

## Creating of word cloud based on computational algebra

Brest State Pushkin University

### **Abstract**

*The article describes a possible application of computational algebra and computer modeling methods for improving methods of linguistic education and linguistic research. The article describes the technology of creating word clouds with the help of Mathematica.*

**Keywords:** *word cloud, Mathematica, Wolfram, computational algebra, modelling, visualization, linguistics*

Modern science is an open system where different areas are combined with the hope for the best results. Linguistics is no exception. And if earlier the main "partners" of linguists were psychologists, sociologists and other humanitarian researchers, today the use of computational methods and computer technologies not only actually simplifies the work with large data sets (previously such work had to be done manually), but also sets new goals of research, that were not possible earlier, helps to update the approach to already known tasks. An example: word clouds as one of the important tools in modern linguistic research.

Word Clouds (also known as wordle, word collage or tag cloud) are visual representations of words (in general terms) [1]. A word cloud is a collection, or cluster, of words depicted in different sizes. The more often the word appears within a given text and the more important it is, the bigger and bolder it is shown [2]. Nowadays it is an effective and representative way to pull out the most pertinent parts of textual data. Interesting and useful is the way of integrating word clouds technology into learning process, particularly – foreign language acquisition.

Word Clouds are used primary for representing text based data in the way we can perceive them better and only then for research purposes. They are shown in various types of presentations; in dictionaries (for example Duden [3]) they represent relations between different types of words and definitions. People typically use word clouds to easily produce a summary of large documents (reports, speeches), to create art on a topic (gifts, displays) or to visualize data (tables, surveys).

Although the word clouds can be manually composed, they are usually generated using special services or internal company's algorithms. Most of them allow you to create a word cloud from the fed in text, give it a certain shape, edit the obtained word list, or import a new one and create a cloud from it, not from the whole text. You can apply user's settings, which depend on the particular service. For example you can set fixed number of words, minimum frequency shown, display word count for every item, group similar words, exclude unwanted words or ignore common words in chosen language. As the tool the following online services might be used: Wordclouds.com, WordItOut, Infogram, TagCrowd etc.

The services mentioned above have a number of limitations in use, such as: it is necessary to have a certain data format, which cannot always be convenient and which differs from site to site; despite the variety of user settings in different services, they a) are in excess in some cases and in lack in others; b) more importantly still do not allow the user to fully control the creation of the cloud – only to set the vector, to create the model of the desired result.

These and other limitations often prevent a linguist or a teacher from realizing the full research or didactic potential of word clouds. They can be used as a way to visualize text information, a way to summarize the meaning of a text, the material for homework (construction, analysis, modeling), a way to work with vocabulary, lexical compatibility and much more. A modern and optimal solution is Mathematica's Computer Algebra System, which can and should improve the effectiveness of using word clouds technologies in linguistic research and language education. This is achieved through the flexibility of the special language for Mathematica.

The Wolfram Language is a general multi-paradigm computational language [4] developed by Wolfram Research. It allows programmers to operate at a significantly higher level than ever before, by leveraging built-in computational intelligence that relies on a vast depth of algorithms and real-world knowledge. [5]. Thanks to integration with computational intelligence, the distinctive feature of word clouds created by

Mathematica is that they can be created not from a given text, but from an automatically generated set of words by search query. For example, "European countries", "African languages", etc.

Wolfram language provides templates, which can be used for creating word clouds. The system allows you to change most aspects: it can be almost fully controlled manually. Changeable parameters can be size, shape, color (as with most services), centering, the place of a word on an axis, set of words, type of view and others. Also it is possible to have post-rendering effects, like clickable words, mouseover effects, etc.

Word clouds models created with Wolfram language are interactive because of the use of CDF. CDF – Computable Document Format – is a published public format created by Wolfram Research [6]. CDF offers content creators easy-to-author interactivity and convenient deployment options – empowering their readers to drive content and generate results live. That's why word clouds generated in CDF with Mathematica are the possibility to develop a new vision on their usage in research and education.

You can see examples of different kinds of word clouds done with Mathematica among Demonstration projects [7], section Textual analysis and School Language Arts: Word clouds, Refrigerator Poetry etc.

Thanks to the user-friendly interface and rich functionality it is convenient to work with Mathematica in a classroom: students can create their own clouds for projects, home and research work, for better understanding of the topic or a summary of the studied materials; clouds can be created by a teacher to sum up the topic, give a hint about a new one, test knowledge or just make teaching more lively and effective. Students should be able to work with the latest technology in various subjects, not only technical. The word clouds in linguistic education will enhance their IT-competencies. Researchers can use word clouds for similar purposes. They can quickly identify key points and efficiently analyze concordances and text arrays. Word clouds can often be found in the presentation of material.

The use of modern advanced technologies is now an integral part of the development of any branch of scientific knowledge. Educations in general and linguistic education in particular are no exception. Technical means allow improving teaching methods and making them more effective. At the same time, their use should be didactically justified. Tools should be specified according to the target group and the purpose.

Having considered an example of using word cloud generation based on Mathematica for teaching and research, we can draw a conclusion: The methods of computational algebra integrated into linguistics and language education are, though rather unexpected, very promising. Specific application and prevalence will depend on the preparatory phase, the technical skills of users and other factors. But in our opinion, the widespread introduction of technologies in the field of linguistics is an important condition for its further development in the context of technical progress.

## REFERENCE LIST

1. Create Live Word Clouds / Mentimeter [Electronic resource]. – Access mode: <https://www.mentimeter.com/features/word-cloud>. – Access date: 03.04.20.
2. Word clouds & The value of simple visualizations / Boost Labs [Electronic resource]. – Access mode: <https://www.boostlabs.com/what-are-word-clouds-value-simple-visualizations>. – Access date: 03.04.20.
3. Typische Verbindungen (computergeneriert) / DUDEN.DE [Electronic resource]. – Access mode: <https://www.duden.de/hilfe/typische-verbindungen>. – Access date: 03.04.20.
4. Wolfram Language. Material from Wikipedia, the free encyclopedia [Electronic resource]. – Access mode: [https://en.wikipedia.org/wiki/Wolfram\\_Language](https://en.wikipedia.org/wiki/Wolfram_Language). – Access date: 03.04.20.
5. Wolfram Language / WOLFRAM [Electronic resource]. – Access mode: <https://www.wolfram.com/language/>. – Access date: 03.04.20.
6. Computable Document Format. Material from Wikipedia, the free encyclopedia [Electronic resource]. – Access mode: [https://en.wikipedia.org/wiki/Computable\\_Document\\_Format](https://en.wikipedia.org/wiki/Computable_Document_Format). – Access date: 03.04.20.
7. Demonstrations Project / WOLFRAM [Electronic resource]. – Access mode: <https://demonstrations.wolfram.com/>. – Access date: 03.04.20.

**Huseva Aliaksandra V.** — Department of Foreign Languages, Brest State Pushkin University, Brest, Belarus, email: [huseva.aliaksandra@gmail.com](mailto:huseva.aliaksandra@gmail.com)

Supervisor: **Kontsevov Mikhail P.** — senior lecturer, Brest State Pushkin University, Brest, Belarus