Electronic Dictionary and Translator of Bilingual Turkish Languages

E. Sh. Nazirova[1(✉)], Sh. B. Abidova[2], and Sh. Sh. Yuldasheva[1]

[1] Department of Multimedia Technologies, DSc, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan
shaxnoza23@mail.ru

[2] Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan

Abstract. The article presents the IDEF0 model, functional module for the implementation of electronic translation in Uzbek and Karakalpak languages, which belong to the family of Turkic languages, as well as software for electronic translation based on these models and modules.

Keywords: Electronic translator · Uzbek and Karakalpak language · IDEF0 model · Software

1 Introduction

Since the development of electronic translators today is based on the formalization of information about linguistic objects in applied linguistics, the successful formation and functioning of electronic translators will need to be studied in terms of the ability to translate linguistic information into machine language.

Analyzing the work done on machine translators [1], the idea of the TurkLang-7 project is to create a data set for Russian-Turkish sources for low-language pairs and a system for translating neural machines. In this paper, a semi-automatic process for creating parallel corpora has been developed.

[2, 10] In this article, a database of phraseological verbs, morphological lexicon, affixes in Uzbek and English, and their morphological and syntactic models are created. The agreement of simple sentence models for automatic translation has been determined. recommendations for combining paradigmatic approaches to the creation of electronic dictionary and software principles for a linguistic database have been developed.

Based on the foregoing, among the urgent tasks are the comparative study of the Turkic languages, the linguistic support of machine translation systems and the development of other special lexicographic resources.

J.-H. Kim et al. (Eds.): IHCI 2021, LNCS 13184, pp. 39–49, 2022.
https://doi.org/10.1007/978-3-030-98404-5_4

Accordingly, this article describes the principle of operation of a software package that provides electronic translation into Uzbek and Karakalpak languages, which belong to the family of Turkic languages.

Before implementing the electronic translator, we will build a mathematical model of sentence construction in Uzbek and Karakalpak languages.

Uzbek and Karakalpak languages belong to the group of agglutinative languages of the Turkic language family. That is, the formation of words in these languages is done by adding word-forming suffixes to the root word.

It is known that in Uzbek language there is a normal order of words. According to this order, the possessive is placed before the possessive, and the participle is placed after it, the definite article and the complement are placed before the participle, which means that the participle usually comes at the end of the sentence.

However, in order to express a particular methodological meaning, the usual order of words in a sentence can be changed, such a changed word order is also called inversion.

In scientific and formal styles, the usual order of words in a sentence is largely maintained. But in conversational, artistic, and journalistic styles, this order change, i.e., inversion order, can be applied.

The Karakalpak language uses the same sentence structure as the Uzbek language.

Mathematical models of simple and complex sentences in Uzbek and Karakalpak languages were developed on the basis of the scheme of simple and complex sentences formation in two languages.

$$G_{\text{сод}} = A(x_i, \dots, x_n) \cup B(x_i, \dots, x_n)$$
$$G_{\text{мур}} = A(x_i, \dots, x_n) \cup D(x_i, \dots, x_n) \cup S(x_i, \dots, x_n) \cup K(x_i, \dots, x_n) \cup B(x_i, \dots, x_n)$$

Here $G_{\text{сод}}$– simple sentence, $G_{\text{мур}}$– complex sentence, $A(x_i, \dots, x_n)$ – possessive, $D(x_i, \dots, x_n)$ – complement, $S(x_i, \dots, x_n)$ – case, $K(x_i, \dots, x_n)$ – determiner, $B(x_i, \dots, x_n)$ – cut.

