International Journal of Language Learning and Applied Linguistics

ISSN: 2835-1924
Volume 2 | No 5 | May -2023

Problems of Understanding and Translating Homonyms in the Artificial Intelligence System

Khalmanova Zulhumor Turdievna [1], Abdurakhmanova Mukaddas Tursunalievna [2]

[1] Doctor of Philology, Professor of the Tashkent State University of the Uzbek Language and Literature named after Alisher Navoi

[2] Doctoral student (DSc) of the Tashkent State University of the Uzbek Language and Literature named after Alisher Navoi

Abstract: In a narrow sense, translation in an artificial intelligence system refers to the process of translating a text by a computer (in whole or in part) from one natural language to another without human intervention. In a broad sense, translation in an artificial intelligence system is a field of scientific research located at the turn of linguistics, mathematics, cybernetics, and aimed at building systems that implement translation in an artificial intelligence system in the narrow sense.

Key words: artificial intelligence system, text translation, types of systems, homonyms.

Relevance of the topic: The effectiveness of automated translation systems in an artificial intelligence system depends on the extent to which they take into account the objective laws of the functioning of language and thinking. The following types of automated translation systems are distinguished:

1) P-systems are direct translation systems. These systems include only the stages of morphological analysis and synthesis, so the result of the work of such systems is a kind of interlinear translation;
2) T-systems - systems with syntactic transformation of the source text;
3) I-systems - systems with semantic and pragmatic analysis. This type of systems is considered the most complex, since it includes not only linguistic information, but also extralinguistic information, that is, the semantics and pragmatics of the subject area [1].

The rapid development of artificial intelligence in recent years has forced us to take a fresh look at the problems of natural language processing, and especially at the problem of understanding. The article provides an overview of a wide range of issues related to understanding in the context of using artificial intelligence systems. It is shown that areas that previously traditionally excluded any machine intervention began to open up to artificial intelligence. The disclosure of the problem of understanding today seems to be the key to further expanding the scope of the use of artificial intelligence, as well as to distinguish between two types of artificial intelligence: in a weak and strong sense.

The purpose of the study is to study the problem of understanding and translating homonyms in an artificial intelligence system.

In accordance with the set goal, the following tasks are solved in the work:

- ✓ define the concept of homonyms;
- ✓ find out the feasibility of using artificial intelligence;
- ✓ consider the effectiveness of translation methods in the artificial intelligence system.

The object of research is the system of artificial intelligence.

The subject of the research is homonyms.

Research methods.

1. method based on rules;
2. method based on statistics;
3. A method based on learning in an artificial intelligence system.

Homonymy complicates automatic text processing. For texts in English, homonymy removal methods based on the use of a probabilistic model are quite widely represented, which give a fairly high accuracy. The problem for texts in Russian is not only part-of-speech homonymy, characteristic of texts in English, but also morphological and lexical homonymy. Due to the fact that the compilation of a mathematical model for the Russian language, which is characterized by a free arrangement of words in a sentence, is difficult, methods based on rules have been more developed to remove homonymy in texts in Russian. In order to identify the results of the work of the support vector machine and the hidden Markov model for removing part-of-speech and full homonymy when processing texts in Russian, an experiment is being conducted, during which a subcorpus with removed homonymy of the national corpus of the Russian language is used. It is shown that the hidden Markov model for removing homonymy in Russian texts works better than the support vector machine.

Since the invention of the computer, man has been striving to ensure that a computer can understand a person. All new systems for recognition of text, speech, gestures are being developed. Automatic text processing is one of the oldest areas of research in this area. The first works on automatic text processing date back to the 1950s.

If tools for grammatical parsing of sentences are developed today at a good level for most common natural languages, then the problem of removing homonymy is still one of the most acute in text analysis. The types of homonymy are discussed, for example, by the authors of [2]. Even such a simple phrase as “Ivanov keeps money in the bank” at the stage of analysis morphology will give at least two meanings, since the sentence can be read with the noun “bank” (masculine) and “bank” (feminine). At the stage of semantic analysis, an additional ambiguity will arise, since "can" can mean not only a container, but also a sandbar, as well as a boat cross-beam. In special dictionaries, the noun "bank" will acquire many more meanings. The verb to keep in this case can have a synonym for “keep”, especially if it is included in the idiomatic expression “keep money”, but it is quite possible to allow the interpretation “Ivanov keeps money in a bank, clutched in his hand”.

