
RULES OF PRODUCTION IN THE PROCESS OF MACHINE TRANSLATION

Anatol POPESCU

Technical University of Moldova

Zinaida CAMENEV

Free International University of Moldova

La traduction comme processus a un rôle important dans la vie quotidienne, elle crée la possibilité de connaître les valeurs universelles. La traduction peut être perfectionnée tant par l'homme que par l'ordinateur (automatiquement). Dans le dernier cas elle représente (provoque) des difficultés considérables.

L'article suit l'intention de proposer quelques règles de production assistées par ordinateur des textes anglais en roumain à l'aide des transformations lexicales, morphologiques et syntaxiques. Les difficultés de traduction sont conditionnées par les similitudes et les divergences entre les unités structurales mises en évidence dans les deux langues.

La qualité de la traduction réalisée doit être bien évaluée parce que les dictionnaires électroniques ne sont pas au niveau de toutes les exigences.

Mots-clés : *traduction automatique, méthodes de traduction, règles de production, transformations lexicales, morphologiques et syntaxiques, traduction directe.*

Traducerea ca proces are un rol important în viața cotidiană, deoarece dă posibilitate de a cunoaște valorile universale. Traducerea poate fi perfectată atât de om, cât și de automat (de calculator), cea din urmă prezentând greutăți considerabile.

Articolul are menirea de a prezenta unele reguli de producere în procesul de traducere asistată de calculator a textelor din limba engleză în limba română în baza transformărilor lexicale, morfologice și sintactice, greutățile la traducere fiind condiționate de similaritatea sau divergența unităților structurale evidențiate a limbilor în cauză. Deoarece dicționarele electronice nu sunt perfecte, este nevoie ca traducerea rezultată să fie supusă unei postredactări.

Cuvinte-cheie: *traducere automată, metode de traducere, reguli de producție, transformări lexicale, morfologice și sintactice, traducere directă.*

We would like to tell some words about the techniques of machine translation (MT) and how these techniques are used in practice. We want to underline that there are often trade-offs and difficult choices among alternative approaches and techniques.

At present MT is developing by two methods: the first method is connected with the modeling of transfer from the meaning to the text and vice versa, the second method is based on the formal application of the translational correspondences not only of separate words but also of the word combinations. We based our investigation on the second method. The basis of it is the typology of translational correspondences classified into the equivalent, variational and transformational. These correspondences are determined by the methods of the contrastive linguistics. They represent the most actual lexical, morphological and syntactical transformations.

There are four models for doing MT: 1. The simplest one is the illustration of the ways in which languages differ. 2. The use of syntactic transformations for overcoming differences in grammar as well as some techniques for choosing target language words. 3. Some ways of exploiting meaning during translation, in particular the use of thematic roles and primitive decomposition. 4. The minimalist “direct” approach.

When you listen to a radio or television program in a foreign language it seems like chaos, completely unlike in your native language or some other language you know, and still there are patterns in this chaos, some aspects of human language which seem to be universal, holding true for every language. Many of these universals arise from the functional role of language as a communicative system between humans. Every language seems to have words for referring to people, for talking about life, for being polite or not.

When languages differ, these differences often have systematic structure. The study of systematic cross-linguistic similarities and differences is called typology. (Croft 115; Comrie 59) We shall sketch some typological facts about cross-linguistic similarities and differences between English and Romanian, because the difficulty of translating from one language into another depends greatly on how similar the languages are in their vocabulary, grammar and conceptual structure.

Similarity of English and Romanian

Syntactically these languages are also different in the basic word order of subjects, predicates, objects. Speaking about English we can say that the order of words is strict, i.e.

He buys a French book.

This sentence may have only such an order of words, we can't change it. Human Romanian translation may be:

El cumpără o carte franceză;

Cumpără o carte franceză.

The MT of it is: El cumpără o/un francez/franceză carte.

Comparing these sentences from the English and Romanian languages we can observe their syntactico-morphological distinction, namely, the distinction between head-marking and dependent-marking languages. (Nichols 70) In the English word combination “a French book” the head-noun is “book” and the determiners are situated before it but in Romanian the modifiers are after the head-noun. Speaking about the lexical organization, we can say that there are interesting problems here too. Many words can be translated relatively directly into other languages, let us take the word *book* in English, in German it is *Buch*; *l'homme* in French, and *om* in Romanian, etc.

Sometimes, rather than a single word there is a fixed phrase in the target language: French *informatique* is translated into English by *computer science*.

