Gohar GHALACHYAN
Yerevan State University

SEMANTIC ISSUES IN MACHINE TRANSLATION

Machine Translators are gaining wider usage in the world nowadays. Along with the growth of the use of Machine Translators, the demand to make them more perfect is also increasing, which in its turn implies growth of the problems and difficulties occurring in the process of development. Semantic, syntactic, cultural and social differences of languages highlight these problems more explicitly and make the solutionseven more difficult.

Key words: *Semantic Machine Translation, Statistic Machine Translation, meaning, context, word-for-word translation*

Machine Translation is a long story of a not yet realized and ambitious goal. The main goal of Machine Translation is to be able to translate from a source language ***S*** to a target language ***T*** /Costa-Jussa, 2012/. Irrespective of the purpose of Machine Translation, be it literary translation of Hamlet from English into Korean or translating a scientific text, one thing is obvious: the translation must be done not merely as well as possible, but correctly and close to human translation. The ability to translate does not by itself generate understanding; it can only transfer understanding from one language into another /Santos, 1992/.

Two main types of Machine Translation arewidely discussed: **Semantic Machine Translation and Statistical Machine Translation**.

The main mechanism of both types of Machine Translationis to convertasource text intoa target text through transformation. The target text will express the same idea and will be equivalent to the source text, i.e. the context of a text will be transferred from one language into another.

Source text	⟶	Target text
e.g. English	⟶	Armenian
Ես գնացի տուն	⟶	I went home.

A statistical translation mode is simply a model of parallel text, that is, a model that knows what sentence pairs are more likely than others to occur as translations of each other. Accordingly, a prerequisite for building a Statistical Machine Translation system for any language pair is to collect texts and their translations into a reference language /Bird, Chiang, 2012/.

Statistical Machine Translation usually develops and decodes all the rules of the language: conjugations, genders, tenses, etc. Millions of sentences are given to Statistical Machine Translation for analysis. After analyzing, a huge database of possibilities is formed, in which the main transition process is built.

The most prominent and widely used Statistic Machine Translator is Google Translator. 200 million people use Google Translator every day. How does Google

Translator work? The use of Google Translator is based on its search engine. There are millions of multilingual texts and websites on the internet. Each time a sentence or an utterance is given to Google Translator to translate, it finds similar source texts with its translated versions and after a fast analysis the target text is presented to the user. Google translator gives its users the possibility to edit the target text and these editions are kept in its database. It means that even the ordinary users play an important role in the refining of Google Translator. There is no need of linguists to create a special database and upgrade it frequently. This Machine Translator is very productive for such languages as English, French, Russian, German, etc as multilingual texts are of great number.

However, it has been claimed since the inception of Machine Translation that a semantic model is necessary to achieve human-like translation /Weaver, 1955/.The claim that semantics should be used in Machine Translation is fairly usual; it expresses the commonly-held view that without understanding you cannot translate, and is frequently used to defend some approaches against others /Santos, 1992/.

Nevertheless, Semantic Machine Translation is actually a difficult and sometimes even creative process in which more linguistsand programmers are engaged than is the case with StatisticalMachine Translation. Semantic Machine Translation is far from perfection. Among the main problemscurrently facing Semantic Machine Translation are:

1. Word Ambiguity

He deposited money in the bank account with a high interest.
Sitting on the bank of Thames, a passing ship piqued his interest.

Most languages will have different translations for the words *bank* and *interest.* In order to translate these sentences correctly human mind takes the context into consideration. But the computer cannot do that.

2. World knowledge

Machine Translator does not have the world knowledge, while humans do.

3. Idiomatic phrases

As Machine Translation does not have the world knowledge, it cannot understand idiomatic phrases either.

Raining cats and dogs.

This idiom will never make any sense in word-for-word translation into Armenian.

4. Preservation of the style

It does not preserve the style. E.g. *When I got up, the morning had already been broken*. While translating this sentence the possibility is great that the style will be lost through which the source text wishes to transfer particular emotions or feelings.

5. Co-references

e.g. When I visit my aunt, I can't help myself playing with my little cousin.
Is cousin male or female?

