Distributional semantics

BHI Youth Awards
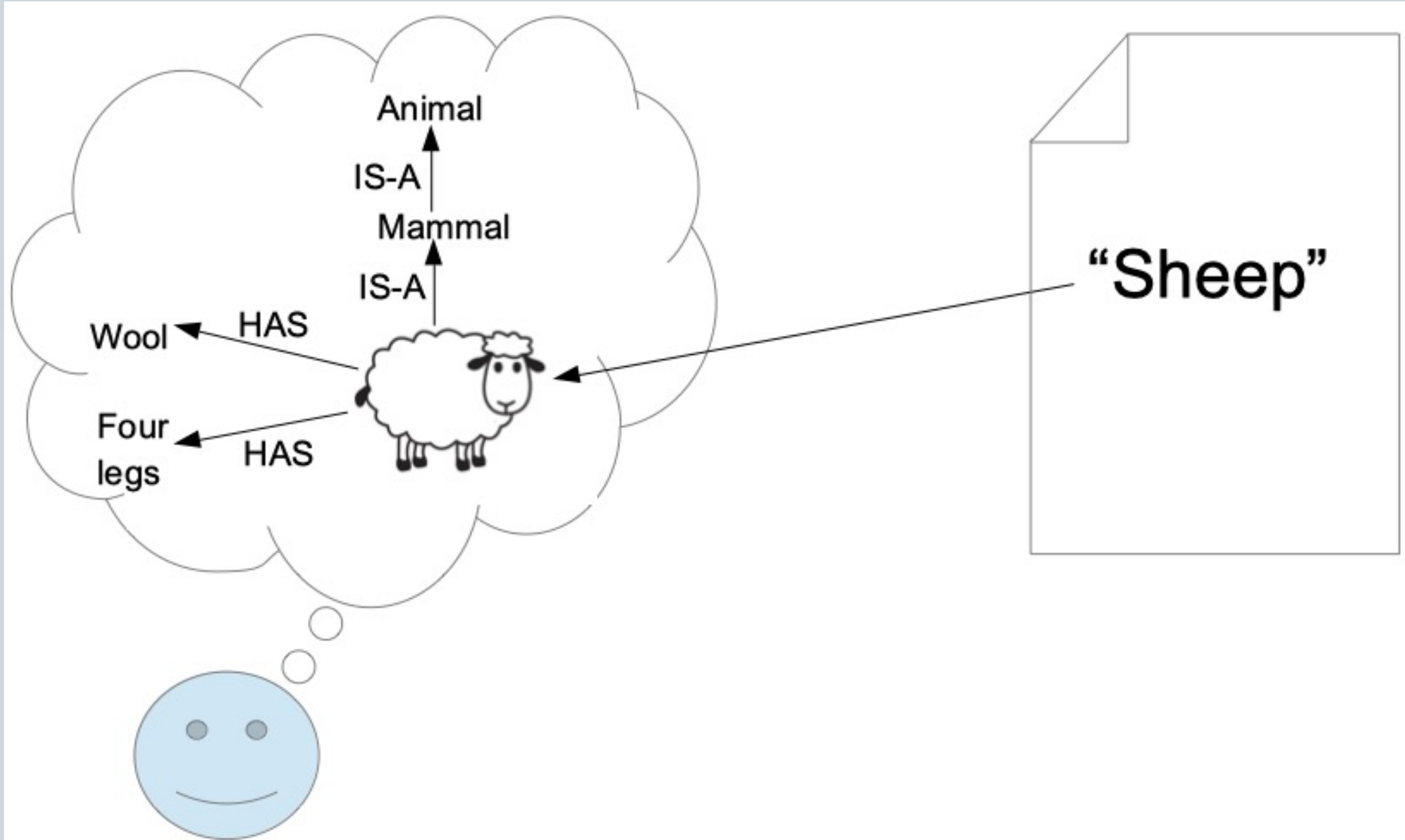
- How can we represent words numerically, so we can build computational models of language?

- How can we account for:
  - Their similarities?
  - Their meaning, or semantics?

- Distributional semantics and context as meaning

- Vector based representations

Lexical semantics

An experiment – wombling and snetches

The Captain's side raked first. Tom staked. The hired sportsmen played so hard that they **wombled** too fast, and were shaky with the rakes. Tom fooled around the way he always did, and all his stakes dropped true. When it was his turn to rake he did not let Captain Najork and the hired sportsmen score a single rung, and at the end of the **snetch** he won by six ladders.

From How Tom beat Captain Najork and his hired sportsmen
by Russell Hoban and Quentin Blake

- "You shall know a word by the company it keeps" (JR Firth)

- The contexts in which words appear correlate with their meaning

- We understand a word by its distribution: the set of contexts in which it is found

- "Don't think, but look!" (Wittgenstein) - i.e. the meaning of a word is the description of its use,

Words in context – concordances and collocations

https://www.webcorp.org.uk/

Thank you

angus.roberts@kcl.ac.uk

https://www.kcl.ac.uk/people/angus-roberts