

**METHODS FOR ADAPTING LEXICAL UNITS TO THE LEARNER'S LEVEL IN
LANGUAGE ACQUISITION (ZIPF-PARETO FRACTAL METHOD)***Saidova Kamola**PhD student at Tashkent State University of Uzbek Language and Literature**E-mail: saidovakamola14@gmail.com*

Abstract: This article examines methods for adapting lexical units to the learner's level in the process of language acquisition, in particular the Zipf-Pareto fractal method. The authors describe the basic principles of this method and analyze the ordering of language elements based on their frequency of use and efficiency. The article reveals the importance of mastering the most frequently used words first in increasing the efficiency of language learning using Zipf's law. It is also shown that in the process of learning based on the Pareto principle, a large part of the language understanding is covered through a small part of the lexical units. The fractal approach, on the other hand, provides an opportunity to optimize learning by taking into account the repetition and mutual similarities in the structure of the language. The article reasonably shows that these approaches together are important in adapting lexical units to the learner's level and increasing the efficiency of learning. As a result, the article explores the prospects for using the Zipf-Pareto fractal method to optimize the language learning process and save resources.

Keywords: language, lexical unit, linguistic competence, speech, psycholinguistic feature, educational materials, fractal analysis, construction.

INTRODUCTION

In native language education, the selection of a text that is appropriate for the student's age, psycholinguistic characteristics, and individual cognitive capabilities is considered a scientific and methodological problem. In school textbooks, the selection of text and vocabulary is carried out based on the teacher's experience based on a traditional approach. This leads to the acquisition of language not consciously, but based on imitation and mechanical memorization. A text that is not appropriate for the student's age and language preparation strains the student's attention, memory, and comprehension processes. An effective solution to the problem in language education is, first of all, to determine the scientific criteria for selecting language material. To do this, first of all, the following:

- 1) Develop criteria for sorting language units appropriate for the student's age and linguistic competence;
- 2) Identify linguistic and cognitive indicators that assess the complexity of the text;
- 3) Develop graded educational material (text graded according to the minimum and maximum lexical unit, grammatical construction, size, and level of complexity);

4) Based on the identified results, tasks such as creating theoretical and methodological guides for textbooks will be on the agenda.

LITERATURE REVIEW

In recent years, foreign and domestic linguistic and didactic research has considered the development of scientific criteria for adapting language material to the level of the learner as an important issue. In world linguistic and didactic research, a solution to this problem is being sought through methods for determining the quality of linguistic speech, the psycholinguistic characteristics of the learner, and the level of complexity of the text.

According to research, (Zakaria, A., Renandya, W. A., Aryadoust, V. A., 2023)¹, 98% lexical coverage, that is, 98% of the words in the text are familiar, is noted as the upper limit of reading comprehension. It is emphasized that to reach this level of text comprehension, the reader's lexical fund should be around 3000 words. It follows from this that the gradual expansion of the reader's lexical fund is the basis of language acquisition. This is called the "i+1" input hypothesis in linguistic research.

S.D. Krashen about this theory The Input Hypothesis: Issues and Implications (S. D. Krashen, 1985) in his work² provided information. According to the input hypothesis, in order to expand the student's lexical fund, educational material should be presented slightly above the student's level of knowledge. Here, "i" is the student's current level of language knowledge (interlanguage), "+1" is new language material that is slightly higher than the current level, but understandable. At the next stage, a new language unit should be presented in context, connecting it with previously learned units. This hypothesis is based on the theory of natural order, according to which language acquisition occurs in a clear, predetermined grammatical or lexical sequence. Scientist O. Saidahmedova (Saidahmedova O., 2022)³ determined how grammatical forms are gradually mastered in the Uzbek language based on the theory of natural order. According to the study, the suffixes -niki, -da, -lar, -(i)m, -ga, -di, -ni are graded in the following sequence. The results of the study can be used as a scientific and methodological basis for compiling a graded grammatical dictionary.

The frequent use of semantically complex words in the text reduces the level of understanding. In the study on the grading of lexical units (Laufer, B., Ravenhorst-Kalovski, G. C., 2010)⁴ based on the frequency of use of words, lexical boundaries such as high frequency (basic vocabulary - 70%), medium frequency (found in special or official texts - 15%), low frequency (abstract, difficult to understand words - 10%), very low frequency (off-list) (highly specialized terms - 5%) were identified. The results of the study make it possible to select lexical units for a level dictionary based on their quantity and cognitive characteristics.

Literature analysis shows that the level of understanding depends on the level of lexical, grammatical, and syntactic perception of the text. The content of the text as a syntactic system is understood depending on the level of comprehension of the lexical unit. This leads to the

¹ Zakaria, A., Renandya, W. A., Aryadoust, V. A. corpus study of language simplification and grammar in graded readers // LEARN Journal: Language Education and Acquisition Research Network. – 2023. – Vol. 16, No. 2. – P. 130–153.

² Krashen S. D. The Input Hypothesis: Issues and Implications. – London; New York: Longman, 1985. – viii, 120 p. – ISBN 978-0-582-55381-0.

