

Polysemy in Machine Translation Exemplified in English and Georgian

Babulia Akhobadze

Language Center, Ivane Javakishvili Tbilisi State University, Tbilisi, Georgia

(Presented by Academy Member Avtandil Arabuli)

ABSTRACT. Some linguistic problems of translation including the machine translation are considered in the paper. In conditions of the modern computerized era, the machine translation is very popular and convenient providing quick translation of huge translation materials necessary in different fields. However, there are number of linguistic problems in connection with machine translation. For instance, the fact that one and the same word has different meaning causes great difficulties in the machine translation and not only. We consider certain systematization of the language vocabulary to be helpful to solve some problems for overcoming the existing difficulties with the translation of polysemic words. © 2019 Bull. Georg. Natl. Acad. Sci.

Key words: business, terminology, machine translation

Today, no field of science can exist and develop without translation. Moreover, people feel the need for translation in everyday life. The translation practice is centuries old and widely distributed throughout the world, however, the science of translation originated only in the twentieth century and developed in different directions and meanwhile it became necessary to research the translation problems in terms of linguistics.

Many scientists share the viewpoint of famous Scottish linguist J. Catford that „translation is an operation performed on languages: a process of substituting a text in one language for a text in another. Clearly, then, any theory of translation must draw upon a theory of language – a general linguistic theory” [1].

N. Sakvarelidze considers that development of a relatively new direction of the scientific study of the translation from the positions of linguistics is conditioned by appearance of new tendencies in linguistics, in particular, by the development of communication theory [2].

Linguistic research of the translation is very important in terms of machine translation. About half a century ago, when there were not so many computers as now, and, the machine translation was not so popular as it is today, American scientists J. Katz and J. Fodor considered the problem of machine translation in the context of semantic theory. According to them, “the dictionary usually supplies more senses for a lexical item than it bears in almost any of its occurrences in sentences. But the machine will not

be able to select the sense(s) which the morpheme actually bears in a given sentence context, except so far as the selection is already determined by the grammatical markers assigned to the morpheme in the derivation of the sentence (Thus the machine will be able to choose the correct sense of *seal* in *Seal the letter* so far as the choice is determined by the fact that in this sentence *seal* is marked as a verb, and the correct sense of *seal* in *The seal is on the letter* so far as the choice is determined by the fact that in this sentence *seal* is marked as a noun. But the machine will not be able to distinguish the correct sense of *seal* in *One of the oil seals in my car is leaking* from such incorrect senses as 'a device bearing a design so made that it can impart an impression' or 'an impression made by such a device' or 'the material upon which the impression is made' or 'an ornamental or commemorative stamp' and so forth, since all of these senses can apply to nominal occurrences of *seal*.)” [3].

Similarly, let us consider a passage from Adam Smith's famous work “On the nature and Causes of the Wealth of Nation”: “The first public stamps of this kind that were affixed to the current metals, seem in many cases to have been intended to ascertain, what it was both most difficult and most important to ascertain, the goodness or fineness of the metal” [4].

In Georgian there is a translation of the Adam Smith's same work translated by Ph. Gogichaishvili [5] where the word 'stamp' is translated into Georgian as "დამღა" (*damgha*), which corresponds to the context of the original. But "Google Translate" the on-line translation program available to us translated it as "postal stamp". The thing is that the word 'stamp' in English and in Georgian has several meanings. Apparently, among different meanings of the word "stamp" provided by the dictionaries the machine was not able to choose the right word for that particular context and, instead, chose the word that does not correspond to the context.

As J. Katz and J. Fodor point out, “What the machine is failing to do is to take account of or utilize the semantic relations between morphemes in a sentence. Hence it cannot determine the correct number and content of readings of a sentence. Nor can it distinguish semantically anomalous sentences from semantically regular ones. Since the machine will associate a dictionary entry with each morpheme in a sentence, it does not distinguish cases in which the sense of a morpheme or string of morphemes in a sentence precludes other morphemes in the sentence from bearing ANY of the senses that the dictionary supplies for them” [3].

Thus, they discussed the problems of machine translation more than half a century ago, and those problems are so relevant to computer translation programs today. The main problem is that in most cases the machine cannot find the context equivalent meaning of the given lexical unit from the number of meanings given in the dictionary entries.

It is very difficult for computer translation programs to identify right equivalent among the numerous definitions given in the database. However, it sometimes happens so that the program is able to select an adequate term. For example, for the sentence -"Junk Bonds are bonds nobody wants to hold" [6] – despite some different kinds of mistakes "Google Translate" suggested quite relevant translation of the terminological unit “Junk Bonds” as „უსარგებლო ობლიგაციები“ (*usargeblo obligaciebi*), that gives hope to find the ways for solution of the problems of machine translation.

It should be noted that in the case of machine translation it is almost impossible to find the correct match for polysemic words in existing dictionary entries. Polysemy is the coexistence of two or more different meanings of the word or phrase. The polysemy is characteristic of any language. Both the Georgian and English languages are rich in polysemic words, homonyms and synonyms, which, on the one hand, show the diversity of

vocabulary of the language, and, on the other hand, sometimes cause uncertainties and difficulties for the linguists, translators and program developers. As we have seen in the example above, polysemic meanings represent a great challenge for machine translation. The phenomenon of polysemy has long been considered by different linguists, philosophers or psychologists. The discussion still continues in the direction of different categories of polysemic meanings.

