Semantic projection: Recovering human knowledge of multiple, distinct object properties from natural word collocations

Gabriel Grand* (Harvard), Idan Asher Blank* (MIT),
Francisco Pereira (NIMH), & Evelina Fedorenko (Harvard, MIT)
iblank@mit.edu

Background. Word meanings (lexical semantics) represent only a subset of our rich and detailed conceptual knowledge (semantic memory). Any theory of lexical semantics should thus specify the kinds of world knowledge captured in the lexicon. Here, we probe the common knowledge captured by a prominent model class of word meanings: Distributional Semantic Models (DSMs), vector-space representations learned from lexical collocations in corpora [1,2].

A DSM captures semantic similarities between words via the proximity between their respective vectors [3,4]. However, such proximity only provides a single, "rigid" measure of overall pairwise similarity. Humans, in contrast, evaluate conceptual similarity flexibly, in a context-dependent manner: for instance, dolphins and tigers are similar in terms of size, but differ significantly in terms of danger or habitat. Can such distinct, multiple relationships be inferred from word collocations? If so, how is such complex knowledge expressed in a DSM?

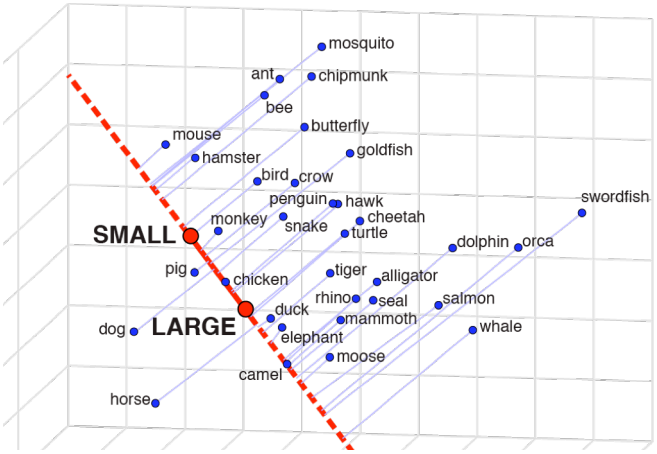
Approach. We introduce a powerful, domain-general method for extracting context-dependent knowledge from DSMs: "semantic projection" of words onto subspaces representing "contexts". To operationalize context-dependent knowledge, we use ratings of concrete entities (e.g., animals) along different semantic properties (e.g., size, danger, or wetness). Fig. 1 depicts an example of rating animal size in a simplified, three-dimensional DSM: first, we construct a "size scale" as the vector difference between antonyms denoting opposite property values—e.g., $\overrightarrow{large}$ and $\overrightarrow{small}$. Then, we project animal word-vectors onto this scale. Similarly, we can project on the line from $\overrightarrow{dangerous}$ to $\overrightarrow{safe}$ (for rating danger), or from $\overrightarrow{wet}$ to $\overrightarrow{dry}$ (for wetness).

Methods. We compared semantic projection against human ratings of diverse objects for different properties. We used (i) 9 categories, each including the 34-50 most frequent nouns from the set in [5]; and (ii) 17 semantic properties. All properties have been produced in feature-elicitation studies and all category items and property antonyms have been used as cues therein [6,7]. Out of 9×17 = 153 possible category/property pairs, we selected 52 based on a norming study and intuitive appropriateness. For each pair, 25 MTurk subjects rated each noun on a separate continuous scale (e.g., for "size": 0="small, little, tiny", 100="large, big, huge"). Semantic projection was applied to the same category-property pairs in a GloVe DSM [4]. We compared the mean human ratings (after within-subject *z*-scoring) to semantic projection via Spearman's *r* and permutation tests [FDR-corrected across pairs; 8].

Results. Overall, semantic projection successfully recovered human knowledge (*Figs. 2, 3*): median Spearman's *r*=0.47 (adjusted for inter-subject reliability: 0.61) (significant for 35 pairs). Items with extreme property values (e.g., "mammoth" or "mosquito" for animals/size) only biased correlations to the same extent that they increased inter-subject reliability. Projection on a line connecting antonyms significantly outperformed projecting on either antonym in isolation.

Conclusions. Whereas flexible object knowledge has traditionally been modeled with symbolic representations (e.g., feature lists, schemata, intuitive theories) [9-11], here we show that it can be constructed bottom-up from word collocations and is easily extractable from DSMs via semantic projection (cf. previous attempts using, e.g., dependency parsed corpora) [12-15]. Our method is robust, generalizing across different categories (animate/inanimate, natural/man-made, common/proper nouns) and different "kinds" of properties (from relatively binary, like animal wetness, to more continuous, like animal size). To the extent that DSMs model lexical semantics, our findings suggest that complex property knowledge is part of a concrete noun's meaning. Moreover, our findings are consistent with the intriguing hypothesis that, like DSMs, humans can use language as a gateway to acquiring conceptual knowledge [16-20].

Further details. For a full write-up, see: https://arxiv.org/abs/1802.01241.



