A common framework for quantifying the learnability of nouns and verbs

Yuchen Zhou
zhouyuchen@cmu.edu
Department of Psychology
Carnegie Mellon University

Daniel Yurovsky
yurovsky@cmu.edu
Department of Psychology
Carnegie Mellon University

Abstract

Across the world's languages, children reliably learn nouns more easily than verbs. Attempts to understand the difficulty of verb learning have focused on determining whether the challenge stems from differences in the linguistic usage of nouns and verbs, or instead conceptual differences in the categories that they label. We introduce a novel metric to quantify the contributions of both sources of difficulty using unsupervised learning models trained on corpora of language and images. We find that there is less alignment between the linguistic usage of verbs and their categories than for nouns and their categories. However, this difference is driven almost entirely by differences in the structure of their visual categories: Relative to nouns, events described by the same verb are more variable and events described by two different verbs are more similar. We conclude that differences between noun and verb learning need not be due to fundamental differences in learning processes, but may instead be driven by the difficulty of one-shot generalization from verbs' visual categories.

Keywords: language learning; distributional semantics; computational models; verbs

Children's early vocabularies are dominated by nouns; children know many more words like "dog" and "cat" than "jump" and "run" (Fenson et al., 1994). Although there is some variability in the size of this effect across languages, children all over the world learn nouns more quickly than verbs and other predicates (Frank, Braginsky, Yurovsky, & Marchman, 2021). Why are verbs so much harder to learn than nouns?

Attempts to understand the unique challenges of verb learning have generally focused on determining whether the primary source of difficulty is conceptual or linguistic. One class of accounts focuses on the difficulty of verb concepts themselves. For instance, verbs may label visual concepts that are more variable than the concepts referred to by nouns or more confusable with the concepts labeled by other words (Gentner, 1982; Golinkoff & Hirsh-Pasek, 2008). For instance, although different `dogs` may differ from each other on a variety of dimensions like color and size, `jumps` differ from each other not only in dimensions like distance and height but also in the identity of the agent doing the jumping. Further, the same agent can perform many different actions, each of which is labeled by a different verb.

Another class of accounts focuses on the structures in language that make learning verbs hard. For example, verbs are fundamentally ambiguous in meaning unless the words for their agent and patient are known (as in "chase" and "flee", Gleitman, 1990). Additionally, there may be more degrees of freedom in how languages can carve up the space of verb meanings than noun meanings. For instance, languages vary in their tendency to lexicalize the manner of an action (as in "stroll" vs. "sprint"), or its path (as in "push" or "pull"; Gentner, 1982; Talmy, 1975). This kind of variability may make verbs harder to learn than nouns because their use in language is less transparently related to the events with which they occur. Accordingly, verbs may require more complex relational reasoning about both unfolding events and the syntactic structure of the utterances in which they occur.

To date, efforts to compare these competing accounts have had to rely on indirect methods such as cross-linguistic comparisons (Talmy, 1975), studies of atypical learners (Snedeker, Geren, & Shafto, 2007; Goldin-Meadow & Feldman, 1977), and artificial language learning experiments that try to approximate the real problems faced by children, sometimes with adults (Gillette, Gleitman, Gleitman, & Lederer, 1999). We propose a novel method for quantifying the sources of difficulty in verb learning directly using unsupervised machine learning models. By formalizing learnability as the degree of alignment between a word's linguistic and visual representations, we ask whether verbs are harder to learn than nouns because of differences in their linguistic usage or because of differences in the structure of the categories that they label.

Alignment as a measure of learnability

According to the natural partitions hypothesis, the mapping between nouns in language and their meanings in the world is transparent: nouns refer to "highly cohesive collections of percepts that are universally conceptualized as objects" (Gentner, 1982). In contrast, because verbs are relational, the mapping between verbs and their visual referents is complex– it requires knowledge of both the objects involved in the event and of the dimensions of meaning one's language tends to lexicalize. One way to formalize this hypothesis is a prediction about how the usage of words relates to their visual categories.

The intuition here builds on the core insight of distributional models of language meaning: Words that are used in similar ways are likely to have similar meanings (Firth, 1957). That is, "dog" and "cat" are probably similar because they are used in the same contexts, whereas "dog" and "table" are not.

