Deception Detection for the Russian Language:

Lexical and Syntactic Parameters

Dina Pisarevskaya [1], Tatiana Litvinova [2] and Olga Litvinova [2]
1 Institute for Oriental Studies of the RAS, Moscow, Russia
2 RusProfiling Lab, Voronezh State Pedagogical University
dinabpr@gmail.com, centr_rus_yaz@mail.ru

Abstract

The field of automated deception detection in written texts is methodologically challenging. Different linguistic levels (lexics, syntax and semantics) are basically used for different types of English texts to reveal if they are truthful or deceptive. Such parameters as POS tags and POS tags n-grams, punctuation marks, sentiment polarity of words, psycholinguistic features, fragments of syntactic structures are taken into consideration. The importance of different types of parameters was not compared for the Russian language before and should be investigated before moving to complex models and higher levels of linguistic processing. On the example of the Russian Deception Bank Corpus we estimate the impact of three groups of features (POS features including bigrams, sentiment and psycholinguistic features, syntax and readability features) on the successful deception detection and find out that POS features can be used for binary text classification, but the results should be double-checked and, if possible, improved.

1 Introduction

In the contemporary world, we get information from diverse sources: news stories and social media posts, product reviews and online communication. Automated deception detection is the field of natural language processing which studies what methods can be used to separate truth from deception, in written texts and in transcripts of oral communication, by identifying verbal predictors of deception. The influence of Internet communications is growing, therefore identification of deceptive information in short written texts is of vital importance.

Deception can be broadly defined as "the intentional misrepresentation of information", and the forms can be different, from direct lie to exaggeration (Hancock et al., 2007a). It can be also understood as "deliberate attempt to mislead others" (DePaulo et al., 2003). In deceptive texts, the content about the given event's factuality is either hidden from the reader or is "uttered in a way that camouflages the factual side of a given proposition" (Heidari et al., 2017). Human ability to detect deception is slightly above 50% (DePaulo et al., 1997). It highlights the importance of automated deception detection methods.

Most papers dealing with this topic were performed using English texts, but, due to the development of data sets for other languages, automated deception detection was recently taken into consideration for other languages too. The evaluation of truthfulness of the narrative has been commonly understood as a text classification task employing the methods of machine learning, data mining and information retrieval (Zhou et al., 2004; Hirschberg et al., 2005; Fuller et al., 2008; Mihalcea and Strapparava, 2009; Zhang et al., 2009). Techniques for detecting deception in text usually use different levels of language analysis: lexics, syntax, semantics, discourse and pragmatics levels. Studies often focus on lexics and semantics and may add some syntactic, discourse and pragmatics principles. As the field of deception detection is methodologically challenging for different sciences, psycholinguistic markers and psycho-social dictionaries are also used in natural language processing methods and models.

The question of deception detection is important for information retrieval, as it could help search engines to give an indication to the user of how reliable or truthful each retrieved document

Proceedings of Natural Language Processing and Information Retrieval Workshop, pages 1–10,
Varna, Bulgaria, Sep 7 2017.
https://doi.org/10.26615/978-954-452-038-0_001

is. Automated deception detection could be included in search engines page ranking algorithms.

2 Related Work

Problem of automated deception detection in oral communication has been the subject of research interest for a long time. This interdisciplinary field is an interesting issue for computational linguistics, speech processing, psychology (polygraph tests for non-verbal cues and psychological markers for verbal cues), and physiology studies. Written texts are also the subject of research for studying deception detection methods. In general, in regard to the computational linguistics and natural language processing, different tools and software already exist which help in detecting deception and lie.

On the lexics level, psycholinguistic lexicons – for instance, the general-purpose psycho-social dictionary Linguistic Inquiry and Word Count (LIWC) (Pennebaker and Francis, 1999) – can be adapted for binary text classifications. As for records of oral communication, it was revealed, using LIWC software, that people show less negative emotions, use less inconsistencies and modal verbs, use more modificators and speak longer if they tell deceptive information (Ali and Levine, 2008). Three data sets of texts, focused on three different topics (opinions on abortion, opinions on death penalty, and feelings about the best friend), were taken for the experiment on written English (Mihalcea and Strapparava, 2009). The classes of words from LIWC are categories relevant to psychological processes (for example, emotion or cognition), and the saliencies of these classes in truthful and deceptive texts were taken as features. An average classifier demonstrated 70% accuracy rate, showing that good separation between truthful and deceptive texts can be obtained (Mihalcea and Strapparava, 2009). LIWC-based analysis of written computer-mediated communication transcripts showed that deceptive texts use more sense-based words, fewer self-reference pronouns but more other-oriented pronouns (Hancock et al., 2007b). Generally, it was also found out, based on LIWC, that deceptive texts use more negative words, fewer exclusive words (for example, "only" and "just"), but more motion words (Bond and Lee, 2005).