Using this model, we can see 8 different ways of creating simple and complex sentences in Uzbek and Karakalpak:

For the Uzbek language

$$G_1 = \begin{cases} A(x_i, \dots, x_n) \\ B(x_i, \dots, x_n) \rightarrow f_{k(\text{ф})} \oplus P_{r(\text{от})} \oplus f_{k(\text{ф})} \oplus a_{j(\text{от})} \oplus P_{r(\text{ф})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_2 = \begin{cases} A(x_i, \dots, x_n) \\ K(x_i, \dots, x_n) \rightarrow a_{j(\text{от})} \oplus P_{r(\text{сиф})} \oplus P_{r(\text{ф})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\text{ф})} \oplus P_{r(\text{от})} \oplus f_{k(\text{ф})} \oplus a_{j(\text{от})} \oplus P_{r(\text{ф})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_3 = \begin{cases} A(x_i, \dots, x_n) \\ S(x_i, \dots, x_n) \rightarrow P_{r(\text{от})} \oplus P_{r(\text{рав})} \oplus P_{r(\text{ф})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\text{ф})} \oplus P_{r(\text{от})} \oplus f_{k(\text{ф})} \oplus a_{j(\text{от})} \oplus P_{r(\text{ф})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_4 = \begin{cases} A(x_i, \dots, x_n) \\ K(x_i, \dots, x_n) \rightarrow a_{j(\text{от})} \oplus P_{r(\text{сиф})} \oplus P_{r(\text{ф})} \\ D(x_i, \dots, x_n) \rightarrow f_{k(\text{рав})} \oplus P_{r(\text{от})} \oplus P_{r(\text{ф})} \oplus P_{r(\text{рав})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\text{ф})} \oplus P_{r(\text{от})} \oplus f_{k(\text{ф})} \oplus a_{j(\text{от})} \oplus P_{r(\text{ф})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_5 = \begin{cases} A(x_i, \dots, x_n) \\ D(x_i, \dots, x_n) \rightarrow f_{k(\text{рав})} \oplus P_{r(\text{от})} \oplus P_{r(\text{ф})} \oplus P_{r(\text{рав})} \\ S(x_i, \dots, x_n) \rightarrow P_{r(\text{от})} \oplus P_{r(\text{рав})} \oplus P_{r(\text{ф})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\text{ф})} \oplus P_{r(\text{от})} \oplus f_{k(\text{ф})} \oplus a_{j(\text{от})} \oplus P_{r(\text{ф})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_6 = \begin{cases} K(x_i, \dots, x_n) \rightarrow a_{j(\text{от})} \oplus P_{r(\text{сиф})} \oplus P_{r(\text{ф})} \\ A(x_i, \dots, x_n) \\ D(x_i, \dots, x_n) \rightarrow f_{k(\text{рав})} \oplus P_{r(\text{от})} \oplus P_{r(\text{ф})} \oplus P_{r(\text{рав})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\text{ф})} \oplus P_{r(\text{от})} \oplus f_{k(\text{ф})} \oplus a_{j(\text{от})} \oplus P_{r(\text{ф})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_7 = \begin{cases} D(x_i, \dots, x_n) \rightarrow f_{k(\text{рав})} \oplus P_{r(\text{от})} \oplus P_{r(\text{ф})} \oplus P_{r(\text{рав})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\text{ф})} \oplus P_{r(\text{от})} \oplus f_{k(\text{ф})} \oplus a_{j(\text{от})} \oplus P_{r(\text{ф})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_8 = \{B(x_i, \dots, x_n) \rightarrow f_{k(\text{ф})} \oplus P_{r(\text{от})} \oplus f_{k(\text{ф})} \oplus a_{j(\text{от})} \oplus P_{r(\text{ф})} \oplus f_{k(\text{ф})}$$