³ Saidahmedova, O. K. Tabiiy tartib gipotezasining o'zbek tili uchun talqini: Filol. fan. bo'yicha falsafa doktori (PhD) dissertatsiyasi avtoreferati. – Toshkent: O'zR FA Til va adabiyot instituti, 2022. – 48 b.

⁴ Laufer, B., Ravenhorst-Kalovski, G. C. Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension // *Reading in a Foreign Language*. – 2010. – Vol. 22, No. 1. – P. 15–30. – URL: <https://nflrc.hawaii.edu/rfl/item/206>

need to level the educational material presented to the learner in language education in accordance with the age characteristics, language competence, and psycholinguistic preparation of the learner. This, in turn, puts on the agenda the development of criteria for sorting language units appropriate to the age and linguistic competence of the learner.

METHODS

In world linguistics, lexical units are ranked mainly according to their frequency.

RoadToGrammar.com/textanalysis/, [VocabProfiler \(nationsonlinetools.org\)](http://VocabProfiler(nationsonlinetools.org)), Lexile Analyzer the main algorithm of online corpora such as is aimed at identifying the lexical core of the text, that is, the most frequently used words. In this case, the position of the lexical unit in the text is determined, mainly if the corpus size is not less than 10,000 words. If the word is most frequently used:

If it is in the range of 0-1000 words, it is very easy (A1-A2);

If it is in the range of 1000 - 2000 words, it is used in everyday life (B1);

If it is in the range of 2000 - 3000 words, it is used in the press and in a wide circle (B2);

If it is in the range of 3000 - 5000 words, it is used in academic and scientific texts (C1);

If it is in the range of 5000 - 1000 words, it is used in special, technical, artistic and poetic texts (C2).

Although the existing electronic corpora in the Uzbek language cannot fully realize the possibility of determining the level of lexical units, it is possible to determine the level of words based on empirical analysis based on private research. In this case, one can rely on a method based on the Zipf-Pareto law.

Zipf's law is a law based on empirical analysis of the distribution of word frequency, according to which a word is inversely proportional to its order of occurrence in a text. That is, the most frequently used word has the highest frequency, the second most frequently occurs about half as often, the third least often, and so on.⁵ Zipf's law is of practical importance in determining the "core" of a vocabulary. Corpus studies show that the few thousand most frequently used words constitute the bulk of communication. For example, in English, the 2000 most frequently used words typically account for about 80% of any text.⁶ Therefore, in lexicography and lexicology, it is precisely these high-frequency units that are paid attention to when determining the basic vocabulary or lexical core. In scientific sources, lists of the most important words in general use are compiled on the basis of this approach, for example, the General Service List (2000 common words of the English language) compiled by West or the Oxford 3000 list, which are based on Zipf's law. Such lists represent the lexical core that defines the minimum requirements of the language. The 80/20 principle, also known as the Pareto law, states that in many processes, 80% of the results come from 20% of the causes. The Italian economist V. Pareto put the inequality in the distribution of wealth into a mathematical formula, and this law was called the Pareto distribution (power-law). The law was later recognized as a universal theoretical model and expresses the idea that resources or results are unevenly distributed in any system⁷. In linguistics, in the distribution of lexical units, Zipf's law provided statistical evidence of this imbalance. That is, it was found that very few words are

⁵ Linders, G. M., & Louwse, M. M. (2022). Zipf's law revisited: Spoken dialog, linguistic units, parameters, and the principle of least effort. *Psychonomic Bulletin & Review*, 30, 77–101.
<https://doi.org/10.3758/s13423-022-02142-9>

⁶ teachingenglishwithoxford.oup.com.

⁷ Koch R. *The 80/20 Principle: The Secret to Achieving More with Less* / Richard Koch. – London: Nicholas Brealey Publishing, 1997. – 278 p. – ISBN 978-1-85788-168-3.

repeated a lot in communication, and a large number of words are used rarely. Zipf's "Principle of Least Effort" proved the application of the Pareto principle to linguistics, social science, and statistical fields.

When ranking lexical units, one can base the theory on these two laws. In this case, a fractal analysis based on a combination of Zipf's and Pareto's laws is performed. According to the analysis, initially the frequency of words is determined based on Zipf's law, and a ranking is made by the number of repetitions. In the next stage, the top 20% of the Zipf list (A1) is allocated based on Pareto's law. The Zipf-Pareto fractal approach is applied to the remaining 80% of units. Accordingly, the remaining part (80%) is accepted as a new 100% within itself, and the Pareto analysis is re-applied to this part in the same 20/80 ratio. In this way, the systematic levels of lexical units are determined based on hierarchical stratification and their distribution according to the learner's competence is ensured.

RESULTS

The scientific and popular text "Bees" was analyzed based on the Zipf-Pareto fractal model. The analyzed text ("Bees") consisted of a total of 357 words, from which 97 recurring lexical units were extracted. The number of repetitions of each word in the text was calculated separately and its statistical position (ranking) in the text was determined.

