



The Structure of Meaning

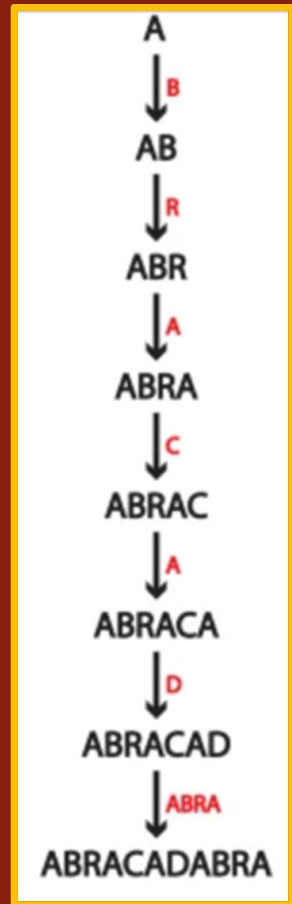
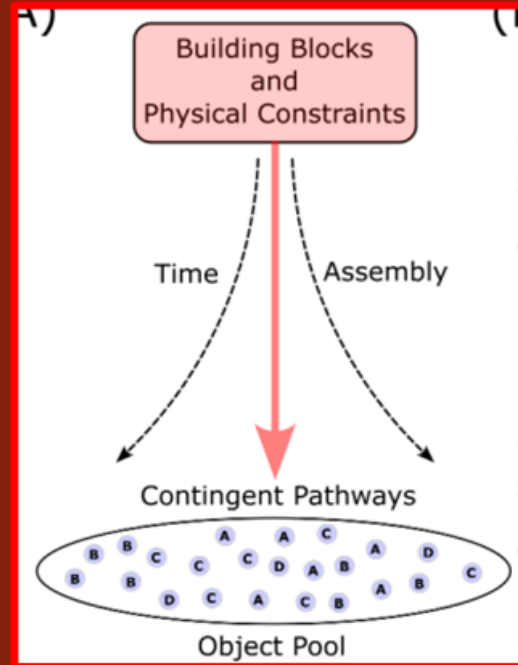
Estelle, Anna & Sam

Overall Ambition

❖ Uncover universal structural properties of meaning as they relate to the **process of abstraction**

❖ **Assembly Theory**: quantifies the complexity of an object based on the minimum number of steps required to build it → captures symmetries, recurring structures, signatures of selection & evolution, defines combinatorial time as a physical feature of the object

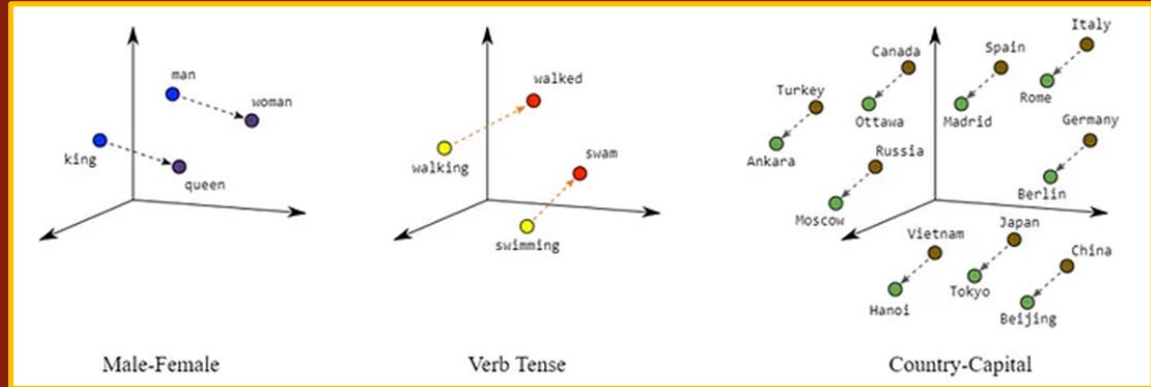
→ Study the emergence of abstraction as an evolutionary process within language



The search for an assembly theoretic object: Word Embeddings

- ★ Key breakthrough of deep learning and natural language processing
- ★ Geometry designed to capture semantic relationships
- Words with similar/related meanings are embedded close together, while words with different meanings are further apart

- Mathematical operations in the embedding space can reflect semantic operations in the language domain, e.g. vector arithmetic to solve word analogies