Lexics and semantics analysis levels can be combined. Predictive lexical cues can be taken from the Statement Validity Analysis (Porter and Yuille, 1996). Such markers can be identified and extracted from texts, and with standard classification methods and algorithms (neural networks, decision trees, logistic regression) 74% accuracy can be achieved (Fuller et al., 2009).

As to facts in the described event, in (Sauri and Pustejovsky, 2012) the model is based on grammatical fact description structures in English and kindred languages. It is implemented in De Facto, a factuality profiler for eventualities based on lexical types and syntactic constructions.

Different types of written texts were investigated to find differences between truth and lie. Fake online reviews, fake social network profiles (Kumar and Reddy, 2012), fake dating profiles, fake news reports were already investigated. LIWC can be also used for fake dating profiles detection: motion verbs and words with "but", "except", and "without" should be taken into consideration (Toma and Hancock, 2012).

As concerns deceptive reviews, most of the works about opinion spam detection focus on the textual content of users' reviews by using word-level unigrams or bigrams as features, along with specific lexicons (LIWC, Word-Net Affect (Strapparava and Valitutti, 2004)), to learn classifiers (Li et al., 2013). The special website, which contains tools for manual and automated detection of fake reviews (https://www.cs.uic.edu/~liub/FBS/fake-reviews.html), suggests to analyze such lexical parameters as n-grams, POS tag n-grams, content similarity and style similarity of reviews written by different authors. Ott et al. (2011) archived nearly 90% accuracy using only n-grams as features. Truthful/deceptive texts classification is connected with texts' sentiment: authors of deceptive texts use more emotions, judgements and estimations (Hancock et al., 2011). Therefore, syntactic patterns can be used to distinguish feelings from arguments based on facts, and there are different classes of argumentations styles. Combining n-grams features and syntactic features derived from Probabilistic Context Free Grammar parse trees, model in (Feng et al., 2012) obtained 91.2% accuracy. Support vector machine classification showed 86% accuracy in deceptive reviews detection (negative deceptive opinion spam), also based on lexical and syntactic features (Ott et al., 2013). As to the semantics level of text analysis, the consistency of a text with similar texts is important: for example, how the definite review about the hotel goes with similar texts in the data

set. Pairs "attribute-descriptor" are created, and the function of the compatibility scores, based on information from the whole corpus, is counted. The precision is about 91% (Feng and Hirst, 2013). Other approach is to use Lexical Chain Based Semantic Similarity algorithm (99.75% accuracy) (Dewang and Singh, 2017). Some consistency features could also be used (Mukherjee et al., 2016). In (Chen et al., 2015) experiments for three features types are held. Shallow syntactic features include bag-of-words features, punctuation (exclamation and question marks), POS-unigrams and present baseline based on LIWC. POS tags bigrams are considered separately, as they show not only information about POS tags frequencies, but also the structure of the sentence. Connectives frequencies, derived from a shallow discourse parser trained on Penn Discourse Treebank, are used as features on discourse and pragmatics level. Frequencies of positive, negative and negation terms from the sentimental dictionary are used as sentiment features. Precision for explicit discourse classifier is 91.2%; precision for different features combinations classifiers is from 98.1% to 89.4%. So we can conclude that psycholinguistic features, sentiment features, POS tags and POS tags n-grams, punctuation marks, lemmas n-grams are helpful in deception detection on lexical and syntactic levels.

Regarding the next specific deception detection field, fake news detection, we can see that on the lexical level stylistic features (POS tags, length of words, subjectivity terms etc.) can be extracted that help to apart tabloid news (they are similar to fake news) with 77% accuracy (Lex et al., 2010). Numbers, imperatives, names of media persons can be extracted from news headlines (Clark, 2014); the numbers of these keywords can be used as features for classification with SVMs or Naive Bayes Classifier (Lary et al., 2010). In (Hardalov et al., 2016) the combined approach for automatically distinguishing credible from fake news, based on different features and combining different levels, is presented: there are linguistic (n-gram), credibility-related (capitalization, punctuation, pronoun use, sentiment polarity), and semantic (embeddings and DBPedia data) features. The accuracy is from 75% to 99% on three different data sets.

