Russian Deception Bank: A Corpus for Automated Deception Detection in Text

Tatiana Litvinova, Olga Litvinova

Voronezh State Pedagogical University
centr_rus_yaz@mail.ru

Abstract. The problem of automatic lie detection in a written text is very urgent particularly with a growing number of Internet communication and thus it has been studied over the last decade but mainly using English-language materials. In order to address this, there have to be special text corpora that are challenging to design. The article presents a specially designed corpus of Russian-language texts - "Russian Deception Bank" - with a "truthful" and a "deceptive" text by the same author as well as metadata with information about the authors (gender, age, psychological testing data, etc.). This is the first Russian corpus of this type. There are also the results of the analysis conducted in order to identify the differences between "truthful" and "deceptive" texts along a range of linguistic characteristics that were extracted using the Linguistic Inquiry and Word Count software.

Key words: deception, text-based deception detection, text corpus, computer linguistics, deceptive text, linguistic cues of deception

1 Introduction

The problem of deception detection in written speech has not only a theoretical, but also, quite obviously, a practical importance. It is well-known that certain typological properties of deception can be identified on different language levels.

Scientists have been focussing on the identification of deception in speech, and particularly written speech. This kind of research has an interdisciplinary character. The computational approach to text-based deception detection has currently gained momentum. Linguists, psychologists and mathematical modelling specialists have been pooling their efforts in order to identify statistically significant differences between truthful and deceptive texts and to build mathematical models to determine whether the given text contains deceptive information [2; 10]. However, such studies have mainly analysed English language material, although there are a few exceptions [1; 2; 5]. Indeed, it is obvious that extensive research into languages other than English is needed.

In order to address lie detection in a written text, what is the most important is text corpora containing deceptive as well as truthful texts by the same author. Comparing such corpora is a daunting scientific task in its own right.

In this article we are presenting a freely available[1] text corpus "Russian Deception Bank" designed for investigating text-based deception detection as well as the results of the studies to identify the differences between truthful and deceptive texts in Russian.

[1] Available by request at: centr_rus_yaz@mail.ru

A. Horák, K. Pala, P. Rychlý, A. Rambousek (Eds.): CBBLR 2016, pp. 1–7, 2016.

2 Related work

Studies into automatic lie detection in a written text involve the use of special text corpora that are processed by means of various software to extract the values of particular text parameters.

Both "real" texts (e.g., trial transcripts) and texts designed in a laboratory particularly for experimental research are employed [14]. As valuable as "real" texts are, they have certain disadvantages. First of all, the veracity or deception of the narratives has to be already determined, and this is not always an easy task. In addition, there is a difficulty in obtaining a control sample of language in which the same speaker tells the truth for the sake of idiolectal comparison however it is very important in developing methods of deception detection in texts is to identify changes in the idiolect of the same individual when they produce both deceptive and truthful texts on the same topic [2]. Instead, the whole set of statements labelled as "false" is contrasted with the set comprising "true" statements.

It is also of importance to control the topic and genre of a text as well as to consider authors' personalities, which can only be accomplished in a laboratory.

The existing text corpora employed in research [2; 10; 11; 12; 15; 16], do not provide detailed metadata providing the authors' personal information (gender, age, education level, psychological testing data, etc.) to establish the effect of personality traits on how deceptive texts are produced.

Besides, as was noted, most corpora contain English texts even though cross-cultural studies are of significance [2]. We are not aware of any Slavic corpus of such a kind.

3 Corpus description

In our paper we have used for the first time a text corpus "Russian Deception Bank". It was launched in 2014 as part of a text corpus called "RusPersonality" [9]. "Russian Deception Bank" currently contains truthful and deceptive narratives (average text length is 221 words, $SD = 15.2$) of the same individuals on the same topic ("How I spent yesterday") (see example in Table 1).

Since it was not a spontaneously produced language, it was deemed necessary to minimize the effect of the observer's paradox by not explaining the ultimate aim of the research to the participants. In addition, to have them highly motivated, the respondents were told that their texts (without information of which of them were truthful and which were not) would be evaluated by a trained psychologist who would attempt to tell a truthful text from a deceptive one. Each respondent whose texts would not be correctly evaluated would be awarded with a cinema ticket voucher.