Linguistic Background: Semantics and Hierarchies

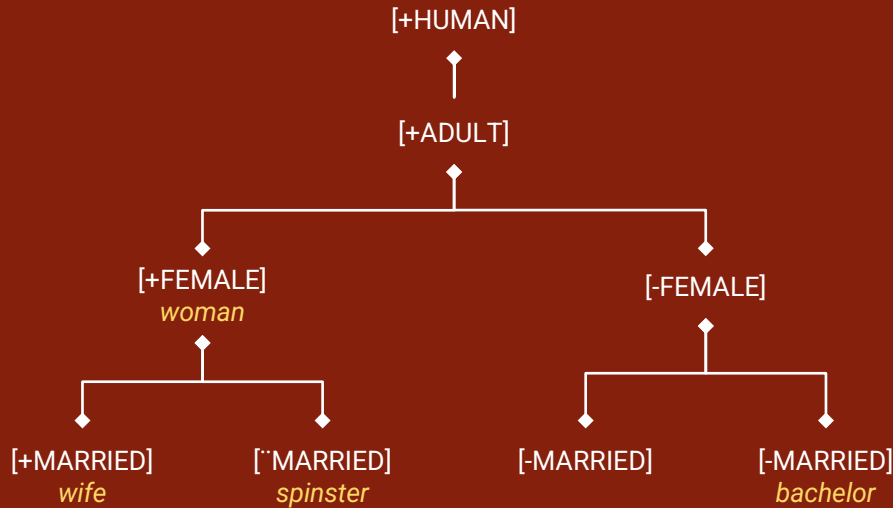
Meaning is traditionally analysed in semantic components, or **primitives**

<i>woman</i>	[+FEMALE]	[+ADULT]	[+HUMAN]	
<i>bachelor</i>	[-FEMALE]	[+ADULT]	[+HUMAN]	[-MARRIED]
<i>spinster</i>	[+FEMALE]	[+ADULT]	[+HUMAN]	[-MARRIED]
<i>wife</i>	[+FEMALE]	[+ADULT]	[+HUMAN]	[+MARRIED]

Linguistic Background: Hyponyms and Hypernyms

A lexical item P can be defined as a hyponym of Q if all the features of Q are contained in the feature specification of P.

Saeed, J. I. (2015). *Semantics* (Vol. 25). John Wiley & Sons.



Representing hierarchy and modularity in language: Poincaré Maps

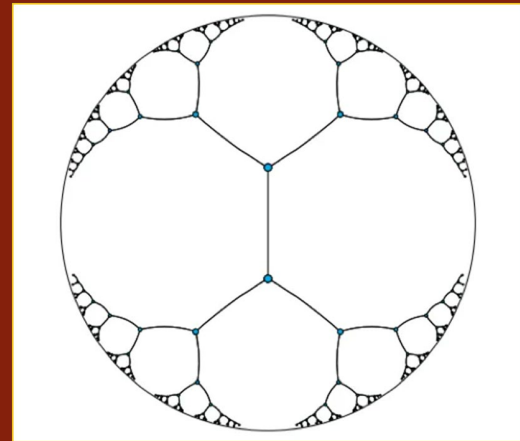
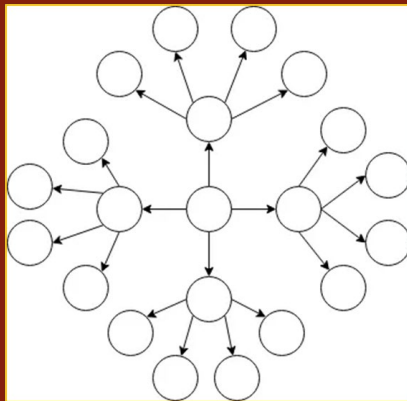
★ Poincaré embeddings map words to vectors in hyperbolic space!!

→ Distances and angles can represent more complex relationships, such as the semantic similarity and dissimilarity between words

