

It turned out, for example, that liberal arts academics have a tendency to use many archaic words, while office workers and speakers with incomplete higher education lean towards etiquette formulas.

In the speech of businessmen, middle-aged speakers, as well as men and the retired, particles were more frequent than in other sociolects.

Ellipsis is typical for the speech of office and creative workers, students and speakers with incomplete higher education. Parcellation prevails in the speech of workers and speakers with high URC. In the speech of speaking middle and senior age groups, liberal arts academics, natural scientists, engineers, law enforcement agents and the retired, a predominance of inverted structures is clearly observed, being more frequent than parcellation and self-correction.

Again, this data can be important for a variety of purposes: from purely scientific tasks of developing systems for the automatic processing of sound signals, speech recognition and synthesis, to forensic linguistics and linguistic didactics.

VIII. FURTHER RESEARCH

Despite the fact that the results of the analysis of the corpus material in all considered aspects (vocabulary, morphology, syntax), obtained at the first stage of the study (2016 [22]) and in this work, confirm each other quite well, the final set of diagnostic signs for the mentioned so far can only be considered potentially diagnostic, requiring another confirmation on the expanded corpus material. This remains the prospect of studying everyday Russian spoken speech in the chosen direction. Traditional linguistic techniques are not sufficient for the analysis of real time spontaneous speech even in transcripts. Therefore most research in the area considers data mining tools from the linguistic domain as main for mining big amounts of data as corpora [20]. The first and central question to the approach if data-mining methods are able to generate and then verify the existing research results. The second one is the ability to lead the linguist to further linguistically interesting patterns emerging from the data of non-standard speech corpora. One more crucial issue in this endeavor is how to discover those linguistic features that are good indicators of sociolinguistic differences, provided they exist. As you can see above, we have explored a set of features potentially distinguishing between the different social classes, on the one hand, but they must be verified, on the other hand. In our future work we plan to carry out more analyses on the basis of aggregated linguistic information using the described approach in order to explore more concrete parameters of variation.

ACKNOWLEDGMENT

The presented research was supported by the Russian Foundation for Basic Research, project 17-29-09175 "Diagnostic Features of Sociolinguistic Variation in Everyday Spoken Russian (based on the Material of Sound Corpus)".

REFERENCES

[1] I. A. Baudouin de Courtenay, *Selected Works on General Linguistics*, Moscow: USSR Academy of Sciences, 1963, Vol. 1, pp. 348-350.

[2] R. Barthes, *Writing Degree Zero, Semiotics [Collection of articles. Translations] / Yu. S. Stepanov (ed.)*. M.: Raduga, 1983, pp. 306-349.

[3] A. D. Schweitzer, *Sociolinguistics // Linguistic Encyclopedic Dictionary / V. N. Yartseva (ed.)*, Moscow: Soviet encyclopedia, 1990, p. 481.

[4] S. Ash, *Social class*. In *The Handbook of Language Variation and Change*, J. K. Chambers, P. Trudgill and N. Schilling-Estes (eds.). Malden, Mass.: Blackwell, 2002, pp. 402-422.

[5] A. B. Hollingshead, *Two Factor Index of Status Position*. New Haven: Yale University Press, 1957, pp. 1-11.

[6] W. Labov, *The Social Stratification of New York City*. Washington, DC: Center for Applied Linguistics, 1966.

[7] *Sociology: Encyclopedia / Comp. A. A. Gritsanov, V. L. Abushenko, G. M. Evelkin, G. N. Sokolova, O. V. Tereshchenko*. Moscow: Knizhny Dom, 2003.

[8] V. I. Karasik, *Language of Social Status*. Moscow: IYa AS of SSSR, Volgograd Pedagogical Institute, 1991.

[9] G. N. Putnam and E. M. O'herm, *The Status Significance of an Isolated Urban Dialect*. *Language*, 31(4). 1955, pp. 5-32.