Lexically, e.g., a word may be translated best by a word of another part of speech in the target language: English *like* is translated into German by the adverb *gern* in the sentence:

She likes to sing
Sie singt gern...

where the syntactic structure is also changed according to the rules of the target language. Sometimes one language puts more grammatical constraints on word choice than on something else. English distinguishes gender only in the pronouns *he/she*, as to the pronoun *they* we see that it is not specified for gender, but translating it into Romanian we must distinguish between *ei/ele* as in French (masculine *ils*, feminine *elles*).

Such differences in specificity also occur on the semantic side; one language may divide up a particular conceptual domain in more details than the other language. English has an impoverished kinship vocabulary: the single word *brother* can indicate either *a younger or older brother* also English *wall* has two equivalents in German: *Wand* (inside) and *Mauer* (outside); the English word *know* has also two words in French - *connaître* (be acquainted with) and *savoir* (know a proposition); German *Berg* has two equivalents in English: *hill* (not very high) and *mountain* (very high).

Dependencies on lexical gap and cultural context also complicate the problems of MT. That is why we cannot get perfect translation because the speakers of the source and target languages have different conceptual systems. A number of translating theorists (Steiner 51; Barnstone 100; Hofstadter 25) referred to a story by Jorge Borges showing that even two linguistic texts with the same words and grammar may have different meanings because of their different cultural contexts.

The Transfer Model

The strategy of performing MT is to translate with a process overcoming these differences, altering the structure of the source language to make it conform to the rules of the target language. This is done by applying contrastive knowledge, that is knowledge about the differences between the two languages, based on the system of the transfer model. In this case MT involves three phases: analysis, transfer and generation. Transfer means the gap between the output of the source language parser and the input to the target language generator.

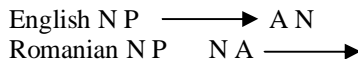
Syntactic Transformations

Speaking about the syntactic transformations we must mention that in English there is an unmarked order in a noun-phrase, when the adjectives precede the noun but in Romanian they follow the noun. How can a MT system overcome such a difference?



In general syntactic transformations are operations that map from one tree to another. In such cases we must transform the nominal combinations, i.e. we must

reverse the order of the noun phrase in the process of translation. According to the rule we write:



We have taken the simplest rule, transformations in the MT system may have more complex conditions, e.g. sentences with *there*, etc. We can formalize the transformations by means of unification-based models.

Lexical Transformations

The process of finding target language equivalents for the content words of the source language, i.e. lexical transfer, is difficult for the reasons that these two languages, English and Romanian, have many divergences because of the fact that they are from different groups of the languages, we have already shown it above.

The foundation of the lexical transfer is dictionary look up in a cross-language dictionary. The translation equivalent may transfer the model and treats translation as a process of changing the structure and words of a source sentence to arrive at a valid sentence of the target language. It is necessary to treat translation as a process of extracting the meaning of the input and then expressing the meaning in the target language. If we could perform this the MT system could do without the contrastive knowledge, relying only on the syntactic and semantic rules.

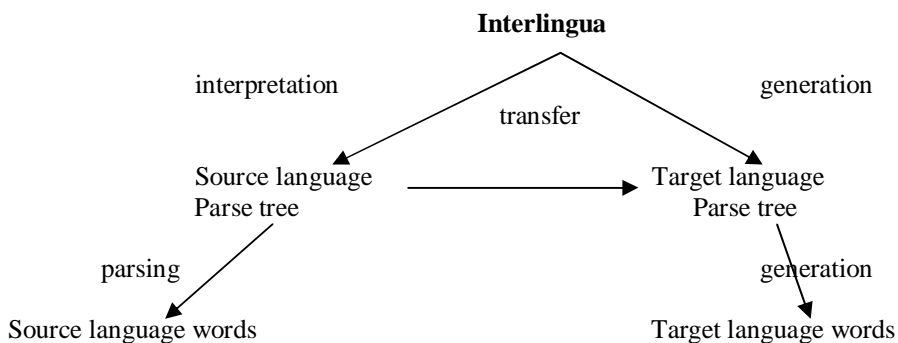
This scheme presupposes the existence of a meaning representation, or interlingua in a language independent form.

The idea of the interlingua is to represent all sentences that mean the ‘same’ thing in the same way, regardless of the language they happen to be in.

The semantic analyzer produces the structure rules by means of an Agent relation. It requires more analysis work than the transfer model which only requires syntactic parsing. Generation then can proceed directly from interlingua without any syntactic transformations.

Interlingua needs an inventory (with two key-words: agent and force and their determiners) in order that the semantic analyzer perform the work. The interlingual idea has implications not only for syntactic transfer but also for lexical transfer. The main idea is to avoid explicit descriptions of the relations between the S L and the T L words.

For simplification of the process we must use decomposition (was –Past Simple and Past Continuous, a - article in the Singular).



