AMERICAN JOURNALS
Open Access, Peer Reviewed Journals

American Journal of Technology and Applied Sciences
ISSN (E): 2832-1766 Volume 31, December - 2024

WEIGHTS OF CONSISTENT AND INCONSISTENT AFFECTIONS IN ENGLISH AND UZBEKISTAN AND THE SIGNIFICANCE OF THE DIFFERENCE IN WEIGHTS IN MACHINE TRANSLATION

Khakimov M. X.1.,
Bekova V. G. 1.,
Omonov A. A. 2
1National University of Uzbekistan, 4, University Str., Tashkent, 100174 Uzbekistan.
2Tashkent Pediatric Medical Institute, Department of "Biophysics, Medical Informatics"
E-mail: muftah@mail.ru https://orcid.org/0000-0001-6937-800X
E-mail: vazira.bekova16@gmail.com. https://orcid.org/0009000865181312
mr.piko.service@gmail.com

ABSTRACT

This article analyzes the word-forming suffixes of the noun, adjective, verb and adverb word classes in English and Uzbek. The weights of noun and adjective suffixes in English are determined by numerical coefficients and the weights of noun and adjective suffixes corresponding to Uzbek are determined by numerical coefficients, the weights of the corresponding suffixes in the two languages are compared, and the affixes of verb and adverb word classes in the two languages and their grammatical differences are discussed with examples.

The comparative analysis of word groups in English and Uzbek is explained in detail in the comments that word affixes play a significant role in the level of abstraction and quality of text translation performed in two languages in computer translation.

KEYWORDS

English, Uzbek, machine translation, affixation, suffix, prefix, noun phrase, adjective phrase, verb phrase, adverb phrase, numerical coefficient, weight, difference in weights.

Introduction

The founders of machine translation were representatives of the field of cybernetics and mathematics, and later linguists also began to actively participate in this work [1]. Thus, the ideas of machine translation have gained great importance in the development of theoretical and applied linguistics around the world. In parallel with this direction, the theory of formal grammar emerged, and attention was paid to creating a model of language and its individual aspects. Today, technologies are being developed and optimized in the field of translation. In the context of globalization, taking into account the trends emerging with the digitization of the entire translation process, translation studies remain one of the priority areas. There are several advantages to using machine translation using computer

science and artificial intelligence. In particular, the input language (EVX) and the output language (EVIX) [2] can be translated instantly by machine translation. The main problem in machine translation is that the content of the result does not correspond one hundred percent. The first programs for machine translation were developed about 70 years ago. The first programs were based on a primitive method of translation - word-for-word translation of texts. According to it, it is limited to determining the equivalent of a unit in the dictionary of one language in another language. However, throughout the history of automatic translation, it became clear that this method is ineffective. Therefore, today it is considered an urgent task to create programs that can deeply analyze the meaning of the source language text, identify its subjectivity, emotionality and stylistic characteristics [3]. To date, a number of theories and ideas have emerged in the field of machine translation. One of these ideas is the creation of a complete database of words and suffixes for word formation by the affixation method of word-forming suffixes added to the front and end of words in the EVX – EVIX languages [2]. The goal of the computer is to teach the translator to think consciously, as if he had translated the translation text in a meaningful way, that is, very close to EVX - EVIX. In this way, we can create high-quality machine translation. Affixation is one of the methods of word formation using affixes in all languages. Affixes are a type of bound morpheme, and in order for affixes to have a meaning, they must be combined with the main word. Affixation is divided into affixes and prefixes. Additional morphemes added before the root of the word are called prefixes, and additional morphemes added after the root of the word are called affixes. Affixes in English and Uzbek form different parts of speech. There are a lot of noun-forming affixes in these languages. In this article, affixes and prefixes that correspond to noun and adjective-forming affixes in English and Uzbek were found. The weights of these corresponding affixes and prefixes in English and Uzbek were calculated. The differences in the weights identified in English and Uzbek were identified. In addition, it was noted that the affixes of the verb and adverb word classes in the two languages were not consistent in terms of grammatical structure.

