LEAD-ME

Funded by the Horizon 2020 Framework Programme of the European Union

SWPS UNIVERSITY

UNIVERSITY OF WARSAW

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Fully automatic vs. human subtitling: comprehension, user experience and perception

Alina Karakanta, Anita Kwiatkowska, Vilelmini Sosoni, Rocío Varela, Claudia Wiesinger

LEAD-ME Summer Training School Warsaw 2021 (5-9 July 2021)
Eye tracking in media accessibility research - methods, technologies and data analyses

Introduction & Background

Fully automatic subtitling = Speech Translation + Auto-spotting + Automatic segmentation

Automatic subtitling is here to stay

- Chrome (Scharff, E. and Kompalli, M. 2021)
- Youtube (Alberti, C. and Bacchiani M. 2009)
- ST integration in subtitling tools: MateSub

Some studies on productivity of MT for subtitles (Bywood et al 2017, Matusov et al. 2019, Koponen et al. 2020)

But is fully automatic subtitling good enough to be used raw?

Research question & hypotheses

RQ1: Do fully automatic subtitles require more processing effort on the part of viewers than human subtitles?

H1: Users spend more time looking at the subtitles instead of the image when the subtitles are produced automatically.

RQ2: What is the impact of fully automatic subtitles on perception and comprehension?

H2: The perception and comprehension scores of the participants in the fully automatic condition are lower compared to the participants in the human condition.

Study design

Independent variables

- fully automatic subtitles vs. human subtitles

Dependent variables

- comprehension > questionnaire
- user experience > fixation duration (subtitle vs image) > questionnaire
- perception > questionnaire

Confounding variables

- level of proficiency in source language
- level of proficiency in target language
- preferred audiovisual translation methods

Method

Procedure

- between-subject design
- 2 studies, one clip
  - Fully automatic subtitles > generated with Matesub
  - Human subtitles > produced by professionals following a three-step QA process
- random assignment of one condition per participant via randomized link

Research materials

- TED talk (0:00-2:44) - German original with English subtitles
  - Stadtlücken - Wem gehört die Stadt? | Christine von Raven | TEDxStuttgart
- Questionnaire
  - 5 general questions
  - 4 comprehension questions
  - 3 attention questions
  - 5 perception questions (1-5 rating scale)

Funded by the Horizon 2020 Framework Programme of the European Union

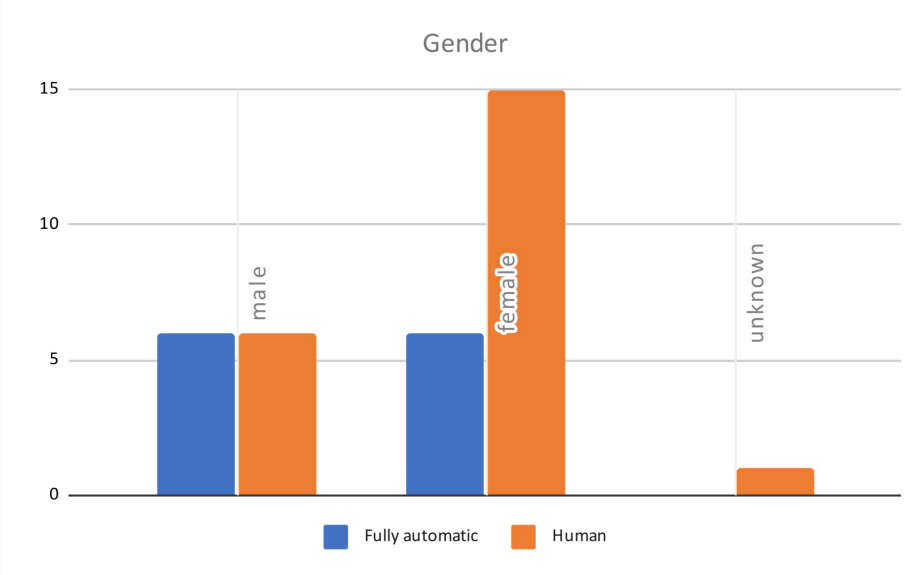
Method

Eye tracking equipment

- experiment conducted in RealEye from 7 July - 8 July 2021
- Subtitle AOI: 1 AOI occupying 14% of the screen, starting from the time the first subtitle appears

