CC BY

MACQUARIE University

Using eye tracking in media accessibility research:
Word-level data in subtitle reading

Jan-Louis Kruger

With thanks to my colleagues for the Language Processing in Multimodal Contexts lab:
Sixin Liao, Erik Reichle, Natalia Wisniewska and Lili Yu

"The soul, fortunately, has an interpreter – often an unconscious but still a faithful interpreter – in the eye".
(Charlotte Brontë 1847)

Background

- Eye tracking has gained a lot of ground in media accessibility research.
- Originated in the 1980s through the work of Gery d'Ydewalle and colleagues, it now includes a range of different measures in diverse research designs, and inspired novel avenues of research and application.
- Early studies focussed mainly on attention distribution between subtitles and images, and amount of visual attention (dwell time) in the subtitles.
- More recently, ET has been used to interrogate (among other things):
  - the impact of translation strategies on subtitle reading,
  - the impact of shot changes, text segmentation, placement and presentation speed on subtitle processing,
  - the impact of subtitles on the effectiveness of visual processing of film,
  - the amount of cognitive load induced by subtitles.

Foveal vision (Tobii.com)

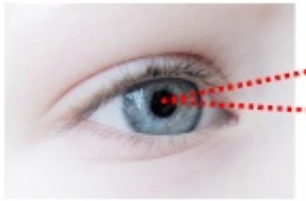
https://www.slideshare.net/AcuityETS/eye-tracking-in-usability

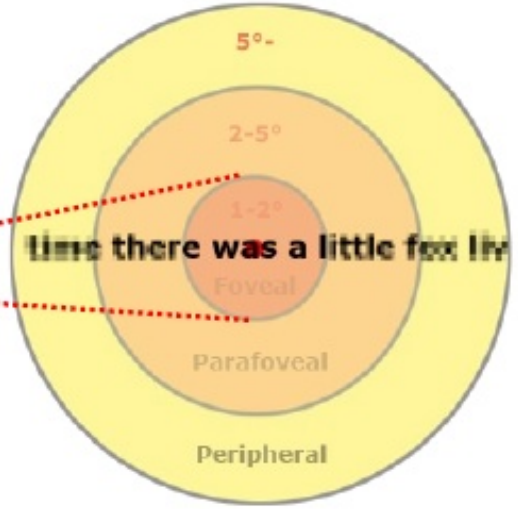
MACQUARIE University

tobii
www.tobii.com

What do we actually see?

- The 1-2° large area of foveal vision is about the size of a thumbnail on an arm lengths distance
- The fovea is less than 1% of the retina but takes up over 50% of the visual cortex in the brain.
- Peripheral vision is mainly good at picking up movements and contrasts





Exercise

MACQUARIE University

Just how much can we see with peripheral vision?

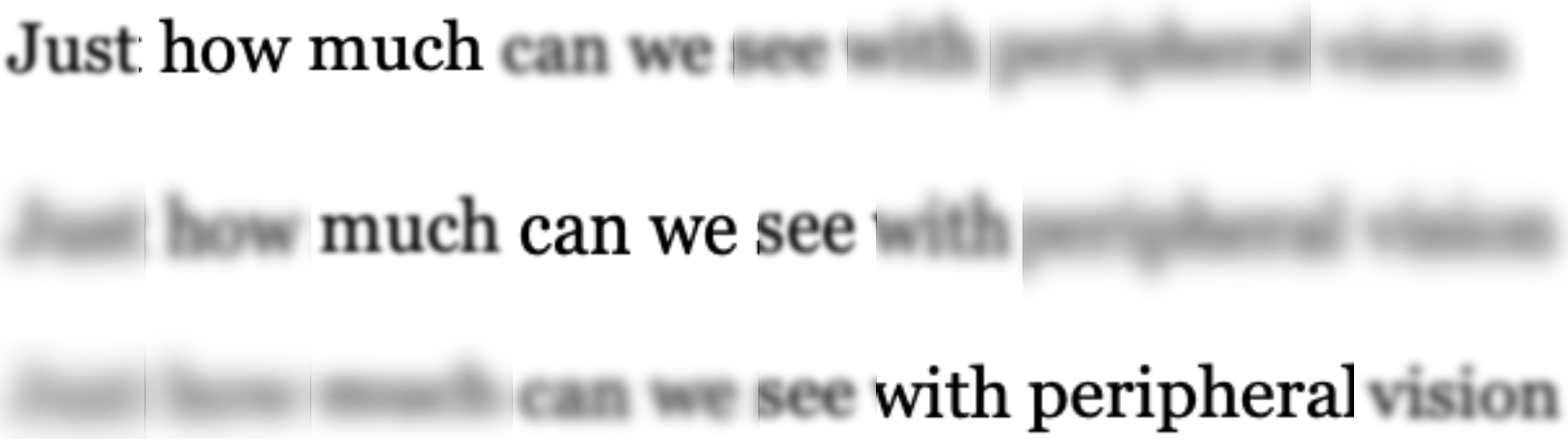


Image and text

Introduction to measures

Some ET measures used to study processing of subtitles:

- Global measures
  - Common (average time spent on subtitles vs. video, fixation counts and durations, dwell time)
  - Less common (word skipping rates during first-pass reading and refixation probabilities)
- Local measures
  - word-level measures to study frequency effects, word-length effects, and wrap-up effects.
- Multimodal Integrated Language Framework provides a theoretical framework.