Comparison with Euclidean Spaces

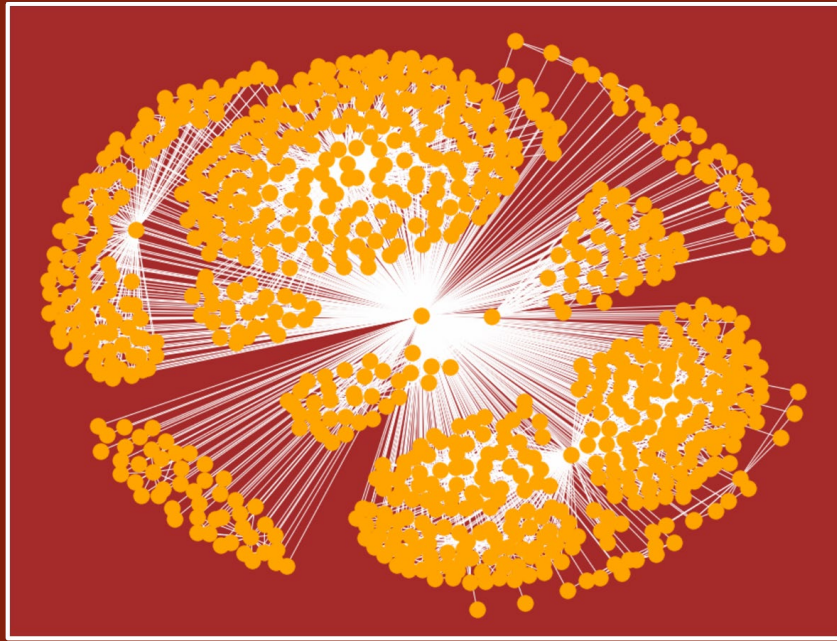
- Captures **linear relationships** between words
- Works well for many tasks, but can struggle with complex hierarchical relationships
- **Higher-dimensional**

e.g. Word2Vec and GloVe

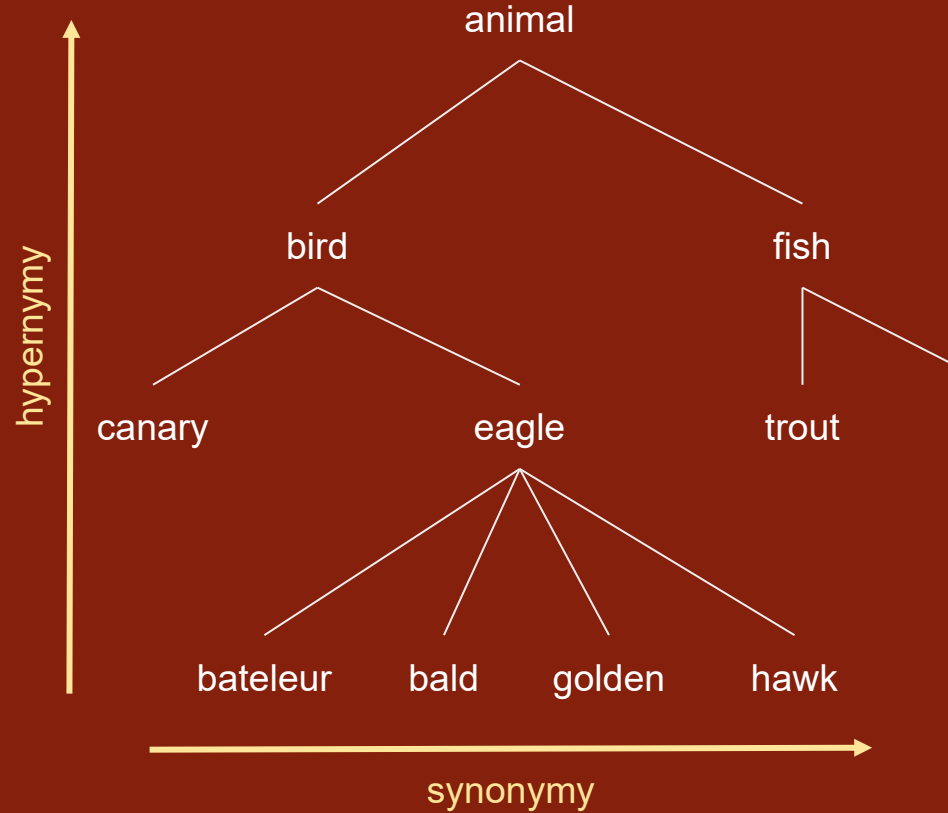


The space from the center increases exponentially and each line are the same Poincaré distance. No line can reach the circumference of the circle → an arbitrary number of levels can fit within relatively few dimensions