Deception detection studies in criminal analysis showed the importance of stylometric and psychological features, if we speak about verbal cues. Stylometry studies text on the basis of its stylistic features and can be used in order to attribute the text to an author (authorship attribution) or to get information about the author, for example, about her/his gender or personality (author profiling) (Fornaciari and Poesio, 2013). It is based on extracting low-level verbal cues from large amounts of text. Features in stylometric analysis can be surface-related (frequencies of function words, frequencies of n-grams of words or POS tags) and content-related (words from grammatical, lexical, psychological lexicons, syntactic features etc.) (Koppel, 2006; Fornaciari and Poesio, 2013). For instance, on the example of DECOUR corpus in Italian, the impact of such features on deception detection was investigated: LIWC features (words and sentences statistics, categories frequencies), lemmas n-grams frequencies and POS tags n-grams frequencies (Fornaciari and Poesio, 2013). Experiments for the Dutch language (on the example of CLiPS Stylometry Investigation (CSI) Corpus) detected, by examining the writing style of the author, that texts can be automatically classified as truthful or deceptive. The only features were the token unigrams; F-score and accuracy reached 72.2% (Verhoeven and Daelemans, 2014). In (Rosso and Cagnina, 2017) the impact of psycholinguistic features is also reviewed. So, in (Parapar et al., 2014) it was revealed, based on LIWC word categories, that use of pronouns, emotion words, markers of cognitive complexity, and motion verbs is more implicated in deceptive texts. In (Cano et al., 2014) six types of similar features were investigated while online behaviour studies: bag-of-words (with unigrams, bigrams and trigrams), POS tags as syntactic features, sentiment polarity, content (complexity, readability, length), psycholinguistic features (obtained with LIWC) and discourse patterns (semantic frames of words).

If we study deception detection for different languages, we should also keep in mind possible linguistic and cultural considerations (Rubin 2014). There is a lack of “real” deceptive texts as they are obviously difficult to collect and label, and studies in text-based deception detection have

mostly been performed for Romance and Germanic languages (T. Litvinova et al., 2017). There are few studies of deception detection in Slavic languages. In (Litvinova and Litvinova, 2016) the Russian Deception Bank corpus was studied with LIWC 2007 software (with Russian dictionary included). It was found out that average deceptive texts in Russian contain more pronouns (particularly personal ones), more singular and plural first-person pronouns but fewer third-person plural pronouns, fewer adverbs but more negations, numerals, emotional words overall. Deceptive texts have more positive and fewer negative emotion terms, which can be explained by the topic of the texts. Deceptive texts generally contain more words describing cognitive processes and contain considerably fewer words describing perception and particularly Seeing, Feeling but more from the subgroup Hearing. The fact that there is a lot of perception vocabulary in truthful texts compared to deceptive ones accords to the theories of reality monitoring, as was noted in other studies (Newman et al., 2003). There are also fewer punctuation marks in deceptive texts. In the latter research, 104 parameters were taken, a majority of which were LIWC parameters related to POS tags (unigrams), lexical-semantic group, and other frequencies. Frequencies of function words, discourse markers, different pronouns types were also taken into consideration, using special dictionaries. On the basis of the chosen parameters the classifier reached the overall accuracy of 68.3%, so the statistically significant difference between truthful and deceptive texts from the same author, written using an identical theme, was discovered for Russian. The accuracy of the model was also tested individually for males and females. The classification accuracy for males was 73.3 % and 63.3 % for females (O. Litvinova et al., 2017).

In (Pisarevskaya, 2017) the framework of the Rhetorical Structure Theory is used to reveal the differences between structures of truthful and deceptive (fake) news in Russian. The corpus of 134 news reports was collected. The best results for text classification were got by using Support Vector Machines with linear kernel (accuracy 65%). Discourse features (rhetorical relations frequencies and frequencies of their bigrams and trigrams) were taken as features, other possible language features were not taken into consideration.

Therefore, several parameters that are usually used for deception detection are already applied to the Russian language: POS tags and psycholinguistic features from LIWC, discourse features based on the Rhetorical Structure Theory. The impact of other psycholinguistic features, sentiment features, bigrams of POS tags, syntactic features, readability features needs to be studied.

