DATABASE OF ENGLISH-UZBEK VERBS AND PHRASAL VERBS FOR MACHINE TRANSLATION

Abdurakhmonova Nilufar

Tashkent State University of the Uzbek Language
and Literature named after Alisher Navoi
abdurahmonova.1987@mail.ru

Abstract. The article deals with the issues about building database for machine translation. There are some difficulties for machine translation due to active positions in all styles of English and the complex characteristic nature of phrasal verbs. The article also discusses all models of English verbs and phrasal verbs to project of database.

Key words: machine translation, database, phrasal verbs, model, computer aided dictionaries, NLP (natural language processing).

In the process of globalization and integration the role of computer-assisted dictionaries has been increasing widely. Reason for this is that it links up the unification terms in different spheres of science and technology and e-learning, information retrieval and others.
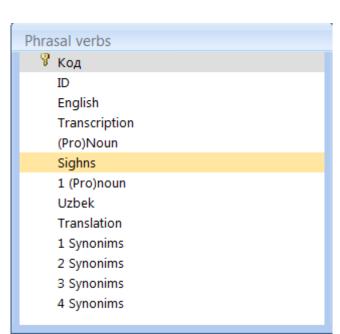
The basic units of language and speech should be included into machine dictionary. Such units may be collocations and phraseological units. Apart from those units main lexemes are within the vocabulary database. In general, natural language is considered to be the most complex system of thinking. So there are different forms of speech. First of all, we should clarify what is database for the dictionary. "The database system-1) systematized collection of data that can be accessed immediately, and manipulated by a data-processing system for a specific purpose; 2) (informal) any large store of information [1].
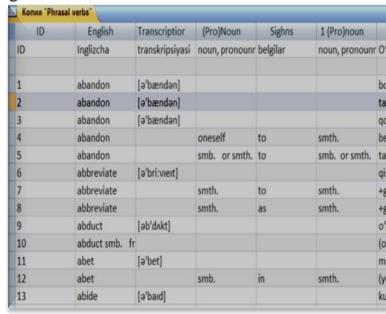
Given this amount of data, it is not surprising that a good number of linguistics have investigated software for managing data. Databases have long been standard repositories in phonetics and psycholinguistics research, but their findings, increasing further use not only in phonology, morphology, syntax, historical linguistics and dialectology but also in areas of applied linguistics such as lexicography and computer-assisted language learning [2]. We agree to this view that this is very crucial now how to build a linguistic database for machine translation as well. There are some database software for complex logical structures: FileMaker Pro, SQL,

Oracle database and for simple databases, software which is good at producing tables: Microsoft Excel, Access or Microsoft Word.

For building machine translation system from English into Uzbek there should be the computer-assisted dictionary in great size that is saved in database. Both English and Uzbek have a very huge database consisting all linguistic layers and they are very dissimilar. One verbal category in English is a phrasal verb. It is a real challenge for structural components of the sentence. "Phrasal verbs are considered to be a very important and frequently occurring feature of the English language. First of all, they are so common in every - day conversation, and non - native speakers who wish to sound natural when speaking this language need to learn their grammar in order to know how to produce them correctly. Secondly, the habit of inventing phrasal verbs has been the source of great enrichment of the language. By means of phrasal verbs it is described the greatest variety of human actions and relations" [3, p 16]. So verb-particle constructions in English are very complex to analyse and describe coherently in synchronic terms.

Database design is considered to be one of the stages of cycle information system and very crucial task is normalization of each unit in the process of database design. We have gathered more than 12 thousand phrasal verbs and their 3 thousand separate main verbs. But every unit placed in detached cell in a database and overall qualities of them is more than 75 thousand phrasal verbs and main verbs. This proposed translation the direction of English-Uzbek-English.