6. It only works well for grammatical sentences.

Another important factor which makes the translating process more complicated when using Machine Translation concerns social and cultural issues, for example, the Armenian utterance ***գլխի ընկնել*** is typical only for Armenians and its word-for-word translation into English ***to fall on head*** will never make sense. Tymoczko argued that meaning assignments are dependent on social and cultural factors, because the important evidence necessary to differentiate among meaning assignments is not truly semantic: two equally plausible meaning assignments can assign different truth values as long as this is compensated for by different theories of pragmatics. According to him, what matters is not only the context, but also the individual perspectives of the speakers, their beliefs. Therefore, to translate between natural languages it is necessary to have some knowledge of the society and beliefs of the speakers of each language /Tymoczko, 1978/. Dyvik comes up with situation schemaas a means of representing the semantics of both languages, recognizing that they do not have to be exactly the same for the two languages /Dyvik, 1990/. Semantic Machine Translation is especially useful for small languages like Armenian, Finnish, Serbian, etc. Surely, the reason for this is not only linguistic, but also economic.

Individual perceptions of the users of Machine Translator also play a crucial role. A clever user, realizing that Machine Translation is not perfect yet, will figure out the general meaning of the target text. The strict users of the Machine Translator have always been the native speakers of the source language. For them each mistake made during the translation process is more explicit.

On this stage of Machine Translation development we are more interested not in the perfection of it, but in a better quality, i.e. as better as possible, as to translate doesn't by itself imply the understanding of context, it can only transfer understanding from one language into another /Santos, 1992/.

e.g. Տունը մեծ է:

While translating this Armenian sentence into English we can come across to many variants. As humans are very smart and human mind is powerful at word modeling and analyzing, we can choose the best one among them.

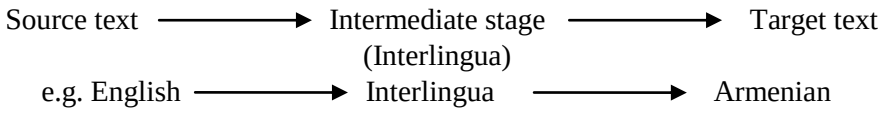
1. The house is big ⟶ good
2. House big is the ⟶ bad, but still understandable
3. The house is XXL ⟶ worst, not understandable and acceptable at all.

At this stage even a grammatically coherent variant will be satisfactory to fulfill some daily translation needs like understanding the headlines of Chinese or

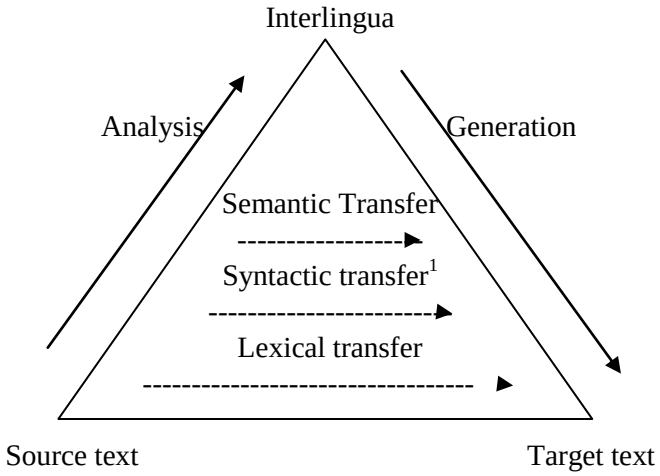
Portuguese news agencies. The demands from Machine Translation are also different at this stage: one person wants a translation, as faithful to the source as possible, the other wants a translation, which sounds right. As prof. Phillip Koehn from University of Edinburgh once said during his lectures, "the target text may not always sound good, or be in a good language, but a smarter user will understand it".

How does Semantic Machine Translation work?

Graph-structured meaning representation lie sat the base of Semantic Machine Translation. There is a striking difference between the main translation processes of Semantic Machine Translation and Statistical Machine Translation. In the case of Semantic Machine Translation, the translation process consists of two levels. The first level is the transformation of the source text into an intermediatestage: graph structured meaning representation. The second level is the transformation of the intermediate stage into target text. This intermediate stage is called interlingua by programmers.



Each Machine Translation processhas its clear plan. Actually, this process is not as easy as one can imagine. Looking at these three stages, it implies many other sub-actions which are shown in the graph below.



