University of BRISTOL

7.2 Word Vectors

Edwin Simpson

Department of Computer Science,

University of Bristol, UK.

bristol.ac.uk

Representing Words

- To process meaning, we have to process the individual words in a sentence
- We need a representation of the words in a document that we cane input to models like neural networks
- Relational view of meaning: encode relations between words into their representations

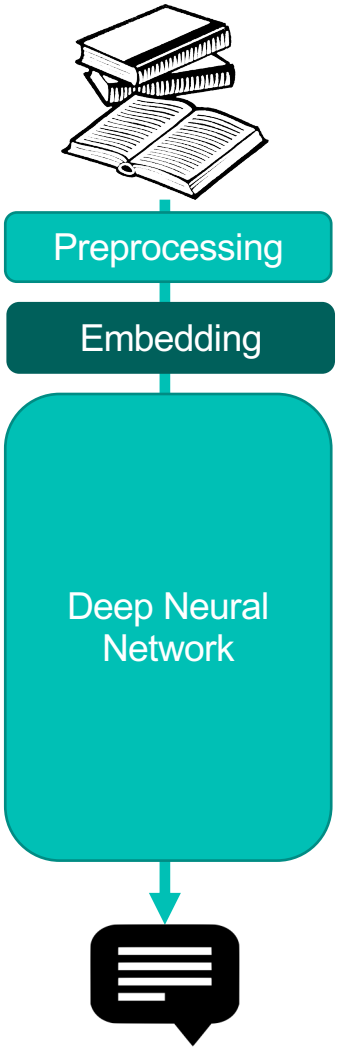
One-Hot Encodings

- How is a word represented?
  - By the text string itself;
  - Or by an index into a vocabulary.
- One-hot encoding:
  - Vocabulary size = *V*
  - Each word is represented by a vector of length *V*
  - All values in the vector are zero…
  - …except the value corresponding to the index of the word in *V.*
- A sparse representation that doesn't allow us to compare words.

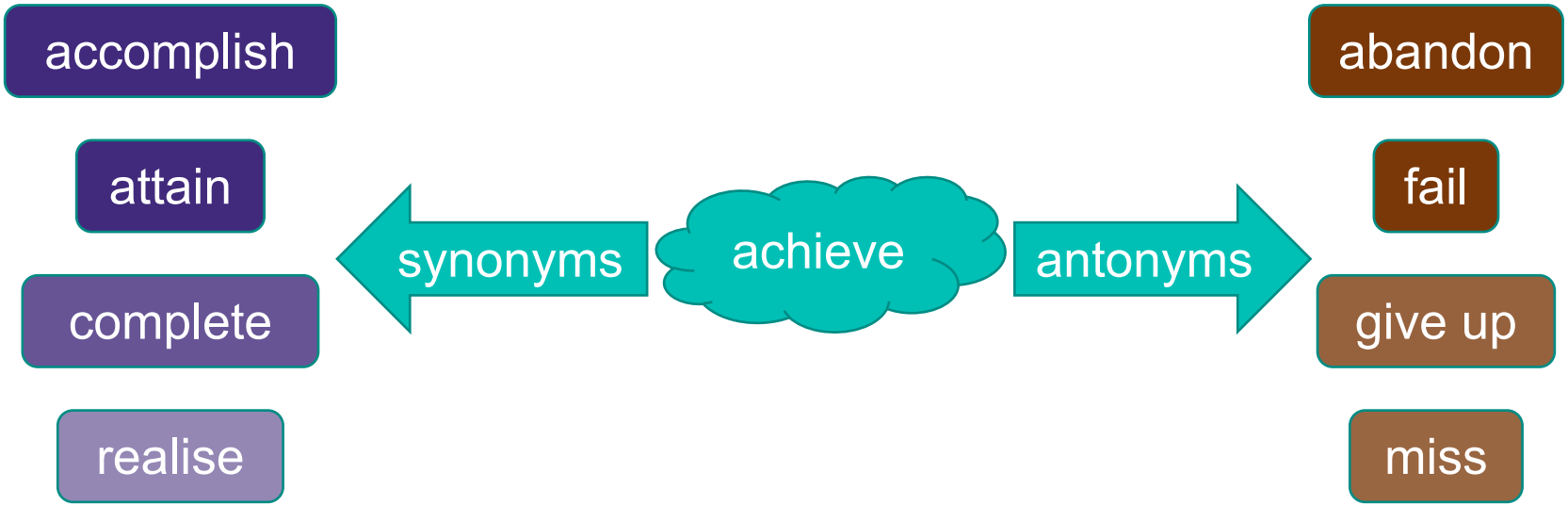
Couch	1	0	0
Elephant	0	1	0
Sofa	0	0	1

Word Representations

- Can we replace the one-hot encoding with a better word representation?
- The representation should capture various aspects of a word's semantics…

