GHENT
UNIVERSITY

LREC 2022
Marseille

13th edition

Le Palais du Pharo

June 20-25

FACULTY OF ARTS AND PHILOSOPHY

DEPARTMENT OF TRANSLATION, INTERPRETING, AND COMMUNICATION
RESEARCH GROUP LANGUAGE AND TRANSLATION TECHNOLOGY (LT3)

GECO-MT: THE GHENT EYE-TRACKING CORPUS OF MACHINE TRANSLATION

Toon Colman, Margot Fonteyne, Joke Daems, Nicolas Dirix, Lieve Macken

GHENT UNIVERSITY

language and translation technology team

LREC 2022 Marseille

EUROPEAN LANGUAGE RESOURCES ASSOCIATION ELRA

THE ARISTOCAT PROJECT

- The present study takes place in the larger **ArisToCAT project:** Assessing the Comprehensibility of Automatic Translations
- Objective: Evaluate the comprehensibility of ‘raw’ (unedited) MT output for readers who can only rely on the MT output

GHENT UNIVERSITY

Macken, L., Fonteyne, M., Tezcan, A., & Daems, J. (2020). Assessing the Comprehensibility of Automatic Translations (ArisToCAT). Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT2020), 485–486. http://hdl.handle.net/1854/LU-8665530.

THE GECO-MT PROJECT

- Compare the reading process of participants reading a full novel in a human translation (HT) vs. a machine translation (MT)
  - How is the reading process affected by MT?
  - To what extent do various MT errors impact the reading process?

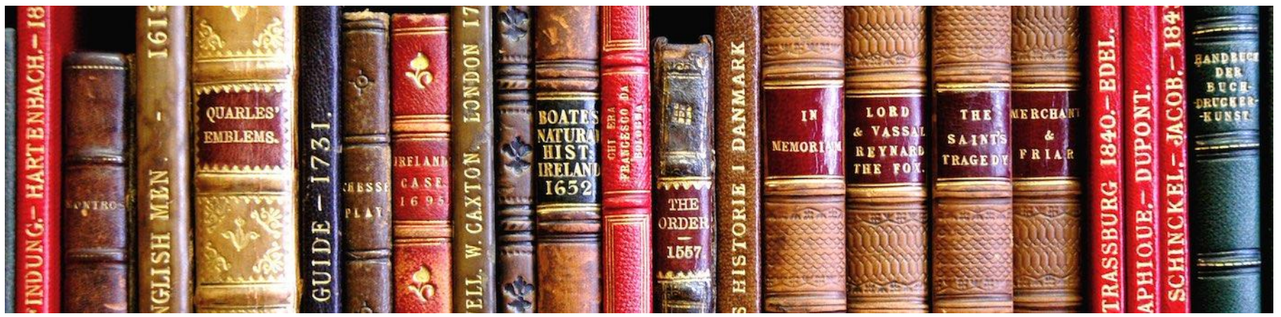
GHENT UNIVERSITY

INTRODUCTION

GHENT
UNIVERSITY

LITERARY MACHINE TRANSLATION

- Quality of MT output has improved significantly (*statistical* systems → *neural* systems)
- Despite overall improvements, using MT for literary texts remains very challenging…
  - Making literary MT perfectly suited for our research goals!

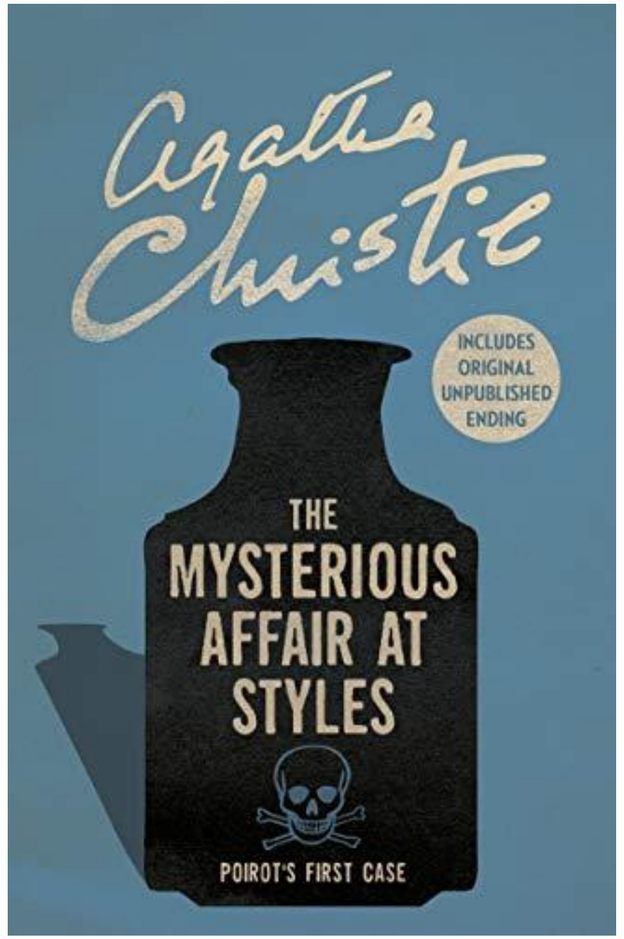


GHENT UNIVERSITY

Van Brussel, L., Tezcan, A., & Macken, L. (2018). A fine-grained error analysis of NMT, PBMT and RBMT output for English-to-Dutch. Proceedings of the Eleventh International Conference on Language Resources and Evaluation, 3799–3804. http://hdl.handle.net/1854/LU-8561558

LITERARY MT UNDER THE MAGNIFYING GLASS

- Case Study: Agatha Christie's *Mysterious Affair at Styles* → Google Translate English to Dutch
  - SCATE MT error taxonomy
  - 56% of sentences contain at least one MT error
    1) 35% mistranslations
    2) 31% coherence errors
    3) 16% style & register errors



GHENT UNIVERSITY

Fonteyne, M., Tezcan, A., & Macken, L. (2020). Literary machine translation under the magnifying glass : assessing the quality of an NMT-translated detective novel on document level. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, … S. Piperidis (Eds.), 12th International Conference on Language Resources and Evaluation Conference (LREC 2020), Proceedings (pp. 3783–3791). Paris, France: European Language Resources Association (ELRA).