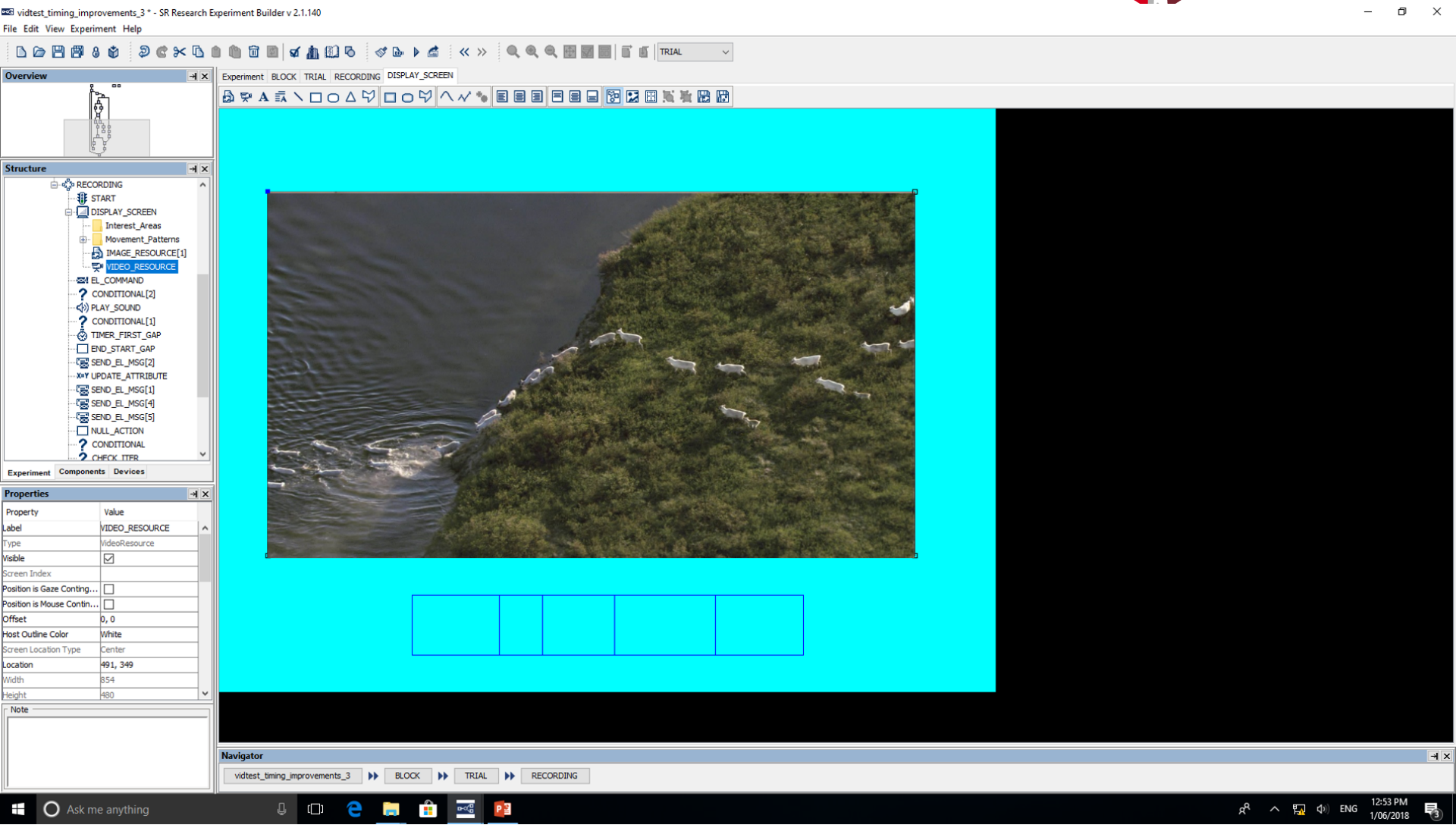
Why eye tracking?
The importance of research questions

MACQUARIE University

- Without an **objective measure like eye tracking**, hard to prove that subtitle reading improves **comprehension**, reduces **CL**, increases **immersion**, etc.
- ET reveals where viewers are looking and how they read subtitles.
- Remains indirect if we only look at global measures:
  - average DT
  - fixation count
  - fixation duration
  - crossovers, etc.
- We need to answer questions such as:
  - How does L1vs.L2/layout/speed impact on global measures and attention distribution?
  - How many of the words are processed? (RIDT, skipping probability)
  - How are words/phrases/subtitles processed? (skimming, refixations, regressions)
  - How does video and soundtrack presence/complexity/redundancy impact on subtitle reading?
- Many questions are hard to answer without word-level data.

