Sketch Engine in Building a Lexical Minimum for Children

Gulshat R. Safiullina, Nuriya R. Davitova, Eva Lelakova

Abstract*: Our article deals with such an aspect of computational linguistics as the construction of lexical minima cases using the Sketch Engine program as an example. The advent of computational linguistics has played an important role in the process of learning foreign languages. Thanks to computer technology, the process of learning foreign languages is greatly simplified and becomes more accessible. Among the many programs for learning foreign languages, we chose the Sketch Engine program, since it is a case manager and a tool for analyzing linguistic buildings, that is, collections of texts selected and processed according to certain rules, which are used as the basis for language research. This resource is software that combines a specialized search engine and a lot of buildings in different languages. We describe the program through the prism of corpus linguistics, consider the functions and capabilities of this program Sketch Engine in drawing up the lexical minimum for primary school age in English, Russian and German. In this paper, we conducted an experiment on drawing up a lexical minimum for schoolchildren, which consisted in selecting 300 most used words of the English language and saturating them with examples from the cases of the Sketch Engine program.

Keywords: lexical, minimum, dictionary, children, Sketch Engine, corpus.

I. INTRODUCTION

In the modern world, automated information technologies play a special role. When it comes to creating advanced information technologies, the problems of automatic processing of textual information presented in natural languages come to the fore. The problems of using natural language in systems of automatic information processing are handled by computer linguistics. This science originated relatively recently, but since then, significant scientific and practical results have been obtained in the field of computational linguistics. In our work we study the definition of computational linguistics. Considering the specificity of the phenomena studied, the main research methods are: descriptive method; scientific observation method; component analysis method. When conducting various linguistic studies, computer linguistics is increasingly being used.

According to Shemakina Yu.I. Computational linguistics is a scientific direction in the field of mathematical and computer modeling of intellectual processes in humans and animals when creating artificial intelligence systems, which aims to use mathematical models to describe natural languages [1:81]. An interesting point of view on the subject of computational linguistics in the Ukrainian researcher and cybernetics A.V. Anisimov. According to the scientist, computational linguistics deals with the search for algorithms and the computer implementation of human interaction with computers in the natural human language. " Thus, "in the case of successful implementation of computational linguistics projects, a person will be able to manage complex technical objects using voice commands or natural language texts [2: 57]. Linguistic means, which are created and applied in the field of computational linguistics, can be divided into two parts: declarative and procedural [3, p. 7]. The declarative part will include, first of all, dictionaries of units of language and speech, texts and various grammatical tables. By the procedural part - a means of manipulating units of language and speech, texts and grammatical tables. On this basis, the computer interface will refer to the procedural part of computational linguistics.

The Sketch Engine program was developed by representatives of the company Lexical computing and Masaryk University in Brno (Czech Republic) and is intended primarily for linguists engaged in the compilation of dictionaries [4]. The program is a product of lexical computing, the founder of which is the lexicographer and researcher Adam Kilgarif [5]. It is also important to note that this program is used when compiling dictionaries by the lexicographers of many well-known publishers, including Oxford University Press, Cambridge University Press, KDictionaries, Cornelsen, Shogakukan, Collins, Macmillan dictionary and Trojina [6]. Domestic and foreign linguists studied the possibility of using the program Sketch Engine. Yu.D.Apresyan, V.Yu.Apresyan, V.Baysa, P.Ryhlyy, T. Botma, R.Gous, V.Pinsloo and others [7-12]

The main purpose of the emergence of this program is to create a tool for studying the lexicographical properties of words. Learning languages can take place on the basis of buildings, as well as on the basis of statistical data. In addition, the researchers emphasize an empirical approach to the study of the behavior of words in a language. In other words, how a word can manifest itself in various texts, phrases or sentences. Some researchers also point out to the need of research in teaching collocations [13-17]

The program interface Sketch Engine offers more than 200 packages in 82 two languages. The largest corpus (TenTencorporafamily) can contain from 2 to 15 billion words.

As we have already noted, one of the features of the Sketch Engine program is a variety of languages, including Chinese, Czech, English (there are two versions of English: American and British), Estonian, Finnish, Japanese, etc. In addition, in the future, this program can be used to work with Turkic languages.