SCATE MT ERROR TAXONOMY

FLUENCY
- coherence
  - logical problem
  - non-exisiting word
  - cultural reference
  - discourse marker
  - co-reference
  - inconsistency
  - verb tense
- lexicon
  - lexical choice
  - wrong preposition
- grammar & syntax
  - agreement
  - verb form
  - word order
  - extra word(s)
  - missing word(s)
- style & register
  - disfluency
  - repetition
  - register
  - untranslated
- spelling
- other

ACCURACY
- mistranslation
  - multiword
  - word sense
  - semantically unrelated
  - part-of-speech
  - partially translated
  - other
- do not translate
- untranslated
- addition
- omission
- capitalisation & punctuation
- other

Mistranslation: “come and be refreshed!” → “Kom en word vernieuwd!” (incorrect word sense…)

Coherence: “We had a good yearn about old times.” → “We hadden een goed garen over oude tijden.” (illogical and confusing…)

Style & Register: “Where is tea today?” → “Waar is thee vandaag?” (not idiomatic…)

Tezcan, A., Daems, J., & Macken, L. (2019). When a `sport' is a person and other issues for NMT of novels. Proceedings of the Qualities of Literary Machine Translation, 40–49. http://hdl.handle.net/1854/LU-8625754

LITERARY MT UNDER THE MAGNIFYING GLASS

– This project resulted in a full annotated MT novel, which can be used for the present eye-tracking study!

WORD	ERROR TYPE
…	
Kom	
en	
word	
vernieuwd	MISTRANSLATION
…	

→ Effect on eye movements?

GHENT UNIVERSITY

THE GHENT EYE-TRACKING CORPUS

- Most importantly, a human translation of this novel was previously used to create an eye-movement corpus of monolingual and bilingual reading

  → GECO

- In the present study we will expand on GECO with eye-movement data of MT reading

  → GECO-MT

GHENT UNIVERSITY

Cop, U., N. Dirix, D. Drieghe and W. Duyck 2017. Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. Behavior Research Methods, 49:602–615.

GECO → GECO-MT

WORD	READING TIME (MS)
…	
Come	312
and	143
be	149
refreshed	369
…	

English

GECO

WORD	READING TIME (MS)
…	
Ga	278
mee	117
en	167
laat	139
je	324
laven	347
…	

Dutch

GECO

GHENT UNIVERSITY

GECO → GECO-MT

WORD	READING TIME (MS)
…	
Come	312
and	143
be	149
refreshed	369
…	

English

GECO

WORD	READING TIME (MS)
…	
Ga	278
mee	117
en	167
laat	139
je	324
laven	347
…	

Dutch

GECO

GECO-MT

WORD	READING TIME (MS)
…	
Kom	345
en	134
word	175
vernieuwd	398
…	

Dutch MT

GECO-MT

GHENT UNIVERSITY

RESEARCH GOALS

- Provide a big corpus of natural eye-movements during the reading of MT literature
- Available at www.lt3.ugent.be/resources/geco-mt

WORD	READING TIME (MS)	FIXATION COUNT
…		
Kom	345	2
en	134	1
word	175	1
vernieuwd	398	2
…		

GHENT UNIVERSITY

RESEARCH GOALS

- Address our own research questions surrounding MT comprehensibility
  - How is the natural reading process affected by MT?
  - To what extent do various MT errors impact the reading process?