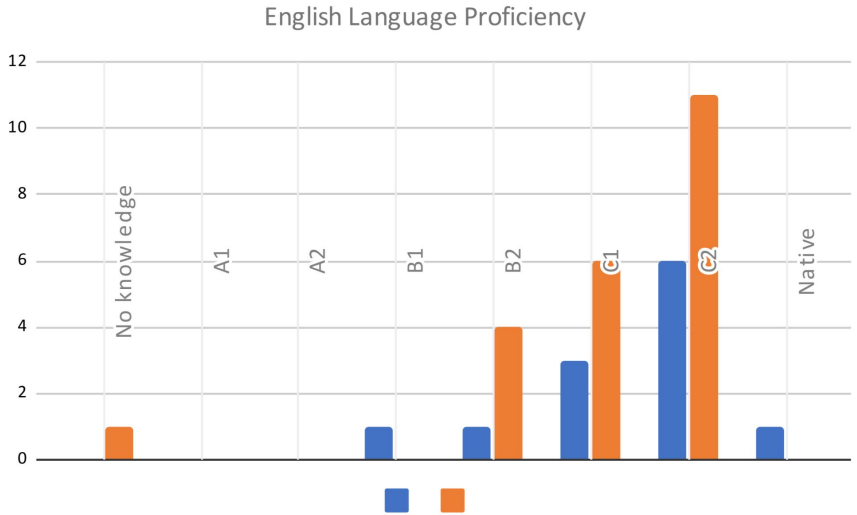
Participants

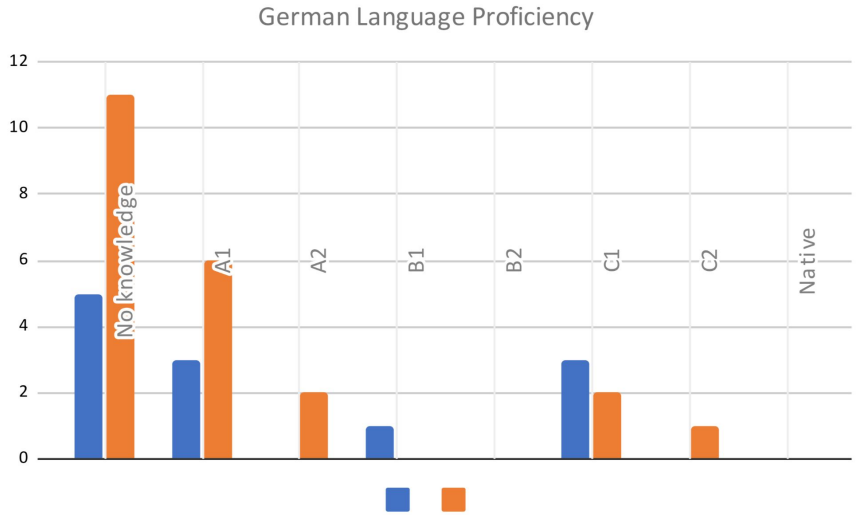
- 12 participants for fully automatic condition
- 22 participants for human condition



Participants

- 12 participants for fully automatic condition
- 22 participants for human condition



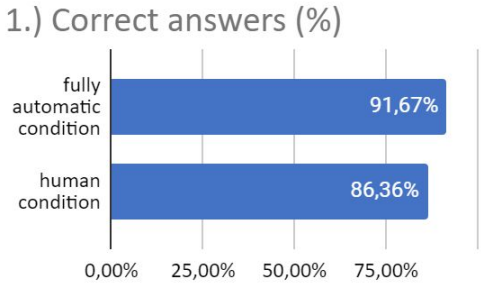


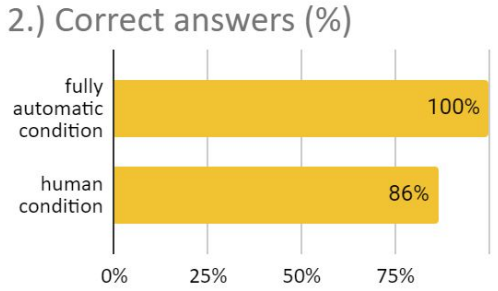
Questions on attention

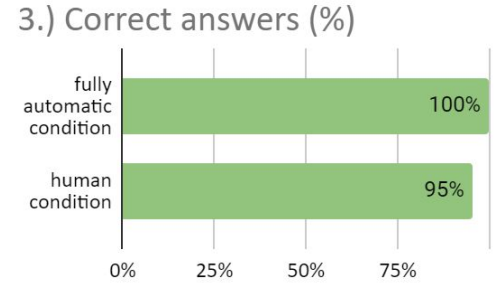
1.) What colour was the speaker's shirt?
a. Red.
b. Blue.
c. Green.
d. I don't know.

