MONIKA SZCZYGIELSKA, ŁUKASZ DUTKA

Live subtitling through Automatic Speech Recognition vs. Respeaking:

between technical possibilities and users' satisfaction

Berlin 2016

Fundacja WIDZIALNI.org strony internetowe bez barier

dostępni.eu

UNIWERSYTET WARSZAWSKI

OVERVIEW

1. OUR BACKGROUND
2. QUALITY ASSESMENT IN RESPEAKING
3. ASR VS. RESPEAKING STUDY
4. RECEPTION STUDY
5. CONCLUSIONS

THE FIRST TESTS OF LIVE SUBTITLING 2013

POLISH LIVE SUBTITLES…

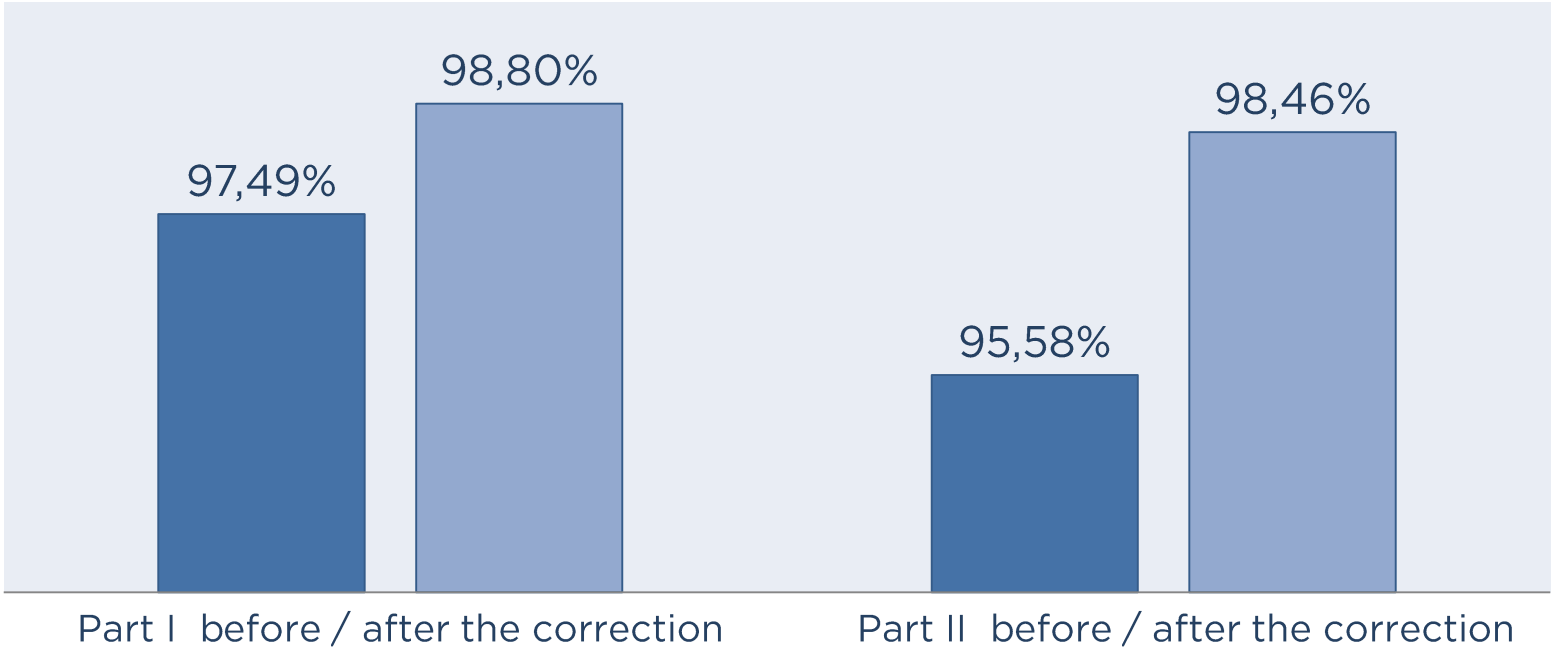
LIVE SUBTITLING QUALITY

- NER model prepared by a team of researchers from Roehampton University, used since 2013 to examine the quality of live subtitling in Ofcom reports.
- **NER > 98%** as a threshold for good quality live subtitling according to Ofcom
- We use the NER model since 2015 to asses the quality of our subtitles.