What is the interlingua?

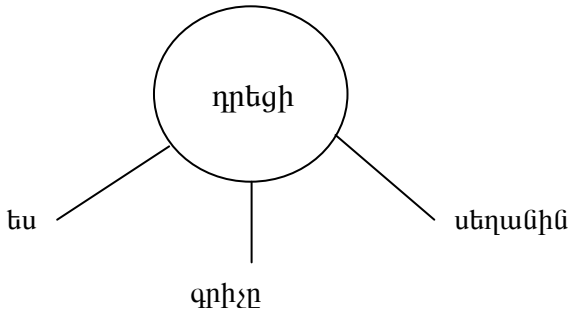
The very earliest successful attempts of Machine Translation were built on large tables ofhand-specified rules. These systems required large amounts of highly

[1] Computational Linguistics has already "learnt" about syntax and syntactic structure, which we cannot say about semantics. Both linguists and programmers face many problems.

specialized human knowledge (and consequently great cost) to produce, but with the investment of enough effort could be coaxed into producing reasonable output. These systems were, with few exceptions, either transfer-based or interlingual. In transfer-based systems, sentences were analyzed in the source language, and rules specified transformations of that analysis to produce analyses of the target sentence, which could then be used to generate the targettext /Andreas, 2012/.

At this stage a very simple analysis of a sentence is carried out. In order to understand this process better, we will take an Armenian sentence ***Ես դրեցի գրիչը սեղանին***. What happens when this sentences is given to Semantic Machine Translation? The sentence is to be translated from Armenian into English.

So, the first step carried out by the Machine is to do a syntactic analysis of the sentence where ***ես*** is the subject, ***դրեցի*** is the predicate ***գրիչը*** is the object and ***սեղանին*** shows the location.



Interlingua is a mathematical graph, the top of which is the predicate of the sentence connected with the other members of the sentence by syntactic relations. The target or final translated text is provided only by means of this graph. As this graph plays an important role in translating a text, aprofound knowledge base is required. Here the unit must not be the word, but the concepts that are expressed through the word. E.g. the words ***apple*** and ***խնձոր*** express the same concept, but in different languages and with different letters. What the concept reflects does not change depending on the letters or language change representing the word. In general we do not accept a word through its letters or letter combination, but through the concept it represents.

One word can represent more than one concept or meaning in one language. In its turn, one of those meanings can express more than one meaning in another language. For example, the words ***table*** and ***սեղան***. They express the same concept only in their general meaning or understanding. But they both also have their secondary meanings.

Table	→	1.a piece of furniture /սեղան/
		2. a chart /աղյուսակ/
Սեղան	→	1. կահույքի մաս /table/
		2. երկրաչափական օբյեկտ /trapezium/

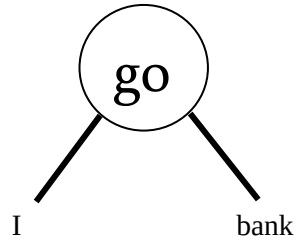
As we can see, the first pairs of concepts or meanings are the same, butthis is not the case with the second pair. While in Armenian, the second meaning of the word ***table***, i.e chart, is աղյուսակ, the second meaning of the ***word*** սեղան, i.e. երկրաչափական օբյեկտ, is trapezium. It is of vital importance for the concepts to be differentiated in order to make the target text better and correct, which is very important.

e.g. I get to the bank.

The Machine Translator will face a difficulty in translating this sentence from English into Armenian. The word ***bank*** has two basic meanings in English: 1. financial institution, 2. the side of a river. So, Machine Translation has two choices: 1. Ես հասնում եմ բանկ, 2 . Ես հասնում եմ գետի ափ:

How does a Semantic Machine Translator choose one from the above-mentioned two? In Semantic Machine Translation, the meaning of each word is closely connected to the connections by which the word is connected to the other members of the sentence. Two basic types of connections are known: heritable and horizontal. These terms are widely used in Mathematical and Computational linguistics.

If we go back to the sentence above, ***I get to the bank,*** Semantic Machine Translation will translate it Ես հասնում եմ բանկ. Actually, not only Semantic Machine Translation, but also humans will choose this version. Why? In fact, **going to the bank (financial institution)** is a more frequent action than ***going to the bank (river side),*** that is why even the human mind does not consider the second version. The version that the computer will choose depends on the possibilities of connections between the members of the sentence.



