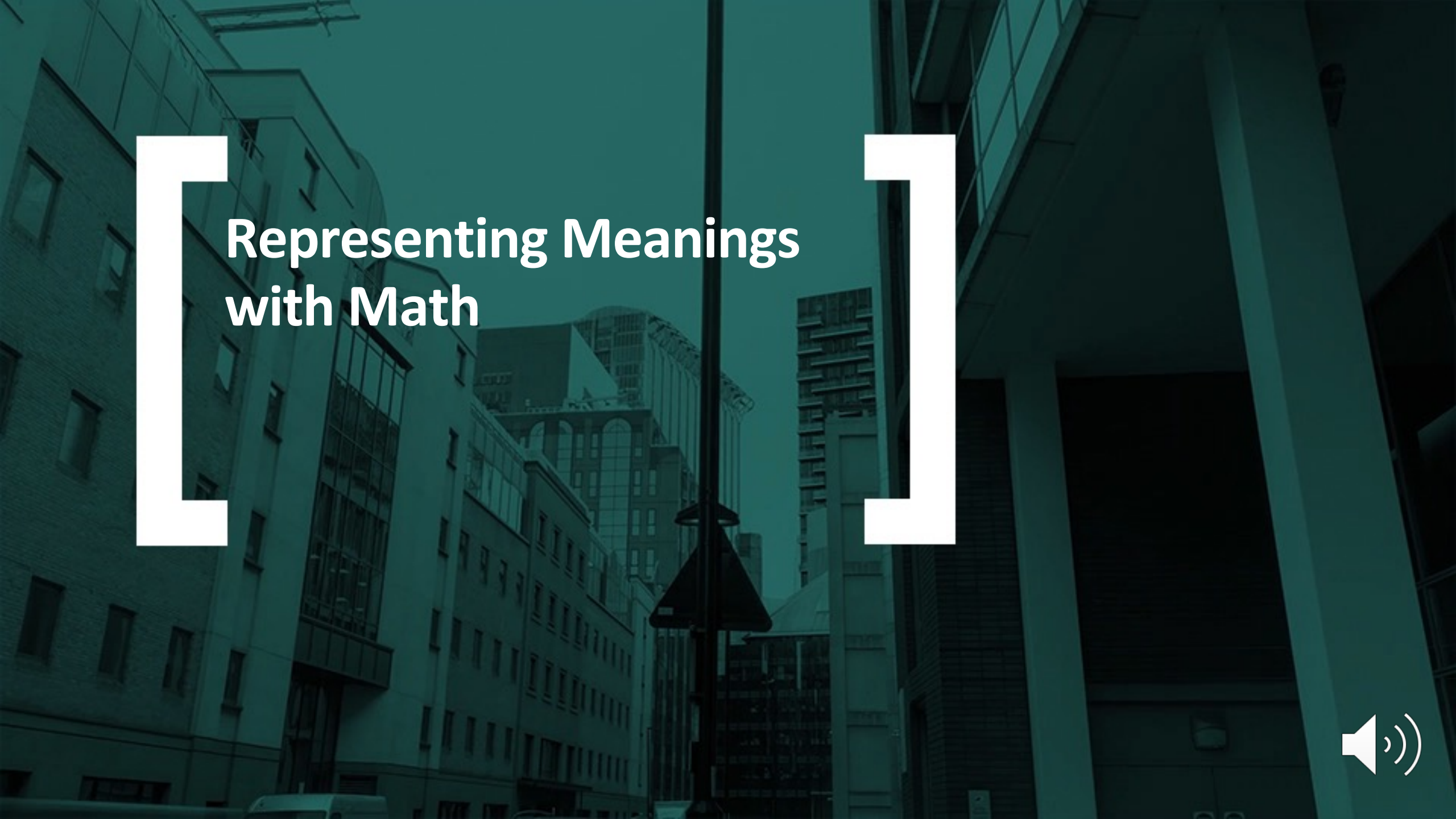