GHENT UNIVERSITY

EYE-TRACKING EXPERIMENT

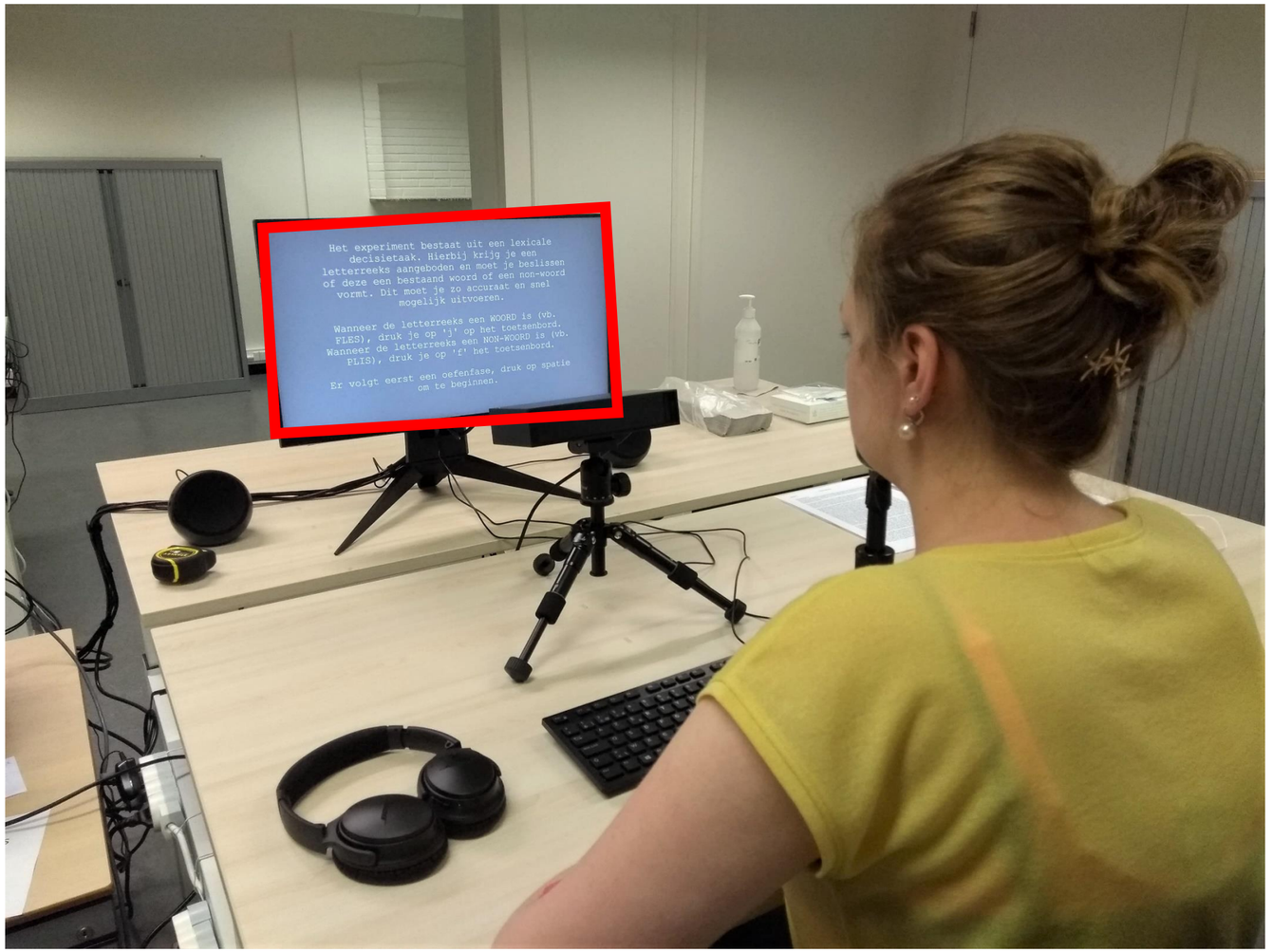
GHENT UNIVERSITY

EXPERIMENTAL DESIGN (N = 20)

Session 1	**Session 2**	**Session 3**	**Session 4**
Part 1 HT / Part 1 MT	Part 2 MT / Part 2 HT	Part 3 HT / Part 3 MT	Part 4 MT / Part 4 HT
Eye-movement data collected during reading + Comprehension questions			
Lexical decision task (ENG & NL)	Spelling task (ENG & NL)	Speeded lexical decision task (NL)	Speeded lexical decision task (ENG)

GHENT UNIVERSITY

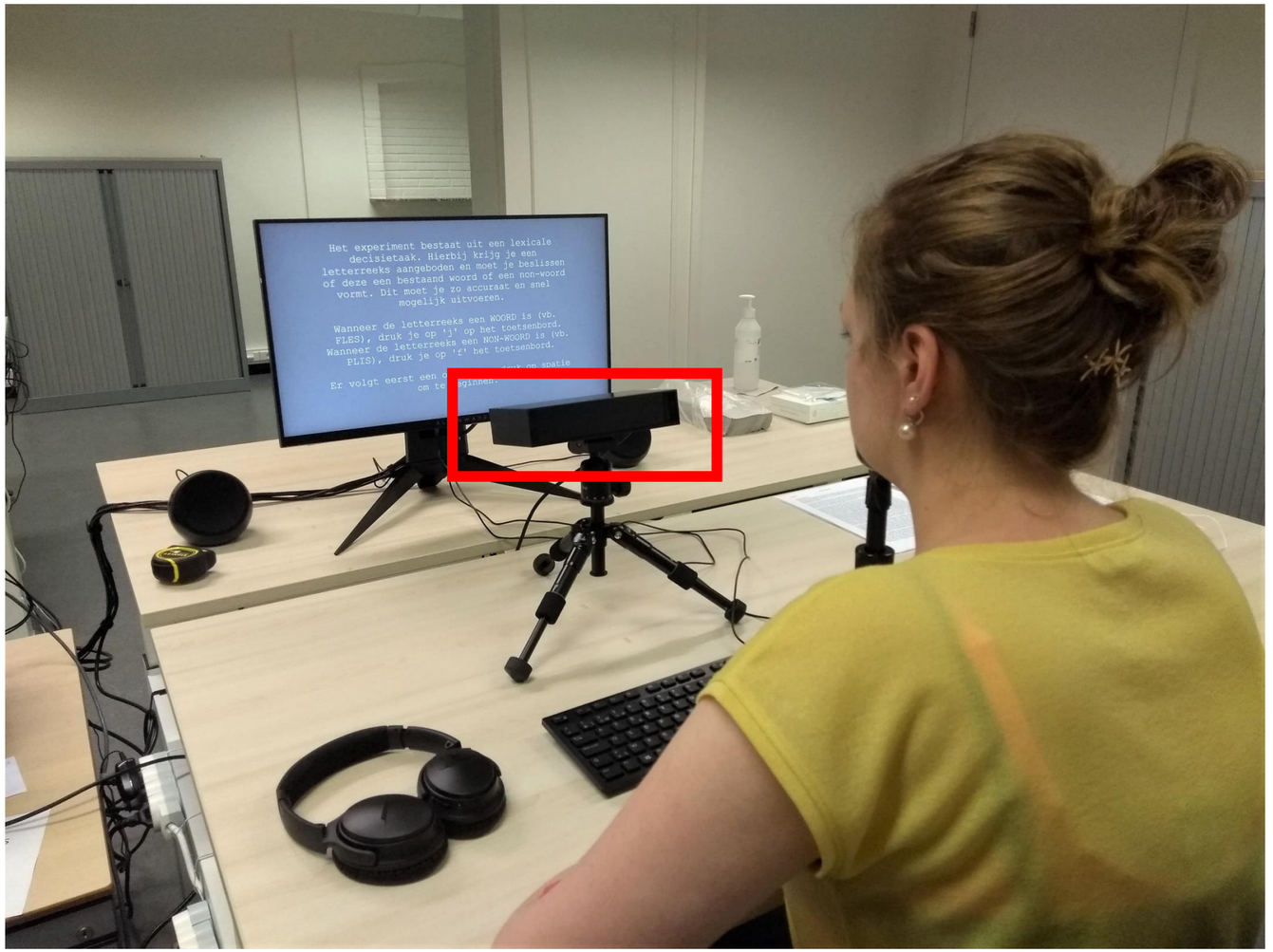
EXPERIMENTAL SETUP

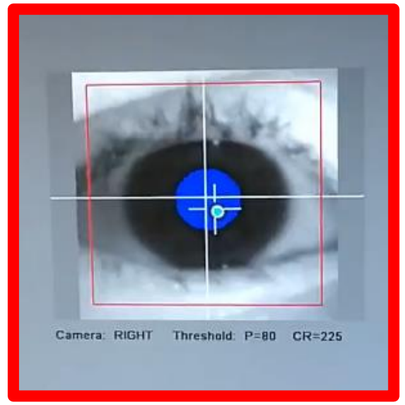


Op dat moment klonk er een bekende stem door de openstaande tuindeuren, vlakbij: "Jij schrijft dus na de thee aan de Prinses, Alfred? Ik zal zelf aan Lady Tadminster schrijven over de tweede dag. Of zullen we daarmee wachten tot we antwoord hebben van de Prinses? Voor het geval zij zou weigeren, kan Lady Tadminster op de eerste dag openen en mevrouw Crosbie dan de tweede. Dan hebben we nog de hertogin... voor het schoolfeest." Ik hoorde een mannenstem wat mompelen en toen die van mevrouw Inglethorp, iets luider: "Ja natuurlijk kan het ook na de thee. Je bent altijd zo attent, Alfred."