/The connections are marked in bold./

Weights or percentages are given to these connections, i.e. how great the possibility is that a person will go to the bank/financial institution/ and how great the possibility is that a person will go to the bank/river side/. These percentages are written into the description of the connections of the Semantic Machine Translation database by linguists. The percentage of the possible connection of going to the bank/financial institution/ is more than 90% and the percentage of the possible connection of going to the bank/river side/ is approximately 80%.

DAPRA GALE project released a special chart determining the quality of the source text./www.darpa.mil/

Mistake percentage	Assessment
0%	
	Publishable
10%	
	Editable (Document is clearly understandable; may require final editing)
20%	
30%	Gistable (Main idea, substance, or content of document is conveyed)
40%	Triagable (Determine whether document is important or not for further analysis)
50%	

Why to use Semantic Machine Translation?

In spite of all the inconveniences caused by Machine Translation, it is still widely used by millions of people all over the words as:

1. It helps to get general idea about the text written in a foreign language. While translating a text from Japanese into English, the target text will be in jumbled English, but the user will be able to figure out the general points.

2. It helps during the online communication when the participants do not speak the same language. E.g. chats, messengers, non-official emails etc.

Semantic Machine Translation is a difficult system which requires more investments and more linguists than Statistical Machine Translation. Yet another difficulty involves the subjective viewpoint of the linguist who describes the percentages of possible connections. This process will take a long time: in order to have high quality Semantic Machine Translation, more than 10 years of database creation are needed. But in spite of all the efforts made by scientists Rose Smith thinks that nobody thinks nowadays, as researchers did back in the 1950s, that high quality machine translation of complex texts is an achievable goal in the short-to-medium term /Smith, 2000/.

REFERENCES

1. Dyvik H. The PONS Project: Features of a Translation System. Bergen, 1990.
2. Andreas J. Towards Semantic Machine Translation. University of Columbia, 2012.
3. Costa-Jussa M. R. An Overview of the Phrase-based Statistical Machine Translation Techniques, Cambridge University Press, 2012.
4. Smith R. English Today. Issue 4, October, 2001.

5. Santos D. Semantics and (Machine) Translation, 1992.
6. Bird S., Chiang D. Machine Translation for Language Preservation, Mumbai, 2012.
7. Tymoczko T. Translation and Meaning, 1978.
8. Weaver W. Translation // Machine translation of Languages. Cambridge, MA.: MIT Press, 1955.
9. www.darpa.mil

Գ. ՂԱԼԱՉՅԱՆ – ***Իմաստային խնդիրները մեխանիկական թարգմանիչում.*** – Աշխարհում գնալով ավելի ու ավելի մեծ կիրառություն են ձեռք բերում մեխանիկական թարգմանիչները: Մեխանիկական թարգմանիչի օգտագործման աճին զուգընթաց, նաև աճում է այն ավելի կատարյալ դարձնելու պահանջը, որին զուգահեռ աճում են խնդիրներն ու ժվարությունները, որոնք խոչընդոտում են այդ գործընթացին: Սույն հոդվածում քննարկվում են լեզուների իմաստային, շարահյուսական, մշակութային և հասարակական տարբերությունները, որոնք ավելի խորն են ընդգծում վերը նշված խնդիրները և դժվարացնում են դրանց լուծումը:

Բանալի բառեր. իմաստային մեխանիկական թարգմանիչներ, վիճակագրական մեխանիկական թարգմանիչներ, իմաստ, կոնտեքստ, բառացի թարգմանություն

Г. КАЛАЧЯН – ***Семантические задачи в машинном переводчике.*** – В мире все большее применение находят машинные переводчики. В связи с этим растет необходимость их усовершенствования. Параллельно растет количество проблем и трудностей, которые препятствуют этому процессу. В данной статье обсуждаются семантические, синтаксические и культурные различия языков, которые еще глубже подчеркивают эти проблемы и осложняют их решение.

Ключевые слова: семантический машинный перевод, статистический машинный перевод, смысл, контекст, дословный перевод