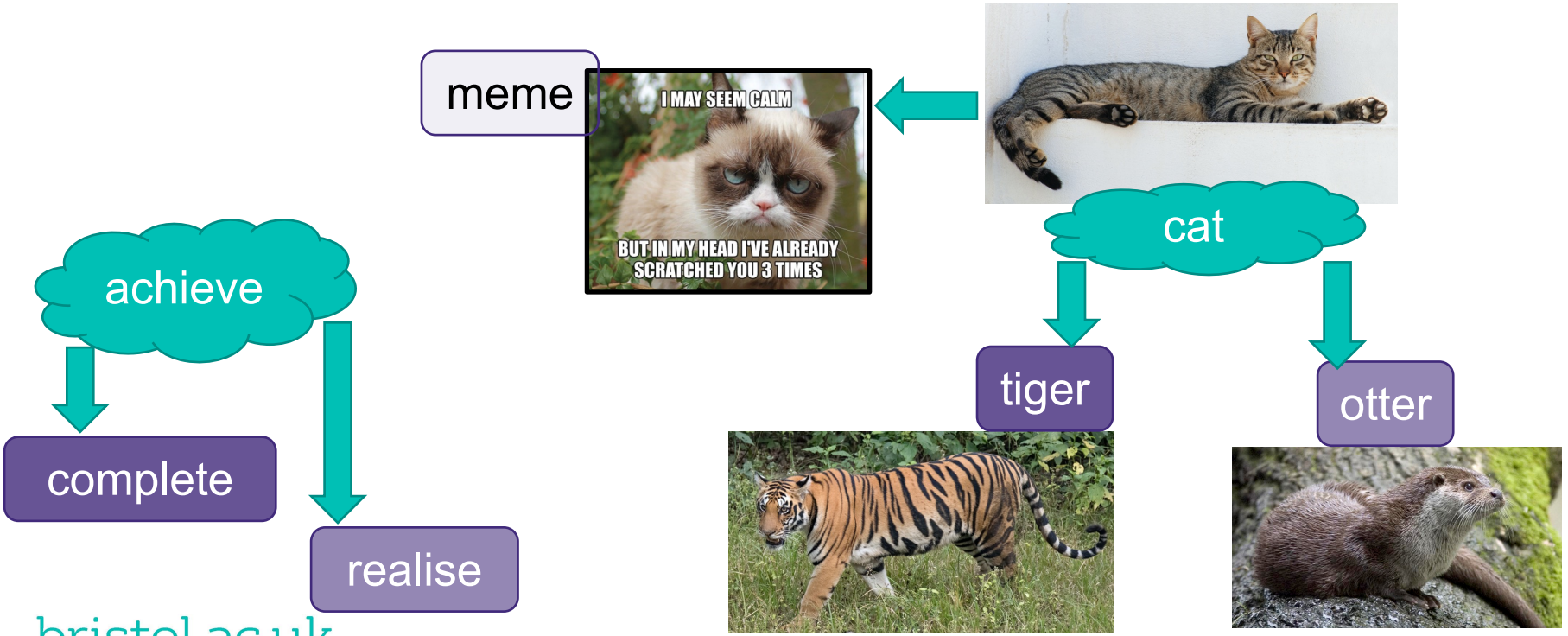


Desiderata for Word Representations: Synonyms and Antonyms



Desirata for Word Representations: Hypernyms, Hyponyms, Meronyms

Desirata for Word Representations: Similarity



Desirata for Word Representations: Associations & Semantic Fields

Semantic field/Topic

coffee

cup

Coffee
Drink
Cafe

mask

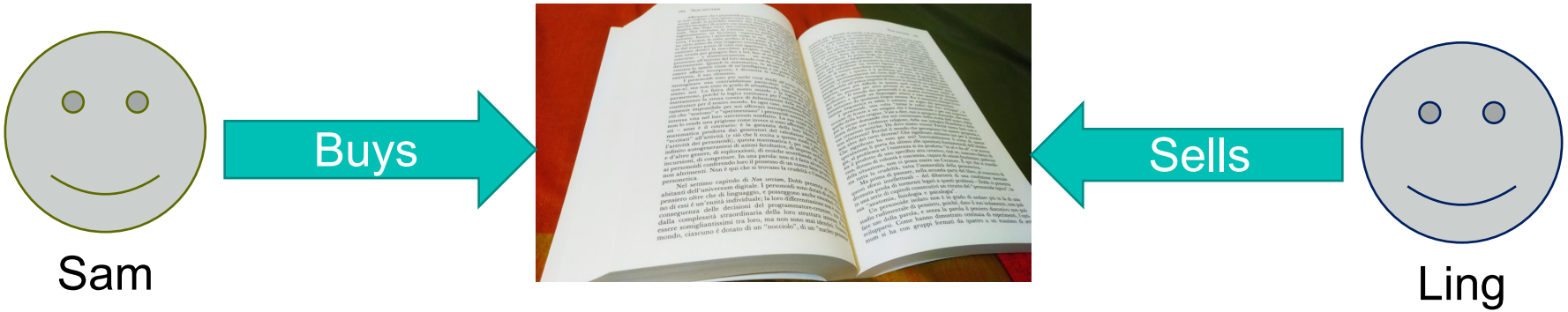
surgeon

Surgery
Operation
Medical

Desirata for Word Representations: Semantic Frames

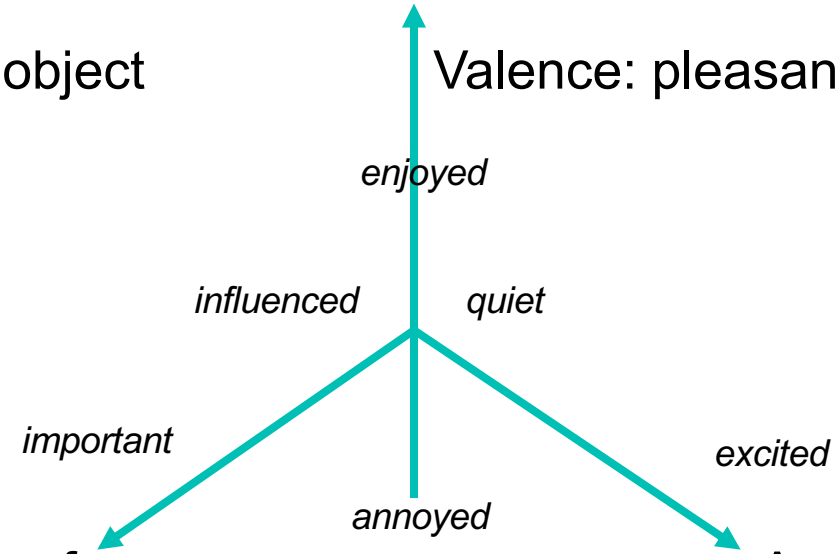
Sam buys a book from Ling

Ling sells a book to Sam



Desirata for Word Representations: Connotation

The stimulus = the object of the discussion.



Vector Semantics

- Represent words as points in a **multi-dimensional space (embeddings)**
  - Different dimensions correspond to different aspects of meaning
  - Compose meaning of multiple words using arithmetic
- **Distributional hypothesis**
  - We know a word by the company it keeps (Firth 1957)
  - Learn a word's vector from the other words that occur near it (its context)

Term-Document Matrix

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

Word vector

Document vector

Counts from Shakespeare plays. Figure 6.3, from Chapter 6, Speech and Language Processing, 3rd edition draft, Jurafsky & Martin (2019).

Summary

- Many aspects of meaning can’t be represented by a bag-of-words
- One-hot encodings and term-document matrices provide one way to represent words numerically
- Better representations would capture relations such as equivalence, opposites, parts of objects, categories, similarity, semantic frames
- The *distributional hypothesis* proposes that meaning can be learned from the context a word is used in
- Word embeddings make use of this hypothesis to provide better word representations