Word-level data using SR Eyelink

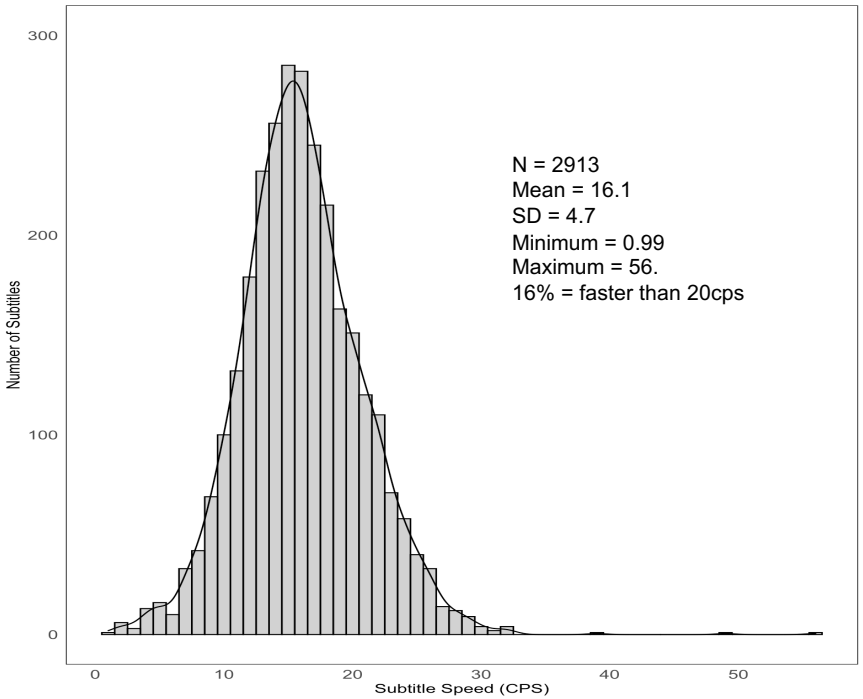
MACQUARIE University



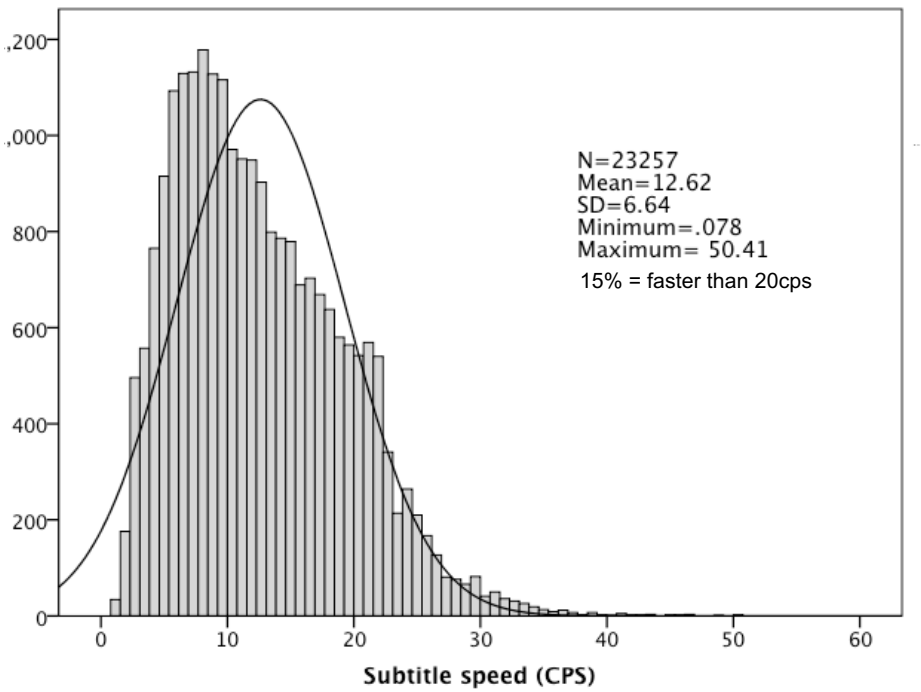
Global vs local measures

MACQUARIE University

- Subtitles are variable in terms of layout, volume, linguistic complexity, presentation speed, and competition with other input.
- This variability means we cannot only look at averaged global data across a full subtitled video, but also need to look at local subtitle- and even word-level data.
- E.g subtitle speed from sample of 11 Coursera clips and 10 Netflix films:



