

# Why observe eye-movements?

We fixate on things we are mentally paying attention to

- > Information on **cognitive processes** involved in language processing (Rayner 1998)
- > Prediction and integration of **semantic information** during **structure-building**, e.g.:
  - word sense disambiguation
  - plausibility
  - dependency processing
- > Online measures  $\Rightarrow$  temporal order of processes

# About the data

- > Several chapters from a Tagalog translation of The Little Prince
- > Parallel corpora in other languages (currently Mandarin and Cantonese; Li et al. 2023)
- > Text presented in short snippets, with occasional comprehension questions
- > Gathered from Tagalog-speaking participants living in Hong Kong with an **Eyelink 1000 Plus**





# The corpora

In progress

Previous studies

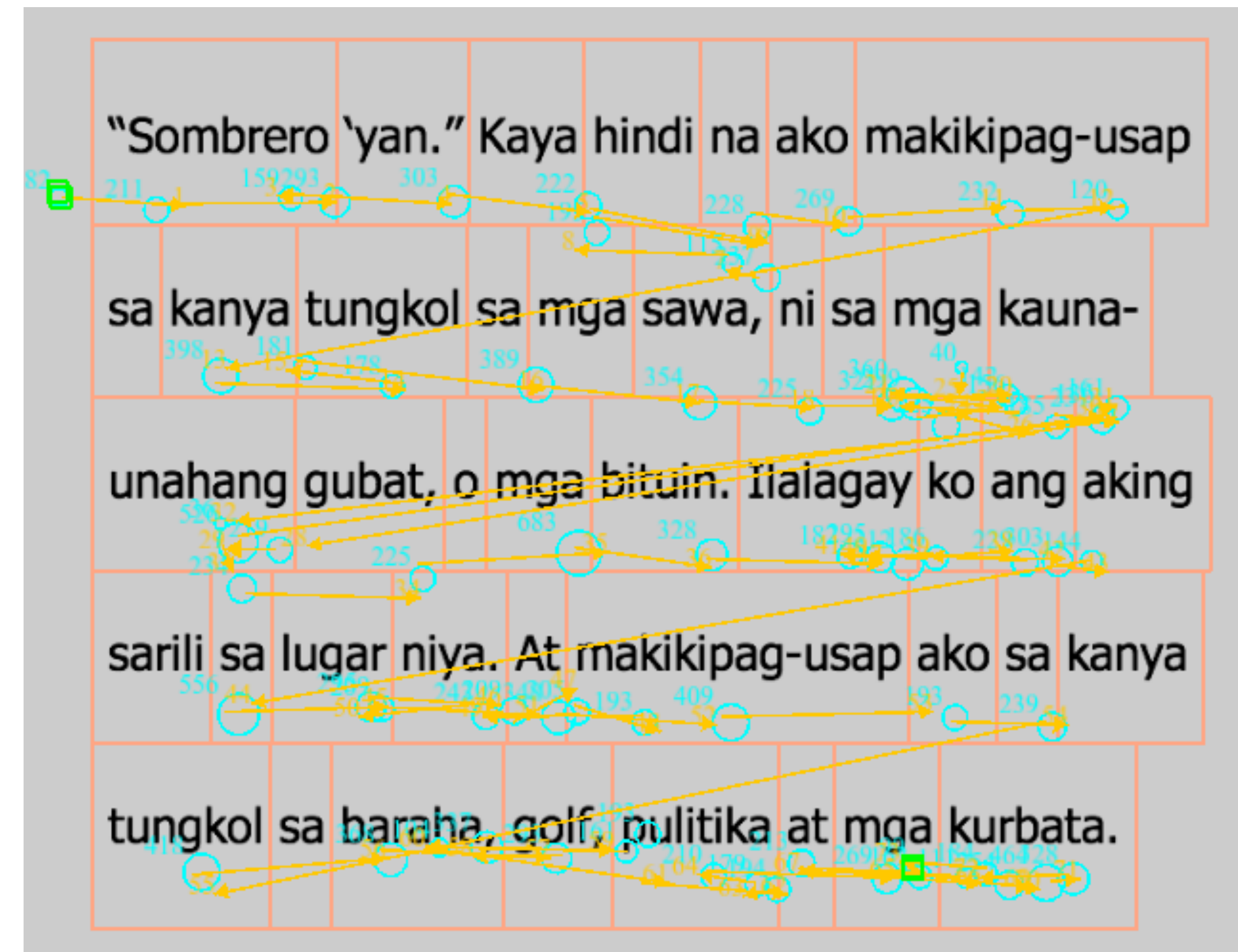
	Tagalog	Mandarin	Cantonese
# tokens	4534	4357	5074
# types	1011	1016	1036
mean # tokens / sentence (sd)	9.51 (6.38)	12.17 (7.66)	15.10 (8.68)
range # tokens / sentence	1-49	1-42	1-46
mean token length (sd)	4.95 (2.80)	1.41 (0.63)	1.34 (0.61)
range of token length	2-23	1-8	1-6
N =	9 (5f; 30.55y)	15 (11f; 25.8y)	15 (10f; 25.27y)



