Machine Translation technologies - where we are at the moment ? - discussion points-

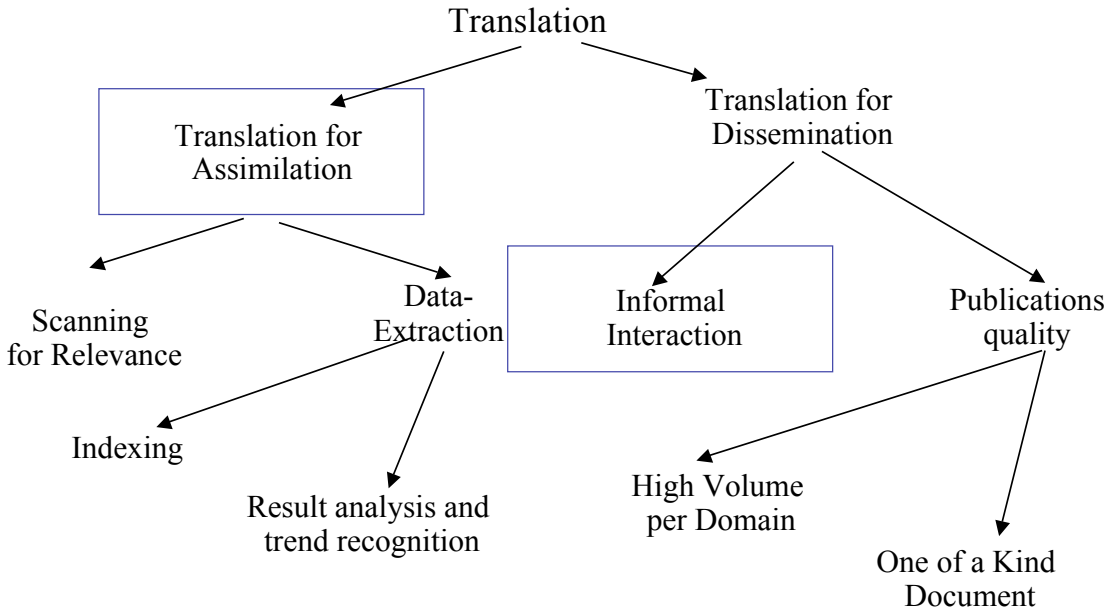
Cristina Vertan, Vladislav Kubon

Requirements of a Machine Translation System

Features, which we expect (at least) from MT systems:

- Semantic adequacy
- Stylistic and pragmatic adequacy
- Cultural adequacy
- Consistency inside a text and between texts
- Reduced costs compared to human translations
- High speed

Functional Typology of MT-Systems



Types of Machine Translation Engines

- Rule-based MT-Systems
- Knowledge-based MT
- Corpus-based MT
  - Statistical
  - Example-based
- MAT technologies, translation memories

Evaluation of Machine Translation Systems

- In contrast to other software there is no “best solution” by human translators, which can be compared with the output of the system
- I.e., for one input sentence there are many different correct translations
- Quality measurement of an MT System depends on its purposes and on the requirements of potential users.

On-line MT-Systems - Requirements

- Must cover a large number of language pairs
- Either cover a broad spectrum of domains, or have to be easy adapted to other domains
- Easy to be manipulated by users without expertise (no parameter tuning) - fully integrated black-box
- Not perfect translation but far beyond „word-to word“ quality

Limitations

- Rule-based systems need high-quality linguistic modules which cannot be integrated in on-line systems
- KB-systems are domain dependent and make use of large ontologies
- Corpus-based MT-systems require training on large domain-dependent parallel corpus (for few languages available)
- Most part of the available systems do not overcome the „word-to-word“ quality
- Embedded lexicons are quite limited

Examples -1-

- *„Verfahren zur Identifizierung von Sprechern spielen in verschiedene Anwendungsgebieten eine wichtige Rolle. Hierzu gehören :.....“*
- Procedures for the speaker identification play an important role in different areas of application. To these belong:...“
- FreeTrans: „Procedures to the identification of speakers play an important role in different use zones. To this belong :...“
- Systran: „Procedures for the identification of speakers play an important role in different areas of application. To it belong:...“