Coursera



Netflix

AVT and multimodality

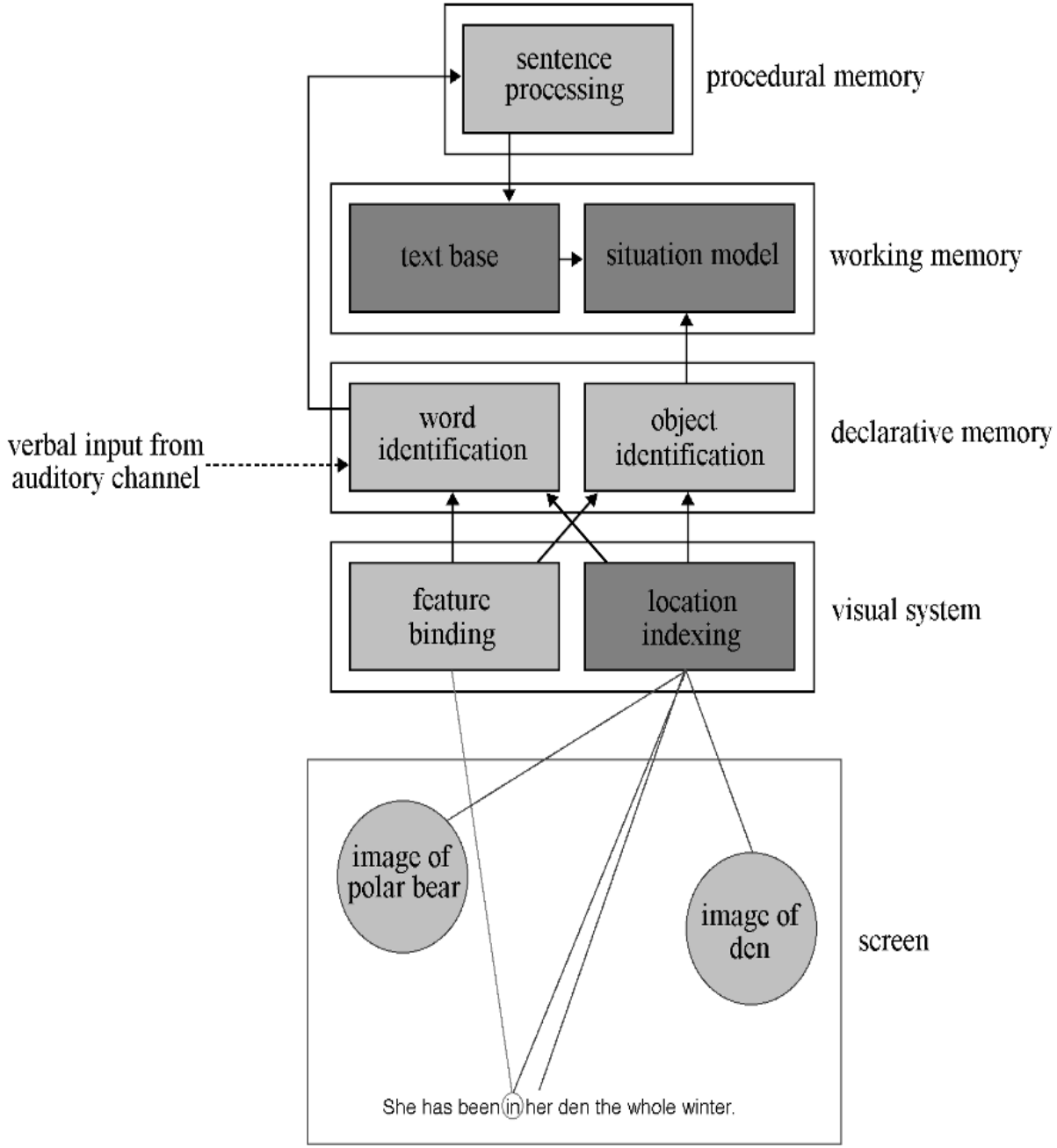
MACQUARIE University

- We engage with spoken and written language in film while also integrating this with the soundtrack and/or the visual information contained in the image.
- Much of the information in subtitles depends on, supplements, or repeats what is available in the other modes.
- Film is also dynamic.
- The viewer has to engage in continuous and dynamic strategic reading while also processing images and the auditory information of the soundtrack.

Processing multimodal redundancy

MACQUARIE University

- Viewers can process multimodal redundancy with astonishing efficiency.
- Often triggers involuntary checking or comparison – errors pop out.



A multimodal integrated language framework (adapted from Liao et al., 2020)

Presentation rate or subtitle speed

MACQUARIE University

- Presentation rate is a contentious issue.