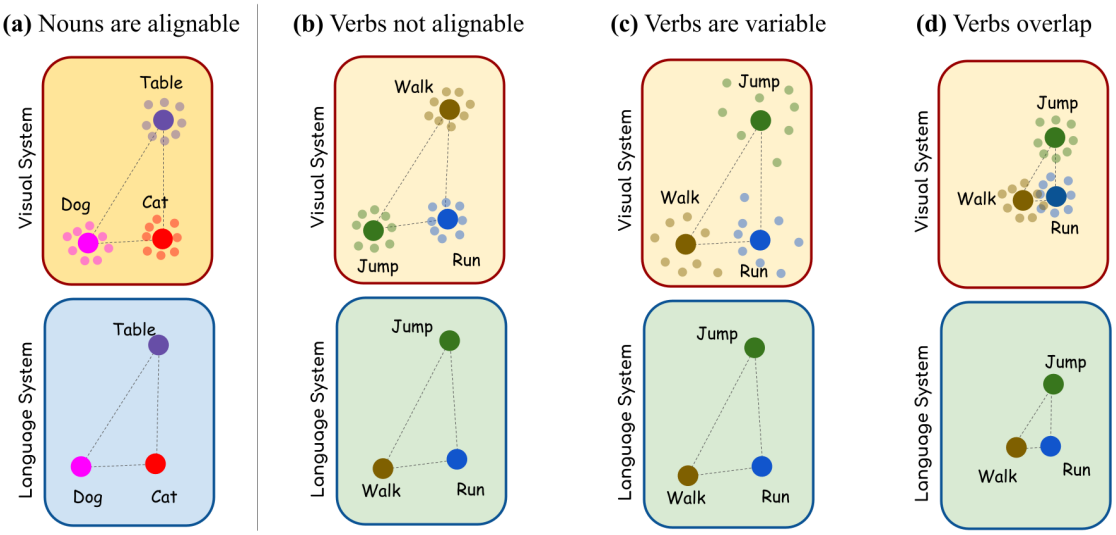

Figure 1: A schematic representation of the possible sources of difficulty in learning verbs. Across panels, large circles represent the average meaning of a word while small circles represent individual exemplars. First, while nouns' linguistic and visual representations are well aligned (a), verbs' representations may not align across modalities (b). For example, events for `run` could be more visually similar to `jump` than `walk`, while "run" is used in language more like "walk" than "jump." Alternatively, individual events for the same verb may be more variable than events for the same noun (c), or events for different verbs may be more similar to each other than events for different nouns (d).

Models that learn these statistical relationships from large language corpora yield similarity judgments for words that are strikingly similar to those produced by people (Landauer & Dumais, 1997; Pennington, Socher, & Manning, 2014). If the natural partitions hypothesis is correct, words' similarities in language should map transparently onto the similarity of their visual categories: `dogs` and `cats` should be *more visually similar* than `dogs` and `tables` (see Fig. 1a).

Recently, Roads and Love (2020) confirmed this prediction by showing that concrete nouns' linguistic and visual similarities are highly alignable. They used off-the-shelf, unsupervised models of language semantics and visual feature extraction to learn representations for ∼400 words and their corresponding visual categories. Compared to alternative hypothetical alignments, e.g. the word "dog" to the visual category `table`, the true mapping system yields the highest correlation between similarities in language and similarities in visual categories.

We adapt Roads and Love (2020)'s method to ask why verbs are harder to learn than nouns. Alignability captures the extent to which a word's meaning as defined by its linguistic usage is similar to its meaning as defined by the visual category to which it refers. One possibility, as predicted by Gentner (1982), is that verbs are less alignable than nouns: similarities in the linguistic usage of verbs does not align with the similarities of the visual categories they label (Fig. 1b). For instance, the linguistic usage of verbs must contain information about the argument structure of these verbs and the contexts in which they may be used. This information might have no analogue in the structure of the visual categories to which they refer, reducing alignability. Alternatively, if the same verb can refer to events that look quite different from each other, the categories labeled by verbs could be more diffuse than those labeled by nouns (Fig. 1c). Finally, events referred to by the same verb could be not just different from each other, but also similar to events referred to by different verbs (Fig. 1d).