3. Corpus

The only existing freely available corpus of truthful and deceptive texts for the Russian language is the Russian Deception Bank (Litvinova and Litvinova, 2016), so we chose it for our research goals. The corpus consists of 113 text pairs. Each pair contains one truthful and one deceptive narrative written by the same person and on the same topic: "How I spent yesterday". All texts have labels if they are truthful or deceptive. The average length of text is 221 words. The corpus can be downloaded from RusProfiling Lab website (http://en.rusprofilinglab.ru/korpus-tekstov/russian-deception-bank /). There are plans to extend it, but it can be already used for deception detection experiments.

Truthful text	Deceptive text
So here we were in Piter and went to the apartment that we had booked, it was not far from the city centre. Having dropped off our stuff, we went on a walk around the city centre and grabbed something to eat. Well, actually every afternoon we spent here was pretty much the same. In the evening we would go to any Pub or Bar and killed time there. Yes, killed time because it was not much fun. Maybe it's because the people around weren't much fun. Of course it was interesting to visit the museums and other sights of the city but I can't say that really left an impression that it was supposed to and all in all, I didn't feel too happy throughout that trip.	Having come to Piter, first thing we went to the apartment that we had booked, it was in the city centre, straight in Nevskiy, our window overlooked the beautiful views of Piter, especially in the evening when the sun went down, it was very beautiful. Of course you can spend ages walking the streets of the city and never get tired, while you are walking, you can't help being happy about everything you see around you. Every evening we would drive around different places in the city and sure thing, we don't have any clubs or pubs like that back home and I don't think we ever will. The way this city makes you feel is just special.

Table 1: Sample true and deceptive texts from the same author, translated into English (O. Litvinova et al., 2017).

The corpus texts were written by university students, native speakers of Russian. To keep them highly motivated, they were told that their texts would be evaluated by a trained psychologist who would attempt to distinguish a truthful text from a deceptive one, without having preliminary information of which texts are truthful and which are not. The respondents did not know the ultimate aim of the research. Therefore, it let make the texts closer to spontaneously produced written texts and minimize the effect of the observer's paradox.

The corpus was launched in 2014 as part of a text corpus called RusPersonality, Hence, the following metadata, with detailed information about the texts' authors, is provided in the corpus: gender, age, faculty and field of studies, dominant hand, psychological testing results – tests on brain lateral profile, test "Domino", questionnaire "Behaviour Self-Regulation Style". This information can be used in studies into the effects of personality on deception production, because these individual differences could be considered as vital factors in creating an objective method of identifying intentionally deceptive information (Levitan et al., 2015). There are 46 male and 67 female authors in corpus.

4. Features and Research Objectives

Our hypothesis is that there are significant differences between truthful and deceptive texts, and we try to reveal them using different types of markers and compare the results. The corpus also enables personal features of authors (psychological and physical), and we can check if they can be considered as a factor contributing to the production of their deceptive texts.

The corpus was already processed using LIWC software that is used in most studies of text-based deception detection (O. Litvinova et al., 2017), so psycholinguistic specialties from LIWC are already investigated and we do not take them into consideration now. However, we take POS tags frequencies, although they were already investigated, and analyze them together with POS tags bigrams frequencies.

There are three groups of markers for each text.

The first group is built by psycholinguistic and sentiment markers. We use four psycholinguistic properties of words from the MRC Psycholinguistic Database: familiarity, concreteness, imagery, average age of acquisition. Familiarity can be undestood as the frequency with which a word is seen, heard or used daily. Age of acquisition means the age at which a word is believed to be learned. Level of concreteness means how tangible the described object is. Imagery means the intensity with which a word arouses images (Paetzold and Specia, 2016). We use the bootstrapped version of these properties presented in (Paetzold and Specia, 2016), it contains information about 85942 words. We take the normalized frequencies in texts as features. As there are no such databases for Russian, we suggested that these properties could be universal for different languages and, hence, translated the corpus texts into English for the study of this group of markers. We use the API of Yandex machine translation service (http://translate.yandex.ru/) for translation from Russian into English.

Sentiment features are normalized frequencies of positive and negative sentiment words, taken separately or together as common sentiment words, so we have three features for them. We use lemmas from the General Russian Sentiment Lexicon (RuSentiLex) (Loukachevitch and Levchik, 2016) (http://www.labinform.ru/pub/rusentilex/rusentilex_2017.txt). It contains more than 12000 sentiment words and phrases, with positive, negative, neutral, positive/negative sentiment orientation. We take only positive and negative words from the lexicon.

