POSTER NUMBER 1

New UK-Based Dominance Norms for Spoken Ambiguous Words

REBECCA A. GILBERT, HANNAH N. BETTS, RACHEL JOSE, JENNIFER M. RODD

University College London

becky.gilbert@ucl.ac.uk

Background

Ambiguous words (e.g., BARK of the tree/dog) have provided important insights into several key topics within psycholinguistics (Vitello & Rodd, 2015).

Dominance (relative meaning frequency) affects numerous tasks and must be carefully controlled in experiments.

Dominance is usually estimated from word association norms, i.e. the % of participants who respond to the word “BARK” with dog-related words (DOG, WOOF, etc.) vs. tree-related words (TREE, BRANCH, etc.).

Currently available large-scale dominance norms for visually presented words:

- Twilley et al. (1994): 566 words, relatively old, North American participants
- Armstrong et al. (2012): 544 words, North American participants
- Maciejewski et al. (2016): 100 words, unrelated meanings only, UK participants

The latter two norms use a new ‘eDom’ procedure: participants explicitly rate frequency of all meanings (Armstrong et al., 2012).

There is relatively little data available for spoken words and from UK participants, especially for related meanings.

Aims

Word association is now routinely used for ambiguous words in test phase of word-meaning priming experiments (e.g., Rodd et al., 2013; 2016).

The current project collates word association data from these word-meaning priming experiments (unprimed conditions only) to build new UK-based norms for spoken words with range of meaning relatedness.

- Summary dominance data (for use in stimulus selection)
- Raw data (to allow recoding of responses for purposes of new experiments)
- Automated coding script (to code new responses on basis of past responses)

Method

14,144 word association responses for 233 ambiguous words (Figures 1-3). Responses were recoded for consistency in meaning definitions and coding values across experiments.

Related word senses were categorised separately whenever they could be distinguished reliably from the responses.

> e.g., “BAR” has separate meaning codes for the ‘pub’, ‘length of solid material’, ‘prevent/exclude’ and ‘legal profession’ meanings, despite being classified by Wordsmyth dictionary as related senses of the same meaning.

This resulted in fine-grained response coding, with up to 8 meanings per word.

Words were excluded if they had fewer than 20 responses, or where it was hard to disambiguate the meaning of many common associates (e.g., “COLD” – WINTER).

Automated coding script developed to assist with coding of future data. Tested on existing data from items with at least 100 responses (N = 50). For each item, 80 training and 20 test responses were randomly selected, and proportion of test responses that could be coded was calculated. This was repeated 100x, and proportions were averaged within items. The mean proportion of coded test responses over items was 0.80 (min = 0.62 for “CRAFT”; max = 0.95 for “PUPIL”).

Results/Conclusions

Existing priming experiments provide a wealth of data about word dominance.

We will provide raw and summary data, and an automated coding script that codes ~80% of new word association responses. This will improve efficiency and reliability of coding in e.g. new word-meaning priming experiments.

Results broadly consistent with those from eDom estimates from UK and US populations (Figure 3), but:

- Some items show clear UK/US differences (e.g., “CRICKET”, “PORT”)
- Word association method may overestimate the dominance of most frequent meaning compared with eDom (and/or vice versa)
- Combining these two approaches may overcome limitations of each

We plan to obtain:

- Additional responses for items with 50-99 responses, resulting in 100+ responses for 122 ambiguous words
- Information on relatedness of meanings

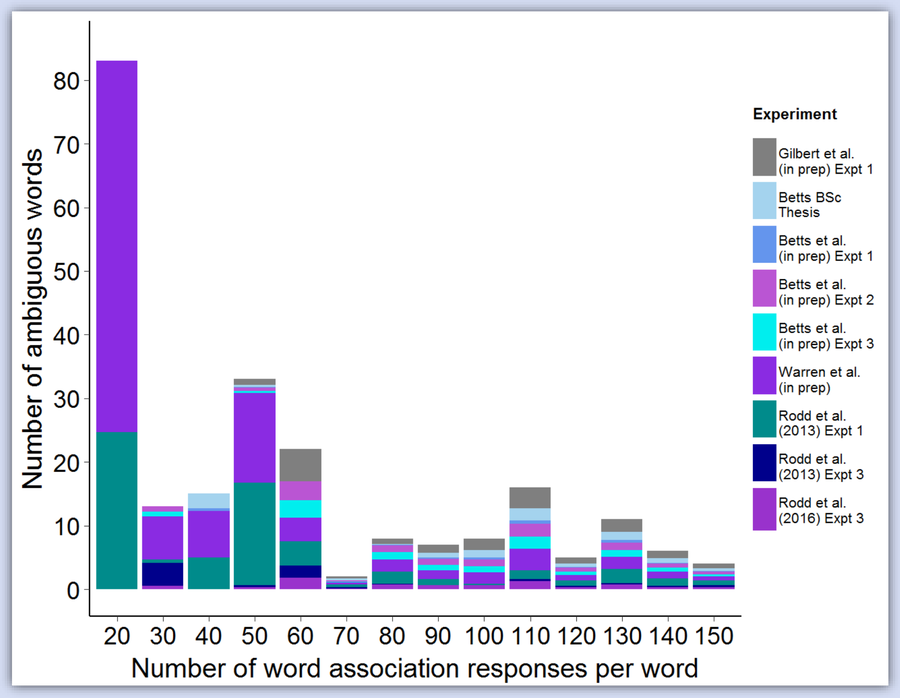

Figure 1. Distribution of words in the database as a function of number of valid word association responses (x-axis) and source of the data (colours). Colours represent the overall proportion of responses from each experiment for items within each bar.