The number of the authors is N = 113 as of now (46 males, 67 females, university students, all native speakers of Russian) and there are plans to extend it. Apart from truthful and deceptive texts on the same topic by each individual, "Russian Deception Bank" (as well as all the texts in "RusPersonality") comes with metadata which provides detailed information about their authors (gender, age, specialty, psychological testing results - tests on brain lateral profile, test "Domino", questionnaire "Behaviour Seда-Regulation Style"). Hence, the annotated "Russian Deception Bank" will enable

Table 1: Sample true and deceptive narratives from the same author

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.

authors' individual differences to be considered as a factor contributing to the production of their "deceptive" texts. We argue that these data are critical in designing an objective method of identifying intentionally deceptive information [6].

4 Results and discussion

Obviously no language category can be a deception cue in its own right. What is important is the combination of frequent characteristics of certain text parameters making up what can be called "a linguistic deception profile". It is recommended that software be used in order to analyze texts along a number of linguistic parameters. For this particular study, we have made use of Linguistic Inquiry and Word Count (with Russian dictionary included) which helped to obtain the numerical values of text parameters [13]. LIWC 2007 has approximately 80 output variables including standard linguistic features (function words, including pronouns, articles etc; verbs, numbers etc.), psychological dimensions (Includes all social, emotional, cognitive, perceptual and biological processes, as well as anything related to time or space), personal concerns (any references to work, leisure, money, religion, etc.), spoken categories (primarily filler and agreement words).

LIWC searches the dictionary file for each word within text and for each word that is found the category to which that word belongs is incremented. Most variables are percentages of certain words in the total document or text set.

This software is extensively employed in studies concerned with language and personality, as well as differences between truthful and deceptive texts (e.g., see [1; 2; 10; 11]).

We extracted the values of all the categories except personal concerns (on the basis that it is too content-dependent, cf. [11]), and spoken categories. Each verbal sample was entered into a separate text file, misspellings being corrected.

Mathematical analysis was performed in two stages. During the first stage, a coefficient of variation for each of the parameters of truthful and deceptive texts was determined:

$$V = \frac{\sigma}{\bar{x}} \times 100\,\% \tag{1}$$

where σ is the average quadratic deviation, $\bar{x}$ is the average arithmetic mean of the selection.

This was done in order to establish which linguistic characteristics remain stable in texts by the same author and which vary. In order to achieve this, we used calculated the deviation of each text parameter from its average value for a particular individual. Furthermore, we averaged a deviation for each parameter in all of the texts and determined a coefficient of variation to enable us to evaluate a range of text parameters and see how large it is in relation to an average value.

The calculations were performed using a statistical software package SPSS.

Statistical analysis showed that the resulting coefficient of variation for the chosen text parameters was wide ranging. For a further analysis we chose only the parameters with the variation coefficient of less than 50 % [7].

In order to see how the text parameters change in relation to the absolute value, the averaged values of each of the parameters were calculated. Table 2 identifies relative changes in each of stable parameters from the “deceptive” texts in relation to “truthful” ones.

Hence on average deceptive texts 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. While deception is often associated with negative emotion terms [11], our deceptive texts have more positive and fewer negative emotion terms. It might be due to the topic of the texts: deceptive texts were commonly descriptions of how the participants spent their day that were meant to be perceived as real.

Deception has also previously been associated with decreased usage of first person singular, an effect attributed to psychological distancing [11]. In contrast, we find increased first person singular to be among indicators of deception, which we speculate is due to our deceivers attempting to enhance the credibility of their texts by emphasizing their own presence in the text as in Ott et al. [12] where a similar research was performed using hotel reviews.

The use of linguistic denials and negative sentences (e.g. no, not, never) in deceptive communication has received certain attention in previous literature. As stated by Picornell, “the easiest way to lie is to deny something” [14, p. 28]. In some other studies there were more negations in deceptive texts compared to truthful ones (see [2] for more detail).

Deceptive texts generally contain more words describing cognitive processes (1.97) but the ratios in the subgroups are different (there are more words describing Insight, Causation, Tentativeness, Inclusion, but fewer words from the subgroups Discrepancy, Inhibition). There are contradictory data regarding this category. Some researchers find

Table 2: Changes in the parameters of deceptive texts in relation to truthful ones (%)

Parameter	Changes in the averaged parameters of deceptive texts in relation to truthful ones, % (>2)
Total pronouns	2.43
Total pers pronouns	6.06
1^{st} pers singular	5.74
1^{st} pers plural	4.76
3^{rd} pers singular	5.22
3^{rd} pers plural	-7.63
Adverbs	-4.54
Negations	2.45
Numbers	13.16
Positive	3.94
Negative	-8.41
Insight	9.07
Causation	4.94
Discrepancy	-7.96
Tentative	2.50
Inhibition	-6.57
Inclusive	3.45
Perceptual Processes	-11.95
Seeing	-10.21
Hearing	10.45
Feeling	-25.47
Time	-4.65
Punctuation marks	-6.86

an increased presence of cognitive processes in deceptive statements, while others identified contradictory data (see review and references therein in [2]). However, most researchers agree on the fact that tentative words in particular are associated with insincerity (see also p. 30).