GHENT UNIVERSITY

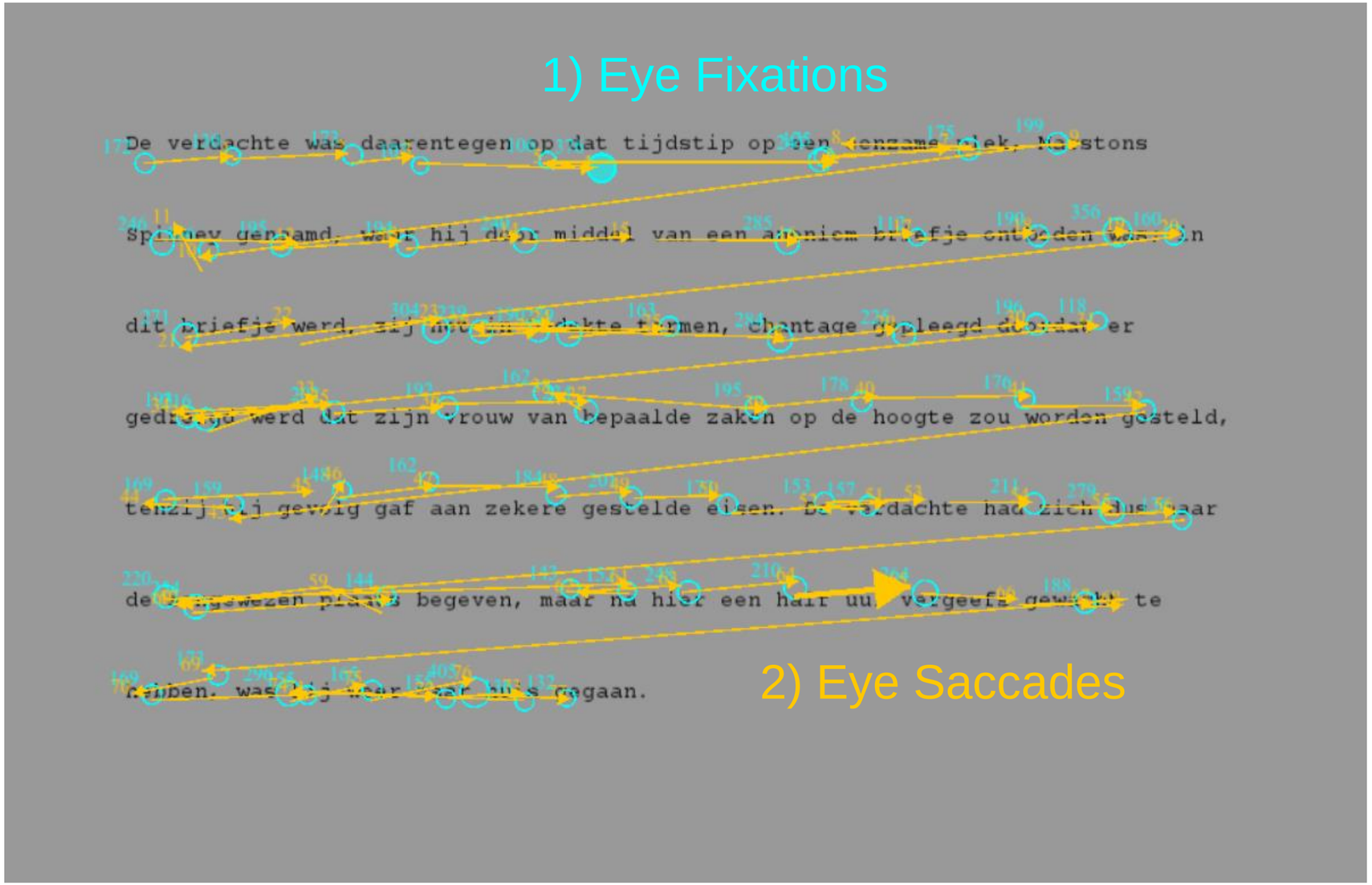
EXPERIMENTAL SETUP





GHENT
UNIVERSITY

COMPONENTS OF EYE-MOVEMENT



GHENT UNIVERSITY

EYE-TRACKING VARIABLES

2 components → many variables:

- Number & duration of fixations
- saccade amplitude
- number of regressive (‘backwards’) saccades
- …

In a nutshell: when text is more difficult to read, fixations get longer, saccades get shorter, and more regressions are made (Rayner 2009)

GHENT UNIVERSITY

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. Quarterly Journal of Experimental Psychology, 62(8), 1457–1506. https://doi.org/10.1080/17470210902816461

EYE-TRACKING FOR MT QUALITY ESTIMATION

- Doherty et al. (2010): Increased eye fixation count and gaze time for MT output rated as bad (human evaluation)
- Stymne et al. (2012): Increased eye fixation count and gaze time on MT errors, compared to correct MT output

→ Using eye-tracking measures to assess the readability of MT output and the 'severity' of errors

