JOURNAL OF MULTIDISCIPLINARY SCIENCES AND INNOVATIONS
VOLUME 04, ISSUE 10
MOPNTHLY JOURNALS
ISSN NUMBER: 2751-4390
IMPACT FACTOR: 9,08

# MODELING THE STRUCTURE OF WORD AND SENTENCE FORMATION IN THE UZBEK LANGUAGE USING MATHEMATICAL-NUMERICAL METHODS, PARTICULARLY LINEAR ALGEBRA TECHNIQUES SUCH AS THE GAUSSIAN METHOD

**Uzakova Mamura Abdurayimovna**
Asia International University.
Lecturer at the Department of
General Technical Sciences

**Abstrakt.** This article pertains to the fields of formal language modeling in linguistics, computational linguistics, NLP, and structural linguistics, and provides a description of the application of numerical methods—specifically the Gaussian method—in word and sentence formation in the Uzbek language. This study focuses on modeling the structure of word and sentence formation in the Uzbek language using mathematical-numerical methods, with a particular emphasis on linear algebra techniques such as the Gaussian method. Uzbek, as an agglutinative Turkic language, exhibits rich morphological and syntactic patterns that lend themselves naturally to formal, quantitative modeling. Words are composed of morphemes—roots and affixes—whose combinations follow well-defined morphological rules, while sentence structures are governed by syntactic relationships among subjects, predicates, and objects.
**Keywords.** formal model, Gaus usuli, strukturaviy lingvistika, NLP.

## 1. Introduction

The study of word and sentence formation in natural languages has long been a central focus of both theoretical and applied linguistics. In recent decades, the rapid development of computational linguistics and natural language processing (NLP) has expanded the methodological toolkit available for analyzing linguistic structures, enabling researchers to incorporate mathematical, statistical, and algorithmic approaches into linguistic modeling. Among these, linear algebra techniques provide powerful means for formalizing and quantifying structural relationships within language [1]. Uzbek, as a Turkic and highly agglutinative language, presents rich morphological and syntactic patterns that are particularly well-suited for mathematical modeling. The productivity of affixation, the regularity of morphotactic rules, and the hierarchical organization of sentence structure offer fertile ground for constructing formal models that capture both combinatorial and transformational aspects of the language. This study explores the use of mathematical-numerical methods, specifically Gaussian methods from linear algebra, to model the structural processes underlying word and sentence formation in Uzbek [2-3]. By representing linguistic units and their interrelations in matrix form, it becomes possible to analyze morphological dependencies, detect structural consistencies, and formally describe the processes through which words and sentences are generated. Such an approach contributes to broader areas of linguistics, including formal language theory, computational linguistics, NLP, and structural linguistics, by offering a quantitative perspective on language structure. The aim of this research is to demonstrate how Gaussian methods can be adapted for linguistic modeling and to illustrate their effectiveness in capturing the systematic patterns of Uzbek morphology and syntax. This framework not only enhances theoretical understanding but also provides practical insights relevant to computational applications such as morphological analysis, parsing, and language technology development.

The idea of mathematizing linguistic structures. Word and sentence formation in the Uzbek language is based on rules (morphological and syntactic rules) [4-7]. To convert these into a mathematical model:

1. Sets of symbols (letters, phonemes, morphemes)
2. Combinations formed from them (root + affix, words, sentences)
3. Relationships and rules among them are represented as mathematical objects.

Linguistic units are typically encoded in the form of graphs, matrices, vectors, or systems.
Mathematical model for word formation.
Each morpheme (root or affix) can be represented as a vector.
For example: o‘qi = $v_1$ , -ma = $v_2$ , -chi = $v_3$ (read)
Word formation:
$v_{word}=v_1+v_2+v_3$
This is a “formal” sum, and when it is transferred into an actual mathematical set, it is modeled using operations such as vector addition and matrix multiplication.
The structure of a sentence (subject, predicate, object, etc.) is represented as a graph model consisting of nodes and connections.
For example, let us take the following sentence:
“O‘quvchi kitob o‘qiydi.” (The student reads a book.)
The structure is as follows:
$x_1$: Subject (o‘quvchi / the student)
$x_2$: Object (kitob / the book)
$x_3$: Predicate (o‘qiydi / reads)
These units are connected to each other through syntactic rules. This relationship can be represented in mathematics as a system of linear equations.

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = s_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = s_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = s_3 \end{cases}$$

Here $a_{ij}$ — syntactic compatibility coefficients.
The Gaussian method is a technique for solving systems of linear equations.
In linguistics, it can be applied in processes such as reverse calculation in morphological analysis. For example, given a word, identifying its root and affixes—this is a reverse problem. A matrix model is constructed for the word structure.

$$A \cdot X=B$$

**A** — the system of affixes (matrix of morphological rules)
**X** — the vector of morphemes to be determined
**B** — the encoded form of the given word

Using the Gaussian method:

$$X = A^{-1}B$$

Morphemes are identified. In syntactic analysis, the above method is also used to determine the connections between words [8-10]. Tasks such as finding the correct order of words in a sentence or correcting an incorrect sentence are considered optimization problems, which are also represented as linear systems. Using Gaussian elimination:

- the syntactic compatibility matrix is simplified,
- the connections that make the sentence “stable” are identified.