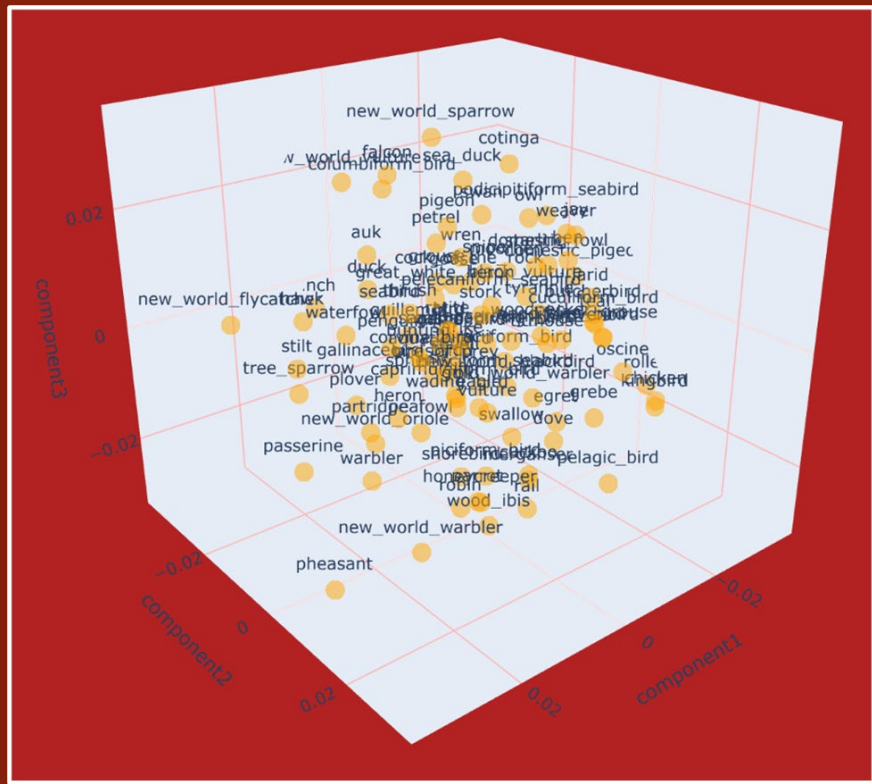
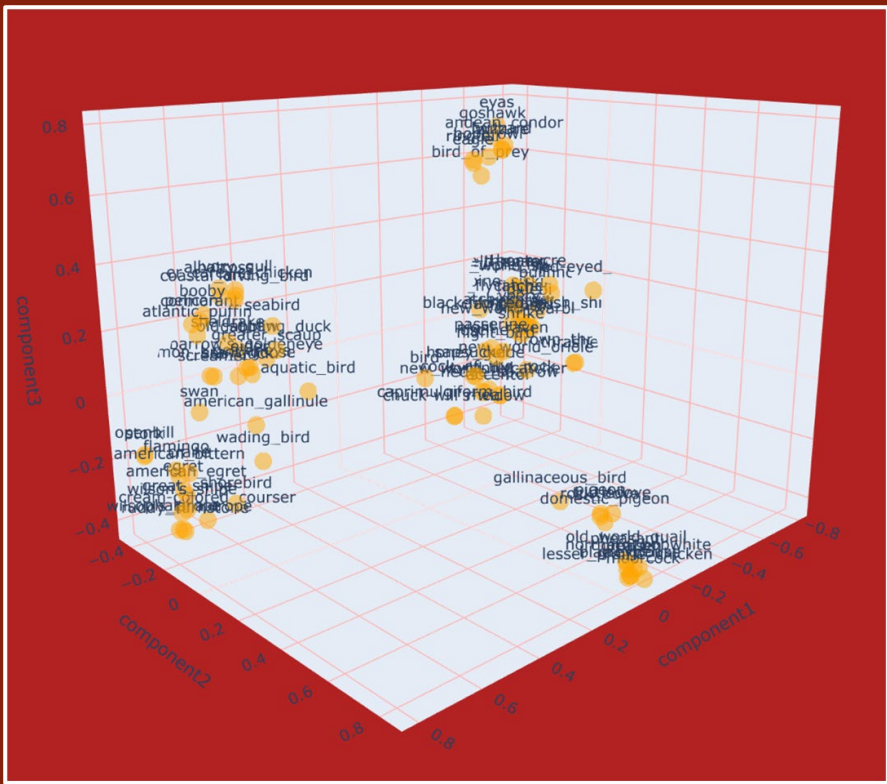
WordNet



bird.n.01



Poincaré vs Word2Vec



Looking ahead: the search continues...

- ★ How well do pre-trained word embeddings (e.g. Word2Vec) map to Poincaré embeddings? How dependent is it on tokenization choices?
- ★ Investigate the number of “downward” connections of a word in a hierarchical framework as an indicator of generalization and abstraction (e.g. using WordNet)
- ★ Project ~1000 words on a **concrete-vs-abstract continuum space**
- ★ Embed different text-based contents into Poincaré spaces: for example a child book vs a philosophy paper
- Gather intuition about **abstraction** and **conceptualisation** processes
- Find the most meaningful structure capturing their **emergence, selection and evolution through combinatorial space**



THANK YOU!