Revised Manuscript Received on November 08, 2019.
***** Correspondence Author
Gulshat R. Safiullina, Kazan Federal University
Nuriya R. Davitova, Kazan Federal University
Eva Lelakova, University of Zilina Slovakia

II. RESULTS AND DISCUSSION

Consider the above provisions with specific examples. We have selected 300 words of the lexical minimum in English for younger students. The next stage was their translation into Russian and German.

Arranging these words in three columns in alphabetical order, we began to work on saturating the lexical minimum with examples. To guarantee the storage of the learned lexicon in a language learning environment, it is necessary to provide students with a context and demonstrate how the words are related [14-19-20]. Sketchengine allows to study words of various lexico-semantic groups in functional aspect [15, 16-18]. We emphasize that at the present stage of drawing up lexical minima, it is important to include in the body of the minimum relevant examples used by native speakers of speech in writing. We turn to the Sketch Engine program described by us and replenish our lexical minimum with examples.

Apple	The fruit of an apple tree is apple.	Яблоко	Фрукты, растущие на яблонях- яблоки.	Der Apfel	Der Apfel ist die Frucht, die an einem Apfelbaum wächst.
Arm	He broke his arm.	Рука	Он сломал свою руку.	Der Arm	Er hat sich den Arm gebrochen.
Air	I was breathing fresh air.	Воздух	Я дышал свежим воздухом.	Die Luft	Ich habe die frische Luft eingeatmet.
Animal	I love all kinds of animals.	Животное	Я люблю все виды животных.	Das Tier	Ich liebe alle Arten von Tieren.
Art	Hapkido is a Korean martial art.	Искусство	Хапкидо - Корейское боевое Искусство.	Die Kunst	Hapkido ist eine koreanische Kampfkunst.
Age	I am the same age as you.	Возраст	Я такого же возраста, что и ты.	Das Alter	Ich bin im gleichen Alter wie du.
All	All was well.	Всё	Всё было хорошо.	Alles	Alles war gut.
Always	I will always love you.	Всегда	Я буду всегда любить тебя.	Immer	Ich werde dich immer lieben.
Alone	She lives alone.	Один/Одинокий	Она живет одна.	Allein	Sie lebt allein.
Be	He wants to be a good man.	Быть	Он хочет быть хорошим человеком.	Sein	Er will ein guter Mensch sein.
Brother	Her brother is an actor.	Брат	Её брат - актер.	Der Bruder	Ihr Bruder ist der Schauspieler.
Body	The human body is a complex system.	Тело	Человеческое тело- это сложная система.	Der Körper	Der menschliche Körper ist ein komplexes System.
Building	The building is of brick.	Здание	Здание из кирпича.	Das Gebäude	Das Gebäude ist aus Ziegeln.
Bee	Bees live in the colonies.	Пчела	Пчелы живут колониями.	Die Biene	Bienen leben in einer Kolonie.
Blue	I have blue eyes.	Синий	Мои глаза синего цвета.	Blau	Ich habe blaue Augen.
Black	I have a black car.	Черный	У меня есть черная машина	Schwarz	Ich habe ein schwarzes Auto.
Brown	He wore a brown coat.	Коричневый	Он носил коричневое пальто.	Braun	Er trug einen braunen Mantel.
Bad	Yesterday was a bad day.	Плохой	Вчера был плохой день.	Schlecht	Gestern war ein schlechter Tag.
Baby	The baby is six months old now.	Малыш	Малышу сейчас 6 месяцев.	Das Baby	Das Baby ist jetzt sechs Monate alt.
Big	This is a big mistake.	Большой	Это - большая ошибка.	Groβ	Das ist ein großer Fehler.
Before	You knew her before she came here.	До	Вы знали до того, как она пришла сюда.	Bevor	Du kanntest sie, bevor sie hergekommen ist.
Because	I was there because I was afraid.	Потому что	Я был там, потому что боялся.	Weil	Ich war dort, weil ich Angst hatte.
Book	Give me this book.	Книга	Дай мне эту книгу.	Das Buch	Gib mir dieses Buch.
Boy	The boy plays football.	Мальчик	Мальчик играет в футбол.	Der Junge	Der Junge spielt Fußball.
But	Yes, but my little brother's here.	Но	Да, но мой младший брат здесь	Aber	Ya, aber mein kleiner Bruder ist hier.