Examples -2-

- *Wenn man ein Referat mit der Seminarleiterin gut genug bespricht, erspart man sich viel Arbeit, Mißverständnisse und Kritik beim eigentlichen Vortrag. Schon kurz nach der Übernahme des Referats sollte man die ersten Arbeitsschritte tun (auch wenn man die eigentliche Arbeit dann erst später anfängt), um mögliche Schwierigkeiten gleich zu erkennen und Rückfragen zu starten. Bei der Übernahme des Referats alles mitschreiben. An viele wichtige Tips kann man sich schon am nächsten Tag nicht mehr erinnern.*
- FreeTrans: If one discusses a report with the seminar leader well enough, one saves himself much work, misunderstandings and criticism in the actual presentation. Already shortly after the acceptance of the report, one should do the first steps (also if one the actual work then first later begins) in order to recognize possible difficulties equally and to start checkback. In the acceptance of the report all with letter. At many important hints, one can remember already the next day no longer.
- Systran : If one discusses a paper with the seminar chief well enough, one saves much work, misunderstandings and criticism with the actual lecture. Already briefly after the assumption of the Referats one should do the first work procedures (even if one begins the actual work only later then), in order to recognize possible difficulties directly and start further inquiries. Write with the assumption of the Referats of everything. Many important tips one cannot remember on the next day more.

How to integrate Semantic Web Technologies ?

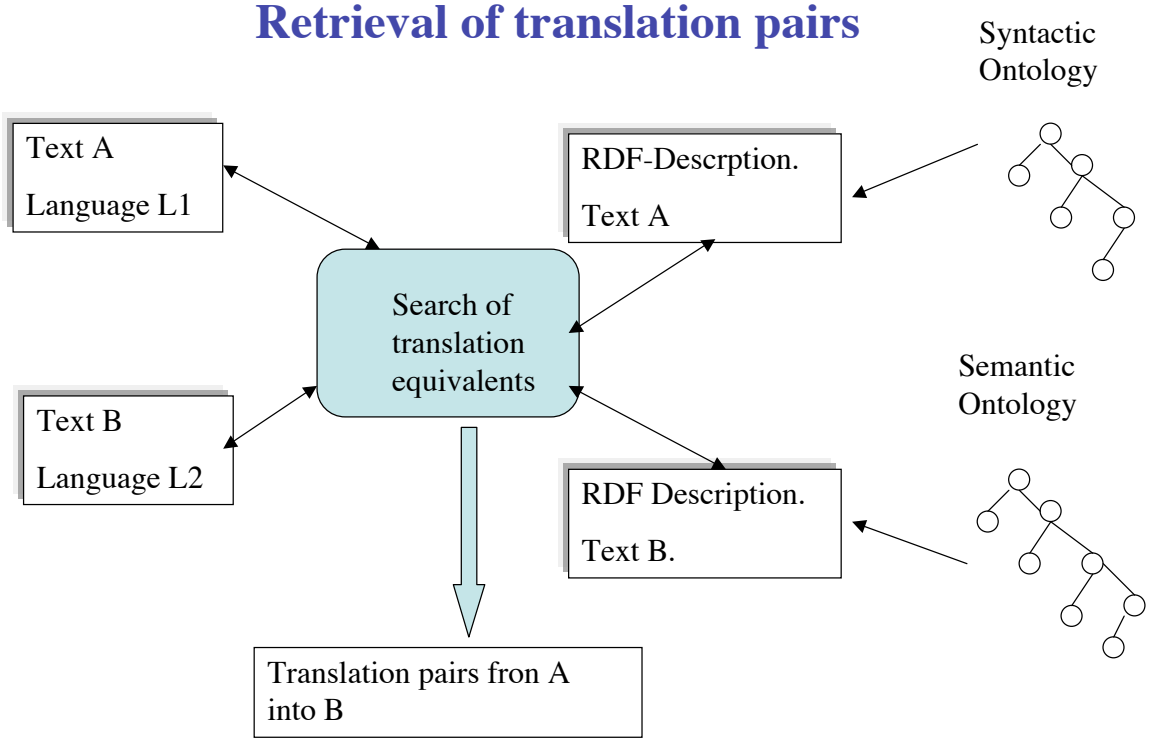
Use of annotated on-line Resources for MT

RDF-Annotation of databases for (corpus)example-based MT Properties related to:

- Translation equivalents of words/expressions
- Transfer rules for syntactic stuctures
- Semantic classes for the candidate solutions

- Such properties can be defined in different languages and then mapped on a linguistic ontology (syntactic or semantic)

Retrieval of translation pairs



Open Questions

- Are there any implications , and which ones, for rule-based approaches ?
- Can the OWL -encoded ontologies improve the KBMT approach ?
- Are Topic Maps more appropriate ?
- Can Statistical MT also benefit from SW?
- How can we integrate SW technologies in on-line MT Systems ?

- What can we do to insure user acceptance of SW ?
  - MT applications can help ???
- OWL vs. other existent KR languages
- To which extent the new approaches to MT already include SW technologies (ontologies, what else) ?
- SW and open source ?