GHENT UNIVERSITY

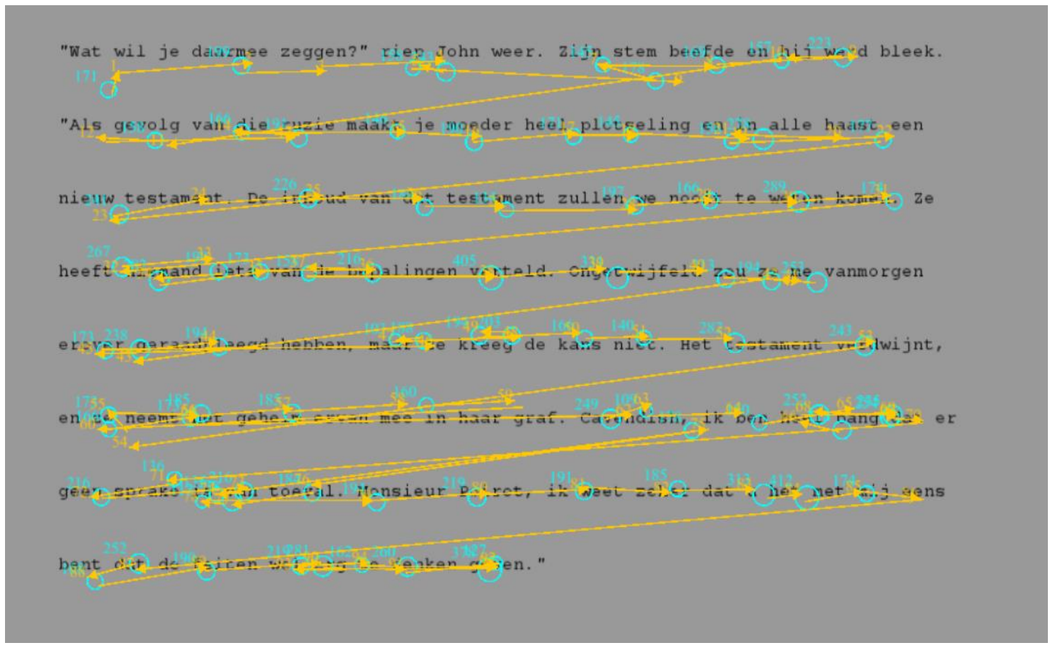
Doherty, S., O'Brien, S., & Carl, M. (2010). Eye tracking as an MT evaluation technique. Machine Translation, 24(1), 1–13. https://doi.org/10.1007/s10590-010-9070-9

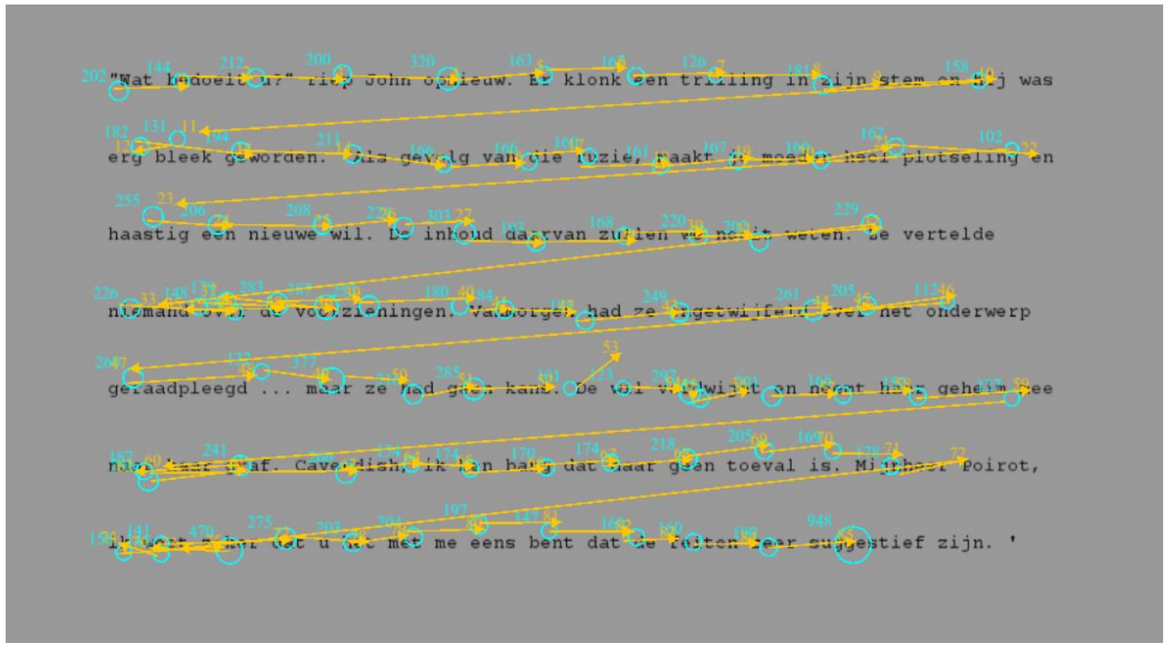
Stymne, S., Danielsson, H., Bremin, S., Hu, H., Karlsson, J., Lillkull, A. P., & Wester, M. (2012). Eye Tracking as a Tool for Machine Translation Error Analysis. LREC 2012—EIGHTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (pp. 1121–1126).

GECO-MT DATA

GHENT UNIVERSITY

A) PARAGRAPH LEVEL





HT **MT**

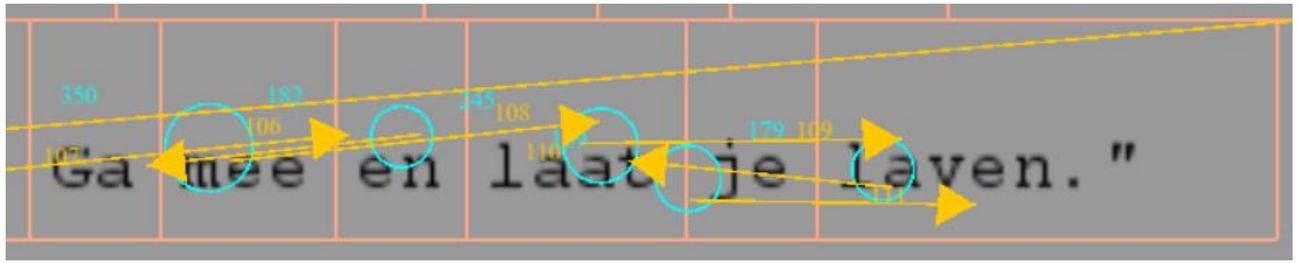
→ Paragraph reading duration, total number of fixations, average fixation duration, average saccade amplitude, …

