How much context is helpful for noun and verb acquisition?

Raquel G. Alhama[1] **Caroline Rowland**[2,3] **Evan Kidd**[2,3,4,5]
[1]Department of Cognitive Science & Artificial Intelligence, Tilburg University
[2]Language Development Department, Max Planck Institute for Psycholinguistics
[3]Donders Institute for Brain, Cognition & Behaviour, Radboud University
[4]The Australian National University
[5]ARC Centre of Excellence for the Dynamics of Language
rgalhama@tilburguniversity.edu,
{caroline.rowland,evan.kidd}@mpi.nl

Introduction

While it is widely accepted that children use distributional information to acquire multiple components of language, the underpinnings of these achievements are unclear. The goal of the current work is to investigate the role of linguistic context in the acquisition of nouns and verbs. In particular, we use a Distributional Semantic Model (DSM) to predict the age of acquisition of nouns and verbs, and we analyse the hyperparameters of the model to find out how much context is helpful for the acquisition of these words.

DSMs have been extensively evaluated against human adult ratings on semantic associations, but less so against children's emerging semantic representations. For reasons of space, we limit our review of prior work to the most recent study that is closest to our goals. In that study, Alhama et al. (2020) propose two methods to evaluate DSMs for children's acquisition of nouns. Their results suggest that the *Skipgram* version of *word2vec* (Mikolov et al., 2013) is most successful in predicting the Age of Acquisition (AofA) of nouns. In our work, we look more in-depth into the hyperparameters of Skipgram that best predict AofA, to find out more about the influence of context in acquisition. In addition, we extend the study to verbs.

Data

We trained the model on transcriptions of child-directed speech from CHILDES (MacWhinney, 2000), for all the English variants, for ages ranging from 0 to 60 months. To evaluate the models on AofA, we used data collected with the MacArthur-Bates Communicative Development Inventory forms (CDI). These forms contain checklists of common words that parents complete, according to whether their child *understands* or *produces* each of those words. The forms are collected at different ages, and thus can be used to estimate the AofA of words. We used the English CDIs from the Wordbank database (Frank et al., 2017) and estimate AofA as the age at which at least 50% of the children in the sample produced a given word.

How much context?

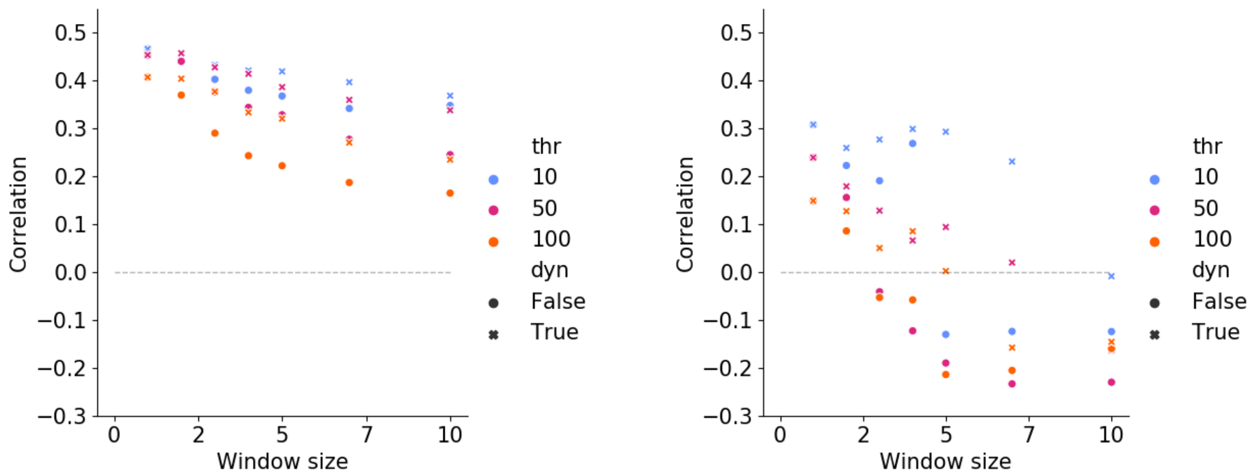
We trained Skipgram on the data described above, in order to derive vector representations for the words. We experimented with several hyperparameters of the model. We put our focus on the following:

- **Window size (win):** defined as the number of context words on each side of a target word (e.g. a window of size 1 includes a context word on each side of the target word). We explore values 1, 2, 3, 5 and 10.
- **Dynamic window size (dyn):** when this hyperparameter is enabled, the window size is dynamic, such that for each occurrence of a target word, the window size is sampled between 1 and **win**. This parameter has no practical effect when **win**=1.
- **Frequency threshold (thr):** words with frequency of occurrence below this threshold were removed, and are assumed to not be part of the vocabulary. Note that this is done after determining which words are in the context of a word, so words under the threshold are not replaced with further words in the context.