English stands out as one of the most commonly spoken languages worldwide. For individuals who are not native English speakers, the investment required to learn a new language is substantial, and attaining fluent communication demands challenges [4]. In formal settings and situations where extensive information exchange is necessary, relying solely on human translation is no longer sufficient to meet the increasing demand [5]. For instance, simultaneous interpretation demands intense concentration from interpreters, limiting their ability to work for extended periods. Consequently, there is a pressing need for translation tools to supplement human efforts [6]. During the communication, lengthy English sentences are conventional [7]. However, the grammatical structure of English differs from that of other languages. When machine translation algorithms translate extended English sentences using a direct one-to-one mapping approach, it often results in grammatically incorrect translations and, in severe cases, translation errors [8]. Fortunately, the advent of intelligent algorithms offers a promising avenue for enhancing English machine translation efficiency and quality. Various studies have explored methods to enhance the efficiency and accuracy of English machine translation [9]. Lin et al. Proposed a neural machine translation approach based on an innovative beam search evaluation function, which notably enhanced English-to-Chinese translation quality. Luong and Manning introduced an end-to-end neural machine translation technique leveraging minimal-risk training, demonstrating its effectiveness through experiments [10]. Choi et al. Contextualized word

embedding vectors using a nonlinear bag-of-words representation of the source sentence. Experimental results showed significant improvements in translation quality using their proposed contextualization and symbolization methods. This study outlines the fundamental framework of intelligent machine translation algorithms and optimizes LSTM-based intelligent machine translation algorithms by incorporating a long sentence segmentation module and a reordering module. Utilizing mathematical models in contemporary high-quality machine translation emphasizes the concept of employing them within translation processes.

If the computer can be programmed to understand this mathematical model in a language that it comprehends, then it is possible to achieve appropriate translation into the target language. At present, machine translation has been advancing in providing us with useful translation in many languages and various fields. If machine translation is used in scientific fields, for example, one of the inevitable issues is to encounter ambiguous words and content that loses its original meaning, despite some improvements in recent years. The research that has been conducted by us have made significant progress in achieving more meaningful machine translation by addressing some of these issues. However, it is essential to admit the challenges that have not been sufficiently resolved in machine translation. The foundation of machine translation is considered computer linguistics, but every natural language (NL) is a complex system composed of mathematically unstructured and unformulated components. In accordance with the scientific work by Y.N. Marchuka [11], various concepts of machine translation models by mathematicians and linguists are illustrated in detail. He also describes one of the machine translation models based on translation writing. According to the author's opinion, "this model replicates the actions of a translator working in a specific language pair." The translator systematically moves from phrase to phrase, forms some approximate idea in his mind about the content of the text, then compares this idea with linguistic tools by selecting translation equivalents, and looks for three types of translation compatibility: equivalent, variant, and transformation. The task of any machine translation system is to decode the meaning of the natural language (NL) input text and present it in a formally understandable manner to the translation system. Indeed, the system needs to translate this text and deliver the result to the user, it is necessary to provide semantic meaning of the text. In order to achieve this, the study of Various NL grammatical constructions that allow formalization and identification is crucial and also confirmed by Novikov I.A. [12]. Morphological analysis is the initial stages of machine translation and is divided into tokenization -> lemming -> stemming [13]. The tokenizer checks words from left to right. As a result, it is determined which word group this or that word belongs to. Since our translation program is conducted in the English-Uzbek direction, initially all the unique features of the English language are compared with the features of the Uzbek language.

METHODS

The problems with the level of study of the topic are that Google Translate is used by a lot of people today. We can hear the same opinion from every user, that the quality of the translation does not meet the requirements. When we study the solution to this problem in two languages, we can clearly state one idea, namely, that in order to perform high-quality computer translation, we must first study the grammar of the two languages according to word classes. Because in computer translation, the incoming EVX text is first translated into words, words into sentences, and sentences are translated

into the text, resulting in the output language EVIX text. In the process of implementing this process on the computer, it has been shown that if the word-forming affixes of the words in the text match each other in EVX, the content of the text resulting in EVIX may match, or vice versa, the content of the translation may be abstract in content. We have proposed that since these two languages belong to two families, the EVIX text should be derived by matching word-forming suffixes or by highlighting incompatible suffixes in red in computer translation. Through the comparisons and numerical coefficients presented below, we can determine in advance, through weights, how much the text given in EVX corresponds to EVIX in terms of content or vice versa.