- The conventional presentation rate in Europe used to be around 145wpm (12cps), which is lower than speech rate in most cases.
- In reality, subtitle speed varies widely from around 12 cps tp 20cps.
- Szarkowska and Gerber Morón (2018) found that viewers can process subtitles at 240wpm (20cps) and still follow the images.

Design

MACQUARIE University

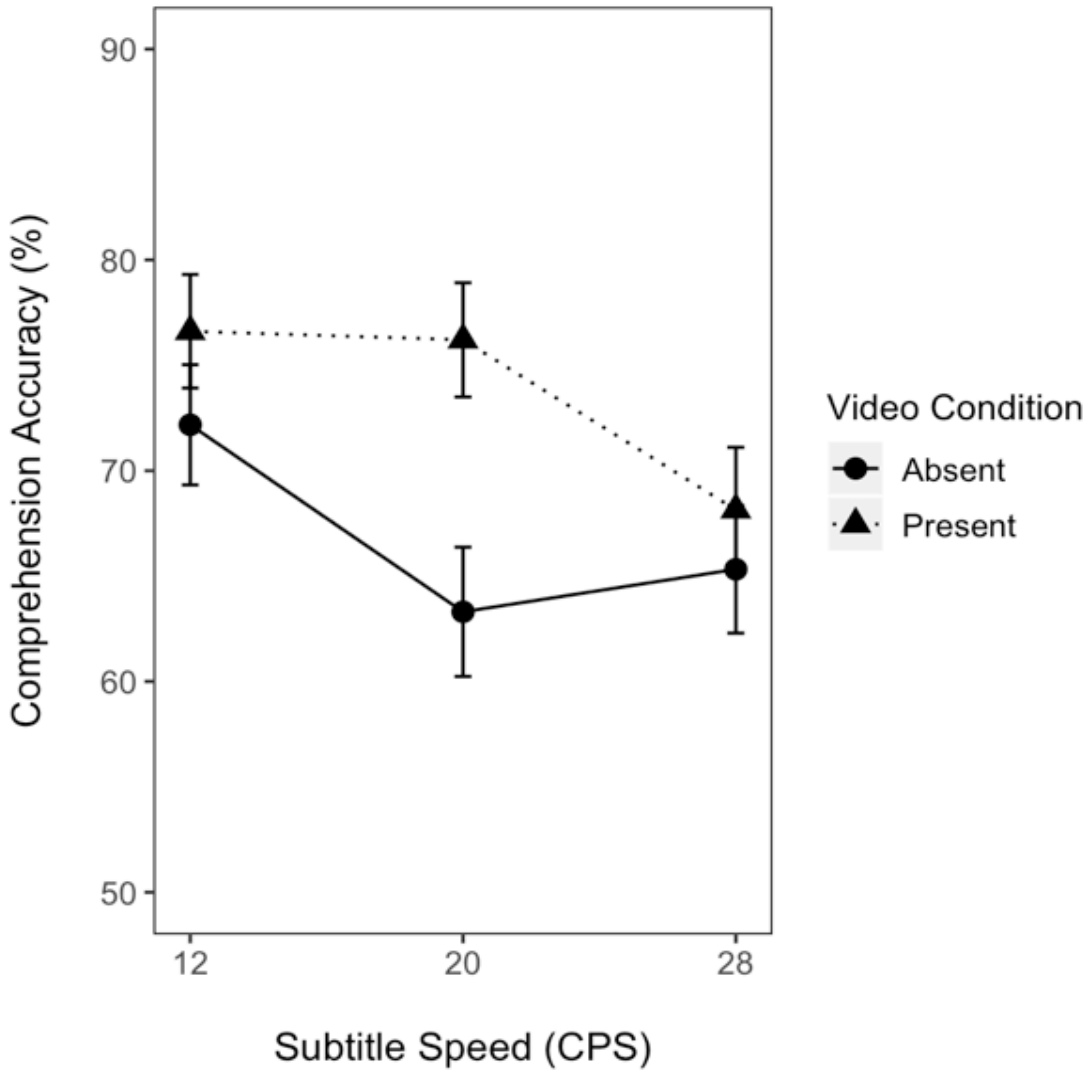
- 2 (**Video conditions:** with video, no video) * 3 (**Subtitle speed conditions:** 12cps, 20cps, and 28cps).
- No sound was used and the volume of text was kept identical across all three speeds.
- 6 videos were used for each experiment, thus, each of the conditions included 1 video/audio and its subtitles (~80).
- Some redundancy of information content between image and subtitles, but also unique elements.
- 8 questions related to the information contained in the **subtitles only** were presented after each video (48 questions in total).