Brushing over numerous important details we can now contrast the transfer model with the interlingua model. Doing the extra work involved by the interlingua commitment is not an easy task. It requires the system designer to perform exhaustive analysis of the semantics and formalize it. Today this is more an art than a science. In some cases the semantics can mostly be given by a database model, as in the air travel, hotel reservation, at a shoe shop or restaurant recommendation domains.

Another problem with the interlingua idea is that, in its pure form, it requires the system to fully disambiguate at all times.

Direct Translation

The models of direct translation are looking fine. They involve: 1) Translating by fragments; 2) Producing elaborate structural analyses and do simple operations that can be done reliably. Such systems are called direct MT systems (with only a pair of languages: English and Romanian).

The direct MT system is composed of several stages, e.g.:

1. Morphological analysis;
2. Lexical transfer;
3. Work relating to prepositions;
4. SVO rearrangements;
5. Miscellany;
6. Morphological generation (post-editing).

MT may be introduced by: transfer, interlingua and direct models. They show what representations to use and what steps to perform in order to translate a text. Still there is another way to approach the problem of translation: to focus on the result, not on the process.

Many MT systems, especially direct ones, have a final phase, in which the system uses local considerations to revise word choices in the output. We can approximate the probability of a sentence being a good translation as the product of the probabilities that each target language word is an appropriate translation of some source language word.

Where do we get these probabilities? Standard bilingual dictionaries do not include such information, but they can be computed from bilingual corpora, i.e., parallel texts in two languages. From bilingual corpora aligned it is possible to count how many times a word, phrase, or structure gets mapped to each of its possible translations. Such alignments are potentially useful not only for MT but also for automatic generation of bilingual dictionary entries for use by human translators. (Dagan and Church 101; Fung and McKeown 65)

Since MT systems are generally run by human operators, the man is available to help the machine. One way to use human intervention is interactively: i.e., when the system runs into a problem, it can ask for help from the user. This is annoying – users do not like to have to answer questions from a computer, or to help the computer get its work done. (Cooper 87) Post-editing is the normal mode of human interaction with M T systems.

One way is to apply MT and see what comes out wrong, and then rewrite those sentences in the original. Another way is to have a model of what MT would

handle, and require input sentences to be rewritten in that sublanguage by disallowing PPs which could have ambiguities. Doing so may also make the source language text more understandable. This leads to the improvement of the process of translation as a whole.

In general the content words are crucial: having the words translated properly is vital. In practice, one of the major advantages of using a M T system is that it handles most of the difficult work of looking up words in bilingual dictionaries. As a result, professional MT users put great value on dictionary size and quality.

It has also become apparent that MT systems do better if the dictionaries include not only words but also idioms, fixed phrases, and even frequent clauses and sentences. Such data can be extracted automatically from corpora. Moreover, in some situations it may be valuable to do this on-line. Good results of translation depend on factors other than the quality of it. They depend on such factors as speed, storage requirements, the ability to run transparently inside the editor, etc. Translation memory, the ability to store and recall previously corrected translations is also preferable.

Language differences are a virtually inexhaustible source of complexity. The task of the linguist together with the programmer is to simplify the materials. Adding more knowledge does not always help, since a working MT system is a large, delicate code. Every M T system is adapted to a domain, to the working habits of its users and to the needs of the consumers of the output, first of all. MT system design is hard work, requiring careful selection of models and algorithms combining them into a useful system. This must be done minimizing the development cost.

Bibliographical references

- Barnstone, W. *The Poetics of Translation*. Yale University Press, 1993.
- Comrie, B. *Language Universals and Linguistics Typology*. Basil Blackwell: Oxford Second Edition, 1989.
- Cooper, D. "Concepts from Semantics as Avenues to Reading Improvement." *The English Journal* 2 (1995): 85-90.
- Croft, W. *Typology and Universals*. Cambridge: Cambridge University Press, 1990.
- Dagan, I. and Church, K.W. "Termight: Coordinating Humans and Machines" *Bilingual Terminology Acquisition*. *Machine Translation* 12 (1997): 89 – 107.
- Fung, P. and McKeown, K. R. *A Technical Word- and Term-Translation and Using Noisy Parallel Corpora*. *Machine Translation* 12 (1997): 53 – 87.
- Hofstadter, D.R. *Le Ton beau de Marot*. New York: Basic Books, 1997.
- Nichols, J. *Head-marking and Dependent-marking Grammar Language* 62.1 (1986): 56
- Steiner, G. *After Babel: Aspects of Language and Translation*. Oxford: Oxford University Press, 1975.
- Горбатиков В. Г., Каменева З. М. *Обучающая диалоговая система, ориентированная на справочное бюро обувного магазина. Школа-семинар по оптимизации преподавания иностранных языков с помощью технических средств*. Кишинев, 1979, p. 29-30.