For the Karakalpak language

$$G_1 = \begin{cases} A(x_i, \ldots, x_n) \\ B(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus a_{j(\text{от})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_2 = \begin{cases} A(x_i, \ldots, x_n) \\ K(x_i, \ldots, x_n) \rightarrow a_{j(\text{от})} \oplus f_{k(\text{сиф})} \oplus f_{k(\text{ф})} \\ B(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus a_{j(\text{от})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_3 = \begin{cases} A(x_i, \ldots, x_n) \\ S(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus f_{k(\text{рав})} \oplus f_{k(\text{ф})} \\ B(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus a_{j(\text{от})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_4 = \begin{cases} A(x_i, \ldots, x_n) \\ K(x_i, \ldots, x_n) \rightarrow a_{j(\text{от})} \oplus f_{k(\text{сиф})} \oplus f_{k(\text{ф})} \\ D(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus f_{k(\text{ф})} \oplus f_{k(\text{рав})} \\ B(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus a_{j(\text{от})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_5 = \begin{cases} A(x_i, \ldots, x_n) \\ D(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus f_{k(\text{ф})} \oplus f_{k(\text{рав})} \\ S(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus f_{k(\text{рав})} \oplus f_{k(\text{ф})} \\ B(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus a_{j(\text{от})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_6 = \begin{cases} K(x_i, \ldots, x_n) \rightarrow a_{j(\text{от})} \oplus f_{k(\text{сиф})} \oplus f_{k(\text{ф})} \\ A(x_i, \ldots, x_n) \\ D(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus f_{k(\text{ф})} \oplus f_{k(\text{рав})} \\ B(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus a_{j(\text{от})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_7 = \begin{cases} D(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus f_{k(\text{ф})} \oplus f_{k(\text{рав})} \\ B(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus a_{j(\text{от})} \oplus f_{k(\text{ф})} \end{cases}$$

$$G_8 = \{B(x_i, \ldots, x_n) \rightarrow f_{k(\text{от})} \oplus a_{j(\text{от})} \oplus f_{k(\text{ф})}$$

The process of translation in electronic translation is carried out on the basis of the following sequence:

- in the first place, each word in a given sentence is singled out;
- the anterior and posterior suffixes and stems of each word are checked and separated;
- the extracted pieces are first checked conditionally, then searched from the database;
- The appropriate translation of the allocated piece in the database is taken and the piece is replaced.

Thus, all the pieces are combined and the result is achieved (Fig. 1).

First, we check and prefix each element of the selected array.

If word have a prefix, it is assimilated as an element of the prefix [j] array. Then the suffix is checked and separated in the same way.

If a suffix exists, it is assimilated as an element of the suffix [j] array. In some cases, a suffix can be multiple words. The rest remains in place as an element of the array.

