Chapter XII
Machine Translation Systems

Athanasios Tryferidis
Electrical and Computer Engineer, MLS SA, Greece

Theofanis Korlos
Aristotle Univerity of Thessaloniki, Greece

INTRODUCTION

Machine translation, sometimes referred to by the acronym MT, is a subfield of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another. At its basic level, MT is performed as a simple substitution of atomic words in one natural language for words in another. Using corpus techniques, more complex translations may be attempted, allowing for better handling of differences in linguistic typology, phrase recognition, and translation of idioms, as well as the isolation of anomalies (Mitkov, 2003).

The European Association for Machine Translation (EAMT) defines machine translation (MT) as "the application of computers to the task of translating texts from one natural language to another."

It is obvious that MT is not a simple process. It has to overcome several problems such as to:

- Analyze the complexity of human/natural languages
- Clarify syntactical/referential ambiguities
- Decide on different word meanings
- Decide on polysemy, homonymy
- Detect and deal with metaphors, anomalies

At this point, it should be noted that MT is not a mere substitution for each word, such as a dictionary lookup.

The rest of the chapter is organized as follows: in Section 2 we give some information about the history of MT. In Section 3 we present the way in which the MT works. In Section 4 we mention the two major applications of MTS. In Section 5 we present the techniques that can make MT work better. In Section 6 we present Systran the worldwide leader and one of the first independent MT developers. In Section 7 we focus on expectations users may have and we present feature MTS applications. Finally, in Section 8 we give some details about the way the dictionaries are used in the process of machine translation, and we focus on the operation of the SEEArchWeb forum.

HISTORY

The history of machine translation generally starts in the 1950s after the Second World War. The Georgetown experiment in 1954 involved fully automatic translation of more than 60 Russian sentences into English. The experiment was a great success and ushered in an era of significant funding for machine translation research. The authors claimed that within 3 or 5 years, machine translation would be a solved problem (Asher, 1994).

However, the real progress was much slower, and after the ALPAC report in 1966 (ALPAC, 1966), which found that the 10 years long research had failed to fulfill the expectations, the funding was dramatically reduced. Starting in the late 1980s, as computational power increased and became less expensive, more interest began to be shown in statistical models for machine translation.

Currently, machine translation continues to progress. Large companies are now using it more, which also increases software sales to the general public. This situation has led to the creation of online machine translation services such as Altavista, which offer rapid e-mail services, Web pages, and so forth, in the desired language, as well as to the availability of multilingual dictionaries, encyclopaedias, and free, direct-access terminology databases.

HOW MACHINE TRANSLATION WORKS

The translation process can be stated simply as:

- Decoding the meaning of the source text
- Re-encoding this meaning in the target language

The diagram that follows in Figure 1 explains the procedure of MT in 5 steps.

It is understandable that MT is not a simple process. It takes advantage of computer power to perform complex procedures based on genetic algorithms and artificial intelligence in order to analyze some basic concepts such as:

- Morphology of words used
- Grammatical structure
- Source text syntax
- Semantics of words
- Word ambiguities, idioms, anomalies

Morphology is concerned with the ways in which words are built up from small meaning-

Figure 1. The procedure of MT