Beyond Islands of Regularity: Evidence against morpho-orthographic decomposition

Roberto Petrosino[1], Diogo Almeida[1], Jon Sprouse[1]

[1] *Division of Science, Psychology Program, New York University Abu Dhabi, United Arab Emirates*

Introduction. Cross-linguistic masked priming evidence shows that (a) bimorphemic (*boneless*) and (b) pseudo-affixed, monomorphemic words (corner) often trigger similar masked priming effects on their stems (*BONE* and *CORN*, respectively). In contrast, (c) monomorphemic words (e.g., *cashew*) trigger negligible effects on the embedded stem (*CASH*). The dissociation between (a, b) and (c) has been taken as evidence that lexical access goes through an obligatory decomposition procedure based on morpho-orthographic "islands of regularity", i.e. the statistically regular letter sequences corresponding to the orthographic realizations of morphemes [1]. This procedure arguably occurs before accessing any lexical (morpho-syntactic, or semantic) information so that, at early stages, (b), but not (c), can be provisionally visually decomposed similarly to (a). Thus, this decomposition process is not an actual morphological or even linguistic process per se (despite often being described as "morphological decomposition"), but a visual process informed by morpho-orthographic statistical regularities. However, recent findings challenge this view. Some studies observe similar priming effects for (a) and (c), or (b) and (c) [2], pointing to pure orthographic processing, while others find differential effects between (a) and (b) [3], suggesting early semantic activation [4].

Methods. (A) We conducted five different replications of the same English masked stem priming experiment both in-lab and online collected at different points in time (N_{total}=804). Five stem priming conditions were tested: (i) identity (*fuss-FUSS*), (ii) transparent morphological (*sharper-SHARP*), (iii) a pseudo-morphological ('*opaque*': *belly-BELL*); and two orthographic control conditions: (iv) non-syllabic (*bark-BAR*) and (v) syllabic (*cashew-CASH*). Items were controlled for orthographic length and word frequency. Semantic (cosine) similarity of prime and target pairs was calculated across five different distributional models (Figure 1). Trials consisted of a 33ms-long prime preceded by a 500ms-long forward mask, and followed by a target, on which participants performed a lexical decision task. After excluding subjects and items with high error rates, and trials with outlying prime durations or extreme RTs, we performed a varying-coefficient meta-analysis [5] on the raw-data estimates across all five replications. **(B)** To eliminate cross-replication sampling error, we then conducted a sixth preregistered confirmatory replication of the same experiment, with about four times larger sample size (N=2,868) and two more control conditions: (vi) orthographic neighbor (*basil-BASIN*) and (vii) associative semantic (*poem-RHYME*). After being pre-processed as in (A), the raw RT data was fitted with a Gamma distribution (identity link) in a GLMM [6], with condition, relatedness and their interaction as fixed factors, and subjects and items as random intercepts.

Results. The results from the meta-analysis (A) and experiment (B) closely matched (Figure 2). Identity and morphological priming effects were equally at ceiling. Pseudo-morphological priming was significantly smaller than both and overlapped with the non-syllabic and syllabic priming (in both A and B), as well as orthographic neighbor priming (in B). Semantic priming was around zero.

Conclusions. Our results are incompatible with the proposal that lexical access includes a *pre-lexical morpho-orthographic decomposition* of the input, since pseudo-morphological priming (iii) patterned with purely orthographic priming (iv-vi). Morphological priming (ii) was instead indistinguishable from identity priming (i), possibly indicating that these priming effects obtain at the lexical, rather than at the prelexical level, even at such brief visual presentations (here, 33ms). At the same time, the null result for associative semantic priming (vii) precludes an early involvement of semantic processes (such as *spreading activation*) at such prime durations, and ultimately suggests an independent, linguistically sophisticated level of morphological analysis.

