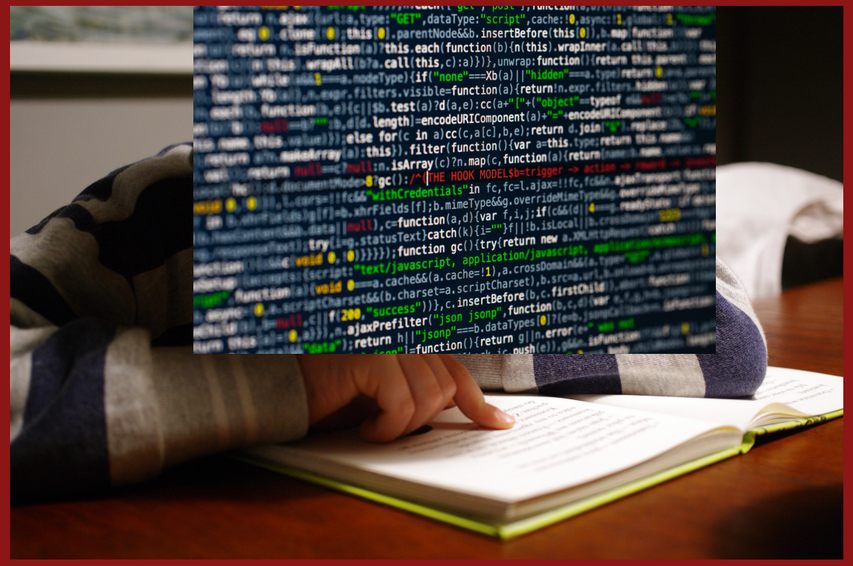
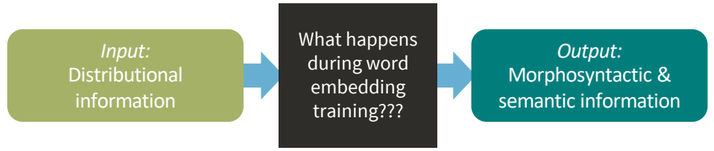
Word2Vec learns to categorise words before homing in on specific semantics.



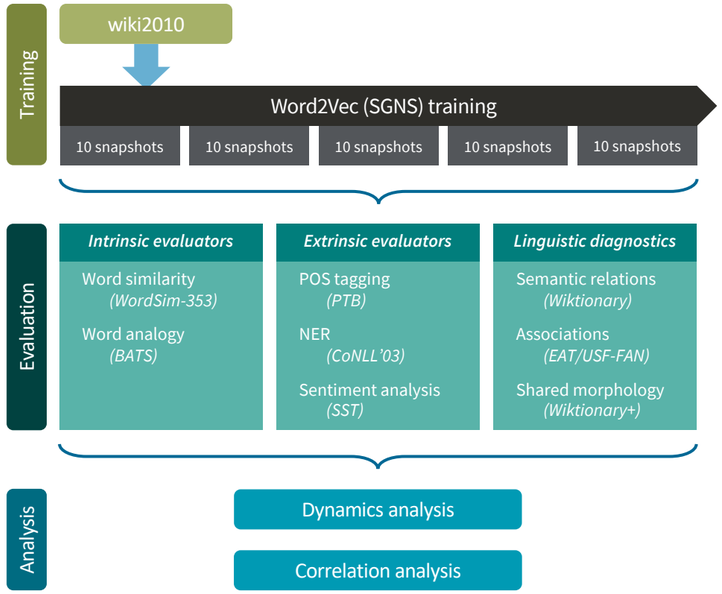
Understanding the learning dynamics of Word2Vec

Alvin Tan | tanawm@stanford.edu

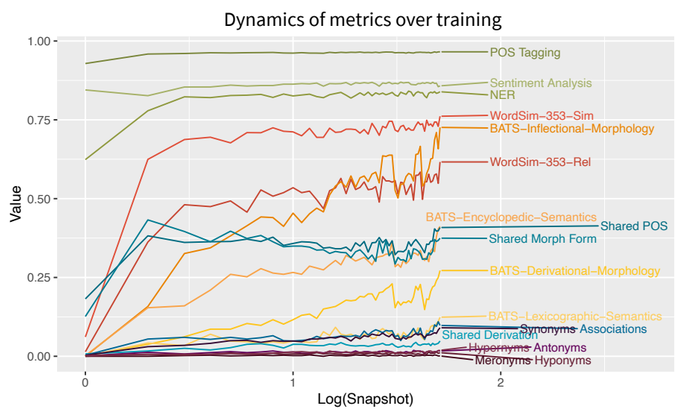
Motivation

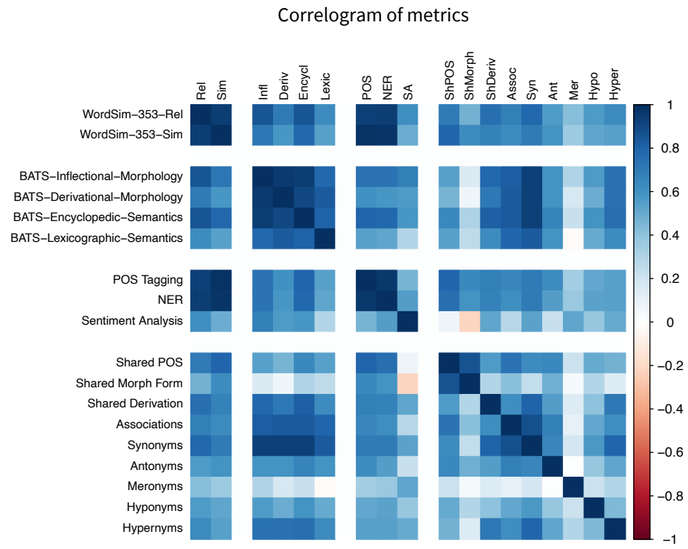


Method



Results





Discussion

- The evaluation and diagnostic metrics form clusters:
  - Word similarity, along with POS tagging and NER, as well as shared POS (suggesting spatial clustering ➞ categorisation)
  - Word analogy, along with synonyms and (to some extent) associations (suggesting direction ➞ specific semantics)
  - The former is learnt earlier and faster than the latter
- The extrinsic evaluators show little change across time ➞ pre-training embeddings may not significantly impact performance for some downstream NLP tasks

Limitations & future work

- Many evaluations are computationally intensive ➞ find efficient simplified measures to better track learning
- "Categorisation" is inferred ➞ develop a novel metric to explicitly measure categorisation (e.g. making use of thematic and taxonomic relations)
- These results apply to one specific way of training word embedding models ➞ investigate the dynamics of other word embedding models (e.g. GloVe)

Key references

Rogers, Anna et al. (2018). "What's in your embedding, and how it predicts task performance." *Proceedings of the 27th International Conference on Computational Linguistics*, 2690–2703.

Wang, Bin et al. (2019). "Evaluating word embedding models: Methods and experimental results." *APSIPA Transactions on Signal and Information Processing,* 8(1): e19.