We applied this alignability metric to nouns and verbs from several large image corpora. We show that verbs are indeed harder to learn than nouns from a single visual exemplar, but become almost as easy to learn as nouns when multiple exemplars can be aggregated together. Further, multiple exemplars help not just because verbs' concepts are more variable, but also because they are less distinctive from each-other. We thus confirm the general qualitative account offered by prior theories of language learning, and also quantify the relative contribution of proposed mechanisms to the overall difficulty of learning verbs.

Study 1: Aligning nouns and their meanings

Since Roads and Love (2020)'s analysis, machine vision researchers have significantly improved the art in unsupervised visual feature extraction. Because we wanted to use a more recent machine vision model than Roads and Love (2020) to compare nouns to verbs, we began by replicating their analysis of nouns with this new model.

Method

To estimate the alignment between nouns' linguistic and visual representations, we asked whether words' linguistic sim-

ilarities in an unsupervised language model were correlated with their visual similarities in an unsupervised vision model. To assess the strength of this correlation, we compared the true system of word-object mappings to simulated alternative systems produced by permuting the true mapping system.

Data. Following Roads and Love (2020), for our visual categories, we used 434 categories from the Open Images V4 dataset (Kuznetsova et al., 2020). Each image was cropped to a bounding box around the target object, and downsampled to 224x224 pixels. To derive a visual representation for each object category, we used an unsupervised vision model called Swapping Assignments between Views (SwAV; Caron et al., 2020). The model we used was based on ResNet-50 architecture and was pre-trained on the ImageNet ILSVRC-2012 dataset (Russakovsky et al., 2015). Because the model learned general visual features of objects, it was not optimized specifically to discriminate images in our dataset. The model was applied to one image randomly sampled from each object category to produce a 128-dimensional vector.

For linguistic representations, we used Global Vectors for Word Representation (GloVe), an unsupervised word embedding model (Pennington et al., 2014). We used pre-trained 300-dimensional semantic vectors derived from the the Common Crawl corpus composed of 840 billion tokens and 2.2 million words. For our analysis, we considered only the words that corresponded to the relevant 434 images.

Procedure. Words with aligned linguistic and visual representations should be similar to the same items in both the linguistic and visual domains (e.g. the words that "dog" is similar to should map onto the concepts that `dog` is visually similar to). In each modality, we compared the pairwise similarities of all words' vectors using cosine similarity. We then computed the correlation between all corresponding linguistic and visual similarities using Spearman's rank correlation coefficient (ρ) because of the non-normal distribution of similarities.

To evaluate the alignment of the true word-object mappings, we compared this alignment to the alignments of other possible systems of word-object mappings by permuting the true mapping. If the alignment of the true mapping (in which "dog"–`dog`, "table"–`table`) is higher than the alignment of other possible mappings (e.g."dog"–`table`, "table"–`dog`), we can conclude that linguistic and visual representations are well aligned. Because most random mapping permutations are very wrong, we sampled 100 possible word-object systems for each level of incorrectness (i.e. proportion of incorrect mappings). This allowed us to understand the relationship between mapping accuracy and alignment systematically.

Results and Discussion

Overall, we found that concrete nouns' linguistic and visual representations were well aligned. The correlation between words' linguistic and visual embeddings was reliably positive ($\rho = 0.214$, $p < .001$), and comparable to the value obtained by Roads and Love (2020) with a different visual embedding model. Thus, nouns that are more similar to each other in linguistic usage map onto visual categories that are more similar to each other in appearance. In addition, the strength of the true mappings' alignment was higher than the strength of 99.93% of mappings produced by permuting the true mapping (Fig. 2). Further, the relationship was monotonic–mappings that were more accurate tended to have higher linguistic-visual alignments.

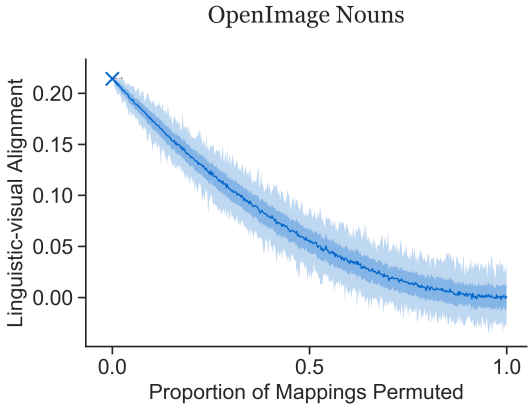

Figure 2: The relationship between degree of permutation in the true mappings of concrete nouns and their resultant system's linguistic-visual alignment. The true mapping in language is represented by the **x**, the mean alignment at each level of permutation is represented by the line. The dark and light regions represent ± 1 SD and minimums/maximums respectively.

