SPECIFIC ASPECTS OF PARAMETRIC STUDY OF LINGUISTIC UNITS

Bobotayev Mirjalol Ibodulla og'li
Researcher, Samarkand State University named after Sharof Rashidov

Resume *This article provides a brief overview of the theory of universal grammar and linguistic parameters proposed by Noam Chomsky in the context of the development of world linguistics in the first quarter of the 21st century. It also highlights, with numerous examples, the history of the parametric study of the lexicon of Uzbek dialects. Reflections are given on the fact that Mahmud Kashgari and his dictionary constitute a multifaceted and in-depth study of the lexicon and semantics of Turkic languages, outlining the laws of their development.*

Keywords *Noam Chomsky, universal grammar, parameter, Mahmud Kashgari,* Devonu lug'otit turk, *language, dialect.*

Аннотация *В данной статье представлена краткая информация о теории универсальной грамматики и языковых параметрах, предложенных Ноамом Хомским, в контексте развития мировой лингвистики в первой четверти XXI века. Также освещена история параметрического изучения лексики узбекских диалектов с использованием многочисленных примеров. Приводятся рассуждения о том, что труд Махмуда Кашгари и его словарь являются многогранным и глубоким исследованием лексики и семантики тюркских языков, определяющим закономерности их развития.*

Ключевые слова *Ноам Хомский, универсальная грамматика, параметр, Махмуд Кашгари, «Девону луготит турк», язык, диалект.*

In the development of world linguistics in the second half of the 20th century and the first quarter of the 21st century, universal grammar (UG) and its component, the theory of parameterization, have played a significant role. The universal grammar theory proposed by Noam Chomsky views linguistic parameters as a key aspect for explaining the variability of languages while assuming the existence of universal principles common to all languages. Several core linguistic parameters can be identified:

Word Order: In Noam Chomsky's universal grammar, the topic of word order is closely related to the theory of parameterization, modularity, minimalism, the principles of subject–predicate structure, mobility, and adaptability of language. Word order is an important feature of a language's syntactic structure, determining the arrangement of syntactic elements (subject, verb, object, etc.). Differences in word order formats account for cross-linguistic variation and serve as one of the main factors ensuring and defining a language's flexibility and adaptability. Word order plays an important role in explaining the syntactic structure of a language, particularly within the framework of UG and minimalism.

Similarly, parameterization theory explains differences in word order among languages through parameters associated with this domain. Each language has its own specific word order parameter. For example, Uzbek generally follows the SOV (Subject-Object-Verb) order: *"Ali kitobni o'qidi"* ("Ali read the book"), while English uses the SVO (Subject-Verb-Object) order: *"Ali reads the book"*. These parameters are governed within the general principles of UG.

SVO refers to the arrangement of syntactic elements in a sentence as follows:

- **Subject (S)**: the agent performing the action (e.g., *John*),
- **Verb (V)**: the predicate expressing the action or state (e.g., *eats*),
- **Object (O)**: the entity or person toward which the action is directed (e.g., *an apple*).

The SVO order is common in many languages, especially Indo-European ones (English, French, Russian). For example, *"John eats an apple"* clearly illustrates the fixed SVO structure in English. Let us note the following word order formats:

- **SVO** (subject–verb–object): e.g., English (*"John eats an apple"* – *"Jon olma yeydi"*).
- **SOV** (subject–object–verb): e.g., Japanese (*"John an apple eats"* – *"Jon olma yeydi"*).
- **VSO** (verb–subject–object): e.g., Classical Arabic (*"Eats John an apple"* – *"Jon olma yeydi"*).

Position of the Adjective (Modifier):

- After the noun: e.g., French (*"la maison rouge"* – "red house").
- Before the noun: e.g., English (*"the red house"* – "red house").

Level of Agglutination:

- High: e.g., Japanese and Turkish, where words are often formed by adding multiple morphemes (root + affixes).
- Low: e.g., English, where word formation is less agglutinative.

Presence or Absence of a Case System:

- With case system: e.g., Russian, where nouns change form according to their syntactic role in a sentence.
- Without case system: e.g., English, where word order determines syntactic roles.

Presence or Absence of Articles:

- With articles: e.g., English (*"a cat"* – "a cat" / *"the cat"* – "the cat").
- Without articles: e.g., Russian.

