Introduction to Computational Linguistics

Frank Richter

fr@sfs.uni-tuebingen.de.

Seminar für Sprachwissenschaft

Eberhard Karls Universität Tübingen

Germany

How to Choose the Best MT Strategy

- If low quality translation is acceptable and if source and target language have similar syntax, then a direct translation system may be acceptable.
- If the system will only translate between two languages and good-quality translation is necessary, a transfer system is all that is needed.
- If the system will have to translate among several languages, an interlingua approach may be preferable, especially if the languages are from the same language family and have similar patterns of word meanings.

The Impossibility of FAHQMT

The Impossibility of Fully Automatic, High Quality Machine Translation (FAHQMT):

Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy.

(Bar-Hillel 1959)

Machine Translation

- full machine translation (MT)

Machine Translation

- full machine translation (MT)
- human-aided machine translation (HAMT)

Machine Translation

- full machine translation (MT)
- human-aided machine translation (HAMT)
- machine-aided human translation (MAHT)

Full Machine Translation

- machine is responsible for the entire translation process

Full Machine Translation

- machine is responsible for the entire translation process
- minimal pre-processing by humans, if any

Full Machine Translation

- machine is responsible for the entire translation process
- minimal pre-processing by humans, if any
- no human intervention during the translation process

Full Machine Translation

- machine is responsible for the entire translation process
- minimal pre-processing by humans, if any
- no human intervention during the translation process
- post-processing by humans may be required

Human-aided Machine Translation (HAMT)

- machine is responsible for translation production

Human-aided Machine Translation (HAMT)

- machine is responsible for translation production
- translation process may be aided by human monitor; e.g. for:

Human-aided Machine Translation (HAMT)

- machine is responsible for translation production
- translation process may be aided by human monitor; e.g. for:
  - part-of-speech disambiguation

Human-aided Machine Translation (HAMT)

- machine is responsible for translation production
- translation process may be aided by human monitor; e.g. for:
  - part-of-speech disambiguation
  - resolving for phrase attachment

Human-aided Machine Translation (HAMT)

- machine is responsible for translation production
- translation process may be aided by human monitor; e.g. for:
  - part-of-speech disambiguation
  - resolving for phrase attachment
  - choosing appropriate word for the target language from a set of candidate translations

Machine-aided Human Translation (MAHT)

- human is responsible for translation production

Machine-aided Human Translation (MAHT)

- human is responsible for translation production
- human translation is aided by on-line tools; e.g. by

Machine-aided Human Translation (MAHT)

- human is responsible for translation production
- human translation is aided by on-line tools; e.g. by
  - a corpus of sample translations

Machine-aided Human Translation (MAHT)

- human is responsible for translation production
- human translation is aided by on-line tools; e.g. by
  - a corpus of sample translations
  - electronic dictionaries for source and target language

Machine-aided Human Translation (MAHT)

- human is responsible for translation production
- human translation is aided by on-line tools; e.g. by
  - a corpus of sample translations
  - electronic dictionaries for source and target language
  - a terminology database

Machine-aided Human Translation (MAHT)

- human is responsible for translation production
- human translation is aided by on-line tools; e.g. by
  - a corpus of sample translations
  - electronic dictionaries for source and target language
  - a terminology database
  - word processing support for text formatting

The History of Machine Translation (1)

1629 René Descartes proposes a universal language, with equivalent ideas in different tongues sharing one symbol.

1933 Russian Petr Smirnov-Troyanskii patents a device for transforming word-root sequences into their other-language equivalents.

1949 Warren Weaver, director of the Rockefeller Foundation's natural sciences division, drafts a memorandum for peer review outlining the prospects of machine translation (MT).

The History of Machine Translation (2)

1952 Yehoshua Bar-Hillel, MIT's first full-time MT researcher, organizes the maiden MT conference.

1954 First public demo of computer translation at Georgetown University: 49 Russian sentences are translated into English using a 250-word vocabulary and 6 grammar rules.

1960 Bar-Hillel publishes his report arguing that fully automatic and accurate translation systems are, in principle, impossible.

The History of Machine Translation (3)

1964 The National Academy of Sciences creates the Automatic Language Processing Advisory Committee (Alpac) to study MT's feasibility.

1966 Alpac publishes a report on MT concluding that years of research haven't produced useful results. The outcome is a halt in federal funding for machine translation R&D.

The History of Machine Translation (4)

1968 Peter Toma, a former Georgetown University linguist, starts one of the first MT companies, Language Automated Translation System and Electronic Communications (Latsec).

1969 In Middletown, New York, Charles Byrne and Bernard Scott found Logos to develop MT systems.

Machine Translation Systems

North America and Canada

- SYSTRAN
  - Originated from GAT (Georgetown Machine Translation project)
  - Founded in 1968 by Peter Toma, a principal member of the GAT project
  - Versions for English, German, Russian, French, Spanish, Dutch and Portugese
  - Purchased by Major Corporations and Government Agencies for further development, including General Motors, Xerox, Siemens, European Commission

Machine Translation Systems

- TAUM-METEO
  - TAUM: Traduction Automatique de l'Universtité de Montreal
  - Fully-automatic MT system METEO
  - Fully integrated into the Canadian Meteorological Center's (CMC) nation-wide weather communications network by 1977
  - Translates appr. 8.5 million words/year with 90-95% accuracy. Mistakes mainly due to misspelled input or unknown words

Machine Translation Systems: Europe

- EUROTRA
- Long-term MT research and development programme funded by the European Commission (1982-92)
- EUROTRA 1 - Research and development programme (EEC) for a machine translation system of advanced design, 1982-1990
- EUROTRA 2 - Specific programme (EEC) concerning the preparation of the development of an operational EUROTRA system, 1990-1992

MT Systems: EUROTRA 1

- EUROTRA 1 - Research and development programme (EEC) for a machine translation system of advanced design, 1982-1990
  - Main Goal: To create a machine translation system of advanced design capable of dealing with all (nine) official languages at the time (Danish, Dutch, English, French, German, Greek, Italian, Spanish and Portuguese) of the Community by producing an operational system prototype in a limited field and for limited categories of text, which would provide the basis for subsequent development on an industrial scale.

MT Systems: EUROTRA 2

- EUROTRA 2 - Specific programme (EEC) concerning the preparation of the development of an operational EUROTRA system, 1990-1992
  - Main Goal: To create, starting from the EUROTRA prototype, the appropriate conditions for a large-scale industrial development, including the development of methods and tools for the re-usability of lexical resources in computer applications as well as the creation of standards for lexical and terminological data.