Together, these results replicate Roads and Love (2020)'s analysis, and are consistent with the natural partition hypothesis (Gentner, 1982). Nouns' meanings are relatively transparent because their linguistic usage and visual categories are well aligned; their representations in the linguistic and visual modality are similar. In Study 2, we ask whether this is also true for verbs.

Study 2: Comparing nouns and verbs

To understand whether verbs are harder to learn than nouns because their language-world mappings are less transparent, we applied the same method we used in Study 1 to a corpus of images containing both nouns and verbs. We asked first whether verbs are less alignable than nouns, and second whether differences in alignability are related to differences in the category structures of nouns and verbs.

Method

We used the same models and procedure as Study 1, but analyzed a new corpus that contains visual representations for both nouns and verbs. In addition, while visual meanings were constructed from a single image in Study 1, we also considered the effects of averaging together multiple exemplars

Figure 3: Examples of nouns and verbs in the Visual Relationship Detection dataset. Compared to nouns (a), images of the same verb category are more visually diverse, in part because of variability in agents and patients (b).

to arrive at a prototype for each category (Rosch, 1973).

Data and Procedure. To compare nouns to verbs, we used the Visual Relationship Detection dataset (VRD), a popular machine learning dataset containing 5,000 real-world images with manually-annotated bounding boxes for 100 object categories, and 70 manually annotated predicate relationships between them (Lu, Krishna, Bernstein, & Fei-Fei, 2016).

From these categories, we excluded 2 object categories for which we did not have GloVe vectors to define their linguistic representation. From the predicates, we selected the 30 that correspond to verbs (rather than e.g. prepositions) for analysis. To construct bounding boxes for verbs, we took the union of the bounding boxes of any agents or objects that were annotated as taking part in them. Figure 3 demonstrates some examples of real images of nouns and verbs from the VRD dataset.

In addition to the procedure used in Study 1, we estimated prototypical representations of visual categories by averaging together feature vectors for multiple randomly selected exemplars for each category. This allowed us to determine whether differences in alignability between verbs and nouns are driven by relationships between linguistic and visual representations, or by differences in visual categories themselves.

Results and Discussion

As in Study 1, nouns' linguistic and visual representations were well aligned ($\rho = 0.176$, $p < .001$). Visual categories learned from a single exemplar of each noun were similar enough to their linguistic representations that the true noun-meaning mappings were more aligned than 99.27% of simulated mappings (Fig. 4a). In contrast, verbs' representations were less well aligned ($\rho = 0.139$, $p < .001$), and the true verb-meaning mapping system's alignment was only greater than 80.97% of simulated mapping systems. These results are consistent with the empirical observation that verbs are harder to learn than nouns, but do not distinguish between the possible sources of difficulty. Are verbs hard to learn because their mappings are fundamentally less alignable, or because of the structures of their visual categories?

To address this question, we constructed prototype representations for each visual concept by averaging together the visual representations of multiple images corresponding to that concept. Figures 4c and 4d show the relationship between degree of mapping permutation and alignment when 25 images were averaged together to form visual prototypes for nouns and verbs respectively. This averaging improved the alignment of both nouns ($\rho = 0.307$, $p < .001$), and verbs ($\rho = 0.345$, $p < .001$). However, because the true mapping for nouns already had higher alignment than almost all of the simulated mappings, averaging multiple exemplars together produced little improvement in its relative strength. In contrast, averaging multiple exemplars together caused the alignment of the true verb-meaning mapping to be greater than 96.62% of permuted mappings. This suggests that the structure of verbs' visual categories is the primary driver of their difficulty relative to nouns.

