# COMP 3225

# Natural Language Processing

## Word Sense and WordNet

Stuart E. Middleton

sem03@soton.ac.uk

University of Southampton

# Overview

- Word Senses
- WordNet

- <break - discussion point>

- Word Sense Disambiguation

# Word Senses

- Words are ambiguous
- A sense (or words sense) is a discrete representation for one meaning of a word

  **mouse**$^{1}$ : .... a *mouse* controlling a computer system in 1968.
  **mouse**$^{2}$ : .... a quiet animal like a *mouse*
  **bank**$^{1}$ : ...a *bank* can hold the investments in a custodial account ...
  **bank**$^{2}$ : ...as agriculture burgeons on the east *bank*, the river ...

- Word sense
  - Human >> Dictionary and thesaurus definitions
  - Machine >> Embeddings
  - Dictionary and thesaurus definitions for words are written for human eyes, but they are themselves sentences which can be used for embeddings

# Word Senses

- A sense is defined through its relationship to other senses

| | |
|---|---|
| right | *adj.* located nearer the right hand esp. being on the right when facing the same direction as the observer. |
| left | *adj.* located nearer to this side of the body than the right. |
| red | *n.* the color of blood or a ruby. |
| blood | *n.* the red liquid that circulates in the heart, arteries and veins of animals. |

'left' is the opposite of 'right'
'blood' is a 'liquid'

- Online resources exist where word senses are defined for words
  - WordNet, FrameNet, VerbNet, Streusle ...
  - https://wordnet.princeton.edu/
  - https://framenet.icsi.berkeley.edu/fndrupal/
  - https://verbs.colorado.edu/verbnet/
  - https://github.com/nert-nlp/streusle

# Word Senses

- Relations Between Senses
  - Synonym, antonym, hyponym, hypernym, meronym ...
- Synonym
  - Lemma (word) with nearly identical meaning

    couch <--> sofa

    vomit <--> throw up  >> includes multi-word phrases

    car <--> automobile
- Antonym
  - Lemma with nearly opposite meaning

    long <--> short

    big <--> little

    up <--> down

# Word Senses

- Taxonomic relations IS-A; part-of ...
  - Hyponym is a more specific lemma (subclass)
    car --> vehicle
    dog --> animal
    mango --> fruit
  - Hypernym is a more general lemma (superclass or superordinate)
    vehicle --> car
    animal --> dog
    fruit --> mango
  - Meronym is a lemma which is part of something (part-whole or part-of)
    wheel --> car
    leg --> chair
  - Holonym is the opposite
    car --> wheel
    chair --> leg

# WordNet

- WordNet 3.0 is a lexical database
  - 117,798 nouns, 11,529 verbs, 22,479 adjectives and 4,481 adverbs
  - On average 1.23 senses per noun, 2.16 senses per verb
  - Available via web, download imported as a corpus via toolkits like NLTK

The noun "bass" has 8 senses in WordNet.

1. bass[1] - (the lowest part of the musical range)
2. bass[2], bass part[1] - (the lowest part in polyphonic music)
3. bass[3], basso[1] - (an adult male singer with the lowest voice)
4. sea bass[1], bass[4] - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass[1], bass[5] - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass[6], bass voice[1], basso[2] - (the lowest adult male singing voice)
7. bass[7] - (the member with the lowest range of a family of musical instruments)
8. bass[8] - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

# WordNet

- Synset is a synonym set
- Supersense is a semantic category
  - 26 categories for noun; 15 for verb; 2 for adjective; 1 for adverb

| Category | Example | Category | Example | Category | Example |
|---|---|---|---|---|---|
| ACT | *service* | GROUP | *place* | PLANT | *tree* |
| ANIMAL | *dog* | LOCATION | *area* | POSSESSION | *price* |
| ARTIFACT | *car* | MOTIVE | *reason* | PROCESS | *process* |
| ATTRIBUTE | *quality* | NATURAL EVENT | *experience* | QUANTITY | *amount* |
| BODY | *hair* | NATURAL OBJECT | *flower* | RELATION | *portion* |
| COGNITION | *way* | OTHER | *stuff* | SHAPE | *square* |
| COMMUNICATION | *review* | PERSON | *people* | STATE | *pain* |
| FEELING | *discomfort* | PHENOMENON | *result* | SUBSTANCE | *oil* |
| FOOD | *food* | | | TIME | *day* |

**26 supersenses for nouns in WordNet**

# WordNet

- WordNet has a number of sense relations

| Relation | Also Called | Definition | Example |
|---|---|---|---|
| Hypernym | Superordinate | From concepts to superordinates | $breakfast^1 \rightarrow meal^1$ |
| Hyponym | Subordinate | From concepts to subtypes | $meal^1 \rightarrow lunch^1$ |
| Instance Hypernym | Instance | From instances to their concepts | $Austen^1 \rightarrow author^1$ |
| Instance Hyponym | Has-Instance | From concepts to their instances | $composer^1 \rightarrow Bach^1$ |
| Part Meronym | Has-Part | From wholes to parts | $table^2 \rightarrow leg^3$ |
| Part Holonym | Part-Of | From parts to wholes | $course^7 \rightarrow meal^1$ |
| Antonym | | Semantic opposition between lemmas | $leader^1 \Longleftrightarrow follower^1$ |
| Derivation | | Lemmas w/same morphological root | $destruction^1 \Longleftrightarrow destroy^1$ |

