Discovering the Lexicon's Statistical Structure in Reading

Jarosław R. Lelonkiewicz and Davide Crepaldi
International School for Advanced Studies (SISSA), Trieste, Italy

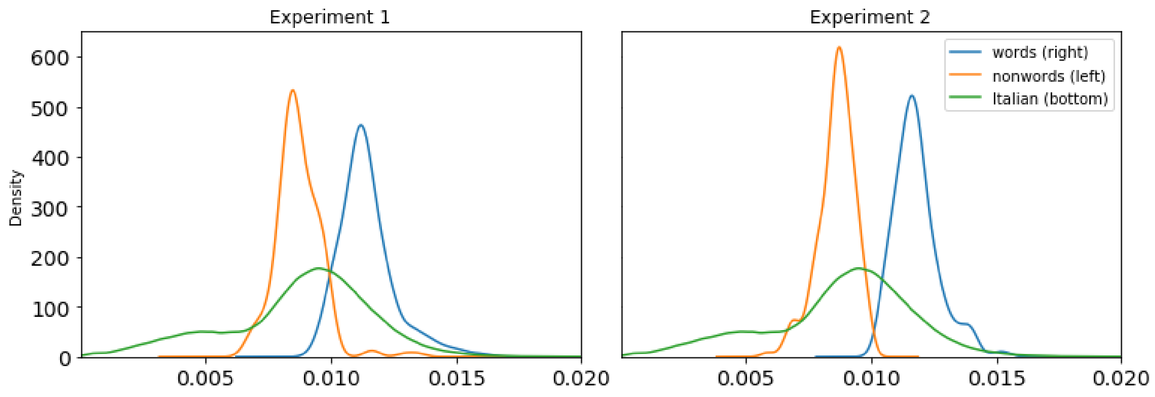
Background

When processing speech, learners discover the probabilistic regularities in the input (1). But it is unclear whether they do so also when reading – current evidence comes from studies of non-linguistic animals (2) and experiments where human readers process toy artificial lexicons (3). To bridge this gap, we challenged readers to learn a large set of items modelled on a natural lexicon, and tested whether in doing so they relied on probabilistic information.

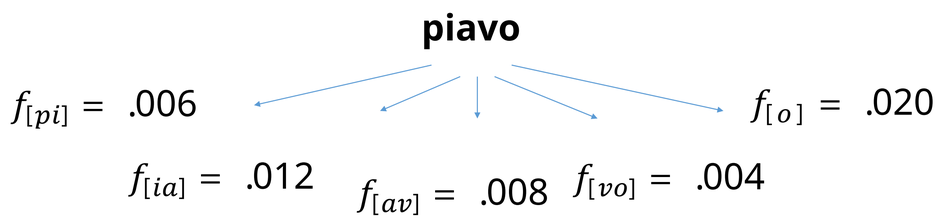
Task

In **two experiments** (n_1=42, n_2=42), Italian monolinguals saw letter strings and judged which were words in a fictitious, unfamiliar language, and which were not. Participants received feedback on each trial.

Materials



words had higher **mean Bigram Frequency** than *nonwords* but either resembled *Italian words* (i.e., word length, letter probabilities, bigram probabilities, transition probabilities; 4)

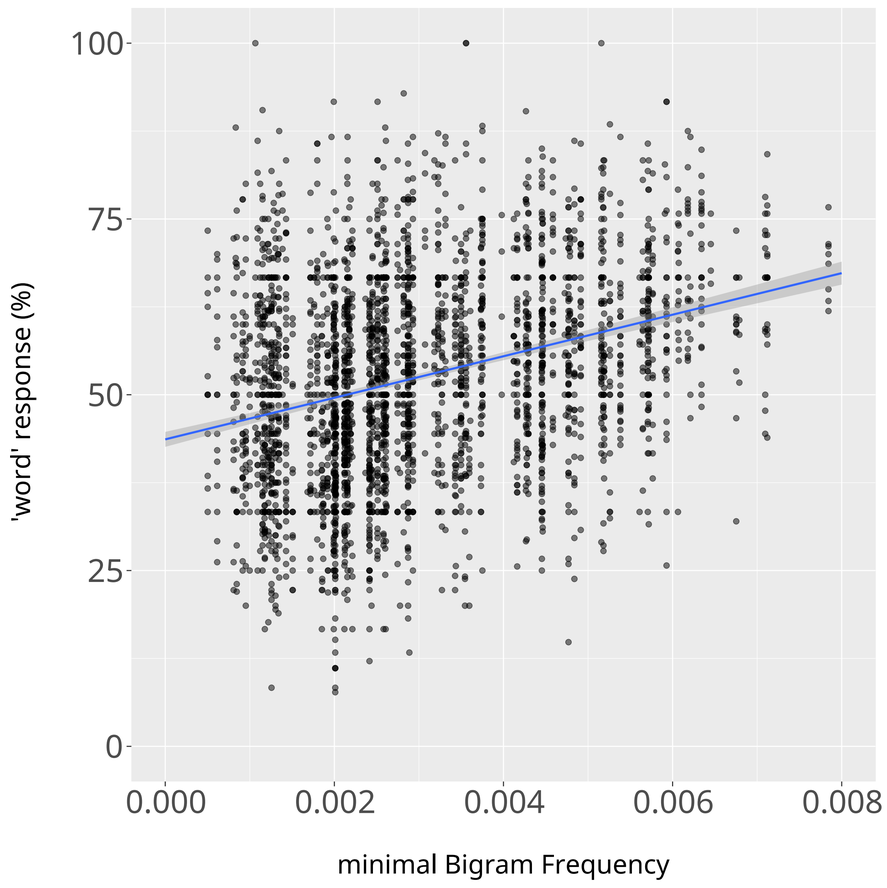


We tested if participants:

- used the local regularity underlying words/nonwords (mean BF)
- spontaneously used another cue based on L1 (minimal BF)

Findings

		B	*SE*	*z*	*p*
mean BF	E1	.04	.08	0.57	.565
	E2	.09	.10	0.90	.369
mean BF * trial	E1	.01	.06	0.22	.822
	E2	-.04	.08	-0.52	.605
minimal BF	E1	.22	.09	2.31	**.021 ***
	E2	.14	.08	1.74	.082 .
minimal BF * trial	E1	.08	.08	1.07	.285
	E2	-.14	.07	-2.04	**.042 ***



Discussion

Participants showed no learning of the local regularity – they instead used another statistic based on their L1. This is consistent with the proposition that languages are represented in a probabilistic manner (5, 6). Further, our data supports the hypothesis that the knowledge of one's native language generates stable, "entrenched" assumptions that determine which regularities are tracked in novel materials (7, 8). We propose that, when unfamiliar materials (e.g., novel words) resemble an already known system (e.g., familiar lexicon), learners may prioritise existing knowledge over exploration of the local statistical regularities.

References

1 Saffran, J. R., & Kirkham, N. Z. *Annu. Rev. Psychol.* 69 (2018) **2** Grainger, J., et al. *Science* 336 (2012) **3** Newport, E. L. *Lang. Cog.* 8 (2016) **4** Crepaldi, D., et al. *Proceedings of the Annual Meeting of the Italian Association For Experimental Psychology* (2015) **5** Marelli, M., et al. Q. J. *Exp. Psychol.* 68 (2015) **6** Siegelman, N., et al. *Cogn.* 177 (2018) **7** Ulicheva, A., et al. *Cogn.* (2018) **8** Potter, C. E., et al. *Cogn. Sci.* 41 (2017)

SISSA 40!
L2R LANGUAGE LEARNING READING
erc European Research Council Established by the European Commission