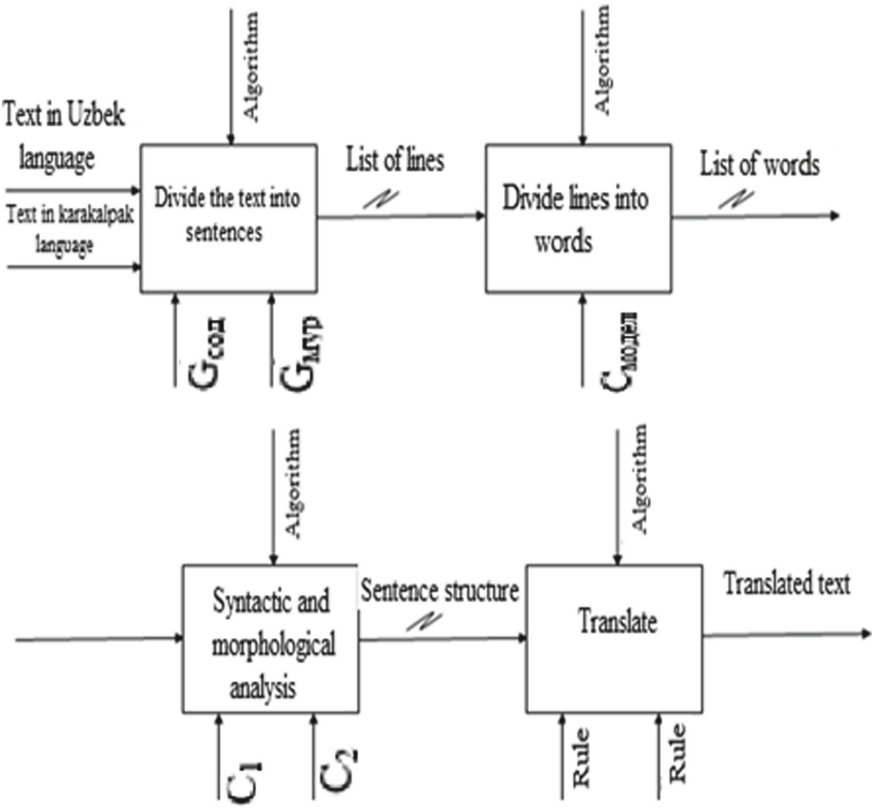

Fig. 1. IDEF0 drawing of electronic translator

The extracted prefix, stem, and suffix are checked from the database (Table 1-2-3), respectively. After verification, a new word is created and printed. If no word is found in the table, it is printed as translated text [3–6].

Based on the above models and algorithms, electronic translator software has been developed. The figure below shows the functional modules of software performance (Fig. 2).

The developed interface of the software "Bilingual electronic translator" has the following appearance.

