



## Composition and Structure of the Russian Deception Bank Corpus Designed for Developing Methods of Text-Based Deception Detection

Deception is intentional attempt to foster a false belief or conclusion.  
People are poor lie detectors (just better than a coin-flip)  
One of the solutions is text-based deception detection methods.

- In recent years deception detection has been commonly addressed as a text classification problem employing the methods of natural language processing and data mining (up to 93 % accuracy for online review spam detection, see Karami, Zhou, 2015).

A very important issue associated with deception detection is designing valid text corpora.

- Few corpora of real world texts · Lack of information about text authors in many existing corpora · Few corpora in non-English languages (DECOUR , CLiPS corpus) · No corpora in Slavic languages ·

We introduce the **Russian Deception Bank**,  
the first Slavic text corpus for deception detection studies

### “Laboratory” texts

- ✓ 113 deceptive texts and 113 truthful texts on the same topic written by 113 Russian university students
- ✓ rich metadata – information about the authors – useful for studies into the effects of personality on deception production (gender, age, personality profile etc.)
- ✓ Classification accuracy of 68 % is achieved using D-LIWC features (Litvinova et. al., 2017)
- ✓ freely available at our webpage <http://en.rusprofilinglab.ru/korpus-tekstov/>

### Real-world texts

- ✓ a collection of dialogues (interviews) in Russian (video recordings of job interviews and crime suspects interrogations)
- ✓ high-stake deception
- ✓ 13 respondents, 19 hours of video
- ✓ the narrator admits lie
- ✓ for those interviewing for a job or already working for the company there is data available on a series of psychological tests
- ✓ currently in the process of labeling
- ✓ transcribed texts without personal information will be available as well as some audio and video-based features
- ✓ Useful for studying correlations between verbal and nonverbal cues

**Are there any other corpora  
for deception detection for non-English languages?**