We also add, as separate features, normalized frequencies of words that reflect 36 different emotions and attitudes: Worry, Friendship, Regret, Sadness, Sincerity, Double-Dealing, Curiosity, Love, Hope, Impertinence, Dissatisfaction, Disfavour, Disbelief, Grievance, Loneliness, Acceptance, Protest, Joy, Disinterest, Courage, Humility, Doubt, Calmness, Fear, Shame, Surprise, Pleasure, Respect, Inspiration, Belief, Call, Selfassumption, Pity, Desire, Cruelty, Anger.

The second group includes normalized frequencies of 11 POS tags and POS tags bigrams. First-person personal pronouns are considered separately. There are 204 possible bigrams collocations in corpus. Punctuation marks and whitespaces are not excluded from the analysis for bigrams and get their own tags.

The third group is presented by syntactic and readability features. We use two readability features: the Automated Readability Index and the Coleman–Liau Readability Formula. Here we also consider average length of tokens for text, type-

token ratio, normalized frequencies of exclamation marks, question marks, and two types of Russian participles (the first type is 'prichastie' and the second type is 'deeprichastie'). Syntactic features consist of 18 parameters that reflect structures of compound sentences and parenthetical constructions. These parameters names are in Russian and are taken from the Russian National Corpus (http://www.ruscorpora.ru).

5. Experiments and Data Analysis

Main experiments. For the whole data set, for each group of markers we run two experiments: with Support Vector Machines (linear kernel) (SVM) classifier and with Random Forest classifier. So we can reveal differences between truthful and deceptive texts, compare the impact of three groups of markers on the deception detection and find out, for each group, which subgroup is more important (Tables 2-4).

As **additional experiments**, we take texts of two groups of respondents (female and male authors separately), as in the previous work based on this corpus (Litvinova et al., 2017), to find out if there are any gender specified differences between truthful and deceptive texts. The experiments are run on the example of Random Forest classifier (Table 5).

The baseline for all experiments is 50%, because there is the equal number of truthful and deceptive narratives in the corpus.

We also try to reveal statistically significant parameters for truthful/deceptive texts for the whole data set and for the texts of female/male authors, taken separately (Table 6), using Student's t-test p-value.

Results for psycholinguistic and sentiment markers can be estimated to be the same as chance results (Table 2). These markers also do not show many statistically significant differences between truthful and deceptive texts. For the whole data set, the only statistically significant feature is Pleasure attitude.

The impact of syntactic and readabililty markers, taken together or separately, on the binary classification task is also not high (Table 3). Most of them do not reveal any statistically significant differences between truthful and deceptive narratives for the whole data set (p-value is <=0.05 for only one feature — 'разъяснительное' syntactical relation). Meanwhile, syntactic features are more important for classification than readability features, and the most important parameters are 'подчинительно-союзное' and 'релятивное' syntactical relations.

POS tags markers and POS tags bigrams markers can help more to detect truthful and deceptive texts, it corresponds to the results in (O. Litvinova, 2017). Accuracy for SVM classification is 0.57. The impact of POS tags features is more important than the impact of POS tags bigrams (Table 4). Conjunctions, interjections and numerals are mostly important for classification. Several features reveal statistically significant differences between truthful and deceptive texts for the whole data set (Table 6).

Texts by female/male authors, analyzed separately, help to reveal more statistically significant parameters for detection if a text is truthful or deceptive (Table 6).

In Table 5, we can see that the results for texts of female and male authors taken separately. Here we can also estimate that classification for the whole data set, based on POS tags markers and POS tags bigram markers, is the best one.

We can conclude that POS tags and POS tags bigrams features are mostly important for the automated deception detection for the Russian language.

	Random Forest Classifier, 5-fold cross-validation			
	Precision	Recall	F-measure	Accuracy
Sentiment features	0.29	0.6	0.39	0.48
Emotion features	0.48	0.45	0.46	0.48
Psycholinguistic features	0.45	0.44	0.44	0.45
Sentiment+ emotion+ psycholinguistic features	0.49	0.44	0.46	0.49
	Support Vector Machines (linear kernel), 5-fold cross-validation			
Sentiment features	0.19	0.4	0.25	0.48
Emotion features	0.47	0.46	0.46	0.47
Psycholinguistic features	0.42	0.46	0.42	0.42
Sentiment+ emotion+ psycholinguistic features	0.47	0.54	0.49	0.46

Table 2: Results for psycholinguistic and sentiment features (the whole data set).