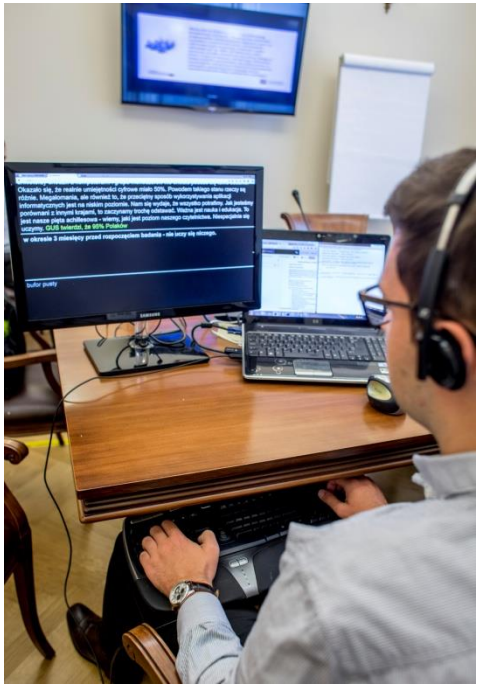
WIDZIALNI FOUNDATION EVENT 2015

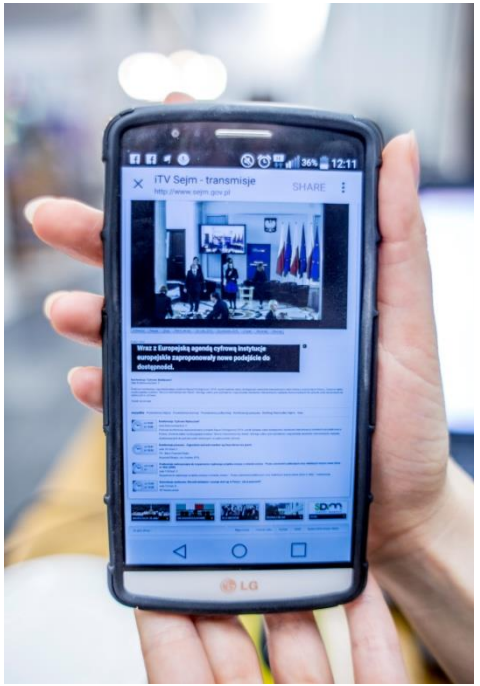
Widzialni Foundation conference "Digitally Excluded" in Warsaw, March 2015

