Prior Reading Experience Explains Individual Differences in Nonword Reading

Ana Ulicheva and Kathleen Rastle
Royal Holloway, University of London, UK

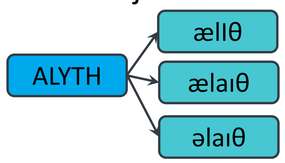
1. Background

Nonword reading in English is variable.

Previous studies focused on differences across nonwords [1, 2, 3].

VASUBE SOTTLE
ALYTH PASAKE

Our study focuses on differences across subjects.



2. General Approach

Differences in nonword reading across subjects reflect differences in reading experience.

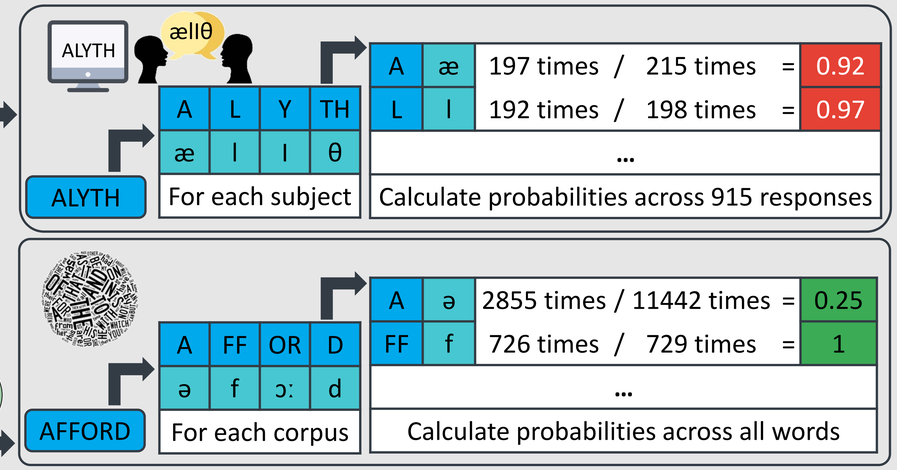
Differences in corpus size = Differences in reading experience → Individual differences in nonword reading

Print-to-sound statistics from corpora were compared to those of individual readers.

HUMAN DATA
Mousikou et al. (2017)
41 participants
915 disyllabic nonwords
Spelling scores [4]
Vocabulary [5]

CORPUS DATA
CELEX
75k words < 9 letters

16 corpora sampled using frequency cut-offs



3. Are Better Readers Closer to Large Corpora?

HYPOTHESIS

Large corpus = Good readers
Large corpus ≠ Poor readers

METHODS

A	æ	0.92
L	l	0.97
...		
For every subject		

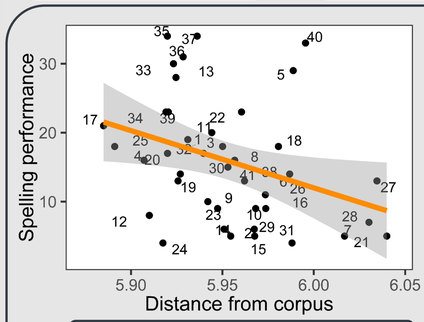
A	æ	0.92
L	l	0.97
...		
Full corpus		

Measured Euclidean distances

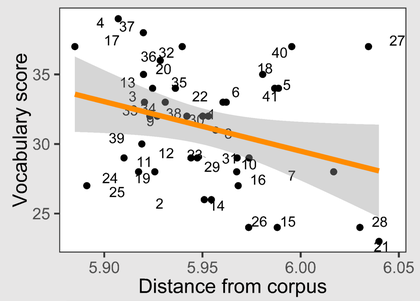
Corpus	Subjects				
	6.1	6.4	6.9	...	6.3

Produced a vector of mean distances between grapheme-to-phoneme correspondences (GPCs) in the corpus vs. each participant.

We investigated the relationship between the vector of distances from the corpus and subjects' spelling and vocabulary scores.



B = -83.02, *t* = -2.38, *p* < 0.05



B = -35.60, *t* = -2.05, *p* < 0.05

RESULTS

The further participants were from the corpus in terms of the print-to-sound statistics that they utilise, the lower their spelling and vocabulary scores were.

4. Variability across Humans and Corpora

HYPOTHESIS

Differences across corpora = Subject 1, Subject 1, Subject 3 ↕ Differences across subjects

METHODS

A	æ	0.07	0.01	...	0.09
L	l	0.97	0.95		0.80
...					
16 corpora					

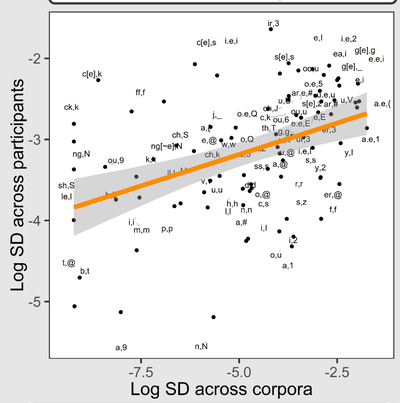
Calculated standard deviation (SD) for each GPC across corpora

A	æ	0.07	0.01	...	0.09
L	l	0.97	0.95		0.80
...					
41 subjects					

Calculated SD for each GPC across subjects

RESULTS

B = 0.33, *t* = 3.33, *p* < 0.01



The more variable the statistics were across corpora, the more variable they were across humans.

5. Conclusions

Varying reading experience can be modelled by diminishing a large corpus of written English.
Good readers' print-to-sound statistics resemble a large corpus, whereas poor readers' statistics are further away from large corpora.
Print-to-sound statistics vary less across individuals if they are less dependent on reading experience (i.e. changes in corpora size).

6. Funding and References

This work was funded by the ESRC Future Research Leader Fellowship (grant number ES/N016440/1) and the Marie Curie Individual Fellowship (747987) awarded to Ana Ulicheva; and ESRC project grants awarded to Kathy Rastle (ES/L002264/1 and ES/P001874/1).

[1] Pritchard, S., Coltheart, M., Palethorpe, S., & Castles, A. (2012). *JEP: HPP, 38*(5). [2] Mousikou, P., Sadat, J., Lucas, R., & Rastle, K. (2017). *JML, 93*, 169–192. [3] Coltheart, M., & Ulicheva, A. (2018). *PeerJ, 6*: e4879. [4] Burt, J., & Tate, H. (2002). *JML, 46*, 518 543. [5] Shipley, W. C. (1940). *Journal of Psychology, 9*, 371–377.