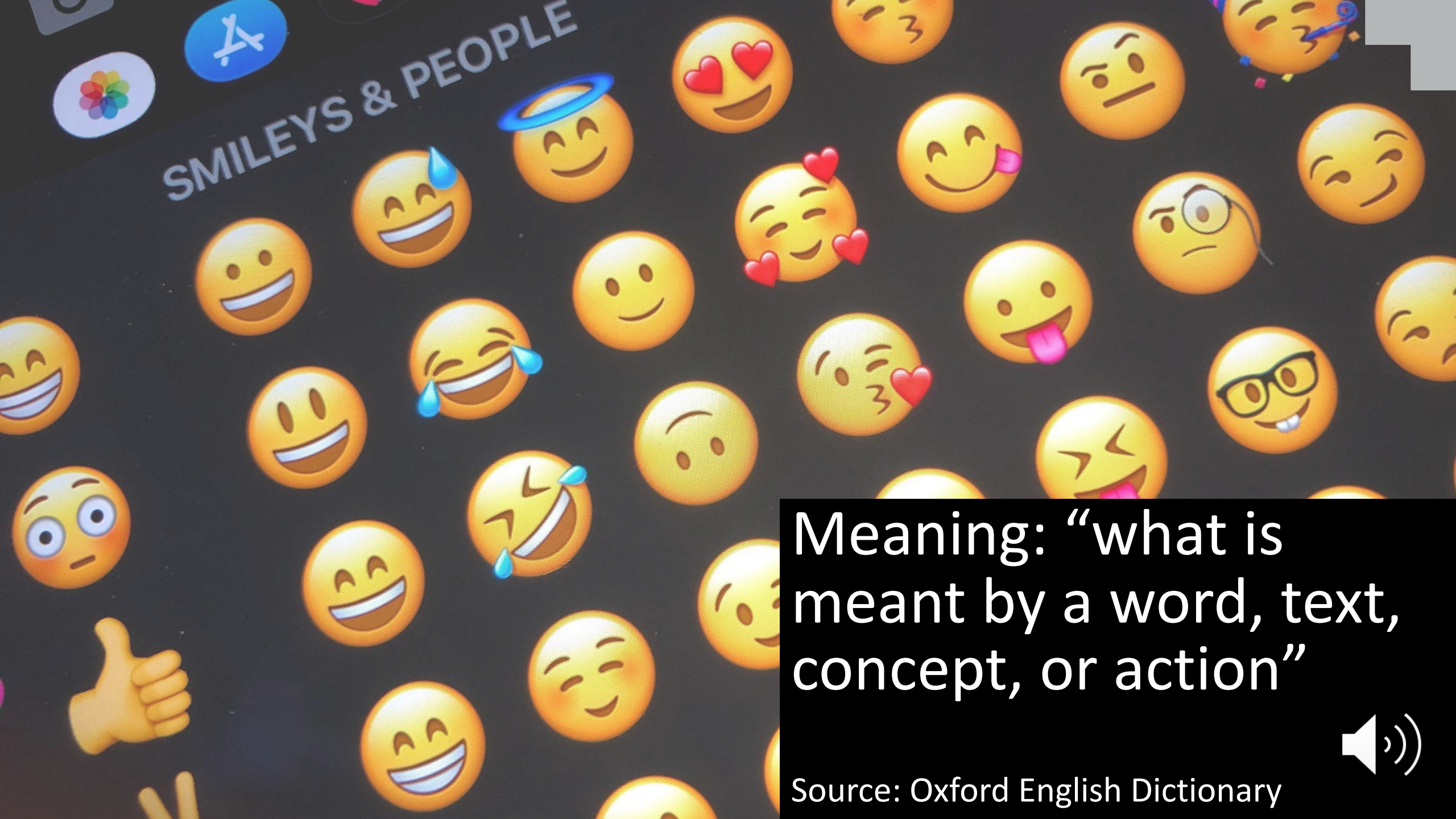