	Random Forest Classifier, 5-fold cross-validation			
	Precision	Recall	F-measure	Accuracy
Syntactic features	0.52	0.49	0.49	0.50
Readablilty features	0.46	0.46	0.45	0.45
Syntactic+ readability features	0.49	0.46	0.46	0.48
	Support Vector Machines (linear kernel), 5-fold cross-validation			
Syntactic features	0.56	0.53	0.54	0.55
Readability features	0.55	0.30	0.26	0.45
Syntactic+ readability features	0.51	0.50	0.50	0.50

Table 3: Results for syntactic features and readability features (the whole data set).

	Random Forest Classifier, 5-fold cross-validation			
	Precision	Recall	F-measure	Accuracy
POS tags features	0.43	0.46	0.44	0.43
POS tags bigrams features	0.51	0.49	0.49	0.50
POS tags+ POS tags bigrams features	0.40	0.39	0.39	0.42
	Support Vector Machines (linear kernel), 5-fold cross-validation			
POS tags features	0.54	0.63	0.57	0.54
POS tags bigrams features	0.50	0.49	0.48	0.48
POS tags+ POS tags bigrams features	0.59	0.56	0.56	0.57

Table 4: Results for POS tags features and POS tags bigrams features (the whole data set).

	Random Forest Classifier, 5-fold cross-validation			
	Precision	Recall	F-measure	Accuracy
Psycholinguistic and sentiment features (the first group)				
The whole data set	0.49	0.44	0.46	0.49
Female authors	0.44	0.48	0.45	0.45
Male authors	0.47	0.53	0.49	0.46
POS tags and POS tags bigrams (the second group)				
The whole data set	0.52	0.45	0.46	0.51
Female authors	0.43	0.50	0.43	0.40
Male authors	0.51	0.46	0.46	0.48
Syntactic and readability features (the third group)				
The whole data set	0.49	0.46	0.46	0.48
Female authors	0.56	0.46	0.48	0.52
Male authors	0.47	0.40	0.42	0.46
The first group+the second group+the third group				
The whole data set	0.45	0.45	0.45	0.45

Table 5: Classification results for texts of female and male authors.

6. Conclusions and Discussion

We studied the impact of basic lexical and syntactic factors, which were not studied and compared before for the Russian language deception detection methods, on the successful deception detection. We revealed that POS tags and POS tags bigrams features are mostly important for binary classification task (the accuracy with SVM is 0.57), and conjunctions, interjections and numerals are the most significant parameters. Syntactic features should be also taken into consideration in further research. The results are preliminary and should be double-checked later on bigger Russian corpora. Other classification methods should be also tried to get better accuracy. The topic should be studied more deeply and intensively, on higher levels of linguistic processing, using semantics, discourse and pragmatics features. It could help search engines to label retrieved documents as reliable or not.

Group of markers	Texts by female authors: statistically significant parameters for truthful/ deceptive texts	Texts by male authors: statistically significant parameters for truthful/ deceptive texts	Texts from the whole dataset: statistically significant parameters for truthful/ deceptive texts
Psycholinguistic and sentiment features	Pleasure, Fear	Curiosity, Loneliness, Inspiration	Pleasure
POS tags and POS tags bigrams (here: excluding whitespaces and punctuation marks)	collocations: adverb-like pronoun+ adjective, numeral+ adjective-like pronoun, particle+ adverb	collocations: conjunction+ conjunction; conjunction+ adjective-like pronoun	numerals, collocations: conjunction+ conjunction, conjunction+adjective, noun-like pronoun+ conjunction, particle+ adverb, conjunction+adjective
Syntactic and readability features	'разъяснительное', 'сравнительное', 'сравнительно-союзное' syntactical relations	-	'разъяснительное' syntactical relation

Table 6: Differences for texts written by male authors and texts written by female authors: p-value is <=0.05.

Acknowledgments

This research is supported by a grant from the Russian Foundation for Basic Research, N 15-34-01221 Lie Detection in a Written Text: A Corpus Study.

References

G.D. Bond and A.Y. Lee. 2005. Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. Applied Cognitive Psychology, 19(3), pages 313-329.

A.E. Cano, M. Fernandez, and H. Alani. 2014. Detecting Child Grooming Behaviour Patterns on Social Media. Social informatics, ed. L. Aiello and D. McFarland, Lecture notes in computer science, vol. 8851, Springer. pages 412-427.

C. Chen, H. Zhao, and Y. Yang. 2015. Deceptive Opinion Spam Detection Using Deep Level Linguistic Features. In NLPCC 2015, LNAI 9362, pages 465–474.