2.) What was the speaker's hair colour?
a. Brown.
b. Black.
c. Blonde.
d. I don't know.

3.) Which picture did the speaker use at the beginning of the talk?
a. A picture of a parking lot.
b. A picture of a highway.
c. A picture of a building.
d. I don't know.

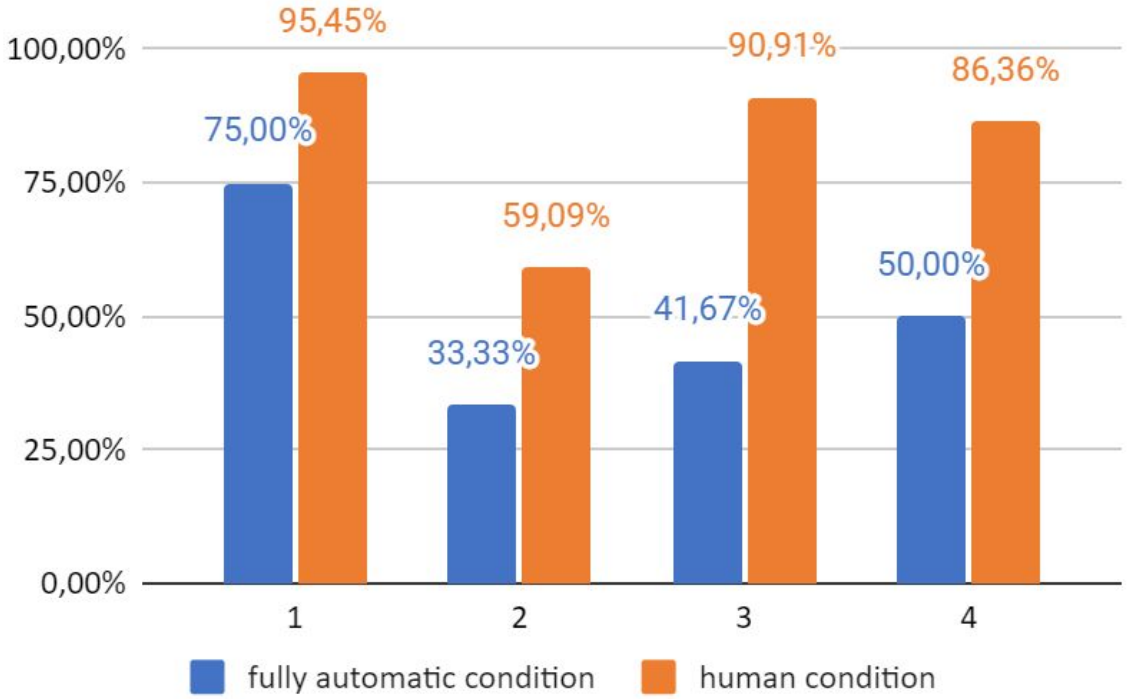






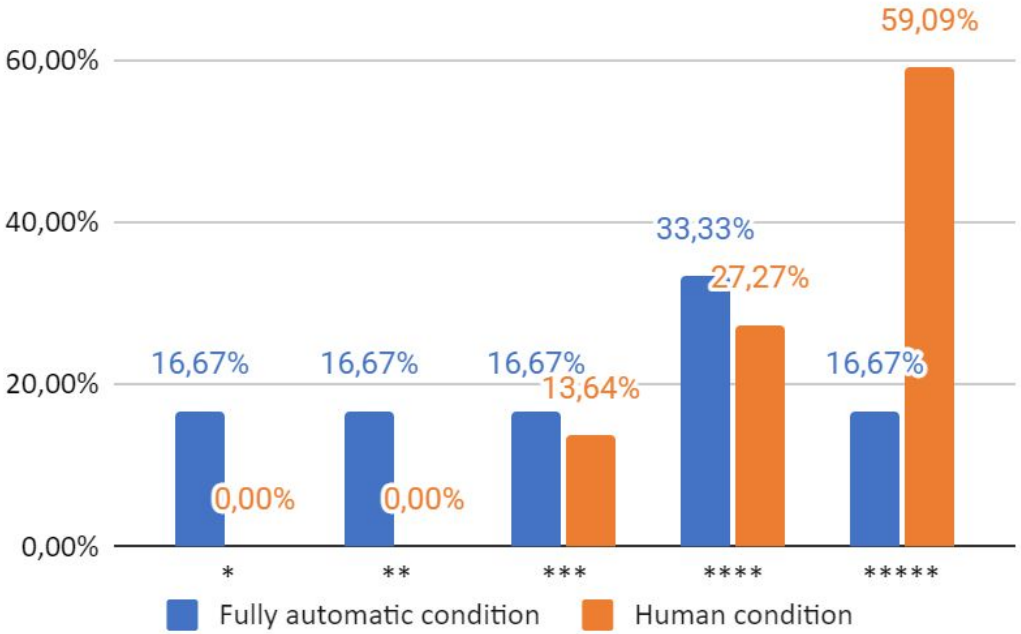
Questions on comprehension (% of correct answers)