GHENT UNIVERSITY

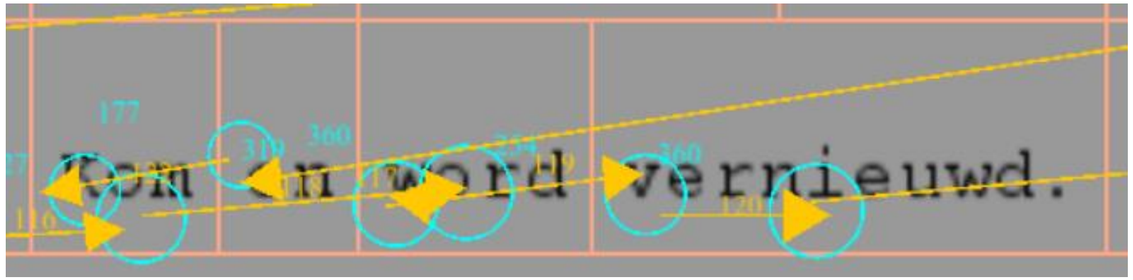
B) WORD / MT ERROR LEVEL

Source text: “Come and be refreshed”

HT



MT



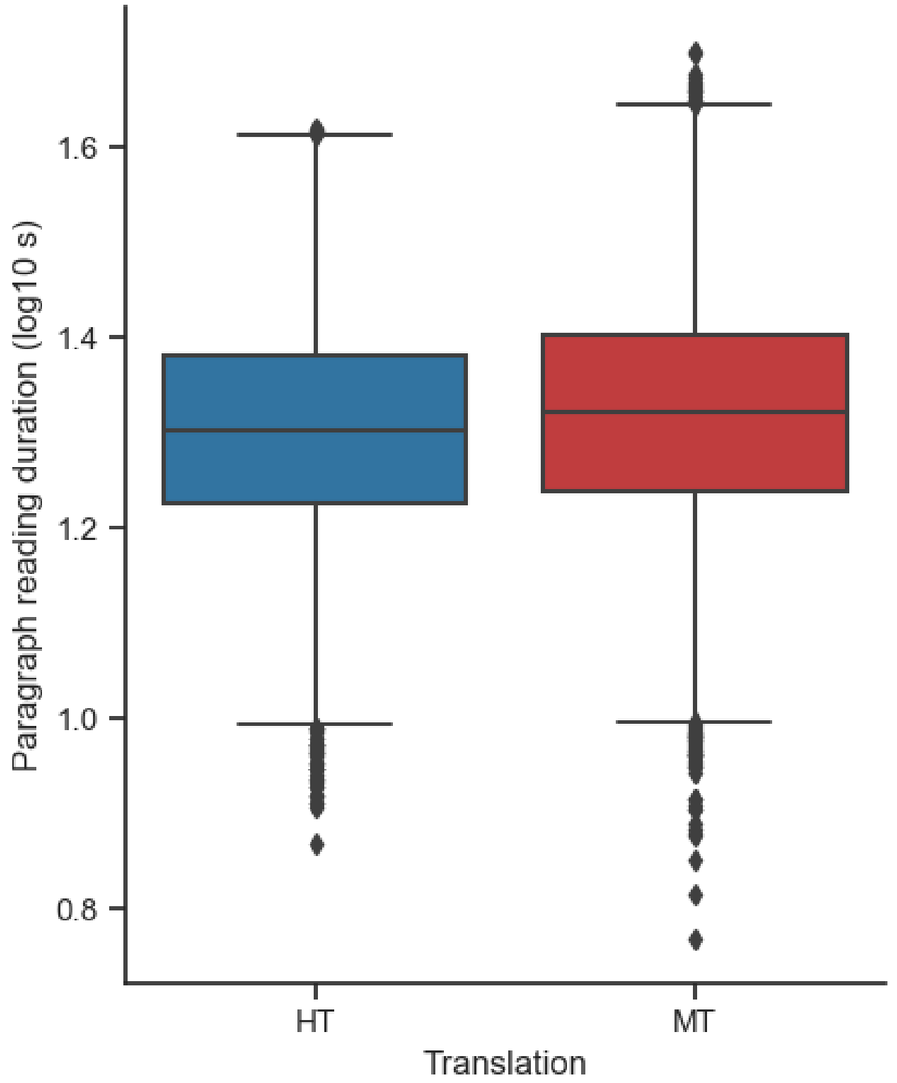
→ Word first fixation duration, gaze duration (sum of first-pass fixations), total reading time, …

GHENT UNIVERSITY

DESCRIPTIVE STATISTICS (CF. PAPER)

Variable	Dutch HT	Dutch MT
RD (s)	20.77 [5.60]	21.70 [6.18]
NF (count)	77.61 [19.62]	80.11 [21.31]
AFD (ms)	221.51 [19.97]	225.10 [19.67]
ASA (deg.°)	3.36 [0.55]	3.25 [0.53]