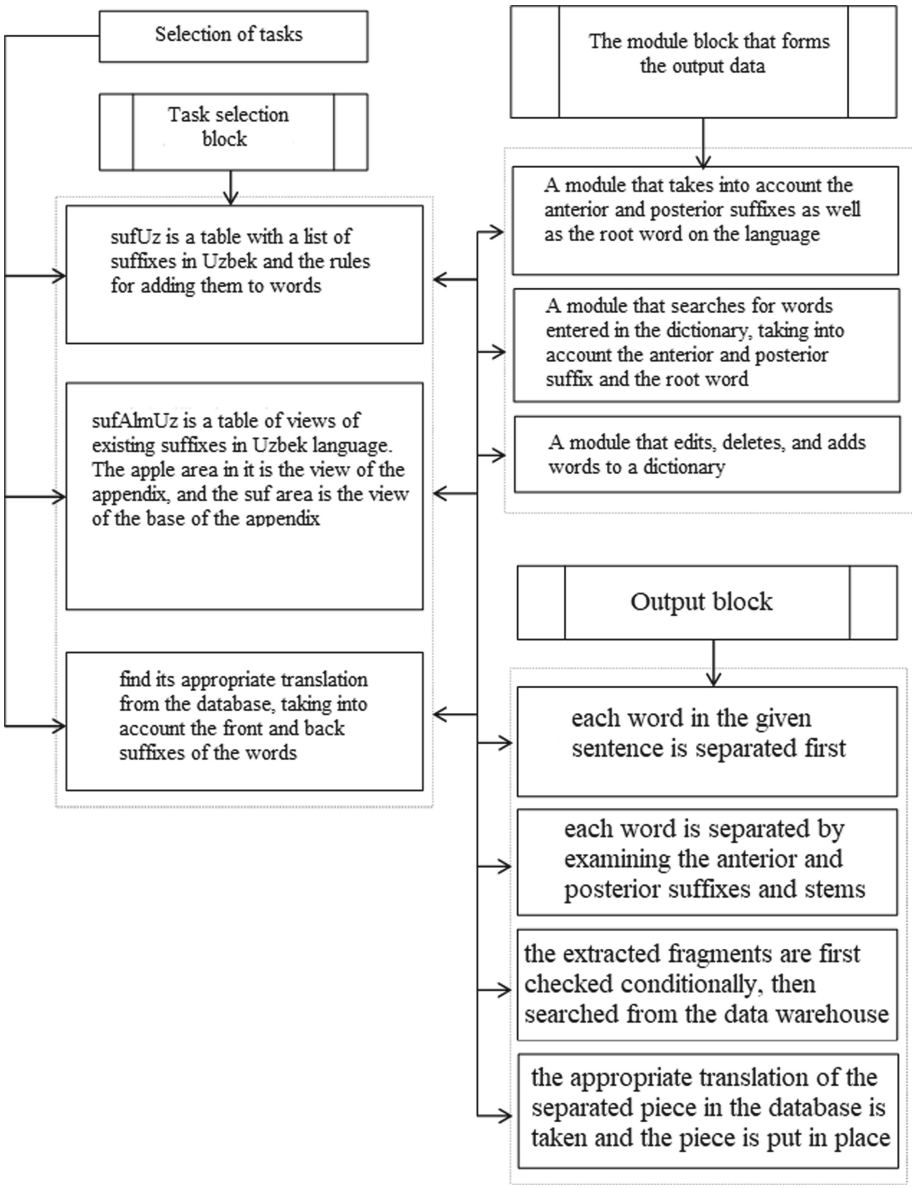

Fig. 2. Functional diagram of software performance

Figure 3 shows a view of the homepage of the program.

The main page consists of the main part of the program performing electronic translation, a translation window, an electronic dictionary, an annotated dictionary, a link, adding a new word, a button to download a file and copying sections.

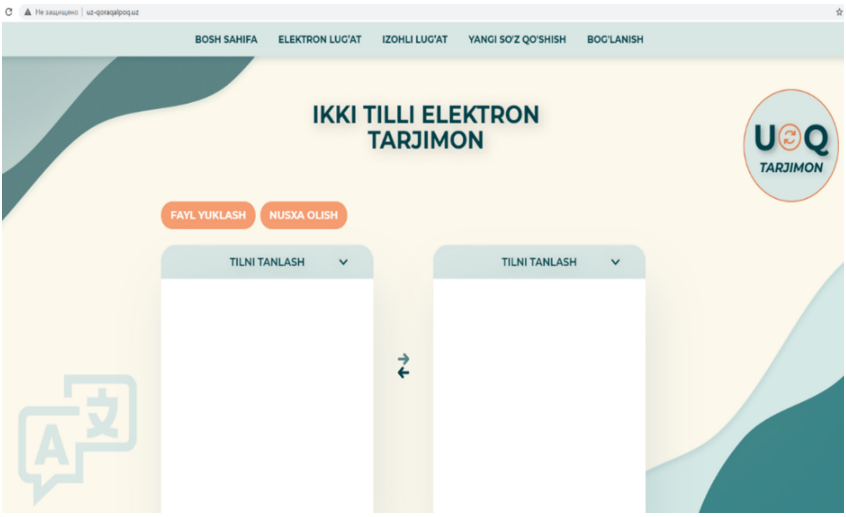

Fig. 3. Home page

In Fig. 4, in the language selection section, the menu for selecting the Uzbek-Karakalpak (Karakalpak-Uzbek) language is shown. The user can translate by selecting the desired language (Fig. 5).

Enter the word in Uzbek and Karakalpak. The meaning of the entered word is indicated in the section "Word Meaning" (Fig. 6).

In the section of the annotated dictionary, the word in Uzbek is entered in the "Enter a word" field, and in the section "Meaning of a word" it is given with comments on word groups.

In the "Add word" section you can add new words to the database in Uzbek and Karakalpak languages. In this case, the word is entered in the "Enter a new word" section, and in the "Select language" section, the language in which the word is entered is selected, and the "Add" button is used to add a new word to the database (Figs. 7 and 8).

Fig. 4. Language selection department when performing electronic translation

In the links section, the user can enter the first name, last name and postal address and leave a message to the developer in the “Message” section.

In the file upload section, the user can upload a.doc file and translate the full text.

In the “Copy” section, the user can make a copy of the translated text and paste it into the desired file [7–9].