Using the corresponding affixes and prefixes in English and Uzbek, the weights of these English and Uzbek languages are calculated as follows. (The weights of word classes in these two languages are strictly defined for weight coefficients in the article cited in [3] and these weights are considered immutable.) The suffixes, prefixes and words in the sentences participating in each sentence type are numbered according to word classes as follows. Since English and Uzbek belong to two families, the word structures of these two languages are different. Therefore, the division of words into word classes (and other languages) was arranged taking into account the use or quantity of word classes in order to correspond to the word classes in other languages):

The weight of the word classes in the optional natural language is defined as follows [3]:

- Words belonging to the noun class (C) – 0.1;
- Words belonging to the adjective class (P) – 0.2;
- Words belonging to the verb class (G) – 0.3;
- Words belonging to the adverb class (N) – 0.4;
- Words belonging to the pronoun class (M) – 0.5;
- Words belonging to the numeral class (F) – 0.6;
- Other word classes (U, D, Y, L) that are not independent word classes (i.e., auxiliary word classes) – 0.07.

RESULT AND ANALYSIS

On this basis, the numbering of word classes helps to calculate the weight of two languages. For example: In the languages being compared, the main function of suffixes is to form one part of a sentence from another, and their secondary function is to change the lexical meaning of this part. Depending on which part of the sentence they form, suffixes can be divided into different groups:

The following English noun-forming and adjective-forming suffixes were calculated and their corresponding weights in Uzbek were written as follows:

Suffix weight – V3;

EVX – input language; EVIX – output language;

C(S) – noun suffix; P(S) – adjective suffix.

EVX(V3) – EVIX(V3) – the difference in the sum of the weights of each suffix in each word in the input and output languages.

Noun-forming Suffixes in English (Noun-forming Suffixes C(S): -eer, -er, -ian, -ics, -or, -ler, -eur, -ist, -agogue, -ant, -let, -acy, -cy, -dom, ancy, -ance, -ese, -hood, -ity, -ite, -ship, -th, -ness, -ty, -y;

Noun-forming suffixes in the Uzbek language: C(A) -chi, -cha, -lik, -li, -siz, -no, -iy;

Table 1: Noun-forming suffixes that are compatible in both languages

№	Noun-forming suffix in Uzbek	V3	Noun-forming suffixes in English	V3	EVX(V3)-EVIX(V3)
1.	-chi	0.10130	-eer	0.10215	0.00085
2.			-er	0.10217	0.00087
3.			-ian	0.10222	0.00092
4.			-ics	0.10223	0.00093
5.			-or	0.10253	0.00123
6.			-ier	0.10226	0.00096
7.			-eur	0.10220	0.0009
8.			-ist	0.10230	0.001
9.			-agogue	0.10203	0.00073
10.			-ant	0.10207	0.00077

The number of noun-forming suffixes in English is 76. The number of suffixes in Uzbek is 94 [2], and the weights of the corresponding suffixes in these two languages and the difference between the weights are given. The Uzbek suffix “-chi” is matched by 10 suffixes in English, and the weights of these suffixes are taken according to the strictly defined place in the table for affixes in the two languages. We can see that the corresponding suffix weight in English is close to the suffix weight in Uzbek. This results in a meaningful translation in which the meaning of the sentences and texts translated with the corresponding suffixes is very close to each other.

Table 2: Noun-forming suffixes that are compatible in both languages

№	Noun-forming suffix in Uzbek	V3	Noun-forming suffix in English	V3	EVX(V3)-EVIX(V3)
1.	-cha	0.10166	-let	0.10238	0.00072

Table 2 above shows the Uzbek noun-forming suffix “-cha” corresponding to the English noun-forming suffix “-let.” The relative weights of these suffixes and the difference between the weights are given.