Within the universal principles outlined by N.Chomsky, applying these parameters helps explain how children can acquire their native language quickly and efficiently. While these parameters differ across languages, they are all based on common principles, making the process of language acquisition intuitive and natural for every child.

It should be noted that the parametric study of language is one of the important approaches in the field of Natural Language Processing (NLP).

According to researcher Pascal Janetzky, human language is a complex system. Through speaking (or writing), we convey to participants in the communication process not only individual words, but also intonation, humor, metaphors, and many other linguistic features. For artificial intelligence or computers, identifying such features is already difficult, and "understanding" them with a machine mind is even more challenging. Addressing such situations in the study of linguistic units requires relying on the following approaches:

Relying on the parametric model: In studying linguistic units, parametric models such as databases containing dictionaries or language corpora require research that incorporates numerous parameters to understand the complex structure and context of language. A parametric model ensures high accuracy and adaptability. Analyzing or processing large volumes of data is therefore an essential condition for analysis and drawing conclusions. Consequently, parametric study makes use of large-scale textual data, enabling precise modeling of a language's statistical features, its dialects, lexical layers, and the relationships between words and the sentences in which those words are used.

Contextual dependence: Contextual dependence refers to understanding the meaning of a linguistic unit (word, phrase, or sentence) in relation to the surrounding words, sentences, or even the broader text. This allows for a more precise interpretation of meaning and the provision of contextually appropriate responses. In linguistics, areas such as the "conceptosphere" and "semantic field" describe how linguistic units are linked to context and to other linguistic units through semantic-associative connections.

Flexibility and versatility: In parametric language study, the inherent flexibility and multifunctionality of language are considered important advantages. These features mean that parametric models particularly those based on in-depth and comprehensive study of linguistic units (for example, transformers such as BERT, GPT, or RoBERTa in computational linguistics) can adapt to different languages, tasks, and specialized domains, and can perform multiple functions. This adaptability allows such models to be applied across a wide range of areas, from translation to automated text and dialogue generation systems.

Self-improvement: Self-improvement refers to the process by which a model enhances its performance over time by absorbing new data, learning from errors, or adapting to new tasks. Parametric models achieve this through various methods, such as retraining, fine-tuning, or self-supervised learning. They can be continually improved by supplementing them with new data or reanalyzing existing data. For example, in the case of Uzbek dialectology, the dialect materials collected during its rapid development in the 1960s-1970s could be re-examined or used to create a database, thereby accelerating the processes of studying linguistic units through modern methods.

If we look at the history of the parametric study of the lexicon of Uzbek dialects, it is natural to first mention Mahmud al-Kashgari and his work *Devonu Lug'otit Turk* (*Compendium of the Turkic Dialects*). This represents the stage of studying the dialects of

the Old Turkic language period, covering the time from the 11th century to the end of the 13th century and the beginning of the 14th century.

Mahmud al-Kashgari's dictionary is a valuable source for the multifaceted and in-depth study of the lexicon and semantics of the Turkic languages, as well as for observing the regularities of their development. Al-Kashgari personally traveled to the regions inhabited by various Turkic tribes of his time, studying and analyzing their lives, customs, and traditions, describing their languages, and systematically examining the similarities and differences between Turkic dialects. Moreover, in the preface of the dictionary, al-Kashgari explicitly states that he conducted ethnolinguistic research:

"I arranged [the words] according to rules and principles… and limited myself to the bases [roots] for each tribe… In it (the work) I gathered examples from the poems they (the Turks) recite during ceremonial and daily festivals, as well as wise sayings created in sorrow and joy and passed down from generation to generation. I also compiled in it (the book) the names of the most commonly used things and famous expressions, so the book attained the highest qualities and utmost refinement."

Because of the aspects described above, Mahmud al-Kashgari's *Devonu Lug'otit Turk* is considered an exceptionally rare and highly important work on the lexicon of Turkic languages for its time, and this has been emphasized repeatedly. In particular, al-Kashgari himself, when speaking about the differences between languages and dialects, states:

"Changes in root words are rare. Variations in words occur in letters - in the substitution of certain letters or in their omission."