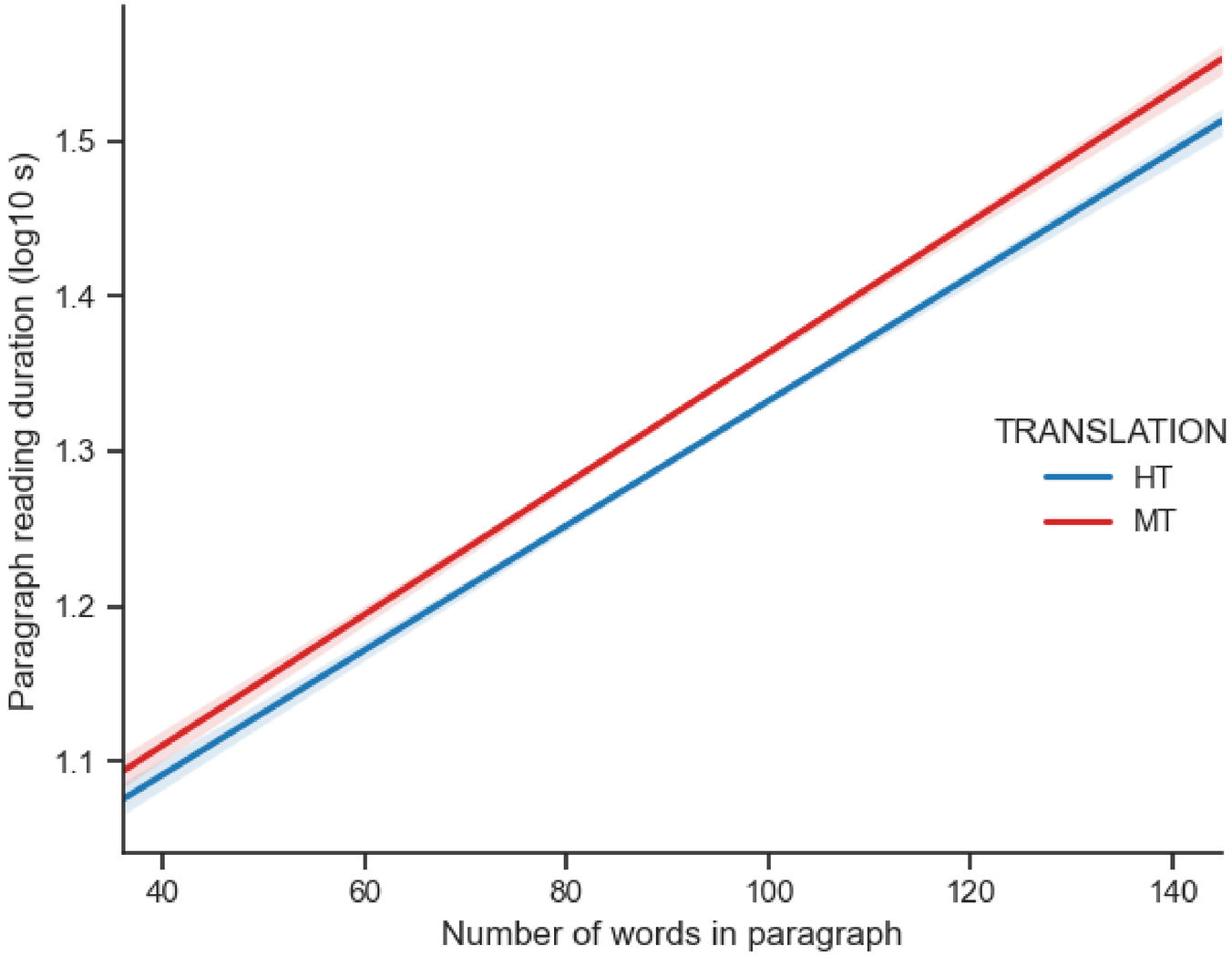
Table 2: *M* [and *SD*] of the paragraph level variables *reading duration* (RD), *number of fixations* (NF), *average fixation duration* (AFD), and *average saccade amplitude* (ASA), after outlier removal.



GHENT UNIVERSITY

EXAMPLE: PARAGRAPH READING DURATION

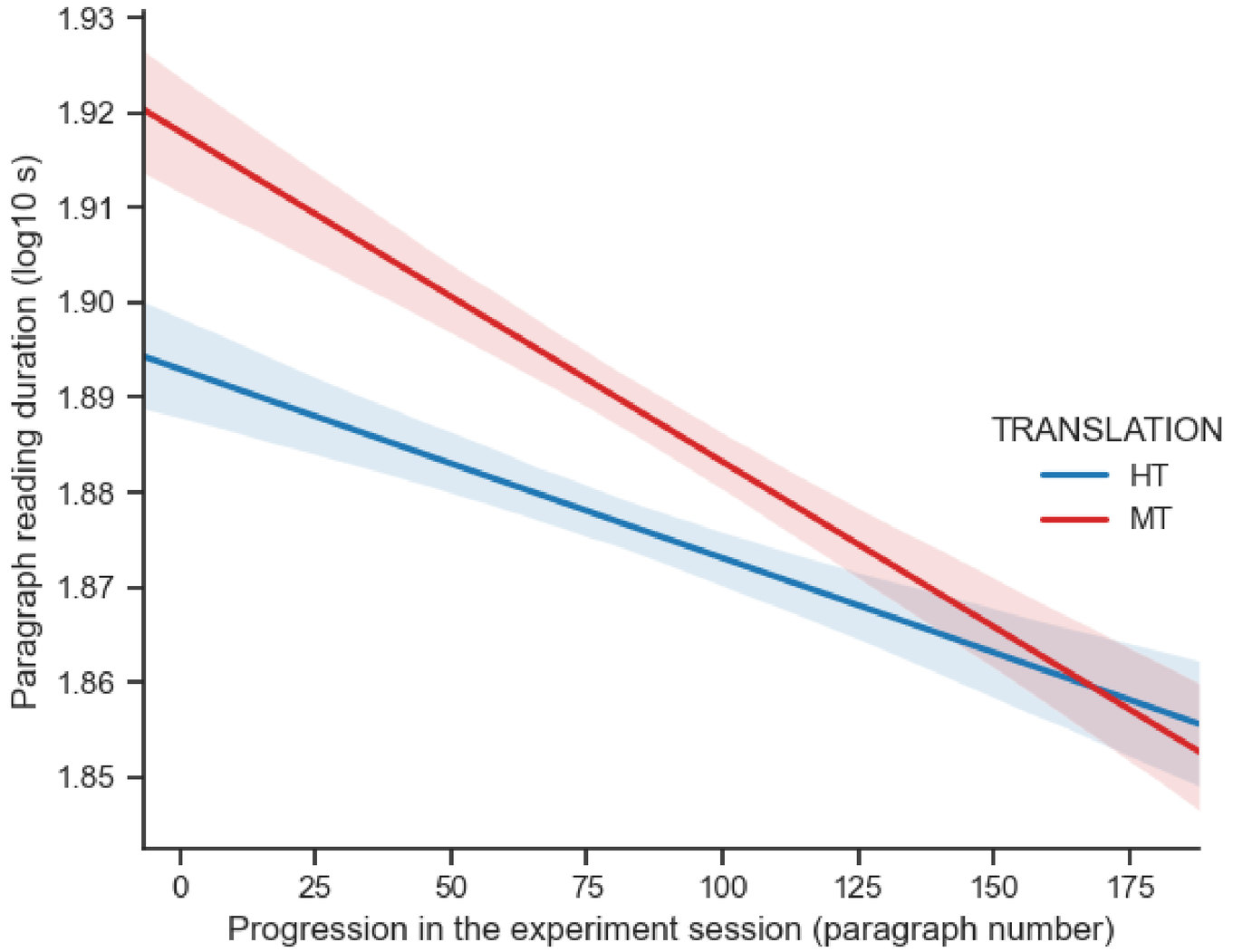
Increased reading duration when reading MT, *controlled for paragraph length*