Table 3: Noun-forming suffixes that are compatible in both languages

№	Noun-forming suffix in Uzbek	V3	Noun-forming suffixes in English	V3	EVX(V3)-EVIX(V3)
1.	-lik	0.10119	-acy	0.10201	0.0008
2.			-cy	0.10210	0.00091
3.			-dom	0.10212	0.00093
4.			-ancy	0.10205	0.00086
5.			-ance	0.10216	0.00097
6.			-ese	0.10218	0.00099
7.			-hood	0.10221	0.00102
8.			-ite	0.10231	0.00112
9.			-ity	0.10233	0.00114
10			-ship	0.10260	0.00141
11.			-th	0.10266	0.00147
12.			-ness	0.10247	0.00128
13.			-ty	0.10272	0.00153
14.			-y	0.10275	0.00156

In Table 3, the Uzbek noun-forming suffix "-lik" was matched with 14 English noun-forming suffixes. The weights of the matched suffixes and the difference between the weights in the two languages were calculated.

The number of adjective-forming suffixes in English is 47 and in Uzbek it is 75 [2]. The corresponding suffixes of adjective-forming suffixes in the two languages correspond to the adjective-forming suffix "-li" in Uzbek and 5 adjective-forming suffixes in English. The adjective-forming suffix "-siz" in Uzbek corresponds to the adjective-forming suffix "-less" in English. We have listed the corresponding suffixes in Table 4 below, and the weight of these suffixes in Uzbek and English and the difference in the weights in the two languages were calculated.

- Adjective-forming suffixes (adjective-forming suffixes (P-S)): -ful, -ant, -ed, -ent, -y, -less, -ese;
- Adjective-forming suffixes in Uzbek: -li, -siz, -no, -iy;

Table 4: Corresponding adjective-forming suffixes in the two languages

№	Adjective suffixes in the Uzbek language	V3	Adjective suffixes in English	V3	EVX(V3)-EVIX(V3)
1.	-li	0.20145	-ful	0.20209	0.00064
2.			-ant	0.20237	0.00092
3.			-ed	0.20220	0.00075
4.			-ent	0.20232	0.00087
5.			-y	0.20247	0.00102
6.	-siz	0.20262	-less	0.20205	0.00057

- Adjective formative prefixes: -in, -il, -im, -ab, -al;
- In Uzbek, adjective formative prefix: no-

There are various ways to create new words in English. One of these ways is to use prefixes. A prefix is a type of affix that is added to the front of a word to change the meaning of the main word (or root). The number of adjective formative prefixes in English is 22 [2]. The number of adjective formative prefixes in Uzbek is 8. However, the adjective formative suffix (-siz) in Uzbek corresponds to the adjective formative prefixes in English (in-, il-, im-), and on the same basis, the adjective formative suffix (-iy) in Uzbek corresponds to the adjective formative suffix (-al-) in English. The morphological correspondence of the adjective formative prefix in English and Uzbek is the (ab-) prefix in English and the (no-) prefix in Uzbek. The weights of the above-mentioned mutually compatible suffixes and prefixes in English and Uzbek were calculated, and the difference in weights between the two languages is presented in Table 5 below.

Table 5: Corresponding adjective-forming suffixes and prefixes in two languages

№	Adjective suffixes and prefixes in the Uzbek language	V3	Adjective suffixes and prefixes in English	V3	EVX(V3)-EVIX(V3)
1.	-siz	0.20262	in-	0.20113	0.00149
2.			il-	0.20111	0.00151
3.			im-	0.20112	0.0015
4.	no-	0.20206	ab-	0.20102	0.00104

In Uzbek, verb-forming suffixes (for example, -la , -lash , -illa , -ulla , -sira , -lan) are used to form new verbs denoting action or state. There are no suffixes in English that directly correspond to the verb-forming suffixes in Uzbek. However, the specific stem-forming suffixes (-ize, -en, -ify) and prefixes (re-, over-, out-, miss-, un-) that are used to express the lexical-grammatical meaning of action, state, or process in English differ in that they have the following suffixes (-ize, -en, -ify) and lexical-grammatical word morphemes. The verb-forming suffixes in these two languages may partially coincide in the translation process.