Representing Meanings
with Math

Goals

1. To appreciate how computational linguistics represent meanings
2. To familiarise with the critical tenets of computational linguistics, ‘denotational semantics’ and the ‘distributional hypothesis’



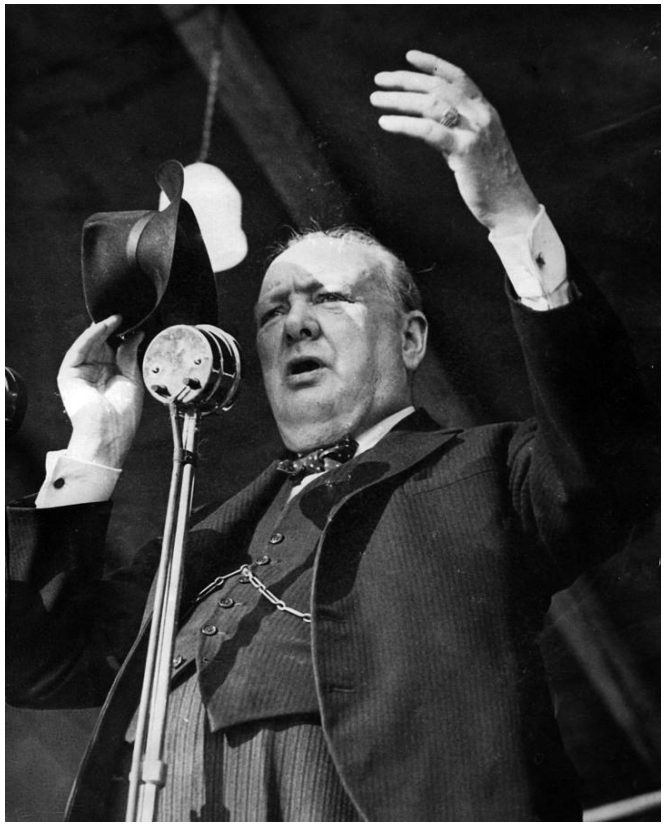
Meaning: “what is meant by a word, text, concept, or action”

Source: Oxford English Dictionary

Meanings are central to natural language

The natural language (human beings speak) presents critical features:

- It conveys meanings to an audience
- Oftentimes, it targets a reference audience
- It uses words that reflect symbols and social categories, that is, culture



Meanings are central to NLP systems

The concept of meaning is the place to start for many natural language processing analyses.

Let us have a closer look at:

- how meanings are represented in computational linguistics
- how machines ‘look’ at meanings

The angle of computational linguistics on meanings modeling

A (computational) linguist's perspective

Two pillars that reflect how linguists' think about meanings:

- Denotational semantics
- Distributional Hypothesis

The intuition behind denotational semantics

Semantics, as the study of meanings, concerns the relationship between signifiers — like words, phrases, signs, and symbols — and what they stand for in their denotation.

Denotations comprise both the salient features associated with an entity (being a concrete instance or a category) and the cognitive and behavioural effects of using a signifier that invokes an entity.