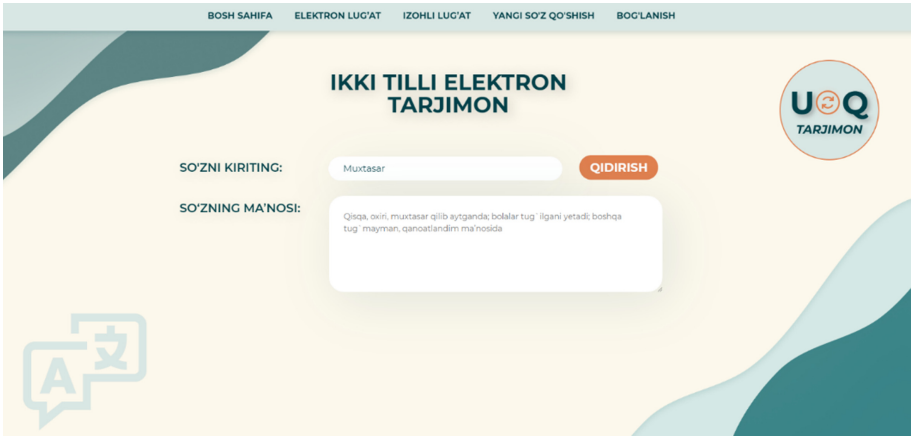

Fig. 5. Electronic dictionary department

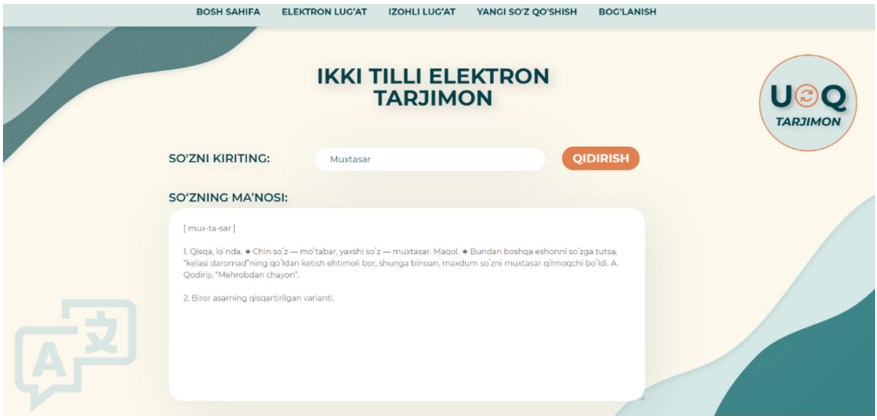

Fig. 6. Annotated dictionary section

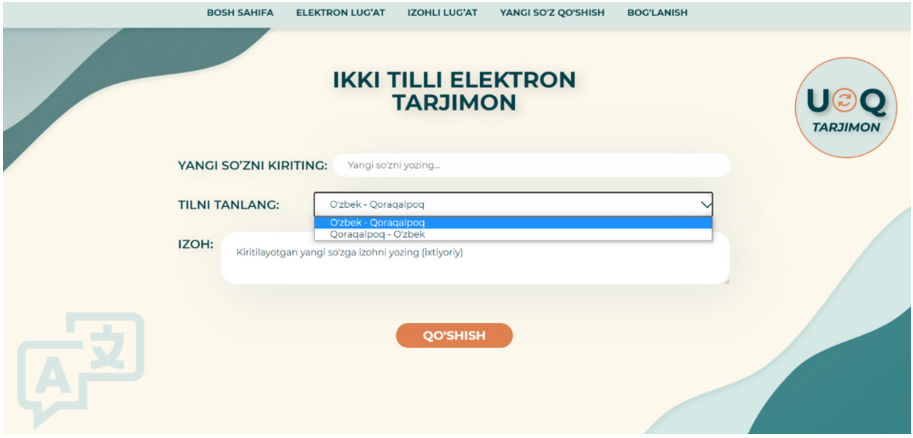

Fig. 7. Add a new word section

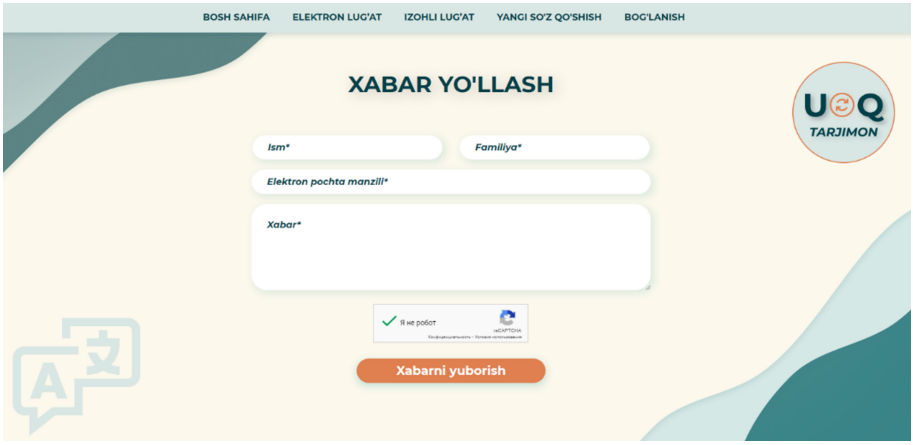

Fig. 8. Communications department

Conclusion. In conclusion, it can be said that the electronic translator was implemented using the mentioned IDEF0 model and the functional module of the software. This model and functional module can be used in languages belonging to the family of Turkic languages and agglutinative languages.

References

1. Khusainov, A., Suleymanov, D., Gilmullin, R., Minsafina, A., Kubedinova, L., Abdurakhmonova, N.: First Results of the TurkLang-7 project: creating Russian-Turkic parallel corpora and MT systems, pp. 90–101 (2020)
2. Abdurakhmonova, N.Z., Teshabayev, A.G.: Modeling IT of discourse analysis and the ussues machine translation, pp. 101–107 (2019)
3. Абидова, Ш.Б.: Электрон таржимон яратишда ўзбек тилининг морфологик таҳлилини амалга оширишнинг инфологик модели. Мухаммад ал-Хоразмий авлодлари **3**(13), 131–136 (2020). (05.00.00; №10)
4. Нуралиев, Ф.М., Абидова, Ш.Б.: Икки тилнинг семантик таҳлили. ТАТУ хабарлари журнали. **3**(40), 113–117 (2017). (05.00.00; №31)
