Predicting semantic priming in Hebrew morphology using word embeddings

Itamar Kastner (University of Edinburgh)
itamar.kastner@ed.ac.uk

Introduction. Vector space representations of word meaning ("word embeddings") have seen a resurgence in recent years, being used not only in NLP but in psycholinguistic studies as well. While it has become increasingly common to evaluate whether these abstract representations can be used to model human behavior, two insights are lacking from the state of the art. The first is that even though word embeddings have been compared with human ratings (Mitchell and Lapata 2010; Hill et al. 2015), and word embeddings have been used to predict behavioral and neurophysiological measures (Ettinger and Linzen 2016; Ettinger et al. 2016; Mandera et al. 2017; Broderick et al. 2018; Jacobs and Erk 2019; Zhang et al. 2019), we do not yet know how well word embeddings compete with human ratings which *themselves* are used to predict behavioral measures. The second is in-depth engagement with a language that has complex morphology, something that might make it more difficult for word embeddings to generalize from divergent surface forms to shared representations. These tasks are taken up here by comparing word embeddings with human ratings as predictors of a semantic priming experiment in Hebrew, a language with complex non-concatenative morphology (see Appendix).

Methods. In a recent cross-modal priming study, Farhy and Veríssimo (2019) demonstrated not only a typical facilitatory effect of shared root priming in verbs, but a novel interaction between template and semantic similarity: their primes and targets were rated for semantic similarity by human raters, and these ratings were found to interact with the template of the prime. These results have interesting consequences for our understanding of the mental lexicon in languages like Hebrew; what is important for present purposes is whether these findings can be replicated using word embeddings. The original raw results from the priming study were thus regressed against different kinds of similarity ratings, and the results compared with those of Farhy and Veríssimo (2019): how close do word embeddings come to human ratings as predictors of participants' behavior? The models investigated included a "simple" word-based WORD2VEC model and two models based on WORD2VEC with dependency parses (Levy and Goldberg 2014), one using UDPIPE (Straka and Straková 2017) and one using YAP (More et al. 2019).

Results. The main finding of Farhy and Veríssimo (2019) consisted of an interaction between prime type on the one hand, and similarity of prime and target on the other hand (see Appendix). This finding emerges in the Human Ratings condition but was not replicated under any of the word embedding conditions (small t-values and large p-values, following the same analysis as in the original paper), or even a combination thereof. Among these conditions, UDPIPE provided a better fit to the data than the word-based model, which in turn performed better than YAP, as revealed by pairwise log-likelihood model comparisons. Yet these differences were small, and in any case did not address the main question of this study.

Discussion. In this reanalysis of the original study, human ratings led to an interaction but word embeddings led to a null result. A number of explanations for this difference come to mind, all of which deserve further investigation. First, perhaps word embeddings do well with limited surface forms (as in English) but cannot learn higher-order relationships between lemmas (as in Hebrew) without additional parameters. Second, it might be the case that there was not enough computational "firepower" in the current design: experimenting with different parameters (e.g. larger window/dimension size) or language models (e.g. BERT) might create a qualitative difference. Third, one could imagine additional evaluation techniques which may unveil similarities between human ratings and word embeddings for Hebrew outside of this manipulation, or for this manipulation outside of Hebrew. And most intriguingly, it could be the case that word embeddings correlate with human ratings and with experimental results, but not with human ratings' prediction *of* experimental results.

Form	Prime			Target
	Unrelated	*XaYaZ*	*XiYeZ*	*hitXaYeZ*
1st Past	mʃktj	ʔbdtj	ʔjbdtj	htʔbd
	/maʃaxti/	/avadti/	/ibadti/	/hitʔabed/
	pulled	was lost	lost	committed suicide
Infinitive	lʃpr	lxlwq	lxlq	htxlq
	/leʃaper/	/laxlok/	/lexalek/	/hitxalek/
	improve	share	divide	share/divide

Table 1: Example stimuli (reproduced from Farhy and Veríssimo 2019)

References

Broderick, Michael P., Andrew J. Anderson, Giovanni M. Di Liberto, Michael J. Crosse, and Edmund C. Lalor. 2018. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Current Biology* 28:803–809.e3.

Ettinger, Allyson, Naomi H. Feldman, Phillip Resnik, and Colin Phillips. 2016. Modeling N400 amplitude using vector space models of word representation. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

Ettinger, Allyson, and Tal Linzen. 2016. Evaluating vector space models using human semantic priming results. In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP*, 72–77.

Farhy, Yael, and João Veríssimo. 2019. Semantic effects in morphological priming: The case of Hebrew stems. *Language and Speech* 62:737–750.

Hill, Felix, Roi Reichart, and Anna Korhonen. 2015. SimLex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics* 41:665–695.

Jacobs, Cassandra L., and Katrin Erk. 2019. Using word2vec to predict human language processing. In *The 32nd Annual CUNY Conference on Human Sentence Processing*.

Kastner, Itamar. 2020. *Voice at the interfaces: The syntax, semantics and morphology of the Hebrew verb*. Open Generative Syntax. Language Science Press.

Levie, Ronit, Orit Ashkenazi, Shirly Eitan Stanzas, Rachel (Chelli) Zwilling, Efrat Raz, Liat Hershkovitz, and Dorit Ravid. 2020. The route to the derivational verb family in Hebrew: A psycholinguistic study of acquisition and development. *Morphology* 30:1–60.

