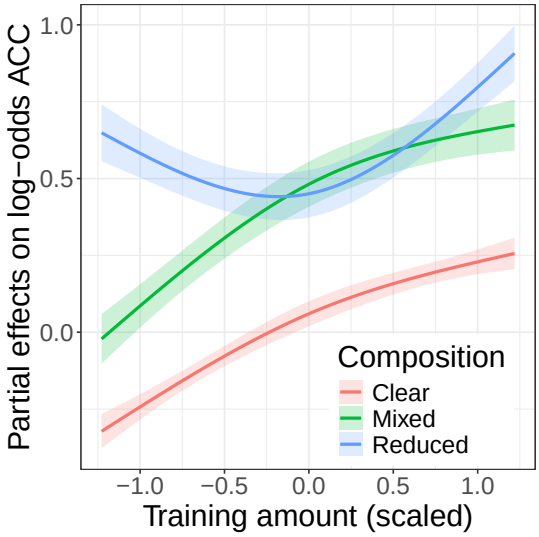
Abstractionism, exemplars, or discriminative learning?: A two-axis test on L2 reduced-word-form recognition

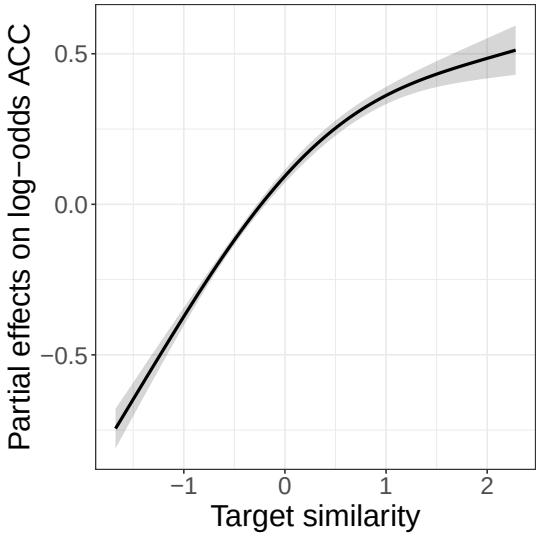
Motoki Saito (motoki.saito@uni-oldenburg.de)[1], Esther Ruigendijk (esther.ruigendijk@uni-oldenburg.de)[1]

[1]Institute of Dutch Studies, University of Oldenburg

Introduction: How phonetic variation is represented in the mental lexicon remains contested. *Abstractionist* accounts assume the lexicon stores only canonical forms [e.g., 4]. *Exemplar* accounts instead store every encountered word-form [e.g., 3]. Both struggle with the prevalence of reduced forms in spontaneous speech [e.g., 2]. Abstractionists must select a canonical form among variable inputs, while exemplar accounts must store forms typically heard only once or twice. The Discriminative Lexicon Model [DLM; 1] occupies an intermediate position, mapping acoustics directly onto meanings via learned weights, with no word-level representations. These accounts make distinguishable predictions about how training quantity and *type* (only clear, only reduced, or mixed forms) shape reduced form recognition. Abstractionism predicts both the highest accuracy and the largest training-amount slope for clear-only, since clear inputs most reliably motivate canonical representations, with reduced-only worst on both axes. Exemplar theory predicts the opposite: highest accuracy and steepest slope for reduced-only, and lowest accuracy and a near-flat slope for clear-only, since exemplars trained only on clear forms cannot match reduced forms precisely. DLM agrees with exemplar on the ordering, but predicts a clearly positive clear-only slope, since every encounter updates the same acoustics-to-meaning weights. Type effects therefore distinguish abstractionism (clear-only best) from the other two, while the clear-only slope distinguishes exemplar (near-flat) from DLM (clearly positive). **Methods:** 81 German native speakers were trained on 10 semantically similar but phonologically distinct Dutch word-pairs and tested on reduced forms (320 trials). Reduced and clear pronunciations were elicited from the same Dutch native speaker in spontaneous conversation and in isolation, respectively. On each trial, participants heard a Dutch word and chose the German translation from two options. Participants were assigned to 1 of 9 conditions (3 training type $\times$ 3 training amount) design. **Results:** Accuracy was modelled with GAMMs, with training type, amount, and covariates as predictors and word as a random effect. Accuracy increased with training for clear-only and mixed conditions (Figure 1a). Reduced-only showed a U-shaped trajectory with accuracy lowest at the middle amount, possibly an artifact due to ceiling performance. To probe the mechanism, we trained three DLM models, one per training type, mirroring the experiment setup, and extracted *target similarity* [5]: similarity between meanings predicted from the acoustic inputs and the correct meanings. A third GAM replaced both training type and amount with this single predictor, and outperformed both prior models in AIC (ΔAIC $= 6.61$ and 6.65). Higher target similarity (more reliable form-to-meaning mappings) predicted better recognition (Figure 1b). **Discussion:** The data contradict abstractionism: clear-only training yielded the *lowest*, not highest, accuracy. Its substantial positive amount slope further rules out exemplar theory's near-flat prediction. DLM is the only account consistent with both findings, and the DLM-derived target-similarity model further outperformed the compositional GAMs in AIC, providing positive support for its mechanism: every encounter updates the same acoustics-to-meaning weights, with reduced-form training yielding a configuration that better supports reduced-form recognition, without word-level representations.



(a) Amount by training type



(b) Target similarity

Figure 1: **Left:** partial effects of training amount (x-axis; scaled: -1, 0, 1 $\rightarrow$ 40, 80, 120, approximately) on reduced-form recognition accuracy (y-axis), split by training type (red: clear-only, green: mixed, blue: reduced-only). **Right:** partial effects of DLM-derived target similarity (x-axis; scaled: -1, 0, 1, 2 $\rightarrow$ 0.10, 0.29, 0.47, 0.67, apprixmately) on reduced-form recognition accuracy (y-axis). Higher target similarity indicates that word meanings are more easily (less ambiguously) predicted from acoustic inputs.

References

[1] R. Harald Baayen, Yu-Ying Chuang, Elnaz Shafaei-Bajestan, and James P. Blevins. The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, 2019(1):4895891, 39, 2019. doi: 10.1155/2019/4895891. URL `https://onlinelibrary.wiley.com/doi/abs/10.1155/2019/4895891`.

[2] Mirjam Ernestus and Natasha Warner. An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(3):253–260, 2011. ISSN 0095-4470. doi: /10.1016/S0095-4470(11)00055-6. URL `https://www.sciencedirect.com/science/article/pii/S0095447011000556`.

[3] Stephen D. Goldinger. Echoes of echoes? an episodic theory of lexical access. *Psychological Review*, 105(2):251–279, 1998. doi: 10.1037/0033-295X.105.2.251.

[4] James L. McClelland and Jeffrey L. Elman. The TRACE model of speech perception. *Cognitive Psychology*, 18(1):1–86, 1986. ISSN 0010-0285. doi: 10.1016/0010-0285(86)90015-0. URL `https://www.sciencedirect.com/science/article/pii/0010028586900150`.

[5] Elnaz Shafaei-Bajestan, Masoumeh Moradipour-Tari, Peter Uhrig, and R. Harald Baayen. LDL-AURIS: A computational model, grounded in error-driven learning, for the comprehension of single spoken words. *Language, Cognition and Neuroscience*, 38(4):509–536, 2023. doi: 10.1080/23273798.2021.1954207. URL `https://doi.org/10.1080/23273798.2021.1954207`.