In Uzbek, the suffixes “-cha”, “-lab”, “-larcha”, “-ona”, “-an”, “-chasiga” are added to nouns and adjectives to form adverbs. The English adverb word class does not contain suffixes that correspond to the Uzbek adverb word class. In the process of translating text in computer translation into two languages, these grammatical rules must be given in the database built for computer translation, taking into account these compatible and incompatible aspects. The English adverb word class does not correspond to the Uzbek adverb word class because it has the following rules.

In English, the adverb indicates the sign of a verb (action) and indicates how the action is performed. In English, the adverb is formed by adding the suffix “-ly” to an adjective. Note the use of adverb in the following examples:

For example:

He speaks English very fluently – U ingliz tilida juda ravon gapiradi. («fluent»-«ravon». «fluently-ravon»(ravish))

In English, the adverb is used after the verb. This is exactly the aspect that should be paid special attention to in computer translation. Not all adjectives in English can be made into adverbs by adding the suffix “-ly”. The following 4 adjectives are never made into adverbs by adding the suffix “-ly”. These adjectives are used in place of both adverbs and adjectives without changing their form.

Fast – tez; hard – qattiq, og’ir; late – kech; early – erta. When these words are clearly identified as belonging to a specific word class in a database built for computer translation, ambiguity in text translation is not allowed.

Pay attention to the examples:

He drives very fast – U juda tez haydaydi («fast» – ravish).

They came very late – Ular juda kech kelishdi («late» – ravish).

I am working hard – Men qattiq ishlayapman.

You must come early – Siz erta (vaqtidan oldin) kelishingiz kerak.

Bundan tashqari “good” (yaxshi) sifatining ravish shakli “well” bo’ladi.

Additionally, the adverbial form of the adjective "good" is (yaxshi).

Pay attention to the example:

Tom plays tennis very well – Tom tennisni juda yaxshi o’ynaydi.

The word "well" is always used as an adverb, this word can only be used as an adjective in the sense of health.

For example:

My brother is not very well at the moment – Ayni vaqtda mening akam juda yaxshi emas (sog’lom emas).

Proverbs express the quality of a verb, an action.

An adjective expresses the quality of something. In computer translation, grammatical rules must also be defined so that the computer can determine which word class the adjectives and adverbs belong to in the sentences in the text.

In English and Uzbek, we can form new words using the affixation method. Words formed from prefixes created using this affixation method are formed in a way that is similar to antonyms. When compared with each other, a comparative analysis of English and Uzbek has shown that affixes are an effective way of forming words in English and Uzbek. English belongs to the Indo-European language family, and Uzbek belongs to the Turkic language family, that is, to an agglutinative language. The number of matching affixes and prefixes in English and Uzbek is not large, which of course makes the difference in the content of the translated text noticeable when we translate from English to Uzbek and from Uzbek to English.

From morphological analysis, we know that despite the fact that the words and word-forming and modifying affixes in the two languages are numbered the same by word class and are defined the same based on the expanding input language, the weights of the words in the examples (in Tables 1-5) did not differ significantly from each other. In order for the difference between the weight coefficients of the sentences in the two languages to not exceed the specified value (2.23), the weight of words belonging to other word classes is changed to 0.07. [7] On this basis, research was carried out. It was found that the difference between the weight coefficients of words and affixes is within the specified value range.

Conclusion

Based on the above results, we can conclude that in order to improve the quality of machine translation from EVX to EVIX in educational and technical fields, we should first pay attention to the following. We can say that if the EVX text affixes given to the computer and the weights of the affixes in EVIX are close to each other, a high-quality machine translation will result. Because in the process of computer translation, the computer translates between two languages by selecting the appropriate one from among the mathematical models and the weights that are close to each other based on this mathematical model. The smaller the difference between the weights, the closer the English text to the Uzbek language will be in terms of content.