1.) What is the topic of this talk?
2.) What is the problem in cities nowadays?
3.) Why is the parking lot not being used?
4.) According to the speaker, what is an activity that might take place at the unused parking lot?



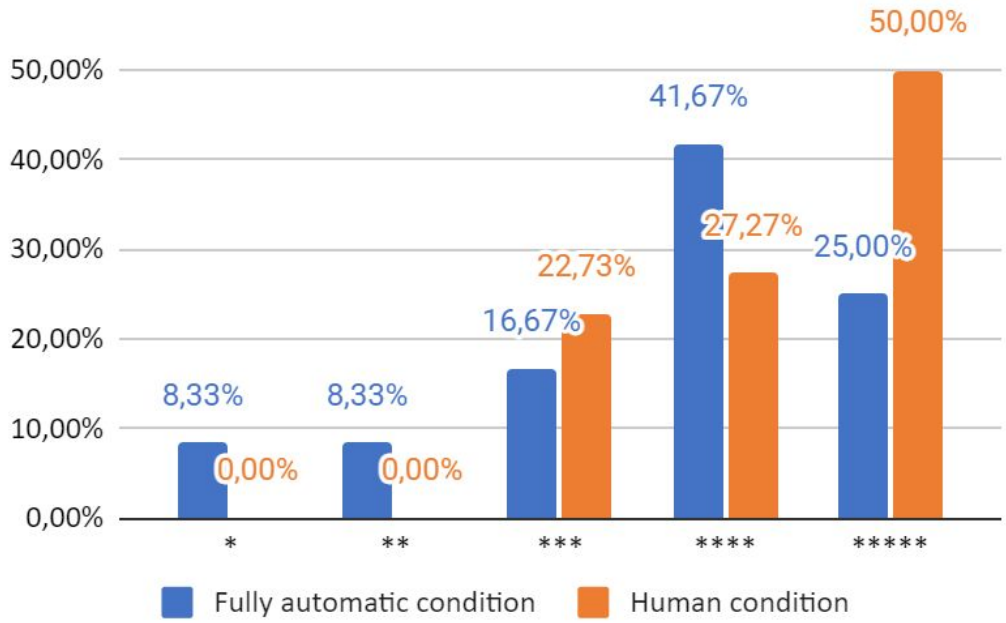
Questions on perception

1.) How would you rate the quality of the subtitles? From 1 (Terrible) to 5 (Excellent).



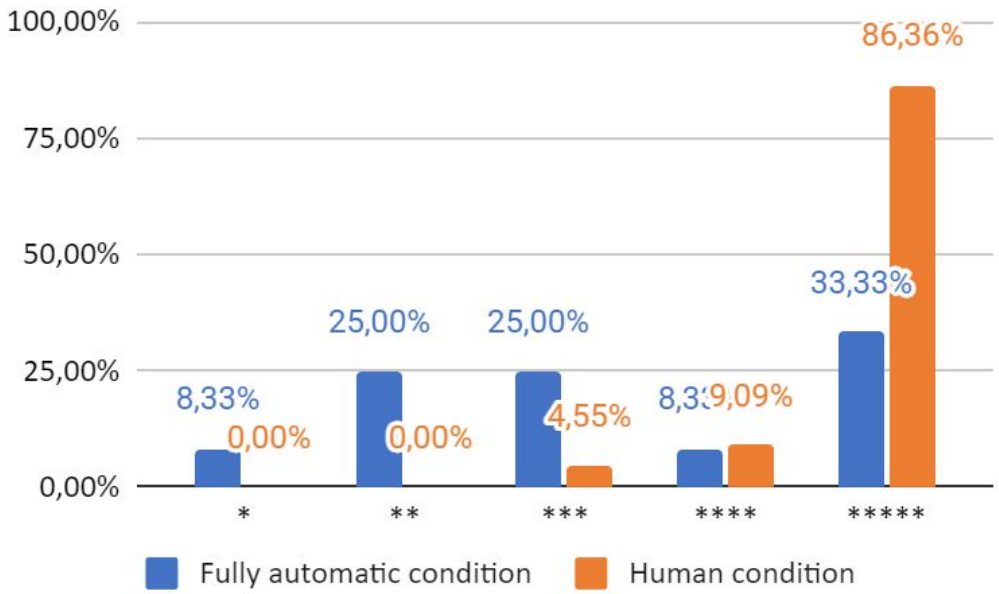
Questions on perception

2.) How would you rate the time available to read the subtitles? From 1 (Insufficient) to 5 (Sufficient).