POLAND: THE QUALITY OF SUBTITLES DURING "DIGITALLY EXCLUDED" CONFERENCE 17.03.2015



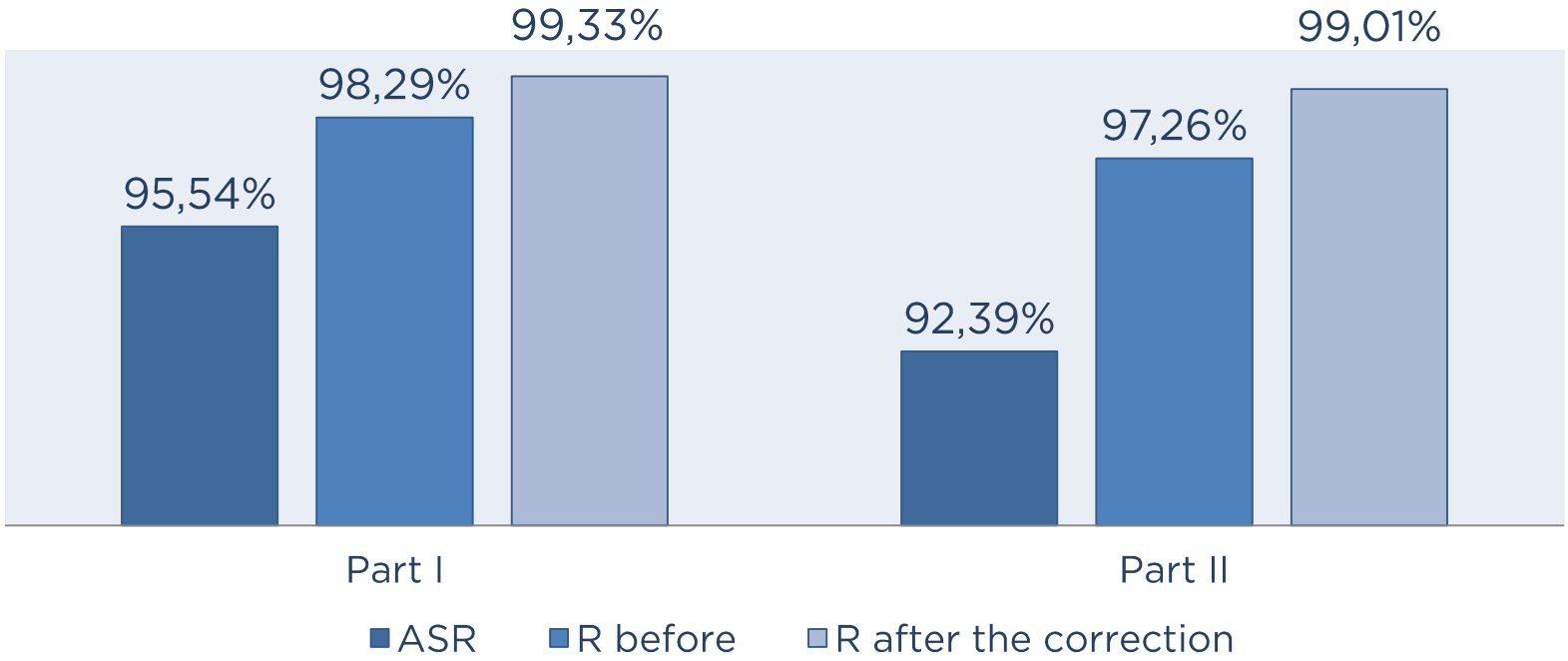
WIDZIALNI FOUNDATION EVENT 2016



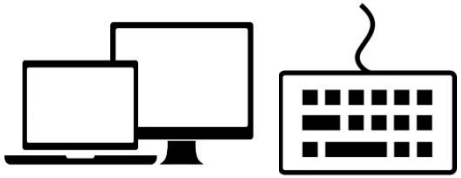


Widzialni Foundation "Digitally Excluded" conference in Warsaw, April 2016

POLAND: THE QUALITY OF SUBTITLES DURING "DIGITALLY EXCLUDED" CONFERENCE 20.04.2016

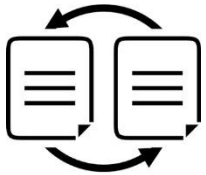


AUTOMATIC SPEECH RECOGNITION VS. RESPEAKING