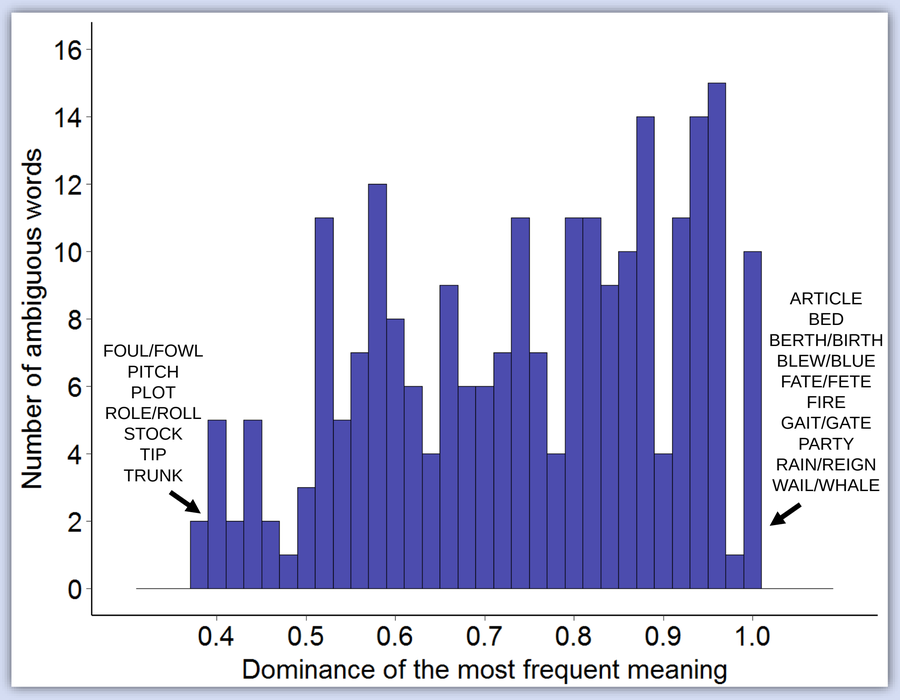

Figure 2. Distribution of dominance scores. Following Armstrong et al. (2012), we calculated the proportion of response for most dominant meaning. Higher values indicate words with single strongly dominant meaning, and lower values indicate relatively balanced ambiguities.

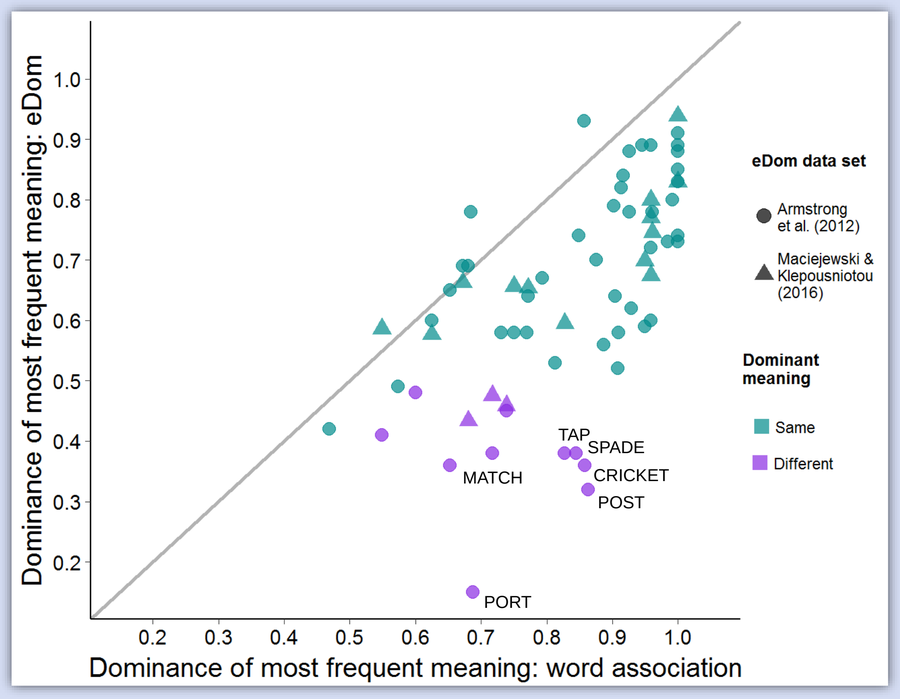

Figure 3. Comparison of our word association dominance results with Armstrong et al. (2012) and Maciejewski & Klepousniotou (2016) eDom estimates. Points represent individual words that overlapped between sets and have a single spelling (e.g. “BARK” included, “BAIL/BALE” excluded).

References

Armstrong, B. C., Tokowicz, N., & Plaut, D. C. (2012). *Behavior Research Methods, 44*(4), 1015-1027.
Maciejewski, G., & Klepousniotou, E. (2016). *Journal of Open Psychology Data*, *4*(1).
Rodd, J. M., et. al. (2013) *Journal of Memory and Language, 68*(2), 180–198
Rodd, J. M., et. al. (2016). *Journal of Memory and Language, 87*, 16-37.
Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). *Memory & Cognition, 22*(1), 111-126.
Vitello, S. & Rodd, J. M. (2015). *Language and Linguistics Compass*, *9*(10), 391-405.