Vector semantics and embeddings unit

- ▶ Lexical semantics, words as vectors (Monday)
- ▶ Cath-up day (Wednesday
- ▶ Word2Vec (**Today**)
- ▶ Applications of embeddings (next week Monday)

Today:

- ▶ Purpose of embeddings
- ▶ The premise of Word2Vec
- ▶ Training Word2Vec
- ▶ Observing results

**Vector semantics** *is the standard way to represent word meaning in NLP . . . The roots of the model lie in the 1950s when two big ideas converged: [using] a point in three-dimensional space to represent the connotation of a word, and the proposal . . . to define the meaning of a word by its distribution in language use, meaning its neighboring words or grammatical environments.*

*The idea of vector semantics is to represent a word as a point in a multi-dimensional semantic space that is derived from the distributions of word neighbors. Vectors for representing words are called* **embeddings***.*

- ▶ **Valence.** The pleasantness of the stimulus
- ▶ **Arousal.** The intensity of emotion provoked by the stimulus
- ▶ **Dominance.** The degree of control exerted by the stimulus

|            | Valence | Arousal | Dominance |
|------------|---------|---------|-----------|
| courageous | 8.05    | 5.5     | 7.38      |
| music      | 7.67    | 5.57    | 6.5       |
| heartbreak | 2.45    | 5.65    | 3.58      |
| cub        | 6.71    | 3.95    | 4.24      |

*Jurafsky and Martin*, 6.1, pg 4 & 5

**Figure 6.1** A two-dimensional (t-SNE) projection of embeddings for some words and phrases, showing that words with similar meanings are nearby in space. The original 60-dimensional embeddings were trained for sentiment analysis. Simplified from Li et al. (2015) with colors added for explanation.
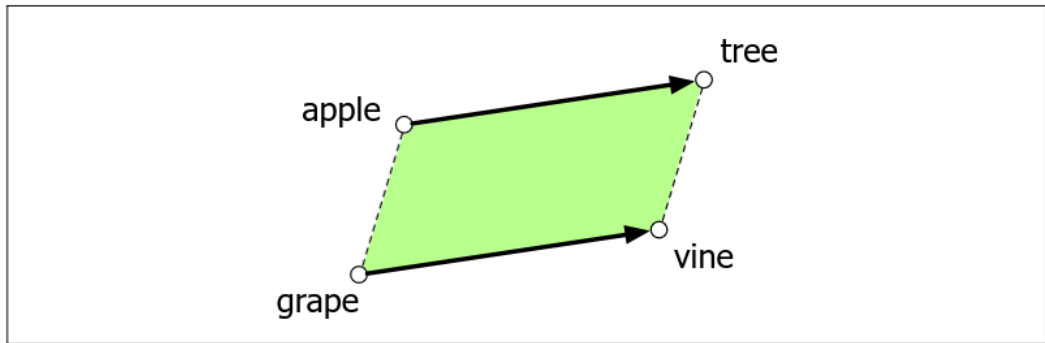
*Jurafsky and Martin*, 6.2, pg 6

**Figure 6.15** The parallelogram model for analogy problems (Rumelhart and Abrahamson, 1973): the location of $\overrightarrow{vine}$ can be found by subtracting $\overrightarrow{apple}$ from $\overrightarrow{tree}$ and adding $\overrightarrow{grape}$.

*Jurafsky and Martin*, 6.10, pg 25

Coming up:

- ▶ Do Semantle exercise (Fri, Nov 10)

- ▶ Read J&M chapter7 (Mon, Nov 13)

- ▶ Work on stylo assignment

Word2Vec assignment coming...