5. Nazirova, E.Sh., Abidova, Sh.B.: Mathematical model and algorithm for calculating complex words in the Karakalpak language. Bull. TUIT Manage. Commun. Technol. **1**(44), 1–7 (2019). (30.07.2020; №283/7.1)
6. Nazirova, E., Abidova, S.: Mathematical model and algorithm for calculating complex words int Uzbek language. Solid State Technol. **64**(2), 4347–4359 (2021)
7. Абидова, Ш.Б.: Лингвистическая схема трансферной системы двуязычного перевода. In: Тезисы международной конференции “Актуальные проблемы прикладной математики и информационных технологий”. Тошкент, 2019 г., 14–15 ноябр, pp. 283–284 (2019)
8. Абидова, Ш.Б.: Туркий тиллар учун электрон луғат яратиш муаммолари. In: Иқтисодиётнинг реал тармоқларини инновацион ривожланишида ахборот-коммуникация технологияларининг аҳамияти. Республика илмий-техник анжумани. Тошкент 2017 й. 6–7 апрел, pp. 267–269 (2017)

9. Абидова, Ш.Б.: Особенности морфологии узбекского языка создании электронного узбекско-каракалпакского словаря-переводчика. In: «Актуальные проблемы математики и механики – CAWMA-2018» тезисы докладов республиканской научно-практической конференции с участием зарубежных женщин-ученых. Хива, 25–26 октябр 2018, pp. 4–5 (2018)
10. Абдурахманова, Н., Хакимов, М.: Логико-лингвистические модели слов и предложений английского языка для многоязычных ситуаций компьютернеого перевода. Компьютерная обработка тюркских языков. Латинизация письменности. 1-я Международная конференция. Астана, pp. 297–302 (2013)