For example: "bee" – 36 times, "hive" – 14 times, "poison" – 12 times...

Based on the Pareto principle, the most frequently used 20% (A1) words in the text were divided into levels (≈ 19 words). Another 20% of the remaining 80% were divided into levels (≈ 16 words). The remaining parts were also divided into layers such as B1, B2, C1, C2 based on the fractal model.

As a result, 97 words were ranked as follows:

Ranking formulas based on the Zipf-Pareto fractal model

Degree	Formula (20% of each part)	Number of words
A1	97×0.20	19
A2	$(97-19) \times 0.20$	16
B1	$(78-16) \times 0.20$	12
B2	$(62-12) \times 0.20$	10
C1	$(50-10) \times 0.20$	8
C2	all the rest	32

In this case, the words of the A1 and A2 levels form the main "lexical core" as the basis of the syntactic construction of the text \

". In the process of stratification, the words of the lower level (for example, C2) appeared as units indicating the contextual, stylistic characterization of the text. For example, stylistic and functionally colored words such as личинка, аммофила, хипча, пейкамак, ковак, разм саломок, шиббаламак, богот were detected at low frequency.

It seems that the Zipf-Pareto fractal method can be effective for the analytical ranking of lexical units, the formation of educational material, and the linguistic-statistical analysis of the text.

DISCUSSIONS

The essence of fractal analysis is based on infinite divisibility. Usually, 6 levels (A1, A2, B1, B2, C1, C2) are distinguished in language teaching. In Zipf-Pareto fractal analysis, the number of lexical units of the last (C2) level differs sharply from other levels. The lexical units identified at this level can be systematized according to the concept of “i+1” based on the theory of natural order. The essence of the concept is to present new language material that is higher (+1) than the student’s existing language level (“i” – interlanguage), but is able to be understood based on the context. At this stage, words identified on the basis of fractal analysis can be presented to increase the student’s lexical fund. So, low-frequency, complex semantic words (C2) can be re-leveled in fractal analysis as follows.

According to the results of the initial analysis, 32 lexical units were identified as the last (C2) level. These units were considered semantic, complex, low-frequency words.

C2 level vocabulary fractal analysis table

Level	Number of initial words	20% is separated	Number of selected words
C2.1	32	32×0.20	6
C2.2	26	26×0.20	5
C2.3	21	21×0.20	4
C2.4	17	17×0.20	3
C2.5	14	14×0.20	3
C2.6	11	11×0.20	2
C2.7	9	9×0.20	2
C2.8	7	7×0.20	1
C2.9	6	6×0.20	1
C2.10	5	5×0.20	1
C2.11	4	4×0.20	1

In this case, 20% of words from each stage can be extracted and included in the educational material as the next new lexical unit.

Words such as lichinka, ammosfera, khipcha, payqamoq, kovak, razm solomoq, shibbalamoq, bokot (Table 2), which are recommended as new units for the educational material and are higher than the student's linguistic competence, are recommended to be included in the dictionary as inactive lexical units due to their semantic complexity, stylistic, territorial, and contextual limitations.

CONCLUSION

The Zipf–Pareto fractal model is an effective tool for the systematic ranking of lexical units, assessment of text complexity, and formulation of educational material based on the competency principle. This method allows for stratification based on the frequency and semantic load of a lexical unit. The step-by-step structure determined on the basis of fractal analysis serves to sequentially present educational material in accordance with the “i+1” concept. The step-by-step mastery of educational material in the linguodidactic process stimulates natural language acquisition.

**REFERENCES:**

1. Zakaria, A., Renandya, W. A., Aryadoust, V. A corpus study of language simplification and grammar in graded readers // LEARN Journal: Language Education and Acquisition Research Network. – 2023. – Vol. 16, No. 2. – P. 130–153.
2. Krashen S. D. The Input Hypothesis: Issues and Implications. – London; New York: Longman, 1985. – viii, 120 p. – ISBN 978-0-582-55381-0.
3. Saidahmedova, O. K. Tabiiy tartib gipotezasining o‘zbek tili uchun talqini: Filol. fan. bo‘yicha falsafa doktori (PhD) dissertatsiyasi avtoreferati. – Toshkent: O‘zR FA Til va adabiyot instituti, 2022. – 48 b.
4. Laufer, B., Ravenhorst-Kalovski, G. C. Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension // Reading in a Foreign Language. – 2010. – Vol. 22, No. 1. – P. 15–30. – URL: <https://nflrc.hawaii.edu/rfl/item/206>
5. Linders, G. M., & Louwerse, M. M. (2022). Zipf's law revisited: Spoken dialog, linguistic units, parameters, and the principle of least effort. *Psychonomic Bulletin & Review*, 30, 77–101.
6. <https://doi.org/10.3758/s13423-022-02142-9>
teachingenglishwithoxford.oup.com.
7. Koch R. The 80/20 Principle: The Secret to Achieving More with Less / Richard Koch. – London: Nicholas Brealey Publishing, 1997. – 278 p. – ISBN 978-1-85788-168-3.