On-line eye tracking measures

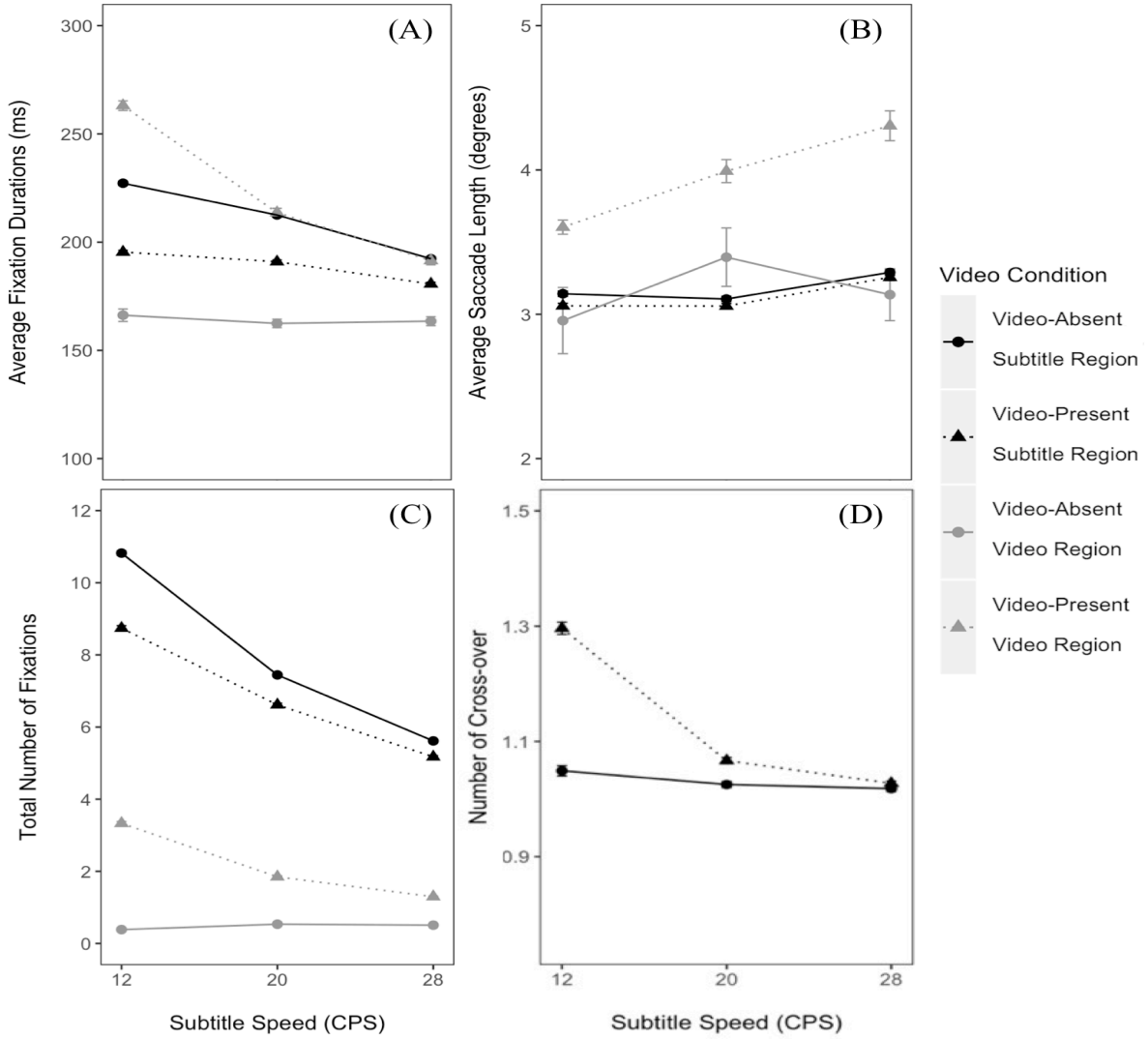
MACQUARIE University

- We examined how video and presentation speed impact both global and local measures.
  - Global measures:
    - Dwell time (first-pass)
    - Total reading times
    - Fixation count
    - Average Fixation duration
    - Saccade length
    - Number of cross-overs (i.e., switching between subtitles and video)
    - Proportion of words fixated at least once (processed visually)
    - Proportion of words re-fixated (regressions are an important part of reading comprehension)
  - Local: (word frequency and length effects)
    - Skipping
    - Revisits
    - Gaze durations
    - Total Viewing times
- We also investigated the impact of the IV on comprehension.