R. Clark. 2014. Top 8 Secrets of How to Write an Upworthy Headline, Poynter, URL: http://www.poynter.org/news/media-innovation/255886/top-8-secrets-of-how-to-write-an-upworthy-headline/

B.M. DePaulo, K. Charlton, H. Cooper, J.J. Lindsay, and L. Muhlenbruck. 1997. The Accuracy-Confidence Correlation in the Detection of Deception. Personality & Social Psychology Review, 1(4), pages 346-357.

B.M. DePaulo, J.J. Lindsay, B.E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. 2003. Cues to deception. Psychological bulletin, 129(1), pages 74-118.

R.K. Dewang and A.K. Singh. 2017. Spam Review Detection through Lexical Chain Based Semantic Similarity Algorithm (LCBSS) for Negative Reviews. International Journal of Engineering and Technology (IJET). 8(6), Dec 2016-Jan 2017, pages 2946-2955.

S. Feng, R. Banerjee, and Y. Choi. 2012. Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Jeju Island, Korea, July, pages 171–175.

V. Feng and G. Hirst. 2013. Detecting deceptive opinion with profile compatibility. In Proceedings of the 6th International Joint Conference on Natural Language Processing. Nagoya, Yapan, pages 14-18.

T. Fornaciari and M. Poesio. 2013. Automatic deception detection in Italian court cases. Artificial Intelligence and Law, 21(3), pages 303-340.

Ch.M. Fuller, D.P. Biros, and D. Dursun. 2008. Exploration of Feature Selection and Advanced Classification Models for High-Stakes Deception Detection. In Proceedings of the 41st Annual Hawaii International Conference on System Sciences, IEEE. Waikoloa, HI, USA.

C.M. Fuller, D.P Biros, and R.L. Wilson. 2009. Decision support for determining veracity via linguistic-based cues. Decision Support Systems, 46(3), pages 695-703.

J.T. Hancock, C. Toma, and N. Ellison. 2007a. The truth about lying in online dating profiles. In Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pages 449-452.

J.T. Hancock, L.E. Curry, S. Goorha, and M. Woodworth. 2007b. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. Discourse Processes, 45(1), pages 1-23.

J. Hancock, M. Woodworth, and S. Porter. 2011. Hungry like a wolf: A word pattern analysis of the language of psychopaths. Legal and Criminological Psychology, 18, pages 102-114.

M. Hardalov, I. Koychev, and P. Nakov. 2016. In Search of Credible News. In Artificial Intelligence: Methodology, Systems, and Applications, pages 172-180.

A. Heidari, M. D'Arienzo, S.A. Crossley, and N. Duran. 2017. Computational Analysis of Lexical and Cohesion Differences in Deceptive Language: The Role of Accordance. In Proceedings of 30th International Florida Artificial Intelligence Research Society Conference. May 2017, pages 270-275.

J. Hirschberg, S. Benus, J. Brenier, F. Enos, S. Friedman, S. Gilman, C. Gir, G. Graciarena, A. Kathol, and L. Michaelis. 2005. Distinguishing deceptive from non-deceptive speech. In Proceedings of Interspeech 2005, Lisbon, Portugal, ACM, pages 1833–1836.

M. Koppel, J. Schler, S. Argamon, and J. Pennebaker. 2006. Effects of age and gender on blogging. In Computational Approaches to Analyzing Weblogs, Papers from the 2006 AAAI Spring Symposium, Technical Report SS-06-03, Stanford, California, USA, March 27-29.

N. Kumar and R.N. Reddy. 2012. Automatic Detection of Fake Profiles in Online Social Networks. BTech Thesis.

D.J. Lary, A. Nikitkov, and D. Stone. 2010. Which Machine-Learning Models Best Predict Online Auction Seller Deception Risk? American Accounting Association AAA Strategic and Emerging Technologies.

S.I. Levitan, M. Levine, J. Hirschberg, N. Cestero, G. An, and A. Rosenberg. 2015. Individual Differences in Deception and Deception Detection. In COGNITIVE 2015: The Seventh International Conference on Advanced Cognitive Technologies and Application.

E. Lex, A. Juffinger, and M. Granitzer. 2010. Objectivity classification in online media. In Proceedings of the 21st ACM conference on Hypertext and hypermedia, pages 293-294.

J. Li, M. Ott, and C. Cardie. 2013. Identifying Manipulated Offerings on Review Portals. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, USA, 18-21 October 2013, pages 1933-1942.