In practical problems, the statistical weight of one or another option is taken into account [3]. From the point of view of the problem of understanding in artificial intelligence, such a solution is not entirely “honest”, since even in machine translation problems, statistics is not able to provide the selection of adequate equivalent synonyms. More complex options involve the use of machine learning [4], markup, or other methods [5, 6].

However, for a more or less adequate "understanding" of natural language and operating with a variety of semantic meanings contained in the text, the machine must not only have a very rich ontology, comparable in volume to human ideas about the world, but also an algorithm that allows working with the actual set of options. In the given example “Ivanov keeps money in the bank”, the car must have a criterion according to which the noun “bank” in the sense of a vessel or a sandbank is not suitable or, on the contrary, is the only correct one.

To resolve morphological homonymy in machine translation, researchers suggest using methods such as 1) a rule-based method; 2) a method based on statistics, 3) a method based on machine learning.

The essence of the rule-based method is that in some situations, context analysis helps to understand the syntactic structure of a part of a sentence, and with its help, the forms of words. This method requires manual compilation of rules, for each of which it is necessary to write an independent program module. Replenishment of the rule system becomes more difficult with each new rule, and therefore such methods are not widely used.

Much more often, statistical methods and methods based on machine learning are used to resolve homonymy in modern morphological processors.

Counting the statistics of different variants of a word is the simplest way to remove morphological ambiguity. For example, the simple probability calculation method allows you to calculate the probability of meeting a certain word form among all the variants of use in the text.

To remove homonymy in machine translation, various classification methods are also used. «The grammatical parameters of a given or neighboring words in a certain window, their lemmas, signs of the presence of punctuation marks, and so on can be taken as classification parameters. The choice of classification method largely depends on the tastes of the developer, for the solution of which such machine learning methods as hidden Markov models, conditional random fields, recurrent neural networks, and others are used».

The above methods are used in such morphological analysis systems as TreeTagger, Pymorghy2, MyStem.

Results of the study and their discussion.

In the course of the study, we compared the homonyms of Turkish words and compared the results of human and artificial intelligence translations.

Here are some examples of homonyms in Turkish:

1. *banka (finansal kurum) - banka (nehir kenarı)*

banka - bench in the park - banka - (slang) mess - (human translation).

2. *baş (vücut bölümü – часть тела) - baş (başlamak - начать)*

baş- Çıban (furuncle) - (human translation).

3. *bay (erkek)- man*

bay- zengin (rich) - (human translation).

bay (koy) - (translation produced by artificial intelligence).

4. *beyaz (renk)- color*

beyaz (flour) - (translation produced by artificial intelligence).

5. *can (yaşam) – soul*

can (kutu - box) - (translation produced by artificial intelligence).

6. *dava (mahkeme olayı) - lawsuit*

dava (çanta) - case (bag) - (translation produced by artificial intelligence).

7. *değişim (farklılaşma) - change (differentiation)*

değişim (para bozdurma) - exchange (money exchange) - (translation produced by artificial intelligence).

8. *düş (hayal) - dream*

düş (düşmek) – fall - (translation produced by artificial intelligence).

9. *el (vücut bölümü) – hand (body part)*

el (ölçü birimi) - unit - (translation produced by artificial intelligence).

10. *ev (konut) – house (residential)*

ev (işareti) – signboard - (translation produced by artificial intelligence).

11. *fakir (yoksul) - poor*

fakir (delikli) - with holes - (translation produced by artificial intelligence).

12. *fatura (faturalama) - invoice*

fatura (kağıt) – paper - (translation produced by artificial intelligence).

13. *garaj (araç park yeri) - garage (car park)*

garaj (müzik aleti) - musical instrument - (translation produced by artificial intelligence).