ASR vs. ASR after correction

Model NER



Users' opinion

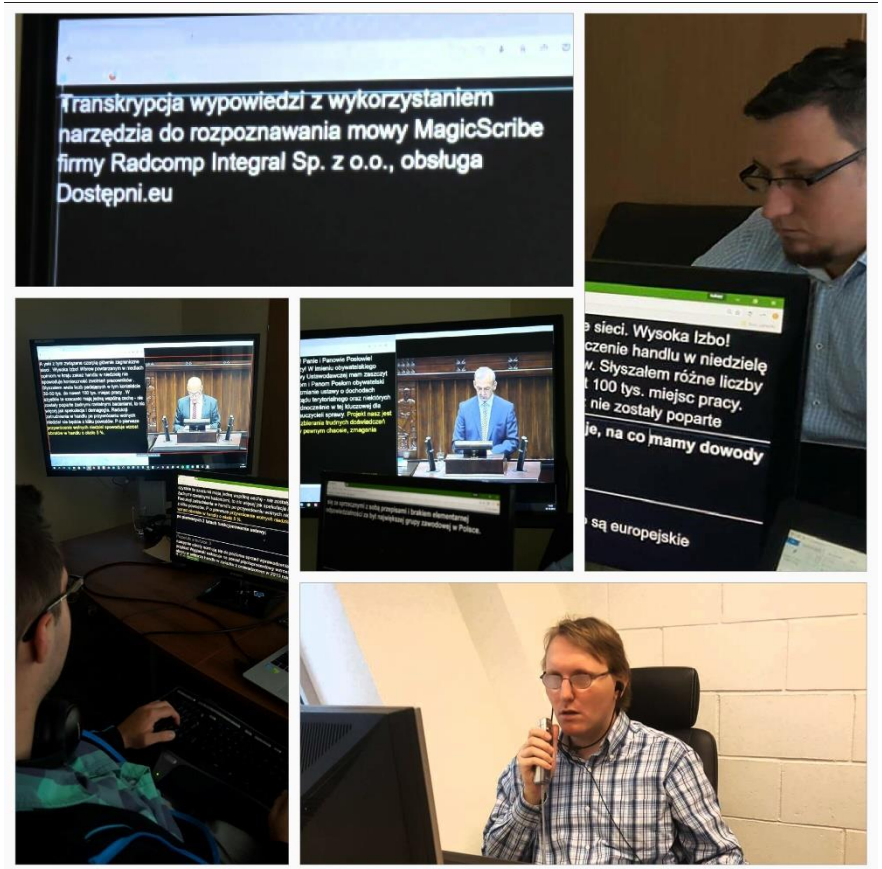
ASR after correction vs. Respeaking after correction

STUDY SET-UP

MagicScribe software developed by Radcomp Integral

The same SR system used for respeaking and by the Polish parliament to prepare transcriptions

Team: 1. blind respeaker (simultanous interpreter), 2. corrector (subtitler, simultanous interpreter)



