Predicting Mandarin and Cantonese Adult Speakers' Eye-Movement Patterns in Natural Reading

Understanding how language knowledge is processed and integrated incrementally has been an important question in language science. The reading of Chinese languages provides an interesting environment for investigating such inquires as its morphosyntactic nature and writing systems require a greater reliance on the lexical context during comprehension (Chen et al., 1992; Jian et al., 2013; Hsu et al., 2023), even in the initial stage of processing (Zhao et al., 2019). Chinese languages encompass varieties in written forms with nuanced visual complexity of words and morpho-syntactic structures. For example, Mandarin and Cantonese differ not only in their grammars and lexicon, but also in their writing systems among different Chinese speaking regions. For instance, Mandarin in Mainland China uses simplified Chinese characters in writing whereas Cantonese in Hong Kong uses traditional Chinese characters. This distinction raises the question of whether the visual complexity of characters influences the processing of characters, despite representing the same meanings (e.g., 书 'book' and 書 'book'). Moreover, simplified characters tend to be more ambiguous than traditional characters, as the former often associates multiple meanings with one character (e.g., 后 means either 'the back' or 'the queen'), while the traditional characters have more one-to-one correspondence between character and meaning (e.g., 后 means 'queen' and 後 means 'back'). Therefore, the contextual influence and predictability of words may differ between writing systems using simplified and traditional characters, in addition to understanding whether each character co-occurs with other characters to form either a word or a phrase. By comparing the processing of Chinese languages, we can gain profound insights into understanding linguistic variations as well as the integration of multilevel linguistic processing.

Eye movement data offer valuable insights into the cognitive processes involved in reading, shedding light on how language is processed across various aspects, such as morphology (Clifton Jr et al., 2007), syntax (Van Schijndel and Schuler, 2015), and semantics (Ehrlich and Rayner, 1981). Many studies have shown that certain characteristics of words can impact language processing, and reading behavior. These features include word position, word length, word frequency, and the number of syllables within the word (Just and Carpenter, 1980). Furthermore, the spillover effect (Rayner et al., 1989) suggests that the cognitive load imposed by a word can affect the processing of subsequent words (Pollatsek et al., 2008). Another important factor that impacts language comprehension is the predictability of a word at the sentence level, based on the preceding context (Kliegl et al., 2004). By measuring and examining eyemovement data, we can gain insights into how these factors influence the cognitive processes during language comprehension.

Therefore, the aim of the current study is twofold. First, adopting a computational approach, and using matched materials read by mature adult native speakers, we introduce a joint benchmark of eye-tracking data that capture natural reading patterns of reading Mainland Mandarin and Hong Kong Cantonese texts. An example fixation heatmap of the two languages is shown in Fig. 1.

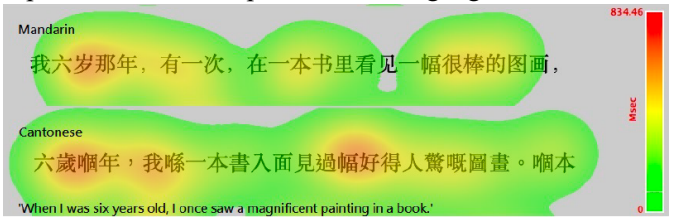

Fig. 1. A fixation heatmap of Mandarin and Cantonese sentences.

This dataset doubles the size of the dataset introduced in Li et al. (2023) and includes several new eye-movement metrics that capture initial lexical processing and later integrated structural processing (e.g., skipping, saccadic landing position, regression out), in addition to the commonly studied measures in eye-movement prediction studies (i.e., first fixation duration, and total reading time). Second, using this dataset, we examine the efficacy and interpretability of recent large language models in predicting early

Proceedings of the 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP (SIGTYP 2024), pages 44–45

and late eye-movement metrics during reading Mandarin and Cantonese texts. Fig. 2 shows an example that visualizes the distribution of predicted and ground-truth fixations over a Cantonese sentence (left) and its corresponding Mandarin sentence (right).

Fig. 2. Distribution of predicted (pred) and ground-true (true) fixations over target sentences. *Note: FFD refers to first fixation duration, SFD refers to second fixation duration, TFD refers to total fixation duration.*

Moreover, it has been shown that language models can be used to generate features capturing human-like behavior in terms of eye-tracking metrics, e.g., surprisal (Hale, 2001; Levy, 2008; Fossum and Levy, 2012; Hao et al., 2020) and token embeddings (Schrimpf et al., 2020; Hollenstein et al., 2021). Therefore, we also focus on whether and to what extent the features generated by language models help to predict various levels of cognitive processing during language comprehension. We conduct a comprehensive evaluation on the prediction of eye-tracking metrics using regression analysis. First, we target at a comparison of features' predictive power on Mandarin and Cantonese reading patterns. Second, we extend the examination to types of language models, specifically, on mono-lingual language model and multilingual language model. For each language, we use both incremental, autoregressive and masked language models, from which we extract features such as token embeddings and surprisals, and we combine them with lexical, orthographic, and syntactic features. In addition to monolingual, pre-trained language models for Chinese (Zhao et al., 2019), we also test multilingual models that have been trained simultaneously for Mandarin and Cantonese (Yang et al., 2022).

Our study presents some highlights on how to account for multilinguality by language models and how they facilitate in-depth investigation of closely related languages in comprehension. Specifically, Mandarin and Cantonese exhibit different advantages in terms of the cognitive effort on word-visual processing (for Mandarin) and sentence contextual processing (for Cantonese). Results of our study will further inform the performance of different types of language models and metrics' usefulness in predicting eye-movement patterns in reading Mandarin and Cantonese texts.

Selected References

Chen, H. C. (1992). Reading comprehension in Chinese: Implications from character reading times. In Advances in psychology (Vol. 90, pp. 175-205). North-Holland.

Jian, Y. C., Chen, M. L., & Ko, H. W. (2013). Context Effects in Processing of Chinese Academic Words: An Eye-Tracking Investigation. Reading Research Quarterly, 48(4), 403-413.

Li, J., Peng, B., Hsu, Y. Y., & Chersoni, E. (2023). Comparing and Predicting Eye-tracking Data of Mandarin and Cantonese. In *Proceedings of the EACL Workshop on NLP for Similar Languages, Varieties and Dialects*.

Zhao, S., Li, L., Chang, M., Xu, Q., Zhang, K., Wang, J., & Paterson, K. B. (2019). Older adults make greater use of word predictability in Chinese reading. Psychology and Aging, 34(6), 780.

Hao, Y., Mendelsohn, S., Sterneck, R., Martinez, R., & Frank, R. (2020). Probabilistic predictions of people perusing: Evaluating metrics of language model performance for psycholinguistic modeling. arXiv preprint arXiv:2009.03954.

Yang, Z., Xu, Z., Cui, Y., Wang, B., Lin, M., Wu, D., & Chen, Z. (2022). CINO: A chinese minority pre-trained language model. Proceedings of COLING.