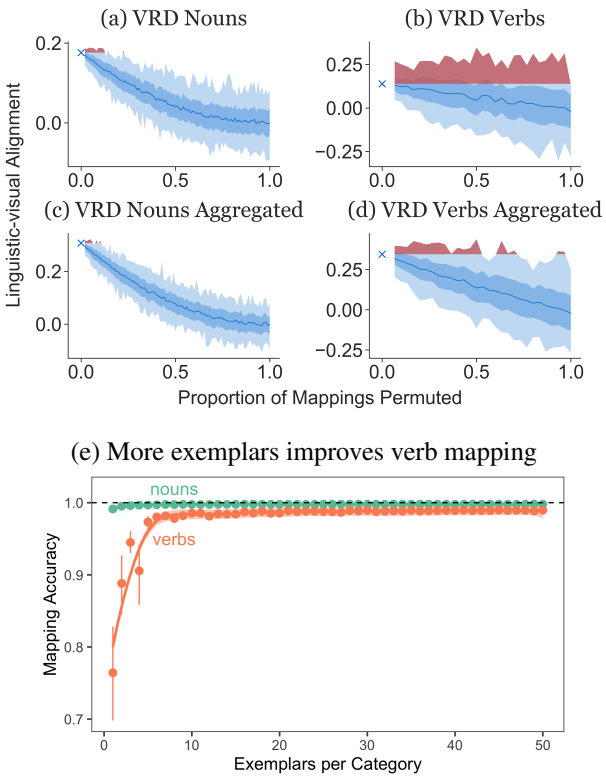

Figure 4: Linguistic-visual alignment was much higher for nouns (a) than for verbs (b). However, when multiple exemplars were averaged together to form a visual prototype, the alignment of the true mapping relative to alternative mappings improved significantly more for verbs (d) than for nouns (c). The red regions are permuted mappings that have higher alignment correlation than the true mapping. As the number of exemplars averaged together to form a prototype increased, the mapping accuracy of the most-aligned mapping for verbs approached nouns (e). Points indicate mean mapping accuracy, error bars indicate 95% confidence intervals computed by non-parametric bootstrapping over 20 simulations.

To further understand the relationship between visual category learning and word-meaning mapping, we asked what proportion of permuted mappings were less well aligned than the true mapping–we call this metric mapping accuracy. As the number of exemplars used to construct visual prototypes increased, there was little benefit for nouns because accuracy was already close to ceiling (Fig. 4e). In contrast, for verbs, aggregating over even a small number of exemplars of a visual category dramatically increased mapping accuracy. Once ∼10 exemplars were averaged together to form a visual category prototype, verbs were mapped nearly as accurately as nouns. We confirmed this statistically with a linear regression predicting mapping accuracy from word type (noun/verb), number of exemplars, and their interaction. Verbs were mapped reliably less accurately than nouns ($\beta = .07$, $p < .001$), and the (log) number of exemplars did not matter for nouns ($\beta = .0001$, $p = .755$), but did for verbs ($\beta = .003$, $p =< .001$).

Taken together, these results suggest that verbs are harder to learn than nouns primarily because of differences in their visual categories. When multiple events can be aggregated to learn a prototype for verbs' visual categories, the relationship between language and the world is nearly as transparent for verbs as for nouns. In Study 3, we sought to replicate this surprising result in a larger corpus and to quantify the structural differences between nouns and verbs' visual categories.

Study 3: Noun and verb category structure

In Study 2, we found that nearly all of the differences in linguistic-visual alignment between nouns and verbs were explained by differences in their visual categories. When 10 visual exemplars of a verb were averaged together, verbs became just as alignable as nouns' meanings. In Study 3, we analyzed a much larger dataset of noun and verb images to replicate this result, and to unpack the contributions of two sources of potential differences between nouns and verbs: the structure of individual visual categories, and the relationship between visual categories.

Data and Procedure. In study 3, we used the Visual Genome (VG), a large densely-annotated dataset containing 108,077 images with 5,154 object categories and 36,550 relationship categories. To exclude rare words, we only considered 355 nouns that appeared least 10 times in the dataset and 356 verbs for comparison.

We used the same method as Study 2 to quantify the relationship between alignment and mapping accuracy for nouns and verbs in a much larger corpus. In addition, we also measured two features of nouns' and verbs' visual categories: (1) the variance of visual representations for exemplars of the same category, and (2) the discriminability of exemplars from different categories.

To compute the variance and discriminability, we computed the centroid using 250 randomly sampled exemplars for each category. Variance was defined as the average Euclidian distance between each exemplar's embedding and the category centroid. To compute discriminability, we computed the average Euclidian distance between all pairs of category centroids within the visual domain.