Levy, Omer, and Yoav Goldberg. 2014. Dependency-based word embeddings. In *Proceedings of ACL 2014*.

Mandera, Paweł, Emmanuel Keuleers, and Marc Brysbaert. 2017. Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language* 92:57–78.

Mitchell, Jeff, and Mirella Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science* 34:1388–1429.

More, Amir, Amit Seker, Victoria Basmova, and Reut Tsarfaty. 2019. Joint transition-based models for morpho-syntactic parsing: Parsing strategies for MRLs and a case study from modern Hebrew. *Transactions of the Association for Computational Linguistics* 7:33–48.

Straka, Milan, and Jana Straková. 2017. Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 88–99. Vancouver, Canada: Association for Computational Linguistics.

Zhang, Linmin, Lingting Wang, Jinbiao Yang, Peng Qian, Xuefei Wang, Xipeng Qiu, Zheng Zhang, and Xing Tian. 2019. Can computers understand word meanings like the human brain does? comparable semantic representation in neural and computer systems. Ms, NYU Shanghai. bioRxiv 843896; doi:10.1101/843896.

Appendix: Roots and Templates. In Hebrew, a word is typically made up of a consonantal *root* and a prosodic and segmental *template*. Verbal templates often stand in certain syntactic and semantic relationships with one another; see Figure 1. The derivational relationships between the templates are a matter of much theoretical debate (Kastner 2020; Levie et al. 2020). Each verb also inflects for person, gender and number, further increasing the total number of surface forms per root.

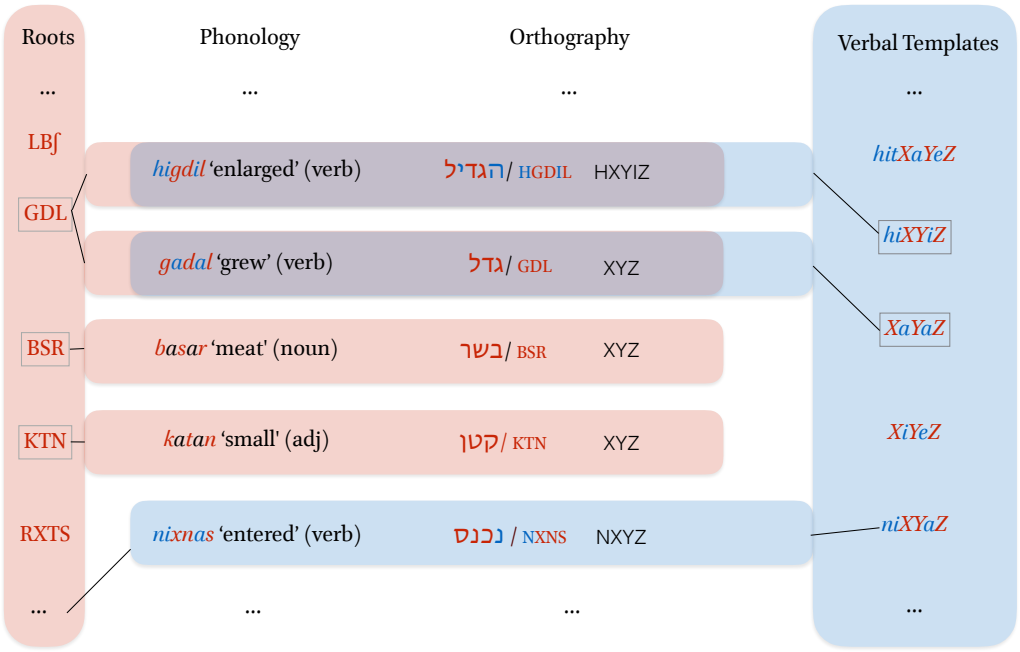

Figure 1: Some roots and templates in Hebrew

Even though certain pairs of templates *often* stand in a transparent relationship with one another, there is no guarantee that they always would. The relationship in (1) is a simple causative one—'enlarge' means 'cause to grow'—but in (2) to 'dictate' is not the same as 'causing to write'.

(1) The root $\sqrt{\text{gdl}}$:
   a. *gadal* 'grew'
   b. *hegdil* 'enlarged'

(2) The root $\sqrt{\text{ktb}}$:
   a. *katav* 'wrote'
   b. *hextiv* 'dictated'

In other roots the semantic similarity itself is questionable, as in (3).

(3) The root $\sqrt{\text{sgr}}$:
   a. *sagar* 'closed'
   b. *hesgir* 'extradited'

If we think of an instantiation of a root in a given template as a lemma, Hebrew (and related languages) showcase a situation in which complex similarity relationships hold between roots and lemmas, between lemmas sharing a root, and between lemmas not sharing a root. Looking at specific templates, it has been argued that some like *XaYaZ* are no longer productive, while others like *XiYeZ* are. Farhy and Veríssimo (2019) demonstrated that response time is modulated by semantic relatedness between prime and target for the template *XaYaZ* but not for *XiYeZ*. This finding is consistent with the claim that within the complex system of Hebrew, non-productive *XaYaZ* forms are looked up as wholes, whereas *XiYeZ* forms get decomposed.