LINGUO MATHEMATICAL MODELS OF SENTENCES BY TYPES OF UZBEK LANGUAGE FOR ELECTRONIC TRANSLATION

Nazirova E. Sh.,
Tashkent University of Information Technologies
named after Muhammad al-Kharazmiy
Uzakova M. A.
Tashkent University of Information Technologies
named after Muhammad al-Kharazmiy
100084, Uzbekistan, Tashkent, Amir temur street, 108
marynmm2622321@gmail.ru

Abstract*. In this paper are given linguo mathematical models of sentences by types of Uzbek language for the electronic translation. Electronic translation is developed as a multilingual system based on the technology of a simulated computer translator designed specifically for machine translation. Linguo mathematical models are described using an extensible input language designed for mathematical modeling of natural languages.*

Keywords: *linguo mathematical model, Uzbek, natural language, computer translator, multilingual situation, machine translation.*

Introduction. Each natural language (NL) is a complex system consisting of mathematically unstructured and non-formalized components. However, the studies conducted on NL show that the lack of structure and formalization of NL can be brought to a structured and formalized form using a linear methodology - identifying the composition of a word and constructing logical-linguistic (semantic) models by types of words and sentences, then constructing mathematical models using the input language [1]. This methodology can be defined as the degree of formalization of a language. The degree of formalization in turn determines the degree of formalization of the semantics of NL and the accuracy of the algorithm. A superficial understanding of the degree of formalization of NL, that a formalized language is an abstract construction with a simple logical structure, completely detached from the content, leads to low technology of machine translation [2]. Formalization allows us to identify various parts of NL, study the dynamics of their connections and mainly provides an opportunity to describe the semantic structure. All these qualities are very important when the common core of the system is used, i.e. when over all the languages included in the given translation environment [3]

Linguistic-mathematical models of sentences in the Uzbek language are formalized representations of sentences using linguistic-mathematical and logical structures. These models are used to analyze syntax, semantics and other aspects of the language from the point of view of formal sciences, such as logic, graph theory and automata theory. Below are examples of linguo-mathematical models for sentences in the Uzbek language.

In Uzbek, after the word formation stage, the sentence formation stage is followed. At the sentence formation stage, we must first study the components necessary for sentence formation. The basis of sentence formation is the possessive and the participle. In this case, the possessive and the participle of the possessive clause are components that express the action in the sentence. From a linguistic point of view, the possessive and the participle are called the primary parts of the sentence [4]. In addition, there are also secondary parts of the sentence. In Figure-1 below, we can see a schematic representation of this information:

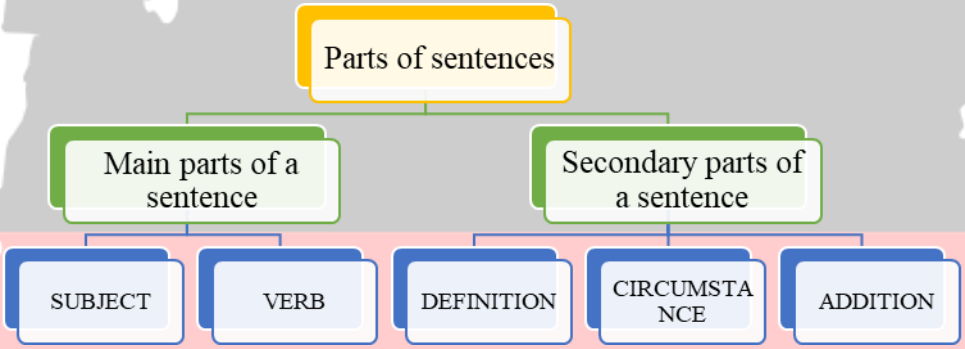


Figure-1. Linguistic classification of the parts sentences that form basis sentences in the Uzbek language

If move on to sentence formation ($S_{uzbsimple}$) after the word formation stage in Uzbek, will be convinced that the grammar of the Uzbek language is very rich.

In the Uzbek language, sentences are divided into simple sentences and compound sentences according to their structure.

