1. ABOUT THE DATASET

----------------------------

Title: Combined eye-tracking and think-aloud data relating to how native-Chinese speakers process coreferential noun phrases in English-language academic texts

Creator: James P Wagstaffe

Organisation: University of Reading

Rights-holder(s): James P Wagstaffe

Year of publication: 2022

Description: This dataset consists of two parts. The first part contains eye-movement data reflecting the reading times gathered when native-Chinese speakers processed coreferential noun phrases which (i) either did or did not contain a grammatical determiner, and (ii) either did or did not contain a semantically incongruent word. The second part consists of the transcripts from think-aloud interviews which were conducted subsequent to the eye-tracking experiment to investigate how readers at different levels of proficiency processed the coreferential noun phrases.

Cite as: Wagstaffe, J.P. (2022) Combined eye-tracking and think-aloud data relating to how native-Chinese speakers process coreferential noun phrases in English-language academic texts. University of Reading. Dataset. <https://doi.org/10.17864/1947.000383>

1. TERMS OF USE

----------------------------

Copyright (2022) James P Wagstaffe. This dataset is licensed under a Creative Commons Attribution 4.0 International Licence: <https://creativecommons.org/licenses/by/4.0/>.

1. PROJECT AND FUNDING INFORMATION

----------------------------

Title: How native-Chinese speakers establish discourse-level mental representations of English-language academic texts: A mixed method study

Dates: 1 October 2017 – 1 October 2021

Funding organisation: SeNSS-ESRC

Grant number: ES/P00072X/1

1. CONTENTS

----------------------------

Eye-tracking data:

ChineseReadEnglishEyetracking.csv

Think-aloud interview transcripts. Initial three-digits indicate the participant number\_ IeltsG1 = Lower-intermediate reading proficiency group, IeltsG2 = intermediate reading proficiency group, IeltsG3 = upper-intermediate reading proficiency group\_ Final number indicates participant Ielts reading score:

004\_IeltsG2\_6.5.docx

005\_IeltsG2\_6.5.docx

006\_IeltsG3\_8.docx

008\_IeltsG3\_7.5.docx

009\_IeltsG3\_7.5.docx

013\_IeltsG1\_5.5.docx

015\_IeltsG3\_7.5.docx

017\_IeltsG1\_6.docx

023\_IeltsG1\_5.5.docx

025\_Ielts G1\_5.docx

028\_IeltsG1\_6.docx

029\_IeltsG3\_8.5.docx

032\_IeltsG1\_5.5.docx

033\_IeltsG2\_6.5.docx

056\_IeltsG3\_9.docx

057\_IeltsG3\_8.docx

060\_IeltsG2\_6.5.docx

071\_IeltsG2\_7.docx

1. DATA COLLECTION METHOD

----------------------------

Ethics and recruitment:

81 participants were recruited opportunistically from among the available university population. Efforts were made to ensure a roughly even range of readers with different levels of reading proficiency were represented in the sample. Prior to taking part in either the eye-tracking experiment or think-aloud interviews all participants were given an information sheet explaining what the study was about and what would be asked of them. Subsequently, they each signed a consent form expressly consenting not only to participate in the study, but for the data thereby collected to be held by the university once anonymised.

Eye-tracking methodology:

The eye-tracking experiment was conducted using an Eye-Link Portable Duo eye-tracker. The participants each read four quasi-academic articles, into which a total of 36 target noun phrases were embedded. Each target noun phrase appeared in one of four possible arrangements. Firstly, the noun phrase either did or did not contain a grammatical determiner (e.g. this/those). Secondly, the noun phrases contained a modifier which was either semantically congruous or semantically incongruous with the preceding dialogue. Eye-movement data was gathered on the head noun in the noun phrase to determine if either (*i*) the presence or absence of a determiner, or (*ii*) the presence or absence of a semantically incongruous modifier in the noun phrases resulted in increased or decreased gaze durations. The texts were counterbalanced so that different participants saw different configurations of the target noun phrases. Altogether, four eye-tracking measures were used, comprising two early and two late measures. The early measures were First Fixation Duration and Gaze Duration. The late measures were Selective Regression Path Duration and Second Pass Reading time (calculated by subtracting Gaze Duration from Total Reading Time).

The think-aloud interviews were conducted between a week and ten days after the eye-tracking experiment with eighteen participants who had also taken part in the eye-tracking experiment. The think-aloud interviews consisted of two parts. In the first part the participants were asked to voice their thoughts as they read one of the texts used in the eye-tracking experiment. During this phrase, they were asked to voice their thoughts in respect of found issues:

|  |  |  |
| --- | --- | --- |
|  | **How they thought the sentences in the text were related to each other** | *used to investigate if they had integrated information in the target noun phrases* |
|  | **How they knew the sentences in the text were related to each other** | *used to investigate what knowledge resources they drew upon if they integrated information in the target noun phrases* |
|  | **What, if any, comprehension difficulties they experienced as they read** | *use to investigate what, if any, comprehension difficulties they faced* |
|  | **What, if any, comprehension regulations strategies they used to overcome any such difficulties** | *used to investigate what, if any, comprehension regulation strategies they favoured* |

Secondly, the participants were asked to engage in an anaphor resolution task, whereby they were asked the following questions about each of the target noun phrases:

|  |  |  |
| --- | --- | --- |
| i)  | **What does this word (the main noun in the target noun phrase) refer to?** | *used to test the participant’s ability to integrate the noun phrase with its antecedent* |
| ii) | **How do you know this?** | *used to explore what knowledge resources the participant drew upon to integrate the noun phrase with its antecedent* |
| iii) | **What does the word mean in Chinese?** | *used to test if the participant had correctly understood the word* |
| iv) | **What adjective did the writer use to describe this word?** | *used to draw the participant’s attention to the incongruent word* |
| v) | **Why do you think they chose this adjective?** | *used to encourage the participant to focus on the meaning of the incongruent word* |

During the interviews, the participants were audio and video recorded, although the video recording only focused on the article they were reading, so that analysis could be conducted on where they pointed as they read. After the interviews were complete, they were transcribed either manually or using an online transcribing tool called Xunfei (<https://www.iflyrec.com>). As the participants were allowed to speak in whichever language was most convenient for them at any give moment, and at times engaged in a high degree of code switching, some of the transcripts contain both English and Mandarin Chinese. As it was felt that translating transcripts of speech containing a high degree of code switching runs the risk of reducing the accuracy of the data, no attempt to do so was made.