Word recognition selects for morphologically structured representations: The case of English *-er*

Jon Gauthier and Canaan Breiss

What kinds of mental representations drive word recognition beyond surface phonetics? We address this question by studying the representations of a speech foundation model at two stages: 1) after pretraining on raw speech audio, and 2) after fine-tuning for word recognition. We focus on English word-final *-er*, which has three surface-identical morphological sources—comparative (*bigger*), agentive (*diver*), and monomorphemic (*bitter*).

We find the model's representational space exhibits a linear geometric relationship between base and derived forms (*big→bigger*, *dive→diver*). This structure is specific to particular morphological sources, indicating that abstract morphology can be learned from unlabeled speech alone. However, the model overgeneralizes this pattern to false-friend pairs (*bit→bitter*).

Fine-tuning for word recognition significantly strengthens this structure for true morphological relations, while suppressing relations in monomorphemes and lexically irrelevant information (ex., speaker identity).

These results suggest abstract morpho-phonological structure is learnable from raw speech input, and plays a central role in word recognition.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*