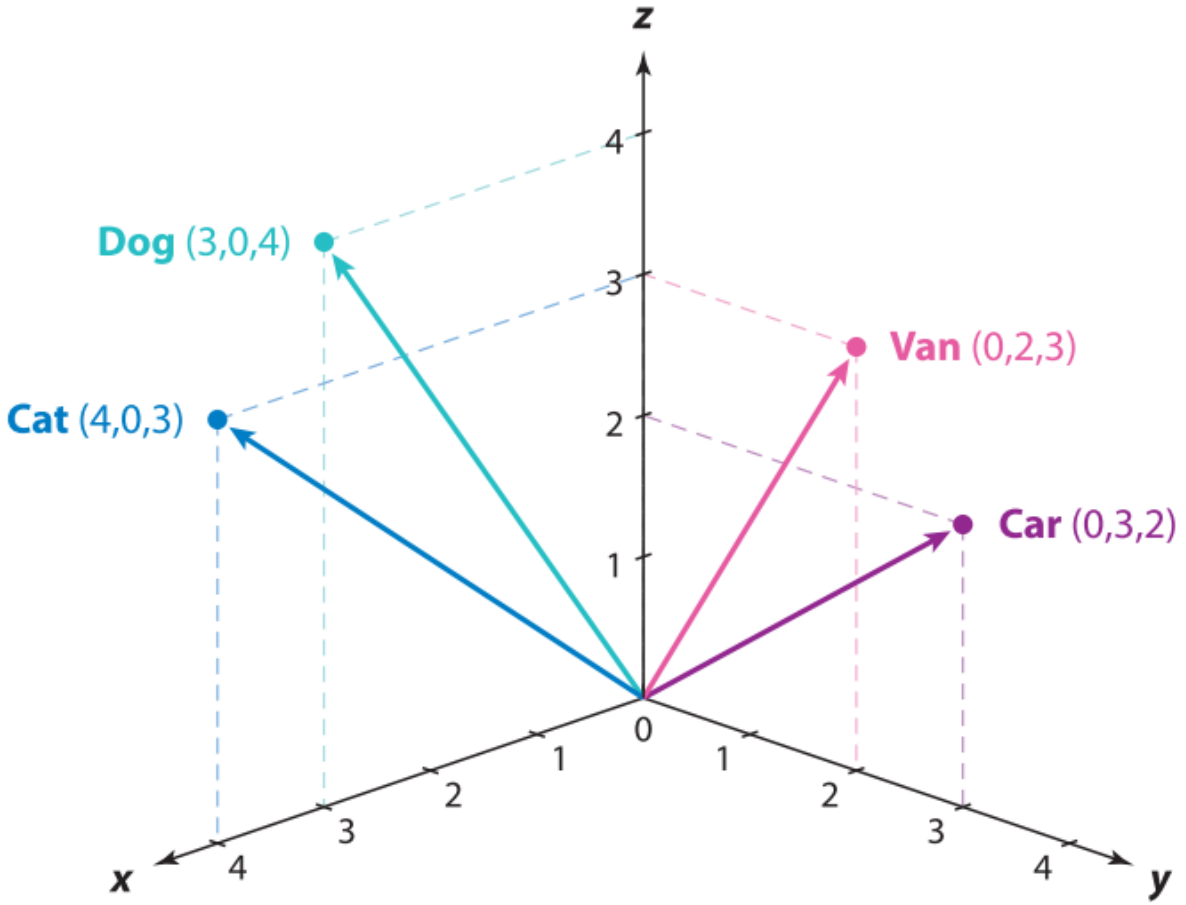
For example, the lexeme ‘hip-hop’ conveys meanings about what constitutes a ‘hip-hop’ song and the values, norms, and beliefs that orient the behaviour of ‘hip-hop people.’

The Distributional Hypothesis (DH)

“Difference of meaning correlates with [a] difference of distribution’’ (Harris, 1954)

”Semantic similarity is a function of the contexts in which words are used”. (Miller & Charles, 1951)

“DS is not only a method for lexical analysis but also a theoretical framework to build computational models of semantic memory’’ (Lenci, 2018)



The ‘standpoint’ of machines

A machine's perspective

Human beings are entrenched in natural language symbols and social categories, while machines are not. Hence, devices are not able to associate meanings with lexemes.

This explains why analysing massive natural language datasets has been traditionally taxing/impossible. Human beings are good at making sense of language, but they are bad at computing (so, hand-curated workflows are not scalable). On the contrary, machines are good at computing, but they're just dull (so computation capacity looks for a workflow to scale up).

Strategies to help machines handle meanings

Mainly, there are two strategies by which machines can handle meanings:

- Human beings can provide devices with ‘pattern-matching’ rules that induce meaningful responses on machines *vis a' vis* natural language inputs.
- With statistical frameworks (e.g., Distributional Representations), machines can discover/learn the semantic similarity between lexemes.