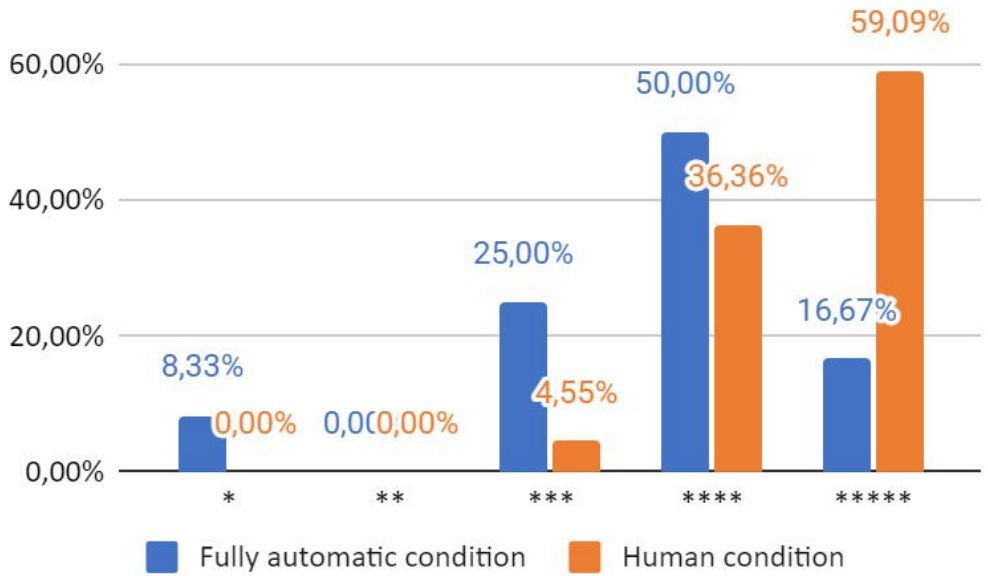
Questions on perception

3.) How comprehensible were the subtitles? From 1 (totally incomprehensible) to 5 (totally comprehensible).



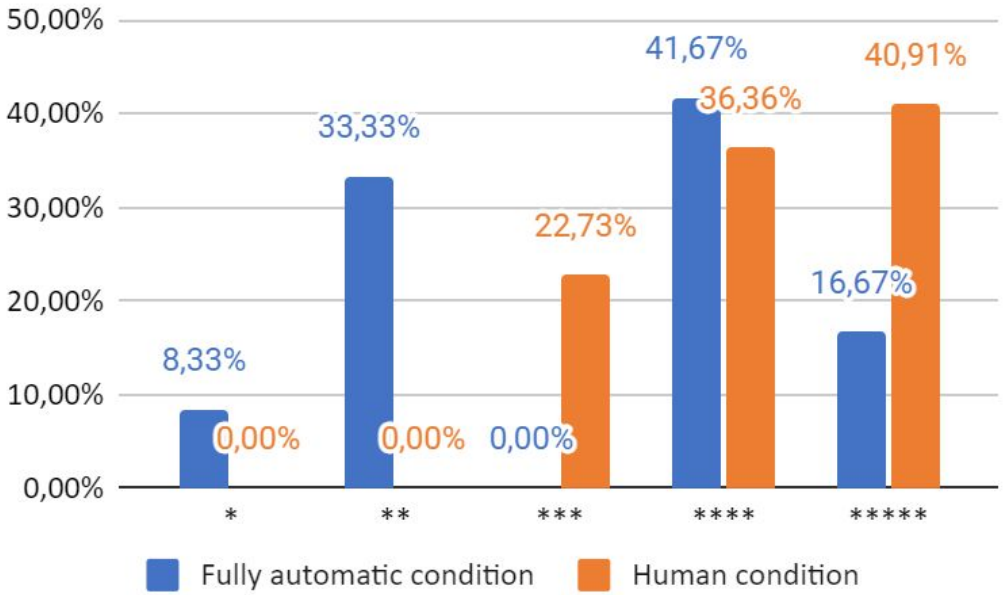
Questions on perception

4.) How synchronised were the subtitles? From 1 (completely unsynchronised) to 5 (completely synchronised).