# Eye movements on text

Patterns of eye movement during text reading can be influenced by the readability and by the task (e.g., Rayner et al. 2006, Hollenstein et al. 2018)

- > **↑ predictable words = ↑ likely skipped, ↓ fixation time**
- > Language-specific reading patterns and effects may exist





Initial features

# Word omission

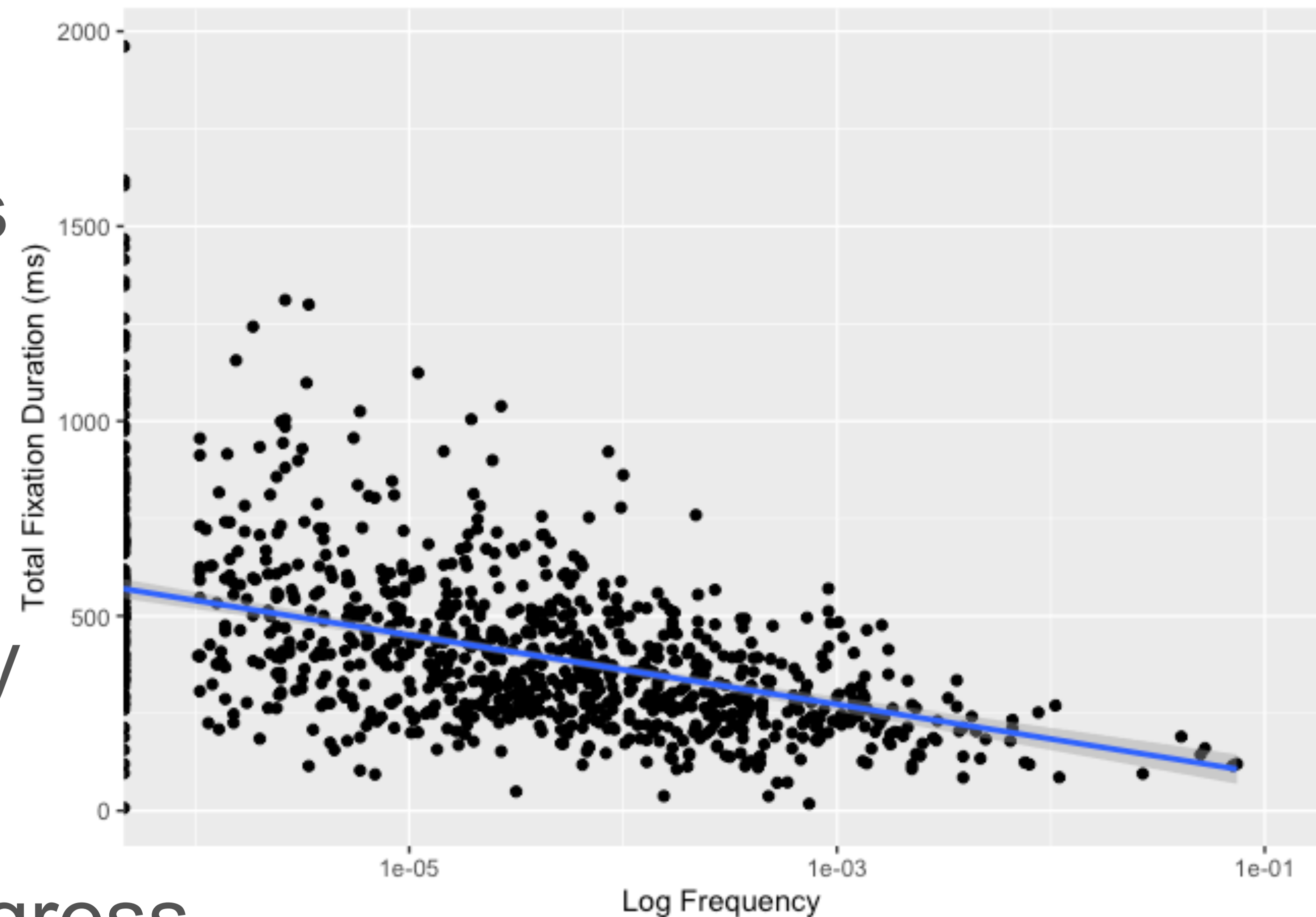
- > How often did participants skip a word?
- > 18 types skipped  $\geq 50\%$   
... of which 6 are N=1: *Mars*, *Venus*,  
*hayop* ‘animal’, *iglap* ‘(an) instant’  
*mundo* ‘world’, *deretso* ‘straight’
- > Mean skipping rates:
  - > Tagalog: 0.2607
  - > Mandarin: 0.4746
  - > Cantonese: 0.5520