Cat	The cat sleeps all day.	Кот	Кот спит целый день.	Der Kater	Der Kater schläft den ganzen Tag.
Clock	The clock has just struck three.	Часы	Часы только что пробили три.	Die Uhr	Die Uhr hat gerade drei geschlagen.
Chicken	The chicken clucks.	Курица	Курица кудахчет.	Die Henne	Die Henne gluckt.
Child	He is still a child.	Ребенок	Он все еще ребенок.	Das Kind	Er ist noch ein Kind.
Color	Green is my favorite color.	Цвет	Зеленый - мой любимый цвет.	Die Farbe	Grün ist meine bevorzugte Farbe.
Come	Come see me tonight after the theater.	Приходить	Приходите ко мне сегодня вечером после театра.	Kommen	Kommen Sie heute nach dem Theater bei mir vorbei.

The purpose of this lexical minimum was to match the examples, so we selected examples with a completely identical value.

The word building is accompanied by an example of The building is of brick. Russian translation of the word building - building. This word in our lexical minimum is presented along with an illustrative example of a brick building. The German version of the lexeme is Das Gebäude, which is reflected in the sentence of Das Gebäude ist aus Ziegeln.

The next example is the word date in the sentence The date is yet unknown. The Russian equivalent of this word is the word date, illustrated by the sentence. The date is still unknown. The sentence Das Datum ist noch unbekannt contains the German version of this word - Das Datum.

In this lexical minimum, we collected the most used verbs of the English language. For example, using the open word as an example, we chose the He opened window. In Russian, this word is translated as open and presented in the sentence. He opened the window. The German equivalent of this word - öffnen is expressed by the sentence Er öffnete das Fenster.

We also considered the most used prepositions. For example, in the sentence I have a present for you we can find the preposition for, which is expressed by the word for in the sentence in Russian. I have a gift for you. In German, the word is translated as für, which is presented in the sentence Ich habe ein Geschenk für dich.

Much attention was paid and numeral. This lexical minimum presents examples of numbers from one to twelve. As an example, we take the number seven in the sentence There Are seven wonders of the world. The German translation of the word seven - sieben, presented in the sentence Es gibt sieben Wunder der Erde. Offer There are seven wonders of the world illustrated by its Russian version.

III. SUMMARY

The program for learning languages Sketch Engine has unique features and functions. After examining each function in detail and illustrating it with examples, we can divide all functions into two large groups: the functions necessary to extract lexicographical data, and the functions with which we can create our own corpus.

The functions necessary for extracting lexicographical data are, in our opinion, most useful for students, because with their help we can find specific words in various contexts, find phrases and phrases with a given word, and, therefore, replenish our vocabulary and improve written and oral skills.

In this paper, we conducted an experiment on drawing up a lexical minimum for schoolchildren, which consisted in selecting 300 most used words of the English language and saturating them with examples from the cases of the Sketch Engine program. Thus, we have shown how this program can be used by lexicographers in compiling dictionaries.

IV. CONCLUSIONS

The Sketch Engine program was created quite recently, which means there may be additional developments and capabilities. One of these innovations will be the creation of buildings in the Turkic languages and in the future the developers plan to create a platform for learning Russian and Tatar languages, which means that further research will be needed on the capabilities and functions of this program.