According to the structure of a simple sentence, it is divided into:

1. Simple collection sentence,

2. Simple spread sentences (Figure 2). In such a classification, the presence or absence of secondary parts is taken into account [4].

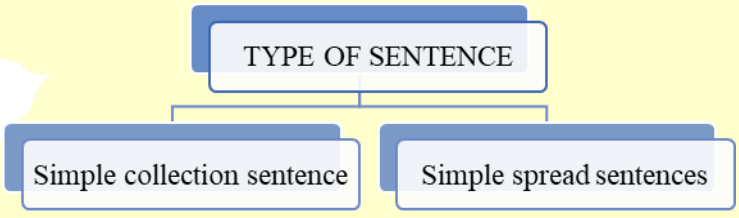


Figure-2. Linguistic classification sentences of the Uzbek language

A sentence made up of only the main parts (subject and predicate) is called a simple compound sentence. For examples of simple compound sentences as follows:

Axborot keltirildi. Kompyuter ishlayapti. .....

S V S V

To explain this example, in the Uzbek language, the possessive always comes first, and the participle always comes at the end of the sentence after the possessive. This order is considered strict.

A sentence composed of primary and secondary parts is called a simple continuous sentence. [4, 38-48 pages].

For examples:

$$\underbrace{Bu\ yilgi}_{D}\ \underbrace{seminar}_{S}\ \underbrace{yanvar\ oyining\ boshida}_{C}\ \underbrace{o'tdi}_{V}$$

$$\underbrace{Jismoniy}_{D}\ \underbrace{mashqlar}_{V}\ \underbrace{organizmni}_{D}\ \underbrace{har\ xil}_{D}\ \underbrace{kasalliklardan}_{C}\ \underbrace{saqlaydi}_{}$$

Scheme of sentence structure in the Uzbek language with the participation of primary and secondary parts

Based on the schemes given above, linguo-mathematical models of the formation of simple sentences in the Uzbek language were develop

$$S_{uzbsimple1}=E1(x_i,\ldots..,x_n)\cup K1(x_i,\ldots..,x_n) \qquad (1)$$

Here, $S_{uzbsimple1}$-sodda yig'iq gap, E1-subject, K1-verb, i and n – change steps each predicate in x set , $\cup$- connection symbol.

Note, $E1 \in x$, $x=\{$ changes from i to n $\}$ $i \geq 1$, $n \geq 1$

$K1 \in x$, $x=\{$ changes from i to n $\}$ $i \geq 1$, $n \geq 1$

For example:

Axborot keltirildi (kel+tir+il+di). (Information provided)

During the analysis of this sentence, the sentence is first divided into pieces (tokens) and the type is determined, each token is divided into morphemes, roots (R_{i1}) and suffixes based on morphological analysis. Here, the possessive root is not divided into words and morphemes, while the participle is divided into morphemes. The following linguo-mathematical model was developed

$$S_{uzbsimple1} = \begin{cases} E1(x_i, \ldots, x_n) \cup \\ K1(x_i, \ldots, x_n) \rightarrow R_{i1} \oplus F_{y1} \oplus F_{y1} \oplus F_{y1} \end{cases}$$

(2)

Here ,

$S_{uzbsimple1}$- simple collection sentence, E1-subject, K1-verb, i and n – change steps each predicate in x set , ∪- connection symbol, R_{i1}-root, F_{y1}-wordforming suffix,

Note, $E1 \in x$, $x=\{$ changes from i to n $\}$ $i \geq 1$, $n \geq 1$

$K1 \in x$, $x=\{$ changes from i to n $\}$ $i \geq 1$, $n \geq 1$

(2) As can be seen from the expression, only two parts are involved in simple compound sentences, namely the possessive and the participle.
In the next step, a lingvo-mathematical model for simple regular expressions was constructed and expressed as follows:

$$S_{uzbsimple2}=E1(x_i,\ldots..,x_n)\cup \downarrow O1(x_i,\ldots..,x_n)\cup\downarrow H1(x_i,\ldots..,x_n)\cup$$
$$\cup\downarrow D1(x_i,\ldots..,x_n)\cup K1(x_i,\ldots..,x_n) \quad (3)$$

Here, $S_{uzbsimple2}$-simple spread sentences, E1-subject, K1-verb, O1-addition, D1-defination, H1-circumtance, i and n – change steps each predicate in x set, ∪-connection sign, ↓- a sign that may or may not come.