GHENT UNIVERSITY

EXAMPLE: PARAGRAPH READING DURATION

Effect of time in the experiment. Influence of MT on reading duration is especially pronounced in the beginning of the experiment

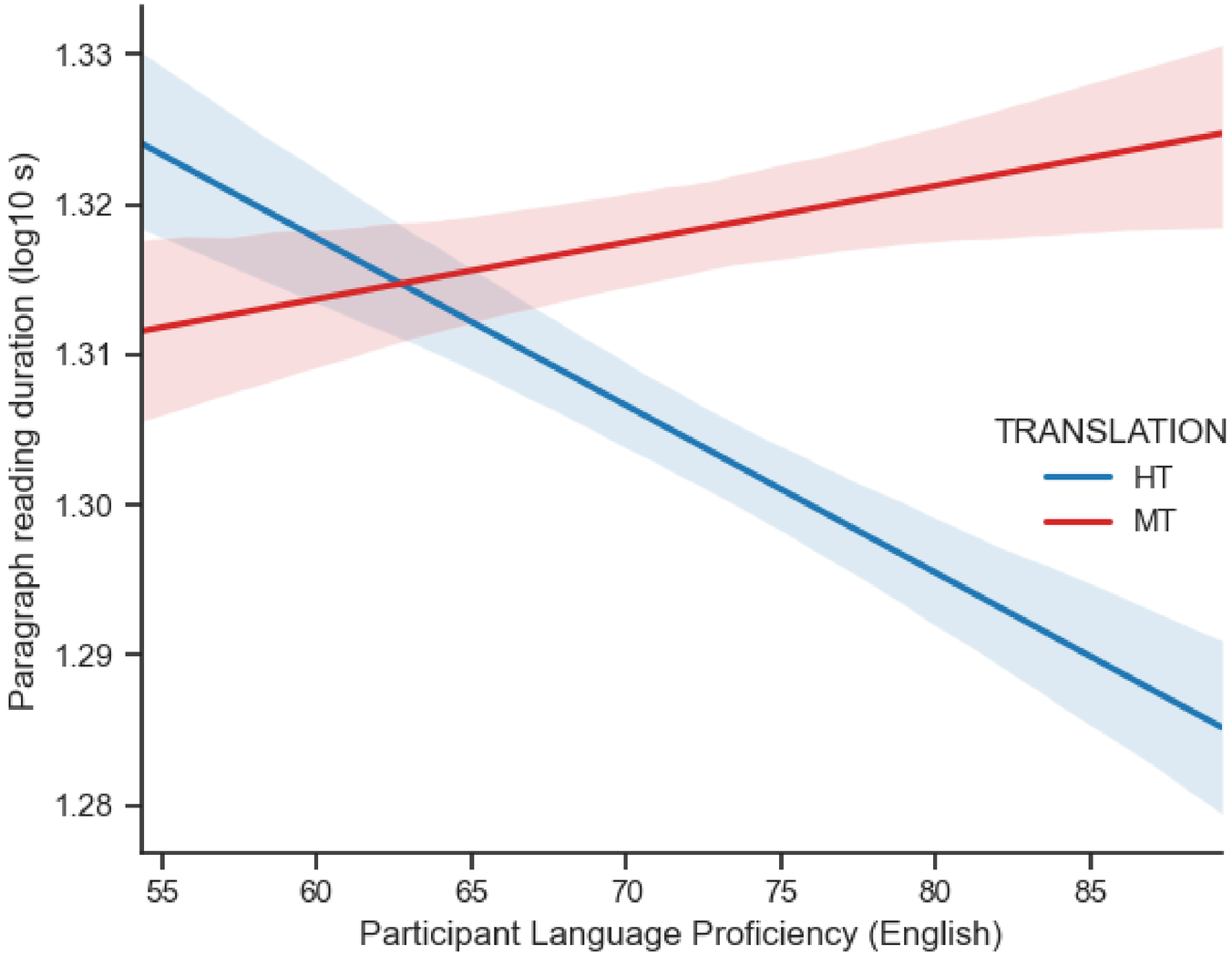


GHENT UNIVERSITY

EXAMPLE: PARAGRAPH READING DURATION

Effect of language proficiency.

Participants with higher proficiency read faster in the HT, but not in the MT



GHENT UNIVERSITY

GECO-MT DATA ANALYSIS

- Conclusion: when analyzing GECO-MT data, it is recommended to account for context variables, participant variables, …
- Linear mixed effects (LME) models are an example of a suitable statistical framework

Dependent Variable *Random Effect* *Random Effect*

```
Reading Duration ~ (1 | Subject number) + (1 | Paragraph number)
    + Translation * Language Proficiency + Paragraph Length
```

Fixed Effect *Fixed Effect* *Fixed Effect*

GHENT UNIVERSITY

WRAPPING UP

GHENT
UNIVERSITY

GHENT EYE-TRACKING CORPUS OF MT

- GECO-MT, a large corpus of eye movement data to study the effect of MT on text readability
- MT error annotations (Fonteyne et al., 2020) can be used to study the effects of specific MT error types
- GECO-MT is freely available at www.lt3.ugent.be/resources/geco-mt

GHENT UNIVERSITY

Fonteyne, M., Tezcan, A., & Macken, L. (2020). Literary machine translation under the magnifying glass : assessing the quality of an NMT-translated detective novel on document level. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, … S. Piperidis (Eds.), 12th International Conference on Language Resources and Evaluation Conference (LREC 2020), Proceedings (pp. 3783–3791). Paris, France: European Language Resources Association (ELRA).

FACULTY OF ARTS AND PHILOSOPHY

LANGUAGE AND TRANSLATION TECHNOLOGY TEAM (LT3)

lieve.macken@ugent.be
www.lt3.ugent.be

Universiteit Gent
@ugent
@ugent
Ghent University

GHENT UNIVERSITY

language and translation technology team

LREC 2022 Marseille

EUROPEAN LANGUAGE RESOURCES ASSOCIATION ELRA