School of  
Linguistics;

email: [toldova@yandex.ru](mailto:toldova@yandex.ru)

**Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.**

© Toldova, Azerkovich, Grishina, Ladygina, Lyashevskaya, Roytberg, Sim, Vasilieva, 2015