radcomp INTEGRAL

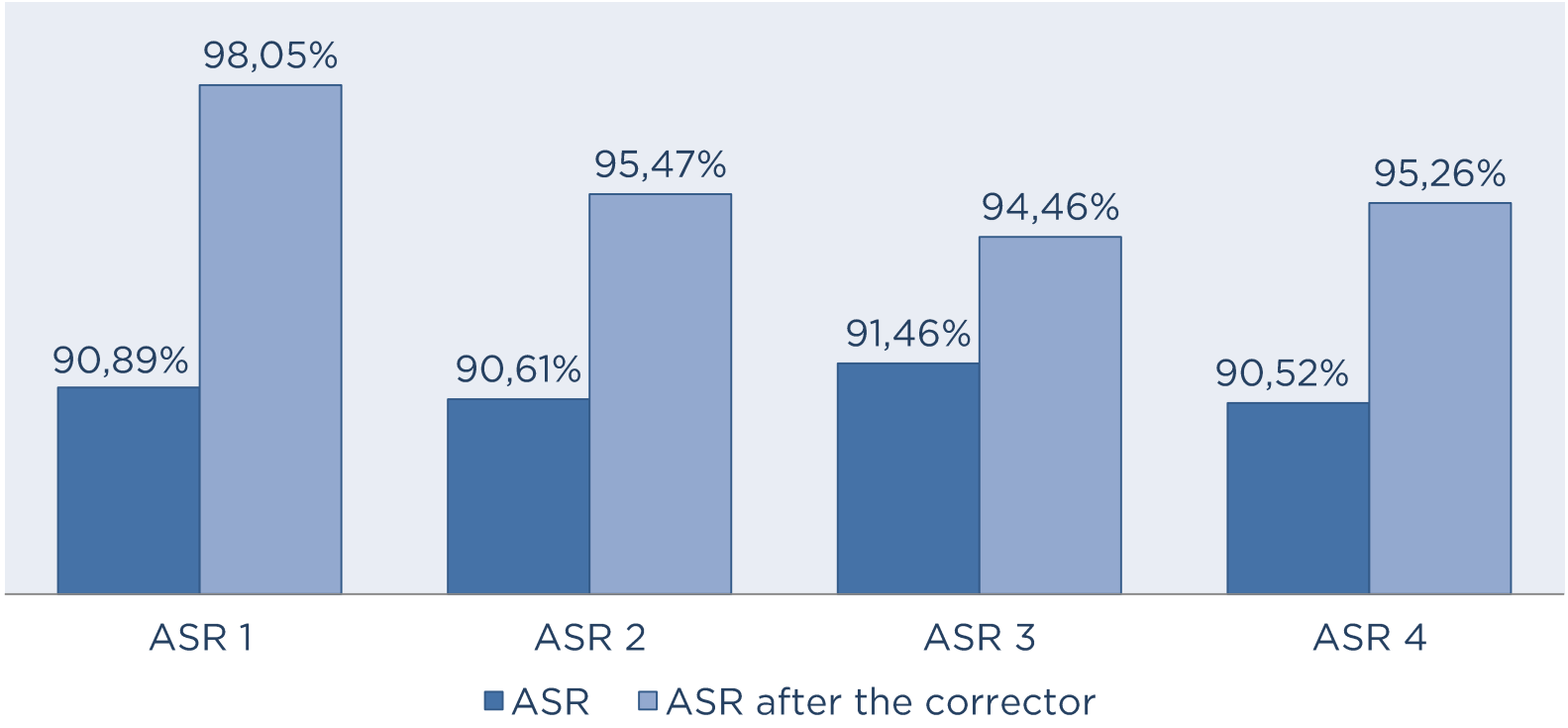
SEJM RZECZYPOSPOLITEJ POLSKIEJ

STUDY MATERIALS

- 2 speeches from the plenary debate at the Polish parliament,
- speeches on topics that weren't controversial or emotional, politicians that aren't well-known
- 2 excerpts from each speech (4 video recordings)
  - same lenght – aprox. 10 minutes each,
  - speed – 2 slow, 2 fast
  - complexity

    out of 7 levels, the easier excerpt was level 3-4 [secondary education]
    the more difficult 5-6 [graduate or postgraduate education]
    (two readability measurements: Logios, Jasnopis – adaptations of FOG index for Polish)