Results and Discussion

As in Study 2, nouns ($\rho = .083$, $p < .001$) were more alignable than verbs ($\rho = .014$, $p < .001$). Again true noun mapping had higher alignment than most of the permuted systems (97.88%). In contrast, the true verb mapping system had higher alignment than only 76.00% of permuted systems. Averaging together multiple exemplars significantly improved both noun ($\rho = .226$, $p < .001$) and verb alignment ($\rho = .156$, $p < .001$). Again, because the true mapping for nouns already had higher alignment than almost all of the permuted systems, this averaging did little to improve it's relative alignment. In contrast, averaging 25 exemplars together improved

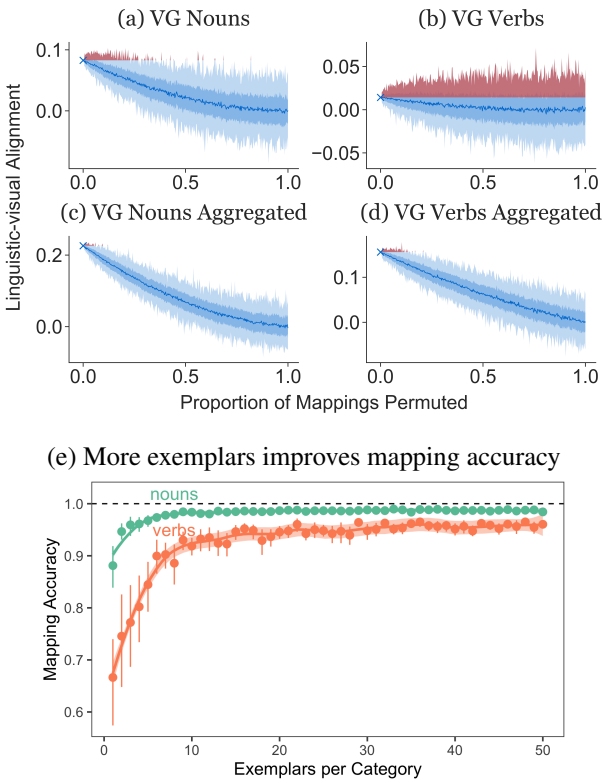

Figure 5: Linguistic-visual alignment was much higher for nouns (a) than for verbs (b). Aggregating together multiple exemplars significantly improved alignment, especially for verbs (c,d). Red regions are permuted mappings that have higher alignment correlation than the true mapping. The relationship between number of exemplars and mapping accuracy is show in (e). Points indicate mean mapping accuracy, error bars indicate 95% confidence intervals computed by non-parametric bootstrapping over 20 simulations.

the true verb mapping's relative alignment to 99.17%, almost the same level as nouns. In this larger corpus, we applied the same linear model, predicting accuracy from word type, (log) number of exemplars, and their interaction. Verbs had lower mapping accuracy ($\beta = -.191$, $p < .001$), and (log) number of exemplars improved accuracy for both nouns and verbs ($\beta = .015$, $p < .001$), but the effect was larger for verbs ($\beta = .046$, $p < .001$).

To understand why averaging exemplars together into a prototype improved mapping accuracy so much for verbs relative to nouns, we analyzed two features of their respective visual categories: (1) variance of individual categories, and (2) discriminability of different categories. Both features were different between verbs and nouns. We found that the variance of verb concepts ($\mu = 0.853$, 95% CI=$[0.844, 0.861]$) was significantly greater than noun concepts ($\mu = 0.764$, 95% CI=$[0.756, 0.773]$). Thus, different events labeled by the same noun are more like each-other than events labeled by the same verb. The discriminability of verbs ($\mu = 0.552$, 95% CI=$[0.551, 0.553]$) was also reliably lower than the discrimanability of nouns ($\mu = 0.627$, 95% CI=$[0.626, 0.628]$). Thus events described by different verbs are more confusable with each other than events described by different nouns [1].

Taken together, the results of Study 3 replicate the main findings of Study 2–verbs are less alignable than nouns when their concepts must be learned from very few exemplars. When a prototypical visual concept can be learned by aggregating over multiple exemplars of the same verb category, the relationship between linguistic representations and visual representations is no less transparent for verbs than for nouns. The primary difference between nouns and verbs, on this analysis, is that nouns' visual concepts are amenable to one-shot generalization of the kind observed in young children (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). In contrast, developing a stable representation of the categories to which verbs refer requires more examples.