Deceptive texts 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 is in full agreement with the theories of reality monitoring [4], as was noted in many other studies (see, e.g., [11]).

Besides, there are fewer punctuation marks in deceptive texts.

Certainly additional work is required, but these findings some of which contradict with other data further suggest the importance of moving beyond a universal set of deceptive language features (e.g., LIWC) by considering both the contextual and motivational parameters underlying a deception as well [11].

5 Conclusions

In order to identify the linguistic features of a deceptive text special corpora are required. In our paper we have presented the first corpus of Russian texts designed for studying text-based deception detection. We have been able to establish the differences between truthful and deceptive texts by the same author using Linguistic Inquiry and Word Count.

There are plans to identify the features of deceptive texts by males and females, individuals with mental disorders, etc. These individual differences have often been neglected in previous research on deception detection due to no relevant information, while our corpus is completely instrumental in this type of research, which will deepen human understanding of the linguistic mechanisms underlying deceit.

Acknowledgements. This research is supported by a grant from the Russian Foundation for Humanities N 15-34-01221 “Lie Detection in a Written Text: A Corpus Study”.

References

1. Almela, Á., Valencia-García, R., Cantos, P.: Seeing through Deception: A Computational Approach to Deceit Detection in Spanish Written Communication. Security, Law and Intelligence 1 (2013)
2. Almela, Á.: Featuring deception in written language: a contrastive study of English and Spanish. PhD thesis. University of Murcia (2012).
3. Burgoon, J. K., Blair, J. P., Qin, T., and Nunamaker, J. F.: Detecting deception through linguistic analysis. Intelligence and Security Informatics 2665, 91-101 (2003).
4. Johnson, M.K., Raye, C.L.: Reality monitoring. Psychological Review, 88(1): 67-85 (1981).
5. Kang, S.-M., Lee, H.: Detecting Deception by Analyzing Written Statements in Korean. Security, Law And Intelligence 2 (2014).
6. Levitan, S. Ita, Levine, M., Hirschberg, J., Cestero, N., An, G., Rosenberg, A.: Individual Differences in Deception and Deception Detection. In: COGNITIVE 2015: The Seventh International Conference on Advanced Cognitive Technologies and Application (2015).
7. Litvinova, T. A.: On the problem of stability of parameters of idiolect. Proceedings of Southern Federal University. Philology, 3: 98-106 (2015).
8. Litvinova, T. A., Seredin, P. V., Litvinova, O. A.: Using Part-of-Speech Sequences Frequencies in a Text to Predict Author Personality: a Corpus Study. Indian Journal of Science and Technology 8 (9) [S. l.], 93-97 (2015).
9. Litvinova, T., Litvinova, O.: Authorship Profiling in Russian-Language Texts, in Proceedings of 13th International Conference on Statistical Analysis of Textual Data (JADT 2016), University Nice Sophia Antipolis, Nice, pp. 793-798 (2016).
10. Mihalcea, R., Strapparava, C.: The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In: Proceedings of the Association for Computational Linguistics (ACL-IJCNLP 2009), pp. 309-312. Stroudsburg, PA, USA (2009).
11. Newman, M., Pennebaker, J., Berry, D., Richards, J.: Lying words: Predicting deception from linguistic style. Personality and Social Psychology Bulletin 29, 665-675 (2003).
12. Ott, M., Choi, Y., Cardie, C., Hancock, J. 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, pp. 309-319 (2011).

13. Pennebaker, J. W., Chung, C. K., Ireland, M. E., Gonzales, A. L., Booth, R. J.: The Development and Psychometric Properties of LIWC2007. The University of Texas at Austin and The University of Auckland, New Zealand (2007).
14. Picornell, I.: Cues to Deception in a Textual Narrative Context: Lying in written witness statements. PhD thesis, Aston University (2013).
15. Rubin, V. L., Vashchilko T.: Identification of truth and deception in text: application of vector space model to rhetorical structure theory. In: EACL 2012 Proceedings of the Workshop on Computational Approaches to Deception Detection, pp. 97-106. Association for Computational Linguistics Stroudsburg, PA, USA (2012).
16. Salvetti, F.: Detecting Deception in Text: A Corpus-Driven Approach: Ph.D. dis-sertation. University of Colorado at Boulder (2012).