Note, $E1 \in x$, $x=\{$ changes from i to n $\}$ $i \geq 1$, $n \geq 1$

$K1 \in x$, $x=\{$ changes from i to n $\}$ $i \geq 1$, $n \geq 1$

For example:

Bu yilgi seminar yanvar oyining boshida o'tdi.
(This year's seminar took place in early January)
Bu yil+gi seminar yanvar oy+i+ning bosh+i+da o't+di.
D S C V

This sentence contains a defination, a subject, a circumtance and a verb.

$$S_{simple2}=\begin{cases} D1(x_i,\ldots,x_n)\rightarrow R_{i1}_R_{i1}\oplus F_{y1}\cup \\ \cup E1(x_i,\ldots,x_n)\cup \\ H1(x_i,\ldots,x_n)\rightarrow R_{i1}_R_{i1}\oplus F_{y1}\oplus I_{f1}_R_{i1}\oplus F_{y1}\oplus I_{f1}\cup \\ K1(x_i,\ldots,x_n)\rightarrow R_{i1}\oplus F_{y1} \end{cases}$$

Here, $S_{uzbsimple2}$-simple spread sentences, E1-subject, K1-verb, O1-addition, D1-defination, H1-circumtance, i and n – change steps each predicate in x set, ∪-connection sign.

Since the languages of the Turkic language family belong to the class of agglutinative languages, the above linguo-mathematical models are also used for the Karakalpak and Kyrgyz languages, respectively.

Mathematical models of the Uzbek language are also presented in the works of Khakimov M., in which mathematical models were developed for the purposeful types of sentences in the Uzbek language.

The role of language corpora in the implementation of electronic translation is incomparable. Currently, the development of natural language corpora and a number of works on them are being carried out [5-7].

In conclusion, the study of simple sentences in the Uzbek language and the development of a linguo-mathematical model based on these studies in electronic translation are very complex issues. Of course, the accuracy and quality of electronic translation depends on the precise implementation of each of the above steps. Above, simple sentences in the Uzbek language were studied from a linguistic point of view, formal models of the formation of simple sentences were built. The structures of simple sentences were determined. All the work done can be applied to other natural languages belonging to the Turkic language family, such as Karakalpak and Kyrgyz, in addition to the Uzbek language.

References

1. Khakimov M.Kh. Rasshiryaemyy vhodnoy yazyk mathematicheskogo modelirovaniya estestvennogo yazyka dlya mnogoyazychnoy situatsii mashinnogo perevoda. UzMU journal. No. 1, 2009. S. 75-80.
2. Khakimov M.Kh. Mathematical model of the Uzbek language. UzMU journal. No. 3, 2010. S. 187-191.
3. Marchuk Yu.N. The model is perevoda. M., Izdatelsky center "Academy", 2010. 176 p.
4. G'. Abdurahmanov, A. Sulaymanov, Kh. Kholyorov, J. Omonturdiev's Modern uzbek literary language syntax Handbook for students of correspondence departments of philology faculties of higher educational institutions. TASHKENT, «УКИТУВЧИ» — 1979
5. B.Elov, Sh.Khamroyeva, Z.Khusainova. Tasks of NLP (natural language processing) and modern approaches. TerSU, philological research: language, literature, education. 2022, 5-
6. O.Abdullaeva. Linguistic annotation in language corpora and its principles. Materials of the international scientific and practical conference on the topic of the national corpus of the Uzbek language: problems and tasks. – Tashkent, May 31, 2022. – P. 130 136.
7. O.Abdullaeva. The issue of syntactic annotation of texts in the corpus of the Uzbek language. // Materials of the republican scientific and practical conference on the prospects of Uzbek applied philology. – Tashkent, October 26, 2022. – P. 122-126.