General Discussion

A long-standing question in the study of language development is why verbs are harder to learn than nouns (Gentner, 1982). Much of the extant work addressing this question attempts to adjudicate between conceptual and linguistic explanations for the challenge of verb learning. We present a novel method for quantifying the sources of difficulty posed in these accounts by formalizing the learnability of words as the degree of alignment between linguistic and visual representations. Using this method, we show that verbs are hard to learn primarily because of the structure of the visual categories to which they refer. If learners can aggregate over multiple exemplars for the same verb, the mappings between word and world are almost as transparent as for nouns.

One important limitation of these analyses is their reliance on static images to represent visual concepts. Verbs in particular tend to refer to actions which unfold over time and may be better represented as videos than pictures. In principle, video exemplars of the same verb could be less variable than image exemplars because they contain more information. On the other hand they could also be more variable. The images representing each verb are not random samples from their temporal sequence–they were posted by people who thought they were good pictures. They might thus contain the most diagnostic information about the meaning of each verb. Also, while we systematically varied the number of visual exemplars our model learns from, we did not vary the linguistic representations analogously to determine how changing linguistic representations contribute to changes in alignment. We plan to explore both of these issues in future work.

The analyses in this paper also point to a potential direction

[1] The analyses yield the same qualitative results in the dataset used in Study 2 (VRD): the variance of verb concepts ($\mu = 0.785$, 95% CI=$[0.745, 0.824]$) was greater than noun concepts ($\mu = 0.738$, 95% CI=$[0.724, 0.753]$), and the discriminability of verbs ($\mu = 0.558$, 95% CI=$[0.540, 0.575]$) was also lower than the discriminability of nouns ($\mu = 0.749$, 95% CI=$[0.744, 0.754]$).

of study for models of language learning. While extant models generally consider learning the meanings of words by establishing cross-modal mappings, our results confirm an additional mechanism that children might bring to bear: alignment between representations learned in different modalities. Cross-modal alignment might be particularly helpful for learning mappings between some words and their meanings even if they never physically occur together. A canonical problem in language acquisition is that most language is not a running commentary on the present moment–a child's caregiver is unlikely to say "I am opening the door!" as they come home from work, but instead something like "It's freezing outside!" (Gleitman, 1990). Cross-modal alignment may give children a way to recover from this problem not just for concrete nouns, but even for challenging words like verbs.

Data and code for analyses are available at `https://github.com/FlamingoZh/verb-alignment`

References

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A. (2020). Unsupervised learning of visual features by contrasting cluster assignments. *arXiv preprint arXiv:2006.09882*.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i–185.

Firth, J. R. (1957). A synopsis of linguistic theory. *Studies in linguistic analysis*.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The wordbank project*. MIT Press.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development: Vol. 2. language, thought, and culture* (pp. 301–334).

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language acquisition*, *1*(1), 3–55.

Goldin-Meadow, S., & Feldman, H. (1977). The development of language-like communication without a language model. *Science*, *197*(4301), 401–403.

Golinkoff, R. M., & Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends in cognitive sciences*, *12*(10), 397–403.

Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., ... others (2020). The open images dataset v4. *International Journal of Computer Vision*, 1–26.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211.

Lu, C., Krishna, R., Bernstein, M., & Fei-Fei, L. (2016). Visual relationship detection with language priors. In *European conference on computer vision* (pp. 852–869).

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543).

Roads, B. D., & Love, B. C. (2020). Learning as the unsupervised alignment of conceptual systems. *Nature Machine Intelligence*, *2*(1), 76–82.

Rosch, E. H. (1973). Natural categories. *Cognitive psychology*, *4*(3), 328–350.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, *115*(3), 211-252. doi: 10.1007/s11263-015-0816-y

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological science*, *13*(1), 13–19.

Snedeker, J., Geren, J., & Shafto, C. L. (2007). Starting over: International adoption as a natural experiment in language development. *Psychological science*, *18*(1), 79–87.

Talmy, L. (1975). Semantics and syntax of motion. In *Syntax and semantics volume 4* (pp. 181–238). Brill.