According to the above examples, the realization of the polysemic meanings of various special terminology, namely business terminology, depends mainly on the context. In general, the context plays a major role in translation. For example, in the Comprehensive English-Georgian dictionary, there are three main meanings for the word "value" in Georgian. Consider the following example: "The things which have the greatest **value** in use have frequently little or no **value** in exchange; and, on the contrary, those which have the greatest value in exchange have frequently little or no **value** in use." [6].

Here it has the meaning of "cost" and in Georgian it is translated by the word of the same meaning [8]. In this case the context of the text is related to the economy and business. But it will not make sense if we translate it with the same word in the context of mathematical text. For example, " $2x + 5 = 15$, where x is an unknown value" – here the word "value" cannot be translated with the word "cost" in Georgian.

Consider four examples of "value" taken from "Economic Principles" by G. Mankiw [7]: "Differences in values"; "economists give different values;" "economists sometimes give contradictory advice because they have different opinions;" "because of differences in scientific judgments and differences in values, some disagreement among economists is inevitable".

In every four cases the word "value" is translated into Georgian by four different words, which are synonymous [8]. And in every four cases "Google translate" translated the word "value" correctly, but with one and same word, unlike the translation given in the Georgian translation.

Examples show that with certain systematization of the language vocabulary otherwise than given in the traditional dictionary entries will be helpful in case of machine translation of polysemic words.

As early as the beginning of the 19th century, German linguist J. Trier developed the "Semantic Field" theory. "Close study of the history of intellectual terminology in Old and Middle High German convinced Trier that it was fundamentally wrong to consider words in isolation, they must be viewed within the context of the lexical field to which they belong. Lexical field is a closely organized sector of the vocabulary, whose elements fit together and delimit each other like pieces in a mosaic. In each field some sphere of experience is analyzed, divided up and classified in a unique way" [9].

From today's perspective, in conditions of great interest and practical necessity of machine translation we consider that classification and systematization of the language vocabulary otherwise than it is provided in traditional dictionary entries can be one of the ways for solution of the machine translation problems. Based on the above examples of lexical units "value" and "mark" we can consider the economic semantic field and suggest that segmentation of different technical terminology, namely, the economic and business terminology into economic and business fields and clusters can be effective for solution of the above considered problems.

ენათმეცნიერება

პოლისემია მანქანურ თარგმანში ქართულისა და ინგლისურის მაგალითზე

ბ. ახოზაძე

ივანე ჯავახიშვილის სახ. თბილისის სახელმწიფო უნივერსიტეტი, ენების ცენტრი, თბილისი, საქართველო

(წარმოდგენილია აკადემიის წევრის ა. არაბულის მიერ)

წინამდებარე სტატიაში განხილულია თარგმანის ლინგვისტური პრობლემატიკის ზოგიერთი საკითხი, მათ შორის ის პრობლემები, რომლებიც მანქანური თარგმანის შესრულებისას გვხვდება. თანამედროვე კომპიუტერიზებული ეპოქის პირობებში განსაკუთრებული მნიშვნელობა შეიძინა მანქანურმა თარგმანმა. ამასთან დაკავშირებით, კომპიუტერული მთარგმნელობითი პროგრამები მრავალი ლინგვისტური პრობლემის გადაჭრას მოითხოვს. მაგალითად, ის ფაქტი, რომ ერთსა და იმავე სიტყვას სხვადასხვა მნიშვნელობა შეესატყვისება თარგმნისას, ხშირად დიდ სირთულეებს იწვევს მანქანურ თარგმანში და არა მარტო მანქანურში. პოლისემიურ სიტყვებთან მიმართებაში არსებული სირთულეების დასაძლევად მიზანშეწონილად მიგვაჩნია ენის ლექსიკის გარკვეული სისტემატიზაცია.

REFERENCES

1. Catford C. J. C. (1965) A linguistic theory of translation, p.1 Oxford University Press.
2. Sakvarelidze N. (2001) The theory of translation, p. 311 Tbilisi University Press. (in Georgian).
3. Katz J., Fodor J. (1963) The structure of a semantic theory. *Language* 39, 2:170-210.
4. Smith A. (2007) An inquiry into the nature and causes of the wealth of nations. Metalibri. Amsterdam, Lousane, Melbourne, Milan, New York, Paris.
5. Smith A. (1938) An inquiry into the nature and causes of the wealth of nations. Tbilisi University Press, Tbilisi.
6. Blanchard O., Amighini A., Giavazzi F. (2010) Macroeconomics. Pearson Education Limited.
7. Mankiw N. G. (2007) Principles of economics, Harward University.
8. Mankiw N. G. (2008) Principles of economics. „Diogene“, Tbilisi (in Georgian).
9. Ulmann S. (1969) Language and style. In: Readings in modern English lexicology.

Received January, 2019