word	gloss	rate
<i>e</i>	PART	0.8333
<i>o</i>	or; PART	0.6042
<i>at</i>	and	0.5981
<i>ay</i>	TOP	0.5903
<i>dito</i>	here	0.5556
<i>ng</i>	GEN	0.5322
<i>ito</i>	this	0.5167
<i>iba</i>	different	0.5079
<i>sa</i>	OBL	0.5062
<i>ko</i>	1sg.GEN	0.5
<i>iyo</i>	2sg.OBL	0.5
<i>gamit</i>	use, thing	0.5

Initial features

# Fixation duration

- > How long did participants look at a word?
- > Potential effects of:
  - > Frequency
  - > Morphological complexity
  - > Part of Speech
- > Metadata annotation in progress....





# Features to be added

## Lexical Features

- > Frequency
- > Part of Speech
- > Morphological Parse?
- > Lemmatization?

## How to handle Tagalog morphology/orthography?

*naka-pag-do~drowing* 'is able to draw'

*drowing* 'a drawing'

*libro='y* 'book + TOP' contraction

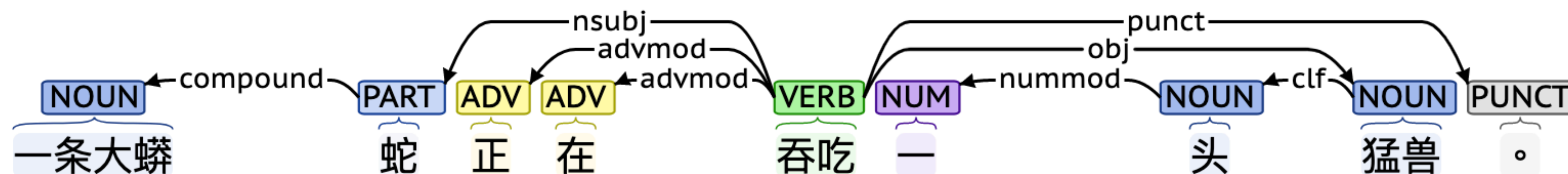
*n'ya=ng* '3sg.GEN contraction + linker'  
vs *niya=ng* vs *n'ya* vs *niya*

# Features to be added

## Text/Syntactic Complexity

- > Lexical/Orthographic neighborhood number?  $\approx$  Predictability
- > Syntactic dependency information  $\approx$  Cognitive load
  - > Linear distance to head / root
  - > Dependency depth from root

Example from Mandarin





- Thanks to Junlin Li, Wenxi Fei, and Yimei Shao at the Hong Kong Polytechnic University Chinese and Bilingual Studies department for their help with data processing and technical support!
- Hollenstein, N., Rotsztein, J., Troendle, M. *et al.* (2018). ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading. *Sci Data* 5, 180291. <https://doi.org/10.1038/sdata.2018.291>
- Hollenstein, N., Chersoni, E., Jacobs, C., Oseki, Y., Prévot, L., Santus, E. (2022). [CMCL 2022 Shared Task on Multilingual and Crosslingual Prediction of Human Reading Behavior](#). In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 121–129, Dublin, Ireland. Association for Computational Linguistics.
- Li, J., Peng, B., Hsu, Y.-Y., Chersoni, E. (2023). Comparing and Predicting Eye-tracking Data in Mandarin and Cantonese. In *Tenth Workshop on NLP for Similar Languages, Varieties and Dialects* (VarDial 2023), 121–132. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.vardial-1.12>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 372–422.
- Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, 21(3), 448–465. <https://doi.org/10.1037/0882-7974.21.3.448>