ID	English	Transcriptior	(Pro)Noun	Sighns	1 (Pro)noun	
ID	Inglizcha	transkripsiyasi	noun, pronounr	belgilar	noun, pronounr	O'
1	abandon	[ə'bændən]				bc
2	abandon	[ə'bændən]				ta
3	abandon	[ə'bændən]				qc
4	abandon		oneself	to	smth.	be
5	abandon		smb. or smth.	to	smb. or smth.	ta
6	abbreviate	[ə'bri:vıeıt]				qis
7	abbreviate		smth.	to	smth.	+g
8	abbreviate		smth.	as	smth.	+g
9	abduct	[əb'dʌkt]				o'
10	abduct smb. fr					(o
11	abet	[ə'bet]				m
12	abet		smb.	in	smth.	(y
13	abide	[ə'baɪd]				ku

Phrasal verbs
- 🔑 Код
- ID
- English
- Transcription
- (Pro)Noun
- Sighns
- 1 (Pro)noun
- Uzbek
- Translation
- 1 Synonims
- 2 Synonims
- 3 Synonims
- 4 Synonims

(Schema of Database)

In the data table with the transcription of main verb will give assistance while

using e-dictionaries. Also, nearly all meanings of the main verb with their synonyms on the same line are included in the database. It helps users to search for all synonyms of verbs not only the main meanings, but also secondary meanings of verbs as well as of that of phrasal verbs. Some separable or inseparable phrasal verbs are taken into account too. This database is formed according to following models of phrasal verbs (*V-verb; P-particles (preposition or adverb); ↓-possible or impossible; smth.-something, smb.-somebody*):

- **V + oneself +P+smth.** => align oneself with smb. or smth.
- **V+ oneself +P=>** arch (oneself) over
- **V + oneself +P+smb.** => attach oneself to *smb*
- **V + P+ smb. or smth.+P smth.** =>arrange with *smb.* about *smth.*
- **V+ smb. or smth. + P+ smb. or smth.** => associate *smb.* or *smth.* with *smb.* or *smth.*
- **V+ smth. +P+ smth.**=> balance *smth.* against *smth.*
- **V+ smth. +P=>** bail *smth.* out
- **V+ smb. +P+ smth.**=> astound *smb.* with *smth.*
- **V+smb. +P+ smb. or smth.**=> bias *smb.* against *smb.* or *smth.*
- **V+smb. +P=>** beat *smb.* up
- **V +P+smb.=>** bet with *smb.*
- **V +P+smb. or smth.=>** attend to *smb.* or *smth.*
- **V +P+ smth.+P+ smth.=>**
- **V +P+ smb.+P+ smb.or smth.=>** book *smb.* through (to some place)
- **V+P+P=>** be in for
- **V+P+P+it=>** be in for it
- **V+P+P+smth.=>** be off for smth.
- **V+P+P+ smb.or smth.=>** bound up with *smb.* or *smth.*
- **V+P+↓P+↓smth**=>bear up (against *smth.*)

  And we also included some symbols to clarify each others. They are followings:
- P1, P2, … , PN –N meanings of main verbs;
- N1, N2, …, Nk –k forms of phrasal verbs;
- B1, B2, …, Bl –l meanings of phrasal verbs;
- P1_, P2_, … , PN_ - synonyms of main verbs;
- B1_, B2_, …, Bl_ – synonyms of meanings of phrasal verbs;
- P – this means it doesn't belong to any main verb.

According to Yorick Wilks "While we agree that is unlikely that the information in machine-readable dictionaries is sufficient for this grand database of facts that will support NLP as a whole, we are optimistic about making use of the information they do provide support the creation of lexical entries for specific natural

language processing systems" [4, 139].

Subsequent our studies will deal with other English parts of speech. In conclusion, it may be said that this database of English-Uzbek verbs and phrasal verbs will be used for e-dictionaries and machine translation. And we can say if such kinds of machine readable dictionaries are built, then the subject will be improved in linguistic databases for any types of translation processes.

REFERENCES

1. Collins English Dictionary. 8th Edition first published in 2006 (e-version).
2. John Nerbonne. Linguistic Databases. 1996 http: // odur. let. rug. nl~nerbonne papersintro-db.pdf.
3. Andreea-Rosalia Olteanu. A Holistic Approach to Phrasal Verbs, Editura Sfântul Ierarh Nicolae 2012, P 16.
4. Yorick Wilks. Machine Translation. Its Scope and Limits. Spring Science+Business Media LLC. 2009 UK, P 139.