In the semantic vector space, minimal connections are isolated. Words are represented as vectors (e.g., Word2Vec or other embeddings). The meaning of a sentence is represented by linear

connections between words. This system of connections is analyzed using the Gaussian method, allowing the identification of redundant semantic elements and the main carriers of meaning.
Model word formation.
O‘qituvchi = o‘qi + t + uv + chi (teacher)
View matrica:

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} v_{o'qi} \\ v_t \\ v_{uv} \\ v_{chi} \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$$

If $s_1$, $s_2$, $s_3$ represent the encoded character system of the word “o‘qituvchi” (teacher), the morphemes can be identified using the Gaussian method.
The Gaussian method and other mathematical-numerical methods are used to model the processes of word and sentence formation in the Uzbek language in a formal, algorithmic manner[10-14].
They are suitable for the following tasks:
- Mathematization of morphological analysis.
- Representing sentence structure as a linear system.
- Modeling semantic connections
- Serving as a basis for developing NLP algorithms in the Uzbek language

Provided a complete, formula-based, step-by-step example of converting word formation in Uzbek into a mathematical model and analyzing it using the Gaussian method.
The example is divided into three parts:
1. Matrix model for word formation and solving it using the Gaussian method
2. Syntactic model of an Uzbek sentence and its analysis using the Gaussian method

Part 1: Linguo-mathematical model of word formation
As an example, let us take the word “kitobxonlar” (readers).
Word structure:
kitob (k o‘k) (book)
-xon (so‘z yasovchi qo‘shimcha)
-lar (ko‘plik qo‘shimchasi)
These three morphemes are represented as vectors:
v1=kitob,
v2=-xon,
v3=-larv
Word formation:
$v_{so'z}=v_1+v_2+v_3$

To encode this, each morpheme is assigned a 3-dimensional feature vector (this is just for modeling purposes):

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

The general form of the word:

$$v_{so'z} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

Therefore:
A · X=B

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

The purpose of this system: to identify the structure of a word (which morphemes are used)
Part 2: Solving using the Gaussian method. Equation:

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 0 & 2 \end{array}\right]$$

Step 1. Elimination using the first row. Subtract the first row from the third row.

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & -1 & 0 \end{array}\right]$$

Step 2. Using the second row, eliminate the “1” below. Row 3 = Row 3 – Row 2.

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -2 & -2 \end{array}\right]$$

Modeling word and sentence formation in the Uzbek language using mathematical-numerical methods, particularly linear algebra and the Gaussian method, is an effective approach in linguistics, computational linguistics, and NLP.

**Conclusion**

Words are decomposed into morphemes and represented through vectors and matrices, while sentence structures are depicted as graphs and linear systems with nodes and connections. Gaussian elimination is used to identify morphemes, analyze syntactic compatibility, and optimize semantic relationships. This methodology enables formal and algorithmic modeling of morphological analysis, syntactic structure, and semantic analysis in the Uzbek language, serving as a foundation for developing NLP algorithms.

**References:**

1. Jabborova, H.Q. Amaliy va matematik lingvistika. Samarqand, SamDChTI, 2016.
2. Nurmonov Abduhamid Struktur tilshunoslik: ildizlari va yo‘nalishlari. Toshkent: Ta’lim, 2009.
3. András Kornai Mathematical Linguistics Canada, 2021 y
4. Geoffrey K. Pullum & András Kornai Mathematical Linguistics UK, 2026 y
5. Matilde Marcolli — Mathematical Models of Generative Linguistics (lectures)
6. Isabella Senturia & Matilde Marcolli — The Algebraic Structure of Morphosyntax
7. M. Kracht — Mathematics of Language (Department of Linguistics - UCLA)
8. M. Heitmeier, T. Pylkkänen va b. — Modeling Morphology With Linear Discriminative Learning

9. L. Schwartz va b. — How to Encode Arbitrarily Complex Morphology in Word Embeddings
10. From morphology to syntax (ilmiy maqola) 15(02), 1177-1180.
11. K Liu va b. — Natural Language Processing methods and systems (metodologiya va nazariy jihatlar)
12. Z. Strakoš Numerical Linear Algebra and Some Problems in Statistics 2019
13. V. Simoncini Computational Methods for Linear Matrix Equations 2021
14. Scott Nelson Introduction to Mathematical Linguistics (kurs darslik) 2023