Machine Translation (MT)

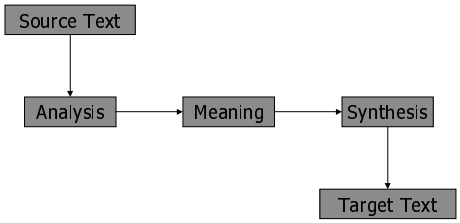
History, Theory, Problems and Usage

Introduction

- Please translate the following sentence into English:
  „In den letzten Jahren wurde die Forschung auf dem Gebiet der maschinellen Übersetzung vorangetrieben, da die Anzahl zu übersetzender Dokumente ständig steigt."
- Now try to imagine how a computer would do this translation without knowing its meaning.

2

Translation Process



3

Historical Overview

- four periods:
  - optimistic beginnings
  - disillusion
  - 70ies: partial successes
  - commercial application
- generations of translation techniques

4

History - Optimistic Beginnings -

- 1933 first machine supported translation systems (France, Russia)
- 1942 first computer -> condition for development of machine translation created
- very optimistic attitude towards the problems, high expectations
- 1952 first MT conference
- 1954 Georgetown Experiment -> enormous success, research intensified

5

History - Disillusion -

- problems got more clearly
- expectations not fulfilled
- 1964 founding of a committee to examine the future of MT
- 1966 ALPAC Report: very negative
- much less research funds
- interest in MT decreased

6

History
- 70ies -

- revival of MT research
- much more realism
- new research field: Computer Linguistics
- development of grammatical formalism

7

History
- Commercial Application -

- amount of documents, that have to be translated, steadily increasing
- MT now taken seriously and used commercially
- successful research especially in Japan
- today: MT systems widely spread
- limits of application are marked clearly

8

Generations of Translation Techniques

- word-by-word-translation
  - simple comparison of dictionaries
- partial inclusion of syntax
  - syntactical rules about the structure of sentences extracted from the dictionaries
- total inclusion of syntax
  - separate rules for syntactical analysis and synthesis
- partial inclusion of semantics
  - extra semantic rules to deal with ambiguities

9

Theoretical Background

- What is an adequate translation?
- Where does the meaning come from?
- now: selection of the many translation problems
- practical exercises

10

Problems of Machine Translation

- polysemy, homonymy
- syntactial ambiguities
- referential ambiguity
- Fuzzy Hedges
- synonyms
- metaphors and symbols
- new vocabulary developments

11

Problems of MT
- Polysemy -

- **Polysem:** A word whose meanings are diverged or radiated.
- the proper translation is difficult to find even for a human translator
- difficult to distinguish from Homonymy
- examples:
  - "walk" might mean "Spaziergang" or "Spazieren gehen"
  - "free" might mean "frei", "unabhängig" etc.

12

Problems of MT
- Homonymy -

- **Homonym:** Several independent words which "share" the same linguistic body.
- difficult to translate, often depends on context and semantics
- examples:
  - "Reif" might mean "ring", "bracelet" or "white frost"
  - "screen" might mean "Schirm", "Leinwand", "Tarnung", "Raster", "Abschirmung"

13

Problems of MT
- Syntactical Ambiguities -

- structure of sentence not only depends on type of words but often also on semantics
- *"Flying planes can be dangerous."* is ambigious, words can be grouped in two ways
  - "(Flying planes) can be dangerous." or
  - "(Flying) (planes) can be dangerous."

14

Problems of MT
- Referential Ambiguity -

- pronouns refer to certain words but it is often not obvious to which
- references might even cross sentence boundaries

15

Problems of MT
- Referential Ambiguity -

- examples:
  - *"I missed my cat. It seemed to have disappeared."* The computer needs to figure out which word is referenced by "it".
  - *"My cat was chasing a mouse. It played with it."*
    Everybody knows that the cat is playing with the mouse. But how is a computer supposed to know?

16

Problems of MT
- Fuzzy Hedges -

- vague words, terms and expressions
- very language dependent
- difficult to translate
- examples:
  - "in a sense", "irgendwie", "very"
  - *"High quality fully automatic machine translation is considered to be virtually impossible."*

17

Problems of MT
- Metaphors and Symbols -

- metaphors and symbols depend on the underlying culture and history
- often cannot be translated (Chinese sayings sometimes just do not make sense to non-Chinese people)
- idiomatic dictionaries may be used to ease translation

18

Problems of MT
- Metaphors and Symbols -

- example:
  - "Mit eiserner Miene feuerte er seinen treuesten Mitarbeiter."
  - corresponding English idiom: "with a stony expression"

19

Problems of MT
- New Developments -

- languages are dynamic
- new words created
- proper names of new technologies
- example:
  - Secure Shell
  - Telnet

20

Problems of MT
- Synonyms -

- there are often several words with almost the same meaning
- it is difficult to choose the right one
- it depends on context, style and semantics

21

Real World Usage

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

22

Human Translation with Machine Support

- fast & easy access to to dicitionaries, thesauri etc.
- system provides suggestions and alternatives for translation
- human translator decides
- extendable dictionaries

23

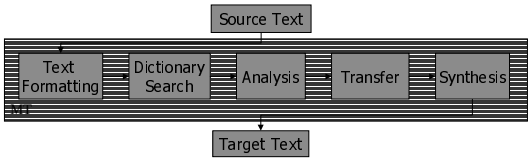
Machine Translation with Human Support

- computer controls the translation process
- computer asks the human if it comes across problems like ambiguities
- interaction takes place when the computer requests it

24

Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents
- pre- and postprocessing usually neccessary



25

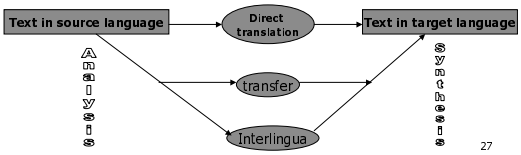
Current Research

- architectures of translation systems
  - rule based paradigm
  - data oriented paradigm
- machine interpreting
- Artificial Intelligence

26

Architectures: Rule Based Paradigm

- basic functionality represented by rules
- translation strategies:
  - direct translation
  - transfer approach
  - interlingua approach



27

Architectures: Data Oriented Paradigm

- statistical MT
  - starting point: each sentence of a language is a possible translation of a sentence in another language
  - works with assigned probabilities
  - no linguistic knowledge necessary
- example based MT
  - basis: former produced translations
  - analogue forming of new translations

28

Machine Interpreting

- very young research field
- approaches from MT and Speech Technology combined
- need to deal with phenomena of spontaneous language ("ähm")
- very high level of difficulty

29

Artificial Intelligence

- field of computer science
- try to develop computer systems that can think and learn by themselves
- Neural Networks
- future: will help to achieve better results in MT

30

Conclusion

Current machine translation systems are already very helpful, but not perfect. There are linguistic problems that cannot be satisfyingly solved by computers unable to think like human beings. Maybe in the future, further progress in Artificial Intelligence will help to solve the remaining problems.