14. *gemi (deniz taşıtı) - ship (watercraft)*

gemi (kristal eşya) – crystal - (translation produced by artificial intelligence).

15. *göl (su birikintisi) - lake (pond)*

göl (yok etmek) – destroy - (translation produced by artificial intelligence).

16. *güneş (gök cisim) - sun (celestial body)*

güneş (gülmek) – laughing - (translation produced by artificial intelligence).

17. *hat (çizgi) – line*

hat (şapka) – hat - (translation produced by artificial intelligence).

18. *hava (atmosfer) – air (atmosphere)*

hava (duygu durumu) - weather (emotional state) - (translation produced by artificial intelligence).

19. *iş (meslek) - work (occupation)*

iş (görev) – task - (translation produced by artificial intelligence).

20. *kap (kaplama) - cover*

kap (kapı) – door - (translation produced by artificial intelligence).

kap - mutfak eşyası – (crockery - kitchen utensils) - (human translation).

kap - üst giyisi – (cap - headdress) - (translation produced by artificial intelligence).

21. *karanlık (ışıksızlık) - darkness (no light)*

karanlık (ekmek çeşidi) - dark (type of bread) - (translation produced by artificial intelligence).

22. *kırmızı (renk) – Red (color)*

kırmızı (et) - red (meat) - (translation produced by artificial intelligence).

Conclusion

Machines are programmed to perform certain algorithms. Machine translators use human-developed cross-language algorithms to translate a range of words from one language to another using dictionaries and grammar rules. However, translation is not just the transformation of words from one language into another. This is a much more complex process. Therefore, no matter how good the algorithm is, machines have one big drawback - they cannot understand the meaning of the text.

As you know, homonyms are words that are the same in spelling and sound, but different in meaning. Here are some examples of homonymous pairs: schedule (work plan) / schedule (artist), environment (set of natural conditions) / Wednesday (day of the week), brush (painting device) / brush (part of the hand), plan (gradually descend during flight) / plan (make plans).

As a rule, we understand the meaning from the context or even intuitively thanks to the linguistic sense. Machine translators can distinguish between some simple homonyms. However, the probability of error increases significantly due to interlingual homonyms - "false friends of the translator". This term refers to words that sound the same but have different meanings in different languages. Such false friends can lead to ridiculous or even offensive translations, as is the case with the English word "preservative", which does not mean a contraceptive at all!

As you can see, a machine translator can easily mess up a translation, while a "animated" translator can easily recognize such controversial points and provide an accurate translation.

References.

1. Shchipitsina, L. Yu. Informatsionnye texnologii v lingvistike: textbook / L. Yu. Shchipitsina. - M. : FLINTA : Nauka, 2013. - 128 p.
2. Anjali, M K., Babu, Anto P. 2014. Ambiguities in Natural Language Processing International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 5, Oct. 2014. p. 392–394.
3. Peter, F. et al. 1991. Word-sense disambiguation using statistical methods. in: 91 Proceedings of the 29th annual meeting on Association for Computational Linguistics, p.264-270.
4. Márquez, L. Machine Learning and Natural Language Processing. Technical Report LSI-00-45-R, Departament de Llenguatgesi Sistemes Informátics, Universitat Politécnica de Catalunya, 2000. URL: http://l2r.cs.uiuc.edu/~danr/Teaching/CS546-11/Papers/marquez-LNLP00.pdf (дата обращения 05.02.2016).
5. Evang, K. & Bos, J. 2013 Scope Disambiguation as a Tagging Task. in: Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Short Papers. Erk, K. & Koller, A. (eds.). Potsdam, Germany: Association for Computational Linguistics (ACL), p. 314–320.
6. Adwait Ratnaparkhi. 1998. Maximum entropy models for natural language ambiguity resolution. University of Pennsylvania Philadelphia, PA, USA.
7. Avtomaticheskaya obrabotka tekstov na yestestvennom yazyke i analiz dannyx : textbook / E. I. Bolshakova [and others]. - M. : Publishing House of the National Research University Higher School of Economics, 2017. - 269 p.
8. Türkçe Sözlük 1.2. Türk Dil Kurumu. Ankara 1988.