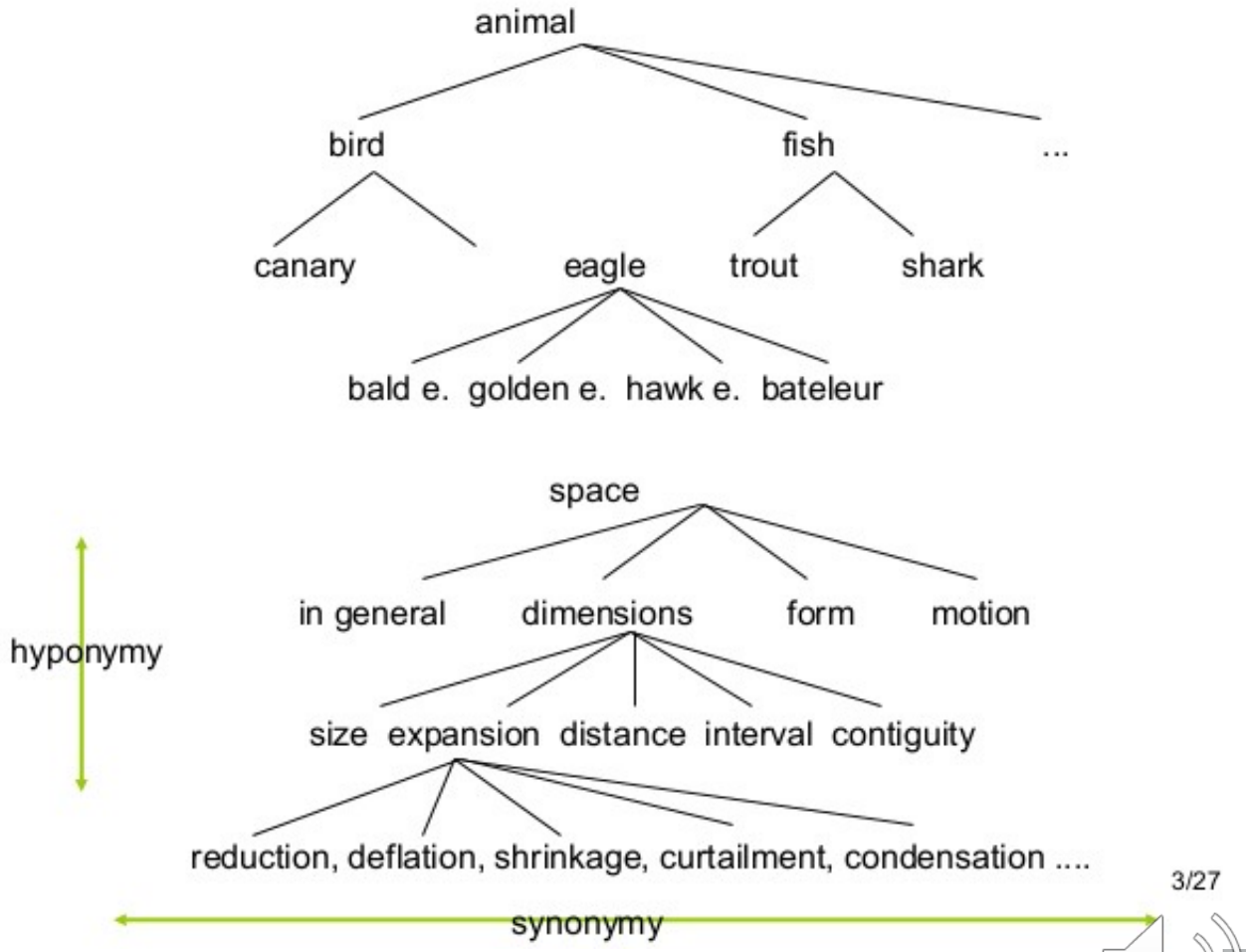
Regular expressions

Regular expressions use a special kind (class) of formal language grammar called standard grammar.

Regular grammars have predictable; provable behaviour yet are flexible enough to power some of the most sophisticated dialogue engines and chatbots on the market. Amazon Alexa and Google Now are mostly pattern-based engines that rely on regular grammar.

Dictionary-based approaches

WordNet® is a human-annotated, sizeable lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked using conceptual-semantic and lexical relations (e.g., hyponymy).



Wrap-up

Wrap-up

- Meanings are central to the field of NLP.
- Computational linguistics has established methods to represent meanings. These methods rely on the idea of ‘denotational semantics’ and the ‘Distributional Hypothesis.’
- Contrarily to human beings, machines do not have the knowledge to understand meanings. However, they can be instructed to behave as if they knew the meaning of words. Furthermore, machines can learn the semantic relationships hidden in natural language data.