REFERENCES

1. Maxmudov M.A. O‘zbek tilidagi gaplarning matematik modellari asosida intellektual sintez qiluvchi ligvistik protsessorini yaratish. Toshkent 2023. 93 b.
2. Хакимов М.Х. Технология многоязыкового моделируемого компьютерного переводчика (Технология ММКП) //LAMBERT Academik Publishing, Монография, г. Рига, 2019. - с. 174.
3. Дроздова К. А. Машинный перевод: история, классификация, методы. Журнал : СЛОВОМОЛОДЫМ// УДК 81’322.4 Гуманитарные исследования 2015 № 3 (7), 156 с.
4. Lin L., Liu J., Zhang X., Liang X. Automatic translation of spoken English based on improved machine learning algorithm // Journal of Intelligent & Fuzzy Systems. – 2021. – Vol. 40, №2. – P. 2385-2395. – doi: http://dx.doi.org/10.3233/JIFS-189234.

5. Zhang G. Research on the efficiency of intelligent algorithm for English speech recognition and sentence translation // Inform An Int J Comput Inform. – 2022. – Vol. 45, №2. – P. 309-314. – doi: http://dx.doi.org/10.31449/inf.v45i2.3564.
6. Wen H. Intelligent English translation mobile platform and recognition system based on support vector machine // Journal of Intelligent & Fuzzy Systems. – 2020. – Vol. 38, №153. – P. 1-12. – doi: http://dx.doi.org/10.3233/JIFS-179788.
7. Dandapat S., Federmann C. Iterative data augmentation for neural machine translation: a low resource case study for English-Telugu // Vol. Proceedings of the 21st Annual Con- ference of the European Association for Machine Translation, (Alacant, Spain), European Association for Machine Translation. – 2018. – may 28-30, P. 287-292.
8. Lin X, Liu J, Zhang J, Lim S. A novel beam search to improve neural machine translation for English-Chinese // Comput Mater Contin (Engl). – 2020. –Vol.65, №1 P. 387-404. doi: http://dx.doi.org/10.32604/cmc.2020.010984.
9. Luong MT, Manning CD. Achieving open vocabulary neural machine translation with hy- brid word-character models // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, (Berlin, Germany), Association for Computational Linguistics. – 2016. –Vol.1, P. 1054-1063. doi: http://dx.doi.org/10.18653/v1/P16-1100.
10. Choi H, Cho K, Bengio Y. Context-dependent word representation for neural machine trans- lation // Comput Speech Lang. – 2017. –Vol.45, P. 149-160. doi: http://dx.doi.org/10. 1016/j.csl.2017.01.007. doi: http://dx.doi.org/10.18653/v1/P16-1100.
11. Marchuk Yu.N. Translation models: textbook.aid for student institutions of higher education // Publishing Center “Academy” . – 2010, P.176.
12. Novikov I.A. Semantics of the and its formalization // : “Science” . – 1983, P.216.
13. Abdurakhmonova N. Representation morphological analysis in machine translation //Ozbekistonda xorijiy tillar ilmiy metodik elektron jurnal. –2017. Vol.40, №2, P.13.
14.Khakimov M. Kh., Sirojiddinov Z. Sh. Computer algorithmization in multi-language modelled translator technology // Modern problems of applied mathematics and information technologies al-Khwarizmi, 2021. – p. 23
15.Khakimov M. Kh., Omonov A. A. Computer algorithmization of models of English sentences in translator system // Modern problems of applied mathematics and information technologies al-Khwarizmi, 2021. – p. 22.
16.Марчук Ю.Н. Модели перевода. Модели перевода: учеб. пособие для студ. учреждений высш. проф. образования / Ю. Н. Марчук. —М.: Издательский центр «Академия», 2010. — 176 с.
17.Хакимов М.Х. Математические модели узбекского языка. ЎзМУ хабарлари. № 3, 2010. с. 187-191.