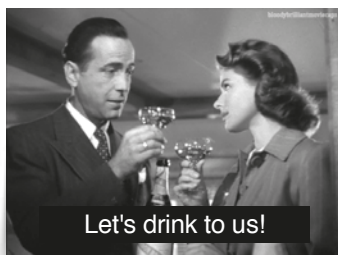




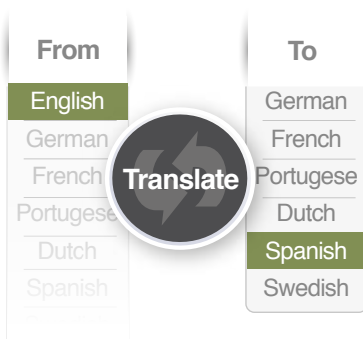
An Online Service for
Subtitling by Machine
Translation

How it works:

Upload subtitle files



Choose language and machine translate



Download translated file & post-edit



Background

Subtitling is the preferred multimedia content translation method in most European countries, and for most genres, ensuring that audiovisual content is widely accessible across languages.

The increasing use of multilingual multimedia through the internet, the popularity of DVDs, and the current European policies promoting linguistic diversity and audiovisual accessibility, have all raised the demand for subtitling in recent years.

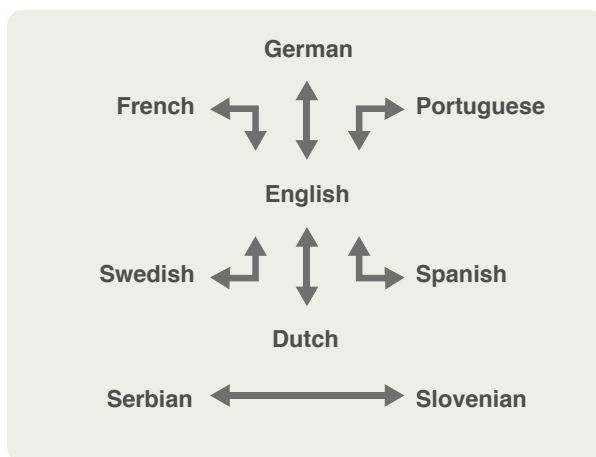
There is a clear need to optimise the productivity of current subtitle translation workflow processes, reducing costs and turnaround times while enhancing the consistency of the translation results.

Goals

SUMAT aims to increase the efficiency of professional subtitle translation through the introduction of statistical machine translation technology.

We are developing an online subtitle translation service for **9 European languages** combined into **14 language pairs**.

The language pairs



Why use MT technology?

Machine translation uses software to translate text from one natural language to another.

Statistical Machine Translation (SMT) is a way of generating translations on the basis of statistical models derived from the analysis of bilingual and monolingual text corpora.

The SMT suits subtitles because:

- Subtitles are short, grammatically sound, textual units, whose linguistic properties fit well with the state-of-the-art SMT models.
- The approach promotes the reusability of existing and new translations as training data.

The rising use of post-editing

The translation industry is embracing post-edited translation in domains where there are enough parallel bilingual corpora to customise machine translation engines.

This means that, for trained human translators, post-editing translation is an increasingly useful method that has been shown to achieve higher productivity than human translation alone.

The SUMAT approach

Build customised SMT engines for subtitles, trained on large professional-quality parallel and monolingual subtitle corpora.

To evaluate the merits of this approach by:

1. Having professional subtitle translators judge the quality of machine-translated subtitles through quality ranking scales.
2. Measuring the productivity gain achieved by post-editing machine-translated subtitles, compared to starting the translation process from scratch.

Project milestones

Corpora

For each of the language pairs in the project, large amounts (ca. 1 million subtitles on average) of professional quality parallel subtitle corpora have been collected and prepared for SMT training purposes.

Experiments

Various technical approaches have been explored with the aim of improving SMT performance:

- Subtitle vs. sentence alignment
- Factored and syntax-based models
- Named Entity Recognition & Compound Splitting
- Augmented phrase-tables
- Mixed models for translation domain adaptation

Online service

A prototype online service has been developed and is currently being refined. The final service will be based on the requirements and specifications provided by professional users in the consortium.

Evaluation

Evaluation by professional subtitle translators is under way. Two evaluation rounds are foreseen:

Round 1: Subtitle translators are scoring individual subtitles and categorising the errors found with the aim of analysing the quality of the SMT outputs. Their feedback is being used to refine the SMT engines.

Round 2: The productivity gain that can be achieved through the use of the SUMAT approach will be measured.

Results

Evaluation results and the Online Service will be finalised by 01 2014.

Project partners



TITEL ► BILD



CAPITA



Subcontractor

text) (shuttle
language technology for the media industry

Contact

Dr. Arantza del Pozo
adelpozo@vicomtech.org
Vicomtech IK4
Donostia San Sebastian
Spain



info@sumat-project.eu
www.sumat-project.eu



 Find us on LinkedIn:
<http://www.linkedin.com/company/sumat>

The work leading to these results has received funding from the European Community under grant agreement n° 270919.