For instance, he points out that nouns and verbs beginning with *k* in some Oghuz and Kipchak dialects are pronounced with an initial *a* or *ch*. He illustrates this with examples such as *jelkin – musofir* (wanderer), *jilig' suv – iliq suv* (warm water), and *jinju – dur, marvarid* (pearl), showing their differences in Oghuz and Kipchak dialects.

He also identifies phonetic changes, such as the tendency in Oghuz, Kipchak, and Suvar dialects to replace an initial *m* with *b*, illustrated by examples like *men bardim – ben bardum* (I went) and *mo'n – bo'n* (soup).

Furthermore, as an encyclopedic scholar of Turkic languages and dialects, he highlights:

"I traveled for many years through the cities, villages, and pastures of the Turks, Turkmens, Oghuz, Chigils, Yagmas, and Kyrgyz, collecting their vocabularies and studying and identifying various linguistic features. I did this not because I did not know the language, but to detect every small difference in these languages… I paid such close attention to them that the languages of the Turks, Turkmens, Oghuz, Chigils, Yagmas, and Kyrgyz became completely ingrained in my mind. I organized them in every respect according to a solid structure."

The work also contains descriptions of changes and distinctive features in the languages of Turkic tribes of that era that can still be observed today. For example, al-Kashgari notes that certain words are pronounced with *a* by some Turkic tribes and with *e* by others – as in

men – man, *yangi – yengi*, *tagida – tegida*, *kema – kama*. He also records that the word *üy* (cow) was pronounced differently with broad and narrow labial vowels.

It is clear that these features are still preserved in modern Uzbek dialects, particularly in Tashkent and Fergana. This indicates that the differences between Tashkent and Fergana dialects existed even in al-Kashgari's time (though not in exactly the same form as today, having undergone some changes), and that these dialects were among the primary ones contributing to the formation of the modern standard Uzbek language.

REFERENCES

1. François Recanati. Contextual Dependence and Definite Descriptions Proceedings of the Aristotelian Society New Series, Vol. 87 (1986-1987), pp. 57-73.

2. Hawkins D.L. A Test for a Change Point in a Parametric Model Based on a Maximal Wald-Type Statistic. Sankhyā: The Indian Journal of Statistics, Series A (1961-2002). Vol. 49, No. 3 (Oct., 1987), pp. 368-376.

3. Huang, Audrey; Foster, Dylan J.; Rohatgi, Dhruv and at. Self-improvement in language models: the sharpening mechanism. The Thirteenth International Conference on Learning Representations ICLR 2025 Singapore Apr 24 2025 https://iclr.cc/Conferences/2025 (April, 2025).

4. Janetzky, Pascal. Natural Language Processing: From one-hot vectors to billion parameter models. Towards Data Science. The World's leading publication for data science, AI, and ML professionals > https://towardsdatascience.com/natural-language-processing-from-one-hot-vectors-to-billion-parameter-models-302c7d9058c6/(January, 2025)

5. Кошғарий, Маҳмуд. Девону луғотит турк. I. – Тошкент: Фан, 1960. – Б. 35-36, 41.

6. Кононов А.Н. Махмуд Кашгарский и его "Дивану лугат ат-турк". – Сов. тюркология, 1972, – № 1. – С. 5-21;

7. Svenonius, Peter. Word Order Universals and Their Relationship to Structure. Annual Review of Linguistics, 11, (2025). pp. 137-162.

8. Орехова Е.С. Язык в аспект лингвистических штудий Н. Хомского: опыт аналитики>https://cyberleninka.ru/ article/n/yazyk-v-aspekte-lingvisticheskih-shtudiy-n-homskogo-opyt-analitiki (November, 2024)

9. Roger E.,L. Rodrigues De Almeida, Loevenbruck H., Perrone-Bertolotti M., Cousin E. and at. Unraveling the functional attributes of the language connectome: crucial subnetworks, flexibility and variability >NeuroImage Volume 263, November 2022, 119672 -https: //www.sciencedirect.com/science/article/pii/S1053811922007935 (April, 2025).

10. Равшанов М. Идеографические описание в «Девону лугатит турк» // Материалы международной конференции, посвященной 930-летию «Девону лугатит турк». – Самарканд: СамГУ, 2002. – С. 46-47.

11. Тенишев Э.Р. Тюркская историческая диалектология и Махмуд Кашгарский // Советская тюркология, 1973. –№ 6. – С. 54-61