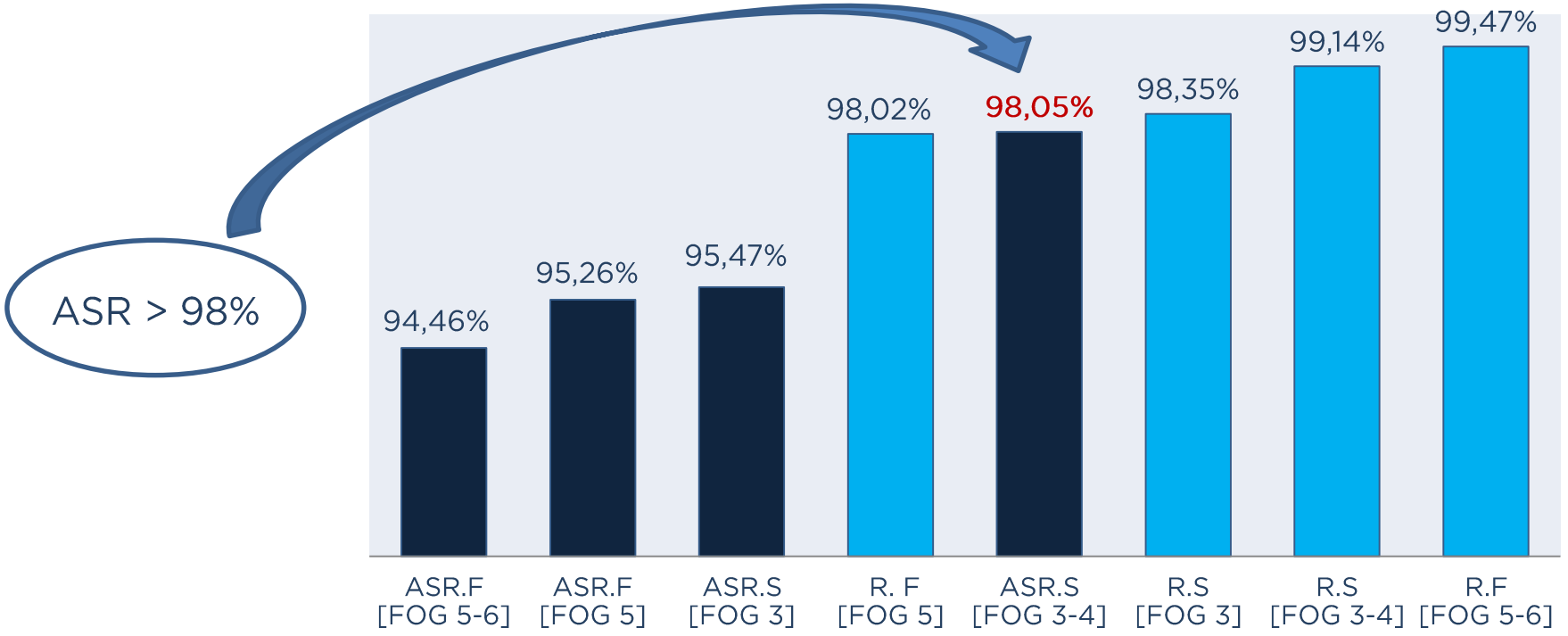
ASR – NER VALUE BEFORE & AFTER CORRECTION



ASR – NER VALUE BEFORE & AFTER CORRECTION

- ASR introduces serious errors in the text (i.e. lies)
- Automatic punctuation is limited to comas, no division in sentences (readability indexes give it highest values)
- Little to no reduction, increases delay
- Correctors can’t keep up with the quantity of errors
- Applying live correction to ASR feasible with NER value > 95%

ASR VS. RESPEAKING
NER VALUE AFTER CORRECTION



ASR VS. RESPEAKING
NER VALUE AFTER CORRECTION

- Higher quality with respeaking than with ASR
  Only in one sample ASR reached NER value > 98%

- ASR with live correction produces good results for:
  simple text (understandable to people with secondary education),
  slow delivery

- The quality of live subtitling through ASR
  is highly sensitive to speed and text complexity

- Respeaking quality is much less sensitive to speed and text complexity
  (respeakers can control reduction level
  to deal with higher speed of delivery)