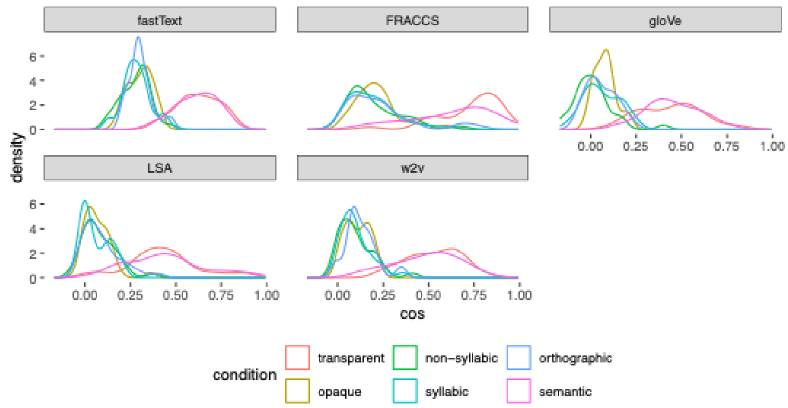
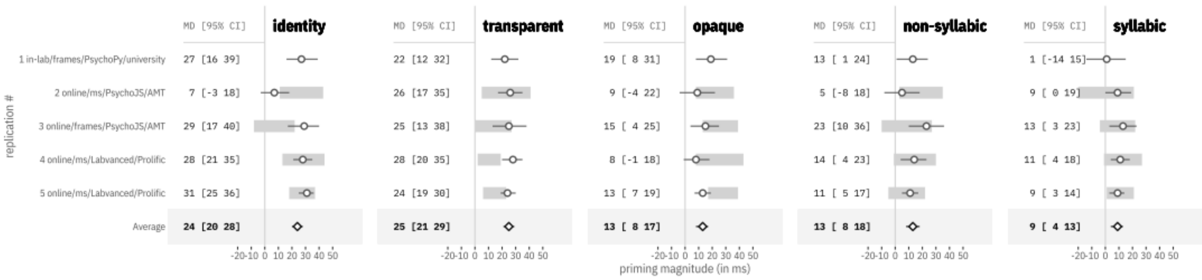


Figure 1: Distribution of semantic (cosine) similarity between the prime-target pairings of the seven conditions calculated across 5 different distributional semantic space models: fastText (Bojanowski et al. 2017), FRACCS [4], gloVe (Pennington et al. 2014), LSA (Landauer & Dumain 1997), word2vec (Mikolov et al. 2013). The identity condition is not plotted to preserve scale, as they yield a uniform similarity of 1.0.

A



B

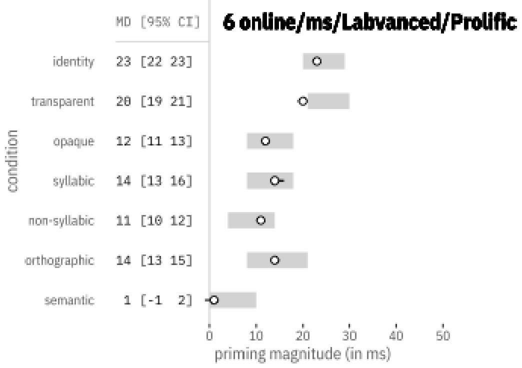


Figure 2: Results. Panel A: Mean and 95each condition in the five replications. The y-axis labels provide further details about each replication: the replication number (1-5), setting (in-lab/online), prime unit (ms./frames), online web app (PsychoPy/Labvanced), and recruitment method (university credit/AMT/Prolific) used. Panel B: Mean and 95% CI of the conditions in the confirmatory replication (exp. 6). The gray bands indicate the 95% PIs calculated from the corresponding meta-analytical averages of the five replications reported in panel A.

References

[1] Rastle, K., & Davis, M. (2008). Morphological decomposition based on the analysis of orthography. *Language and cognitive processes*, *23*(7-8), 942–971. https://doi.org/10.1080/01690960802069730

[2] Morris, J., Porter, J., Grainger, J., & Holcomb, P. (2011). Effects of lexical status and morphological complexity in masked priming: An ERP study. *Language and Cognitive Processes*, *26*(4-6), 558–599. https://doi.org/10.1080/01690965.2010.495482

[3] Feldman, L., O'connor, P., & Martín, F. D. P. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic bulletin & review*, *16*, 684–691. https://doi.org/10.3758/pbr.16.4.684

[4] Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological review*, *122*(3), 485. https://doi.org/10.1037/a0039267

[5] Bonett, D. (2009). Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychological Methods*, *14*(3), 225–238. https://doi.org/10.1037/a0016619

[6] Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171