Proseminar Language and Computer

lecture: Machine Translation

– Handout –

speakers: Marion Riedel. Tino Schwarze

TU Chemnitz, November 2, 2000

1 Structure

1. Introduction
2. The Translation Process
3. Historical Overview
4. Theoretical Background
5. Problems of Machine Translation
6. Real World Usage
7. Current Research
8. Conclusion

This structure of today's lecture will help you to follow the lecture more easily.
The present handout does not exactly follow the structure.
Section 2, the theoretical part, contains the most important facts of each part of the lecture and the necessary knowledge for the practical exercises.
Section 3, the practical part, can be seen as a worksheet, where you can write down the results of your practical exercises.

2 Theoretical Part

History

Optimistic Beginnings

In France and Russia, the first machine supported translation systems were already built in the early 30ies of the 20th century. In 1942, the first real computer was built, which created the condition for the development of Machine Translation (MT). In those years, the attitude towards the known problems of MT was very optimistic and the expectations were very high. In 1954, the so-called Georgetown Experiment took place: a public performance of a MT system, that had an enormous success. As a result, many gouvernments spent money on an intensified MT research.

Disillusion

In the 60ies, the problems of MT and the fact, that it would not be easy to solve them, got more and more clearly. The high expectations expressed ten years ago had not yet been fulfilled. So, in 1964, a committee was founded to examine the current state of MT and the possibilities for the future. After two years, this committee published the ALPAC Report, that judged MT in a very negative way. This lead to a decreasing in research funds and interest in MT.

The 70ies

In the 70ies, there was a revival of MT research characterized with much more realism. A new research field was formed: Computer Linguistics, to try to solve the MT problems by developing grammatical formalisms.

Commercial Application

Until today, the amount of documents that have to be translated is increasing steadily. Since the 80ies, MT is therefore taken seriously and used commercially. Today, MT systems are very wide spread. But the limits of application are marked clearly.

Generations of Translation Techniques

Today, you can distinguish between four generations of translation systems:

1. **word-by-word-translation:** only a simple comparison of dictionaries, no syntactic or semantic rules included
2. **partial inclusion of syntax:** sentence structure determined from the morphological information in dictionaries
3. **total inclusion of syntax:** separate rules for syntactic analysis and synthesis introduced
4. **partial inclusion of semantics:** extra semantic rules to deal with ambiguities

Problems of MT

- Syntactical Ambiguities
  - structure of a sentence often depends on semantics, not only on the type of words
- Polysemy:
  - Polysems are words which have several similar meanings.
  - Proper translation is difficult to find even for an human translator.
- Homonymy:
  - Homonyms are several independent words which "share" the same linguistic corpse.
  - also difficult to translate, often depend on context and semantics
- Referential Ambiguity:
  - pronouns refer to certain words, but it is often not clear to which
  - references might even cross sentence boundaries
- Metaphors and Symbols:
  - metaphors and symbols depend on the underlying culture
  - often impossible to translate
  - idiomatic dictionaries help a bit
- Synonyms:
  - there are often several words with almost the same meaning
  - it is difficult to choose the right translation since it depends on context, style and semantics
- Fuzzy Hedges:
  - vague words, terms and expressions like "in a sense", "irgendwie"
  - language dependent and therefore difficult to translate

See also the practical part!

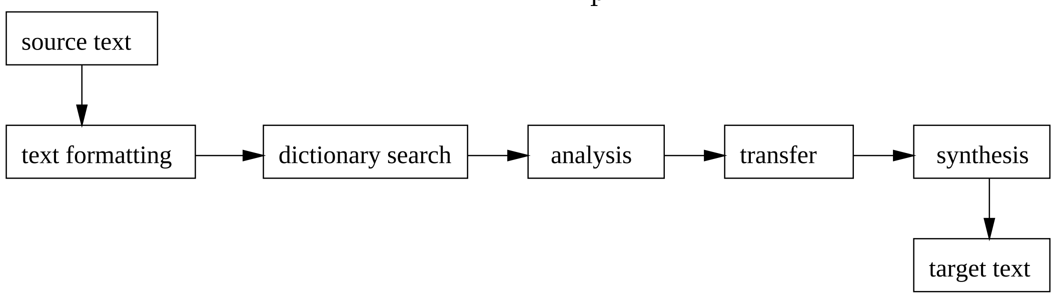
Usage

categories:

- human translation with machine support
- machine translation with human support
- fully automated translation

Fully Automated Translation

The translation process



text formatting: isolation of the sentences of a text, sorting of all words of the text in alphabetical order

dictionary search: searching of the words in different dictionaries: high frequency dictionary (contains the most frequent used words), main dictionary and idiomatic dictionary

analysis: syntactic analysis of the source sentence, determination of its structure

transfer: assigning of the words in target language to those in source language

synthesis: creating the sentences with the rules of the target language grammar

Research

Architectures of Translation Systems

- rule based paradigm:
  - basic functionality represented by rules
  - strategies: direct translation, transfer approach, interlingua approach
- data oriented paradigm:

- statistical MT
- example based MT

Machine Interpreting

This very young research field combines approaches from Machine Translation and Speed Technology. The level of difficulty is very high, because Machine Interpreting systems have to deal with phenomena of spontaneous language.

Artificial Intelligence

In this field of computer science, computers that can think and learn by themselves are tried to be developed. Maybe in future, if there are computers thinking like human beings, there will be Machine Translation systems whose results are of an equal quality like human translations.

3 Practical Part

Necessary WWW Addresses

Group 1: `http://www.systranlinks.com/systran/voila`
Group 2: `http://www.FreeTranslation.com/`
Group 3: `http://www.elingo.com/text/text.php3`

You will find links to these web sites at
`http://www.tu-chemnitz.de/~tisc/LaCo/examples.html`

Please do not forget to set the correct language combination (English → German)!

Exercise 1: Homonymy

example	I screened myself to escape discovery.
human translation	
translation group 1	
translation group 2	
translation group 3	

Exercise 2: Syntactical Ambigiuty

example	Flying planes can be dangerous.
human translation	
translation group 1	
translation group 2	
translation group 3	

Exercise 3: Referential Ambiguities

example	I missed my cat. It seemed to have disappeared.
human translation	
translation group 1	
translation group 2	
translation group 3	

Exercise 4: Fuzzy Hedges

example	High quality fully automatic machine translation is considered to be virtually impossible.
human translation	
translation group 1	
translation group 2	
translation group 3	

Exercise 5: New Developments

example	To access a remote computer, please use Secure Shell instead of Telnet.
human translation	
translation group 1	
translation group 2	
translation group 3	

Exercise 6

Think out a text of about five (not too easy) sentences (English or German), translate them by each of the three translation machines and write down the funniest (or most impressive) result.

original text:

best result:

Exercise 7

Please try to classify the three used online translation machines in the presented four groups (theoretical part).

FreeTranslation:

Voilà:

Elingo:

References

[1] D. Ufert / A. Harder (1994). *Einführung in die Datenverarbeitung für Linguisten und Literaturwissenschaftler.* Aachen: Verlag Shaker.

[2] M. Schwanke (1991). *Maschinelle Übersetzung*, Berlin Heidelberg: Springer-Verlag

[3] B. Schmitz (1997). *Pragmatikbasiertes Maschinelles Dolmetschen.* Dissertation. Berlin.

[4] W. Fleischer / W. Hartung / J. Schildt, P. Suchsland, ed. (1983). *Kleine Enzyklopädie Deutsche Sprache.* Leipzig: VEB Bibliographisches Institut.