[10] L. S. Harms, *Listener Judgments of Status Cues in Speech*. *Quarterly Journal of Speech*, 47(2), 1961, pp. 164-168;

[11] D. James, *Listener judgments of status cues in speech: A replication and extension*, *Speech Monographs*, 39:2. 1972, pp. 144-147.

[12] M. M. Bakhtin. *Questions of Literature and Aesthetics*. Moscow: Khudozhestvennaya literature Publ, 1975.

[13] *Speech Corpus as a Base for Analysis of Russian Speech*. Collective Monograph. Part 1. Reading. Retelling. Description [Zvukovoj korpus kak material dl'a analiza russkoj rechi: kollektivnaja monografija. Chast' 1. Chtenie. Pereskaz. Opisanie]. N. V. Bogdanova-Beglarian (ed.). St. Petersburg, 2013.

[14] A. Asinovsky, N. Bogdanova, M. Rusakova, A. Ryko, S. Stepanova, T. Sherstinova, *The ORD Speech Corpus of Russian Everyday Communication "One Speaker's Day": Creation Principles and Annotation*, in: Matoušek, V., Mautner, P. (eds.) TSD 2009. LNAI, vol. 5729. Springer, Berlin-Heidelberg, 2009, pp. 250-257.

[15] N. V. Bogdanova-Beglarian, T. Yu. Sherstinova, O. V. Blinova, G. Yu. Martynenko, *An Exploratory Study on Sociolinguistic Variation of Spoken Russian*. SPECOM 2016. Lecture Notes in Artificial Intelligence, LNAI, vol. 9811. Springer, Switzerland, 2016, pp. 100-107.

[16] N. V. Bogdanova-Beglarian, T. Yu. Sherstinova, O. V. Blinova, O. B. Ermolova, E. M. Baeva, G. Ya. Martynenko, A. I. Ryko, *Sociolinguistic Extension of the ORD Corpus of Russian Everyday Speech*, in: Ronzhin, A. et al. (eds.) SPECOM 2016, Lecture Notes in Artificial Intelligence, LNAI, vol. 9811. Springer, Switzerland, 2016, pp. 659-666.

[17] N. V. Bogdanova-Beglarian, O. V. Blinova, T. Yu. Sherstinova, G. Ya. Martynenko, *Corpus of Russian Everyday Speech "One Day of Speech": present state and prospects*, in: Proceedings of the V. V. Vinogradov Institute of Russian language. Vol. 21. Russian National Corpus: Research and Development / A. M. Moldovan, V. A. Plungyan (eds.). Moscow, 2019, pp. 101-110.

[18] T. Sibata, *Study of Language Existence within 24 Hours] // Linguistics in Japan*. Moscow: Raduga, 1983, pp. 134-141.

[19] N. Campbell, *Speech & Expression; the Value of a Longitudinal Corpus*, LREC 2004. Lisbon, 2004, pp. 183-186.

[20] Han and M. Kamber, *Data Mining: Concepts and Techniques*. 2006.

[21] *Reference Guide for the British National Corpus (XML edition)*. Published for the British National Corpus Consortium by Oxford University Computing Services. L. Burnard (ed.), 2007. [Web: <http://www.natcorp.ox.ac.uk/docs/URG/2007>].

[22] *Everyday Russian Language: Functioning Features in Different Social Groups*. Collective Monograph. N. V. Bogdanova-Beglarian (ed.), St. Petersburg, 2016.

[23] *Mystem Web*: <https://tech.yandex.ru/mystem/>

[24] G. Ya. Martynenko, *The Syntax of a Live Spontaneous Speech: the Symmetry of Linear Orders // Corpus Linguistics*. 2015. Proceedings of the International Conference, in V. P. Zakharov, O. A. Mitrofanova, M. V. Khokhlova (eds.). Saint Petersburg: SPBU, 2015. pp. 371-378.

[25] *ELAN Web*: <https://tla.mpi.nl/tools/tla-tools/elan/>.