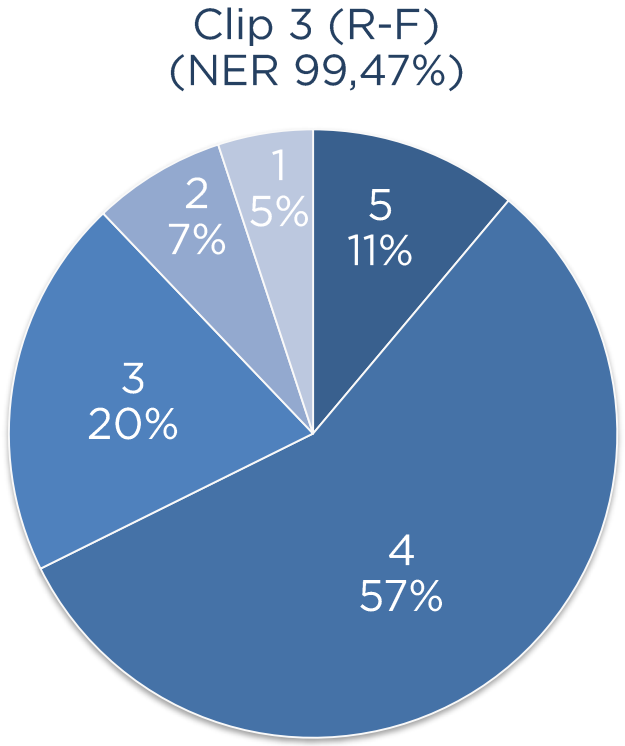
QUESTIONNAIRE

Combined recording of video and live subtitling,
as included in the questionnaire

N = 55 [Hearing 24, HoH 15, Deaf 16]

Only 30% participated
in an event with live subtitles
67% use subtitles on TV
53% use subtitles
wherever they're available

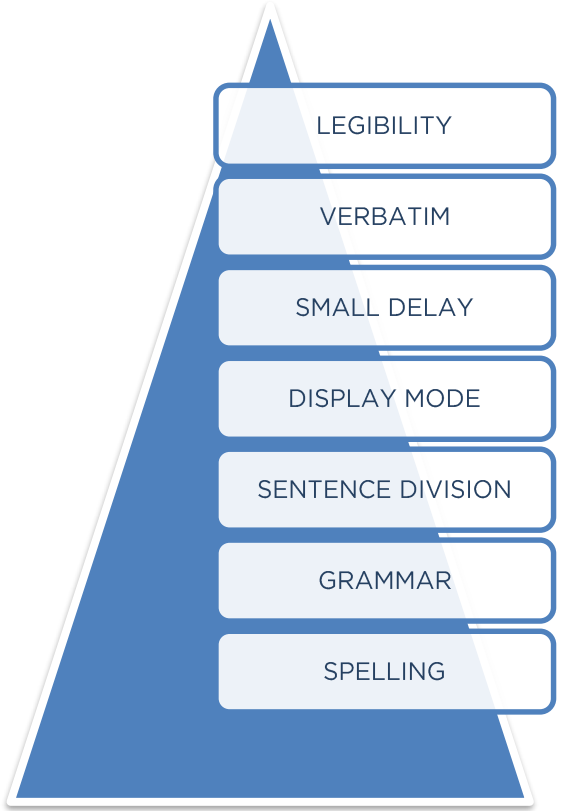
USERS'OPINION



Live subtitling quality rating on a scale of 1 to 5
(1 – very bad, 5 – very good)

- NER value is in line with users' opinions
- Clip 3 - highest NER value (99.47%) - highest user rating (3.60)
- Clip 1 - lowest NER value (95,26%) – lowest user rating (2.33)
- BUT, 15% said that subtitles for clip 3 were highly edited and simplified (the clip had a reduction of 0.5%)
  12% evaluated the subtitles with the highest NER value as the worst

USERS' OPINION



Users were asked to order the items according to their importance (starting from the most important at the top)

Live subtitling priorities acording to users

Contradictory expectations

Deaf users expected verbatim subtitles with little or no delay and emphasised the importance of legibility.

CONCLUSIONS

- Higher quality with respeaking than with ASR
- Live human correction makes sense if NER value for ASR > 95%; human correction necessary for ASR to achieve NER > 98%.
- ASR < 95% with live correction produces NER > 98% only for slow speech of moderate complexity
- NER value is in line with users' perception of quality
- Live subtitling through ASR with human correction is still useful to some users: 55% for the best ASR clip, 44% for the worst one (84% for the best respeaking)

Contact us:
monika@widzialni.org
lukasz.dutka@uw.edu.pl