We fixed the values of the rest of hyperparameters to common default values (vector size: 100, initial learning rate: 0.025, negative sampling: off, context distribution smoothing: off, 'dirty' subsampling: off). Our code is available at: `https://github.com/rgalhama/public_ICCM2021`.

We then computed semantic relations between words as the cosine similarity between the corresponding vectors. As done in Alhama et al. (2020), we established a threshold *theta*, such that only words with cosine similarity larger than the threshold are considered to be neighbours. We then compute the *neighbourhood density* (ND) as the number of neighbours of each word. For reasons of space, we report results for $\theta = 0.7$, which led to highest correlations.

Figure 1 shows the results. We first focus on nouns (left graph). A very clear trend is evident for window size: given the same value of **dyn** and **thr**, a smaller window size predicts a larger correlation. Not surprisingly, the use of dynamic windows increases the fit (relative to the same fixed window size), as it decreases the amount of context available to a number of words; nevertheless, the minimum window size of 1 still performed better. We found that a small frequency threshold (**thr**=10) improves performance, indicating that even words with relatively small frequency have a role in shaping the semantic connections. In addition, the positive correlations indicate that words acquired earlier by children (i.e. smaller AofA) are those that have more semantic neighbours. This has interesting implications for language acqui-



(a) (b)

Figure 1: Correlation between AofA and ND in Skipgram, for nouns (left) and verbs (right). The hyperparameters window size (**win**), **thr** and **dyn** are defined above (in the text).

sition, as we discuss later. Overall, the results suggest that Skipgram holds promise for modelling word learning, with the best model (win=1, thr=10) having a correlation indicative of a medium effect size of 0.47. The results from these simulations suggest that restricting the influence of context to a very small window size consistently leads to a better fit, and that words with low frequency shape the semantic space in ways relevant to acquisition.

In order to see whether the good fit of Skipgram model extends to other syntactic categories, we evaluated its performance against AofA of verbs. As can be seen, the model shows a similar trend as for nouns, but also notable differences. For lower window sizes, results are fairly similar to nouns, albeit with smaller effect size. However, for the models with **thr**=10 (which overall performs better for verbs, as it did with nouns) there is not such a strong tendency for performance to decrease with window size, especially up to a window of size 5. As in the case of nouns, the correlations with greater effect size are positive (though this trend disappears as window size increases, specially for models with **thr**$>$10), indicating that having fewer semantic neighbours is beneficial for learning.

Discussion

In the case of nouns, the window size that best fits the AofA data is very small (**win**=1), suggesting that children attend to very local context, at least at an early age. Such a result makes intuitive sense in the context of children's small verbal memory spans, which only improve as they acquire more language. The positive correlation between ND and AofA, which very consistent in the case of nouns, indicates that nouns with fewer semantic neighbours are learnt earlier. This suggests that semantic neighbours may be acting as competitors during the process of noun learning, and nouns with more competitors are therefore less favoured.

Interestingly, we saw that the pattern of results of Skipgram is to some extent replicated for verbs, although with relevant differences. A dynamic window with a maximum size of 5 resulted in almost as good fit to the data as a window of 1 (provided **thr**=10). One potential interpretation is that larger windows allow the model to reach distant content that may include a verb's arguments, which is likely a helpful source of information about verb meaning (Gleitman, 1990). Thus, one reason why verbs are acquired later than nouns may be the need to learn to use more distant contexts, although more simulations are needed to support this explanation (in particular, simulations with adaptive window size that depend on age and/or syntactic category). We leave this to future work.

References

Alhama, R. G., Rowland, C., & Kidd, E. (2020). Evaluating word embeddings for language acquisition. In *Proceedings of the workshop on cognitive modeling and computational linguistics* (pp. 38–42). Association for Computational Linguistics.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of child language*, *44*(3), 677–694.

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

MacWhinney, B. (2000). *The CHILDES project: Transcription format and programs*. Lawrence Erlbaum Associates.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).