Comprehension questions



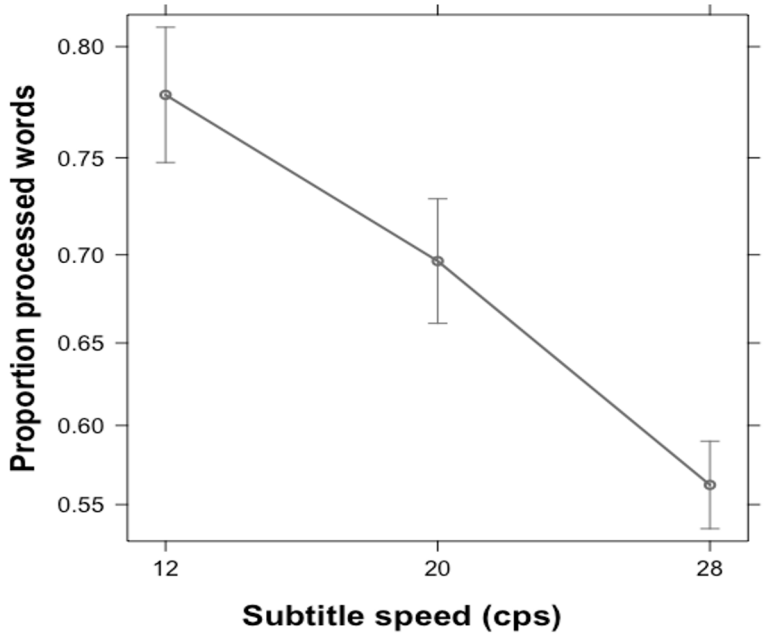
Global subtitle reading

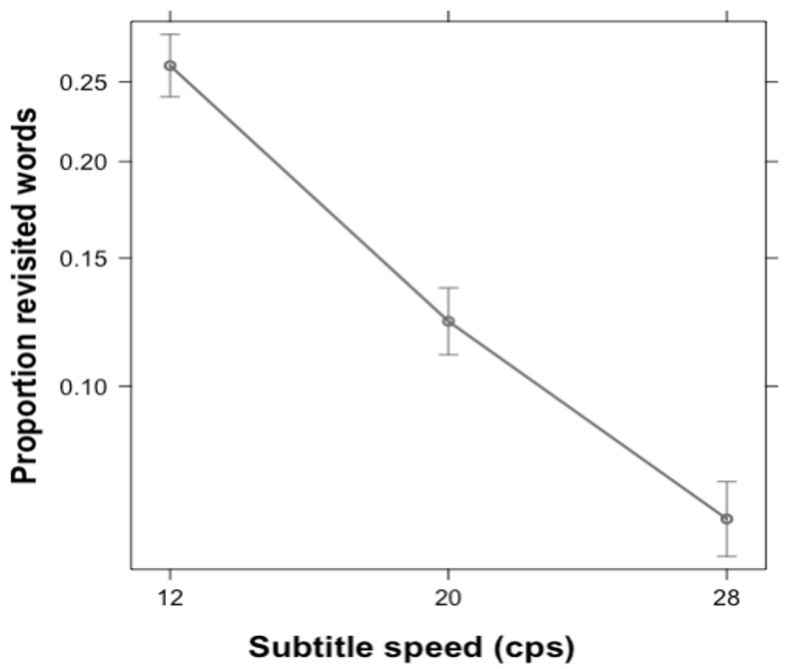


Processing of words

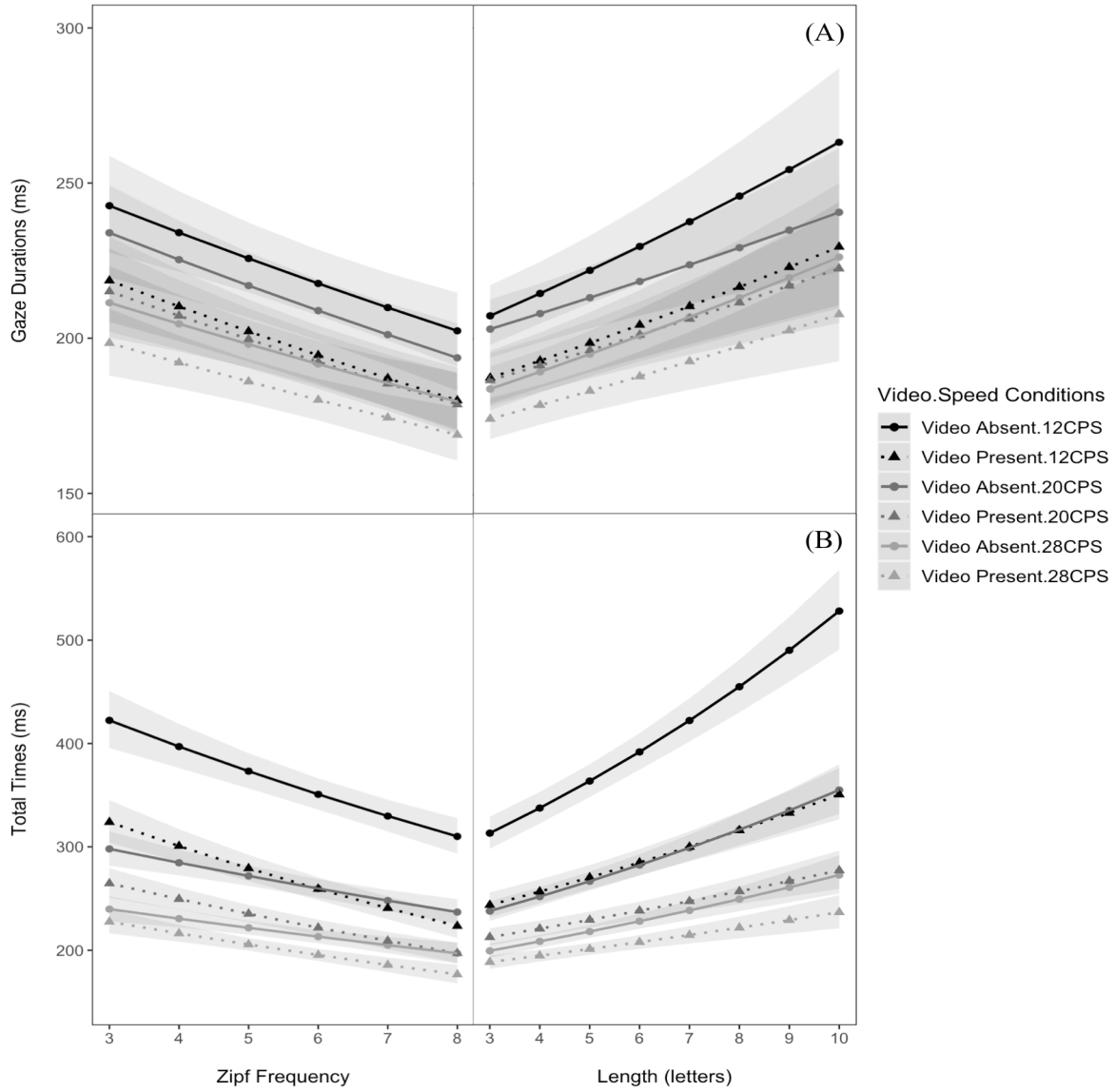
MACQUARIE University

Subtitle speed	Proportion of words processed	Proportion of words revisited	Proportion of subtitles not read to completion
12 cps	0.74 (0.43)	0.28 (0.45)	0.06 (0.24)
20 cps	0.69 (0.46)	0.15 (0.36)	0.10 (0.30)
28 cps	0.60 (0.49)	0.09 (0.28)	0.17 (0.37)