Questions on perception

5.) How easy were the subtitles to read? From 1 (not easy at all) to 5 (very easy).



Results - RQ1

H1: Users spend more time looking at the subtitles instead of the image when the subtitles are produced automatically.

Results:
In general, participants spend slightly less time looking at the subtitles compared to the image (47%)

Participants spend the same time looking at the subtitles regardless of method of production

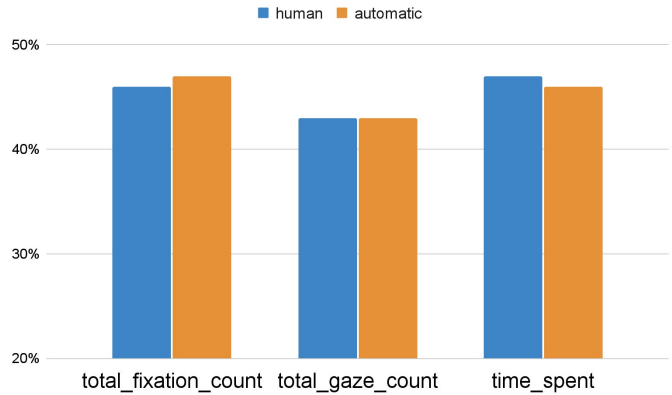


Figure: Percentage of fixations, gazes and time spent on subtitles compared to image

Results - RQ1

H1: Users spend more time looking at the subtitles instead of the image when the subtitles are produced automatically.

Human condition

Automatic condition

Results - RQ2

H2: The perception and comprehension scores of the participants in the fully automatic condition are lower compared to the participants in the human condition.

Results:

Rating:

- Excellent/very good quality
  - Human subtitles: 86.3% of users
  - Fully automatic subtitles: 50 % of users
- Sufficiently/well timed
  - Human subtitles: 77.3% of users
  - Fully automatic subtitles: 66.7% of users
- Totally/very comprehensible
  - Human subtitles: 95.5% of users
  - Fully automatic subtitles; 41.6%
- Not easy at all/not very easy
  - Human subtitles: 0% of users
  - Fully automatic subtitles: 41.7% of users
- Fully/very much synchronised
  - Human subtitles: 96% of users
  - Fully automatic subtitles: 66% of users

The hypothesis appears to be confirmed.

Comprehension questions:

The comprehension scores are higher in the human condition for all four of the comprehension questions.

Discussion & Limitations

Discussion

Automatic subtitles do not distract viewers more from the image than human subtitles.

It seems that viewers ignore subtitles when their quality is very low.

Limitations

- different number of participants in the two conditions
- 1 AOI for all subtitles

Conclusion & Future work

Conclusion

Fully automatic subtitling performed better than expected but is still not ready to be used raw.

Future work

More detailed analysis of eye-tracking data

Correlation of eye-tracking data with

- User profiles
- Translation quality (BLEU, TER)

Analysis of crisis points (translation errors, synchronisation, segmentation, reading speed)

References list

- Maarit Koponen, Umut Sulubacak, Kaisa Vitikainen,and J ̈org Tiedemann. 2020. MT for subtitling: User evaluation of post-editing productivity. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT 2020)*, pages 115–124
- Lindsay Bywood, Yota Georgakopoulou, and Thieery Etchegoyhen (2017). Embracing the threat: machine translation as a solution for subtitling. *Perspectives: Studies in Translatology*, 25(3): 492–508.
- Evgeny Matusov, Patrick Wilken, and Yota Georgakopoulou. 2019. Customizing neural machine translation for subtitling. In *Proceedings of the Fourth Conference on Machine Translation* (Volume 1: Research Papers), pages 82–93, Florence, Italy.Association for Computational Linguistics
- Scharff, E. and Kompalli, M. (2021). Chrome Can Now Caption Audio and Video. Google – The Keyword. Retrieved 8 July 2021 (https://blog.google/products/chrome/live-caption-chrome/).
- Alberti, C. and Bacchiani M. (2009). Automatic Captioning in YouTube. Google AI Blog. Retrieved 8 July 2021 (http://ai.googleblog.com/2009/12/automatic-captioning-in-youtube.html).