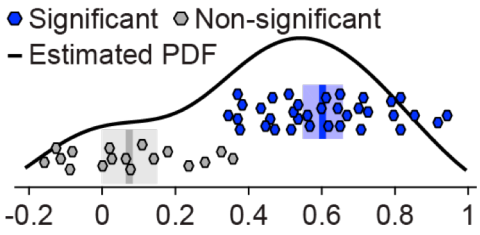


Fig 2: Distribution of Spearman's *r* values (*x*-axis) for 52 category/property pairs. Mean and 95% CI are plotted for significant (blue) and non-significant (grey) values.

Fig 1: Illustration of semantic projection. Word-vectors in the category "animals" (blue circles) are orthogonally projected (blue lines) onto a linear scale for "size" (red line), defined as the vector difference between $\overrightarrow{large}$ and $\overrightarrow{small}$ (red circles). Dimensions are arbitrary and chosen to enhance visualization.

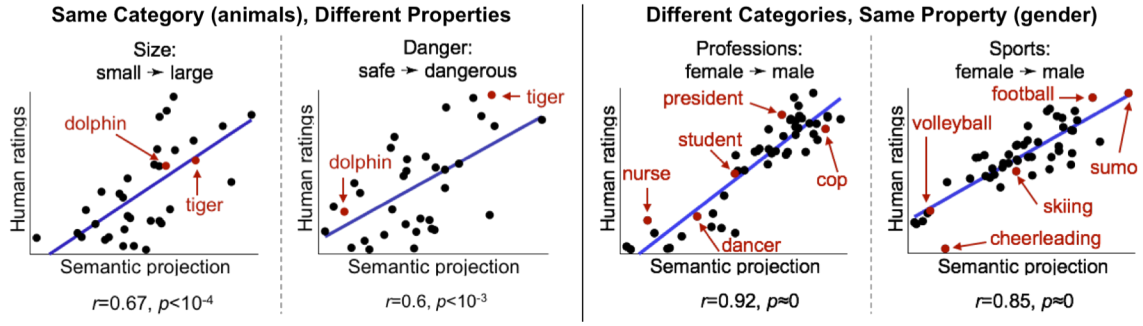


Fig 3: Example scatterplots of semantic projection (*x*-axis) predicting human judgments (*y*-axis). Example items are labeled in red for illustration. Blue lines are linear fits to the data.

References

1. Lenci A. (2008). *Italian J. Ling.*, 20(1), 1-31.
2. Clark S. (2015). In Lappin S. & Fox C. (Eds.), *Handbook of Contemporary Semantics* (pp. 493-522). Blackwell.
3. Mikolov T., Sutskever I., Chen K., Corrado G.S., & Dean J. (2013). *Proceedings of NeurIPS (prev. NIPS)*.
4. Pennington J., Socher R., & Manning C. (2014). *Proceedings of EMNLP*.
5. Mahowald K., Isola P.E., Fedorenko E., Oliva A., & Gibson E. (2014). *Proceedings of AMLaP*.
6. McRae K., Cree G.S., Seidenberg M.S., & McNorgan C. (2005). *Beh. Res. Methods*, 37(4), 547-559.
7. Nelson D.L., McEvoy C.L., & Schreiber T.A. (2004). *Beh. Res. Methods*, 36(3), 402-407.
8. Benjamini Y. & Yekutieli D. (2001). *Ann. Stat.*, 29(4), 1165-1188.
9. Rosch E. & Mervis C.B. (1975). *Cog. Psych.*, 7(4), 573-605.
10. Rumelhart D. & Ortony A. (1977). In C. A.R., J. S.R., & E. M.W. (Eds.), *Schooling and the Acquisition of Knowledge* (pp. 99-135). Lawrence Erlbaum.
11. Gopnik A. (2003). In L. A. & N. H. (Eds.), *Chomsky and his Critics* (pp. 238-254). Blackwell.
12. Poesio M. & Almuhareb A. (2005). *Proceedings of ACL-SIGLEX Workshop on Deep Lexical Acquisition*.
13. Baroni M., Murphy B., Barbu E., & Poesio M. (2010). *Cog. Sci.*, 34(2), 222-254.
14. Rubinstein D., Levi E., Schwartz R., & Rappoport A. (2015). *Proceedings of ACL*.
15. Kelly C., Devereux B., & Korhonen A. (2014). *Cog. Sci.*, 38(4), 638-682.
16. Rumelhart D.E. (1979). In Ortony A. (Ed.), *Metaphor and Thought* (pp. 71-82). Cambridge University Press.
17. Landauer T.K. & Dumais S.T. (1997). *Psych. Rev.*, 104(2), 211-240.
18. Elman J.L. (2009). *Cog. Sci.*, 33(4), 547-582.
19. Lupyan G. & Bergen B. (2016). *Topics in Cog. Sci.*, 8(2), 408-424.
20. Marmor G.S. (1978). *J. Exp. Child Psych.*, 25(2), 267-278.