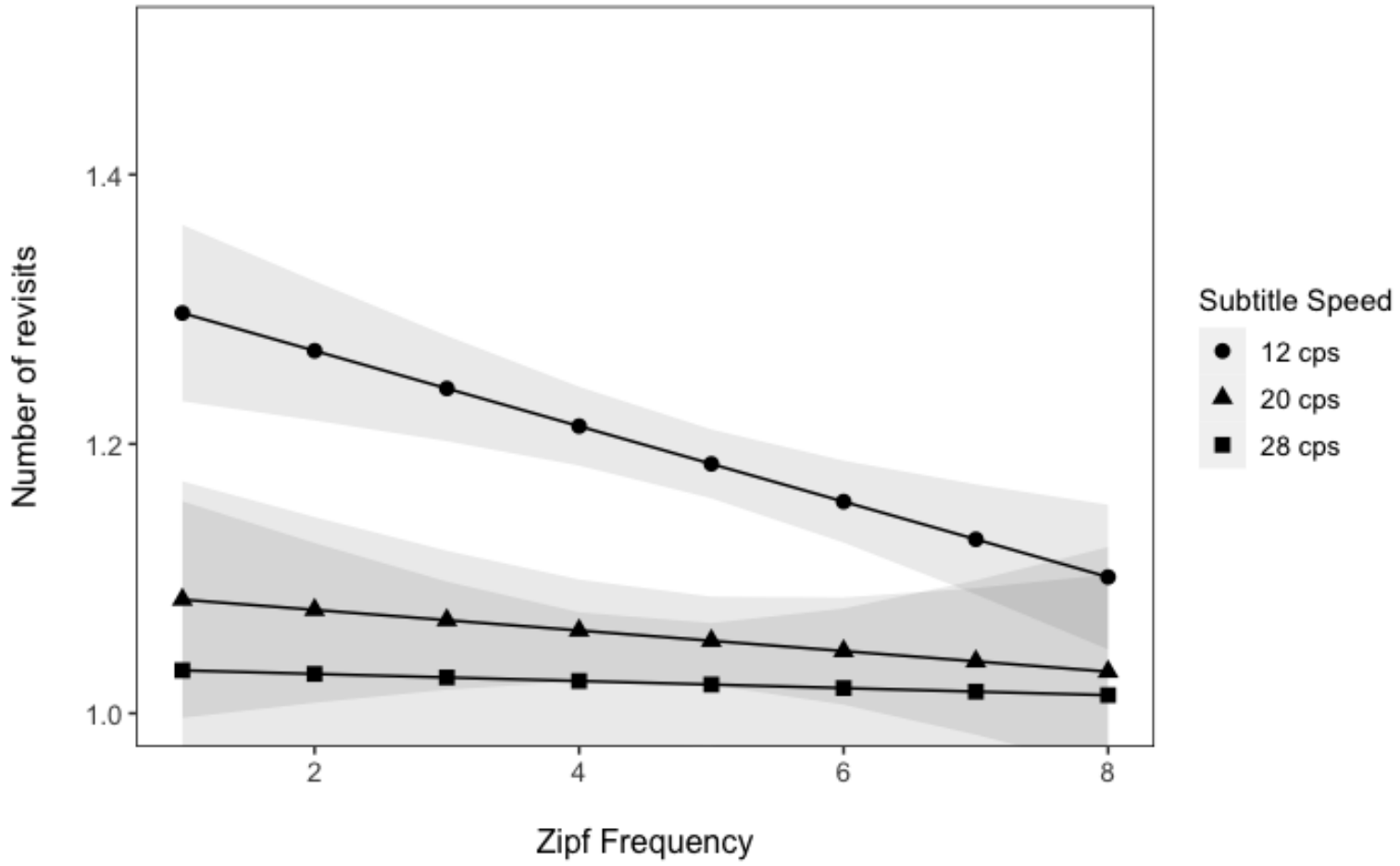




Local word reading: Word frequency and word length



Words frequency effect: Number of revisits to words



12cps

MACQUARIE University

28cps

Findings and conclusions

MACQUARIE University

- In presence of video, and as the subtitle speed increases, subtitles are read more superficially:
  - Fewer, shorter fixations, more words skipped, fewer words re-fixated (fewer regressions), fewer subtitles read to completion.
- As the speed increases, the differences in local measures marking lexical processing (word frequency and length effects) start to disappear – uncommon or low frequency words are not fixated longer or revisited more.
- The combination of global and local measures therefore helps us to illustrate the full impact of subtitle speed on reading.

Future projects and conclusion

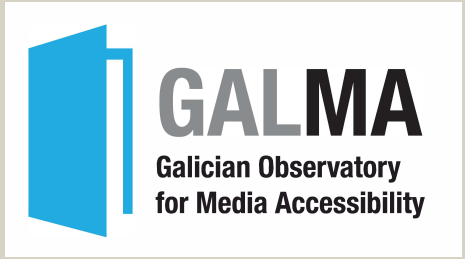
MACQUARIE University

- In our current project we are investigating the impact of different levels of secondary task demand on subtitle reading (no task, easy task [colour change], and difficult task [vowel-change].
- Sixin Liao's PhD is looking at video presence, visual redundancy, as well as the presence of L1 or L2 audio on L2 subtitle reading.
- Erik Reichle is also testing models in our attempt to show the value of the multimodal integrated language framework and has already shown through modelling that high subtitle speed results in more skimming and the skipping of 3 and 4-letter words.
- We have entered a new era in the use of global and local eye movement measures to understand media accessibility and language processing.

Select references

- Chan, W., Kruger, J.L. & Doherty, S. (2019). Comparing the impact of automatically generated and corrected subtitles on cognitive load and learning in a first- and second-language educational context. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*. 18: 237-272.
- Kruger, J.L. (2018). Eye tracking in audiovisual translation research. In Luis Perez-Gonzalez, ed. *The Routledge Handbook of Audiovisual Translation*. London: Routledge. Chapter 22. pp. 366-382.
- Kruger, J.L., Doherty, S., Fox, W., and de Lissa, P. 2018. Multimodal measurement of cognitive load during subtitle processing: Same-language subtitles for foreign-language viewers. In I. Lacruz and R. Jääskeläinen (eds.), *Innovation and expansion in Translation Process Research*. London, UK: John Benjamins. Chapter 12. pp 267-294. doi:10.1075/ata.18.12kru
- Kruger, J.L., Soto-Sanfiel., M. T., Doherty, S., and Ibrahim, R. 2016. Towards a cognitive audiovisual translatology: Subtitles and embodied cognition. In. Ricardo Muñoz (ed.), *Reembedding Translation Process Research*. pp.171-194. London: John Benjamins Publishing Company.
- Liao, S., Kruger, J.L., Doherty, S. (2020). **T**he impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*. Issue 33: 70-89 .
- Liao, S., Yu, L., Reichle, E.D., Kruger, J.L. (2020) Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading*. DOI: 10.1080/10888438.2020.1823986.
- Szarkowska, A., & Bogucka, L. (2019). Six-second rule revisited: An eye-tracking study on the impact of speech rate and language proficiency on subtitle reading. *Translation, Cognition & Behavior*, 2(1), 101-124.
- Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *Plos One*, 13(6).

MACQUARIE University

Language processing in multimodal contexts

Asia-Pacific Translation and Cognition research group (APTRAC.org)