The PASSAGE project: Standard German Subtitling of Swiss German TV content

Pierrette Bouillon, Johanna Gerlach, Jonathan Mutal, Marianne Starlander

(pierrette.bouillon, johanna.gerlach, jonathan.mutal, marianne.starlander)@unige.ch
Faculty of translation and interpreting, University of Geneva

Motivation
Swiss German
- A collective term for all the regional dialects
- Spoken everyday language
- No standard written form

Swiss TV
- Swiss German is mainly used in Swiss TV
- Contents in Swiss German are subtitled into Standard German

Data
Syntactic transformation

Human Normalised Transcriptions
98k segments

Original Subtitles
101k segments

ASR
78k segments

Synthetic training data
18k segments

Automatic Alignment

Post-edited Transcriptions
21k segments

Manual Alignment

Post-editing

Test Data

Human Normalised Transcriptions
- Post-Edited Transcriptions

Test Suite
50 ASR segments with syntactic phenomena

Results
For NMT, the addition of targeted synthetic training data improves the result

NMT system was slightly better than the APE system (BLEU 64.91 vs 61.49)

Motivation

60% of the population speak Swiss German

Swiss German
- A collective term for all the regional dialects
- Spoken everyday language
- No standard written form

Swiss TV
- Swiss German is mainly used in Swiss TV
- Contents in Swiss German are subtitled into Standard German

PASSAGE is a collaboration between the University of Geneva, SRF (Schweizer Radio und Fernsehen) and recapp IT. This project has received funding from the Initiative for Media Innovation based at Media Center, EPFL, Lausanne, Switzerland.

"Und d'Regierig hät no wiiteri Idee zum d'Stüüre abetue."