O. Litvinova, T. Litvinova, P. Seredin, and J. Lyell. 2017. Deception Detection in Russian Texts. In Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain, April 3-7, pages 43–52.

T. Litvinova and O. Litvinova. 2016. Russian Deception Bank: A Corpus for Automated Deception Detection in Text. In Proceedings of CBBLR 2016, Tribun EU, pages 1-7.

T. Litvinova, E. Ryzhkova, O. Litvinova, E. Larin, J. Lyell, and P. Seredin. 2017. Building a corpus of real texts for deception detection. In IMS'17, June 21–23, 2017, St. Petersburg, Russia.

N. Loukachevitch and A. Levchik. 2016. Creating a General Russian Sentiment Lexicon. In Proceedings of Language Resources and Evaluation Conference LREC-2016.

R. Mihalcea and C. Strapparava. 2009. The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In Proceedings of the Association for Computational Linguistics (ACL-IJCNLP 2009), ACM, pages 309-312.

S. Mukherjee, S. Dutta, and G. Weikum. 2016. Credible Review Detection with Limited Information using Consistency Features. Machine Learning and Knowledge Discovery in Databases, pages 195-213.

M. Newman, J. Pennebaker, D. Berry, and J. Richards. 2003. Lying words: Predicting deception from linguistic style. Personality and Social Psychology Bulletin, 29, pages 665-675.

M. Ott, Y. Choi, C. Cardie, and J.T. Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA, June, pages 309–319.

M. Ott, C. Cardie, and J. Hancock. 2013. Negative Deceptive Opinion Spam. In Proceedings of NAACLHLT, pages 497–501.

G.H. Paetzold and L. Specia. 2016. Inferring Psycholinguistic Properties of Words. In Proceedings of NAACL-HLT 2016, San Diego, California, June 12-17, pages 435–440.

J. Parapar, D.E. Losada, and Á. Barreiro. 2014. Combining Psycho-linguistic, Content-based and Chat-based Features to Detect Predation in Chatrooms. Journal of Universal Computer Science, 20(2), pages 213-239.

J. Pennebaker and M. Francis. 1999. Linguistic inquiry & word count: LIWC. Erlbaum Publishers.

D. Pisarevskaya. 2017. Rhetorical Structure Theory as a Feature for Deception Detection in News Reports in the Russian Language. In Computationa Linguistics and Intellectual Technologies. Papers from the Annual International Conference 'Dialogue' (2017). Moscow, May 31-June 3, 16(1), pages 191-200.

S. Porter and J.C. Yuille. 1996. The language of deceit: An investigation of the verbal clues to deception in the interrogation context. Law & Human Behavior, 20 (4), pages 443-458.

P. Rosso and L.C. Cagnina. 2017. Deception Detection and Opinion Spam. Chapter in Book A Practical Guide to Sentiment Analysis, ed. E. Cambria et al. Springer International Publishing AG 2017, pages 155-171.

V.L. Rubin. 2014. Pragmatic and Cultural Considerations for Deception Detection in Asian Languages. In ACM Transactions on Asian Language Information Processing, Vol. 13, No. 2, Article 10, Publication date: June 2014, pages 10:1-10:8.

R. Sauri and J. Pustejovsky. 2012. Are You Sure That This Happened? Assessing the Factuality Degree of Events in Text. Computational Linguistics, 38(2), pages 261-299.

C. Strapparava and A. Valitutti. 2004. WordNet-Affect: an Affective Extension of WordNet. In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, May 2004, pages 1083-1086.

C.L.Toma and J.T. Hancock. 2012. What Lies Beneath: The Linguistic Traces of Deception in Online Dating Profiles. Journal of Communication, 62(1), pages 78-97.

B. Verhoeven and W. Daelemans. 2014. CLIPS Stylometry Investigation (CSI) Corpus: a Dutch Corpus for the Detection of Age, Gender, Personality, Sentiment and Deception in Text. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), isbn 978-2-9517408-8-4, ELRA: Reykjavik, Iceland, pages 3081-3085.

H. Zhang, S. Wei, H. Tan, and J. Zheng. 2009. Deception detection based on SVM for Chinese text in CMC. In Proceedings of Sixth International Conference on Information Technology: New Generations (ITNG '09) (Las Vegas, NV, USA, April 27-29, 2009), IEEE, pages 481–486.

L. Zhou, J. Burgoon, J. Nunamaker, and D. Twitchell. 2004. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. Group Decision & Negotiation, 13(1), pages 81-106.