ACKNOWLEDGEMENTS

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

REFERENCES

1. Shemakin Yu. I. the Beginning of computer linguistics Moscow: Publishing house of Moscow state University, A/O "Rosvuznauka", 1992. – 181 p.
2. Anisimov A. Computational linguistics for all: Myths. Algorithms. Language. - Kiev: Naukovadumka, 1991. - 208 p.
3. Belonogov G. G., Kalinin Yu. P., Khoroshilov A. A. Computer linguistics and perspective information technologies. Moscow: Russian world, 2004. - 240 p.
4. Site Sketch Engine [Electronic resource]. - Mode of access: https://www.sketchengine.co.uk free (date accessed: 10.01.2018)
5. Kilgarriff A., Rychly P., Baisa V. Extrinsic Corpus Evaluation with a Collocation Dictionary Task [Electronic resource]. - Режим доступа: http://www.lrec-conf.org/proceedings/lrec2014/pdf/52_Paper.pdf, свободный (date accessed: 25.05.2019)
6. Kirpichnikova A. scientists of KFU will cooperate with developers of the Sketch Engine program [Electronic resource]. - Mode of access: https://kpfu.ru/philology-culture/struktura-instituta/otdelenie..., free (date accessed: 20.02.2018)
7. Apresyan Y. D. Selected works, vol. The integrated description of language and systemic lexicography. - Moscow: School "Languages of Russian culture", 2005. - 379 p.
8. Apresjan V., Baiza V. RuSkELL: Online Language Learning Tool for Russian Language [Electronic resource]. - Mode of access: http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202016/euralex_2016_030_p292.pdf free (date accessed: 05.05.2019)
9. Apresjan V. Dictionary as an Instrument of Linguistic Research [Electronic resource]. - URL: http://euralex.org/wp-content/themes/euralex/ proceedings/Euralex%202016/ euralex_2016_020_p224.pdf, (date accessed: 14.05.2018)
10. Baiza V. Multilingual CPA: Linking Verb Patterns across Languages [Electronic resource]. - URL: http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202016/euralex_2016_044_p410.pdf, (date accessed: 18.05.2019)
11. Kilgarriff A., Rychly P., Baisa V. Extrinsic Corpus Evaluation with a Collocation Dictionary Task [Electronic resource]. - URL: http://www.lrec-conf.org/proceedings/lrec2014/pdf/52_Paper.pdf, (date accessed: 21.05.2019)

12. Theo J.D. Bothma, Rufus H. Gouws, Danie J. Prinsloo The Role of E-lexicography in the Confirmation of Lexicography as an Independent and Multidisciplinary Field [Electronic resource]. - URL: http://euralex.org/wp-content/themes/euralex/proceedings/ Euralex%202016/ euralex_2016_008_p109.pdf, (date accessed: 26.05.2019)
13. Varlamova, E.V., Naciscione, A., Tulusina, E.A. (2016) A study on the phenomenon of collocations: Methodology of teaching English and German collocations to Russian students // International Journal of Environmental and Science Education. 11(6), ijese.2016.116, pp. 1275-1284.
14. Noskova A.I. Development of lexical competence in learning Spanish as a foreign language // EDULEARN17, the 9th annual International Conference on Education and New Learning Technologies, July 3rd-5th, 2017. – Barcelona, Spain, 2017. – Pp: 331-335.
15. Tarasova F.K, Tarasov A.M., The phraseosemantic group of «family relations« in the system of proverbs and sayings with «food« component in languages with different structures//Analele Universitatii din Craiova - Seria Stiinte Filologice, Lingvistica. - 2018. - Vol.40, Is.1-2. - P.191-201.
16. Garipova Elza V, Sadykova Aida G., Tarasova Fanuza Kh, Functional Aspect Of Linguistic Identity In Russian Blogosphere//MODERN JOURNAL OF LANGUAGE TEACHING METHODS. - 2018. - Vol.8, Is.10. - P.378-383
17. Sazesh, A., & Siadat, S. A. The Relationship between Quantum Management and Organizational Agility in Ministry of Roads and Urban Development of Golestan Province, Iran. Dutch Journal of Finance and Management, (2018). 2(2), 51. https://doi.org/10.29333/djfm/5827
18. Piteira, M., Costa, C. J., & Aparicio, M. Computer Programming Learning: How to Apply Gamification on Online Courses?. Journal of Information Systems Engineering & Management, (2018). 3(2), 11.
19. Sulistyaningsih, D., Purnomo, & Aziz, A. Development of Learning Design for Mathematics Manipulatives Learning based on E-learning and Character Building. International Electronic Journal of Mathematics Education, (2019).14(1), 197-205. https://doi.org/10.29333/iejme/3996
20. Kozhabergenova, G. E., Taubaeva, S., Bulatbayeva, A. A., Kabakova, M. P., & Asanov, N. A. The Stewardship of School Counselor Education in Higher Educational Establishments. Opción, (2018). 34(85-2), 386-414.