**Some relations for nouns in WordNet**

| Relation | Definition | Example |
|---|---|---|
| Hypernym | From events to superordinate events | $fly^9 \rightarrow travel^5$ |
| Troponym | From events to subordinate event | $walk^1 \rightarrow stroll^1$ |
| Entails | From verbs (events) to the verbs (events) they entail | $snore^1 \rightarrow sleep^1$ |
| Antonym | Semantic opposition between lemmas | $increase^1 \Longleftrightarrow decrease^1$ |

**Some relations for verbs in WordNet**

# Break

- Panopto Quiz - discussion point

- Open up WordNet using the link below and explore the lemma 'bark' in the word sense 'a dog barked at me'.
  What are the hypernym lemma for this word sense?
  http://wordnetweb.princeton.edu/perl/webwn

utter, emit, let out, let loose
noise
yelp, yip, yap
talk, speak, utter, mouth, verbalize, verbalise

# Break

- Panopto Quiz - discussion point

- Open up WordNet using the link below and explore the lemma 'bark' in the word sense 'a dog barked at me'.
  What are the hypernym lemma for this word sense?
  http://wordnetweb.princeton.edu/perl/webwn

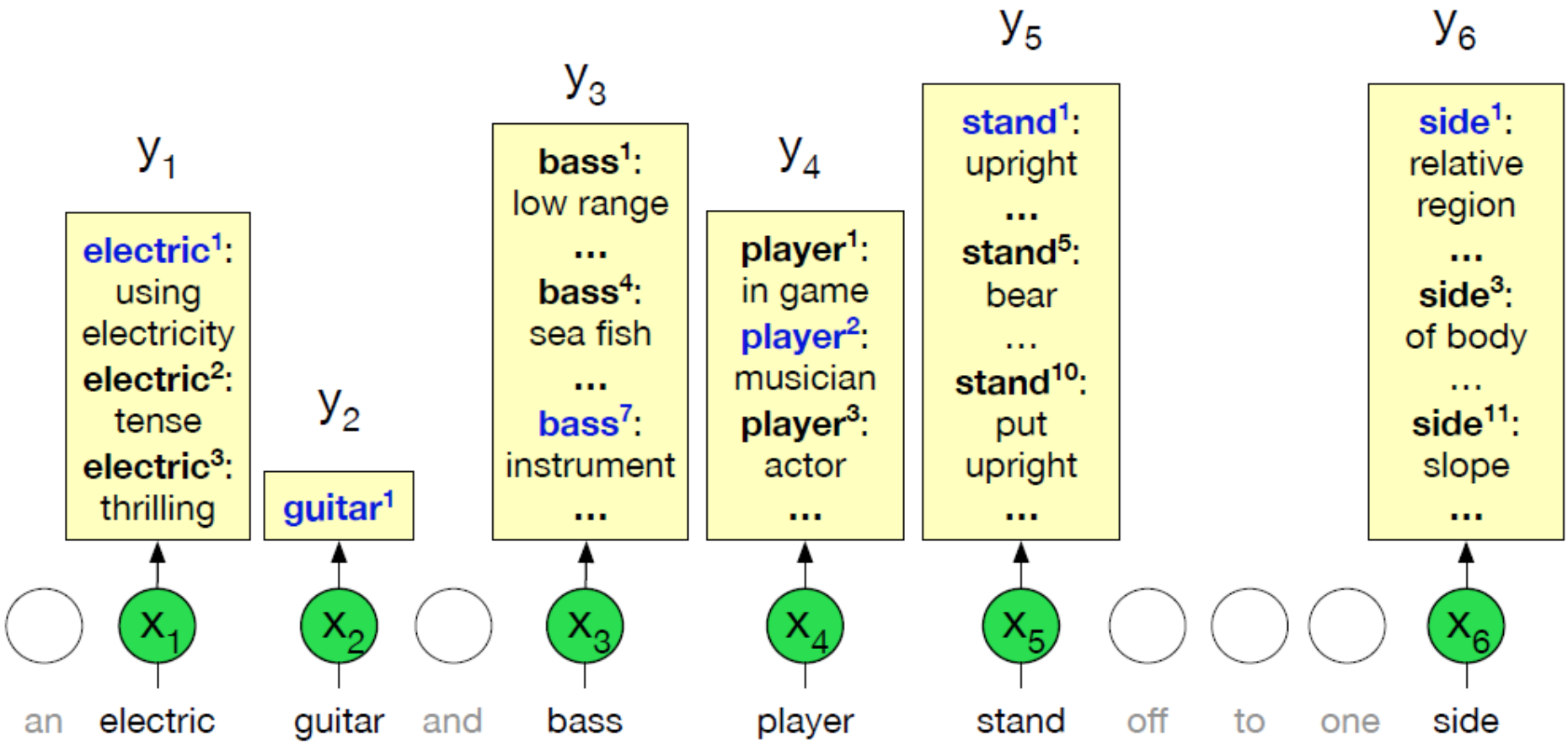
utter, emit, let out, let loose >> hypernyms (more general lemma) for 'bark.v.04'
noise >> hypernym of noun word sense 'bark.n.02'
yelp, yip, yap >> these are the hyponyms (more specific lemma)
talk, speak, utter, mouth, verbalize, verbalise >> hypernym for 'bark.v.01' which is the wrong sense (not a dog bark its an unfriendly tone)

# Word Sense Disambiguation

- All-Word Word Sense Disambiguation (WSD) task
  - Given a sentence select the correct word sense for all words
- Training data
  - Dataset with word sense annotations
  - SensEval and SemEval WSD task corpus using WordNet sense annotations
  - http://web.eecs.umich.edu/~mihalcea/senseval/
  - https://www.cs.york.ac.uk/semeval-2013/task11/

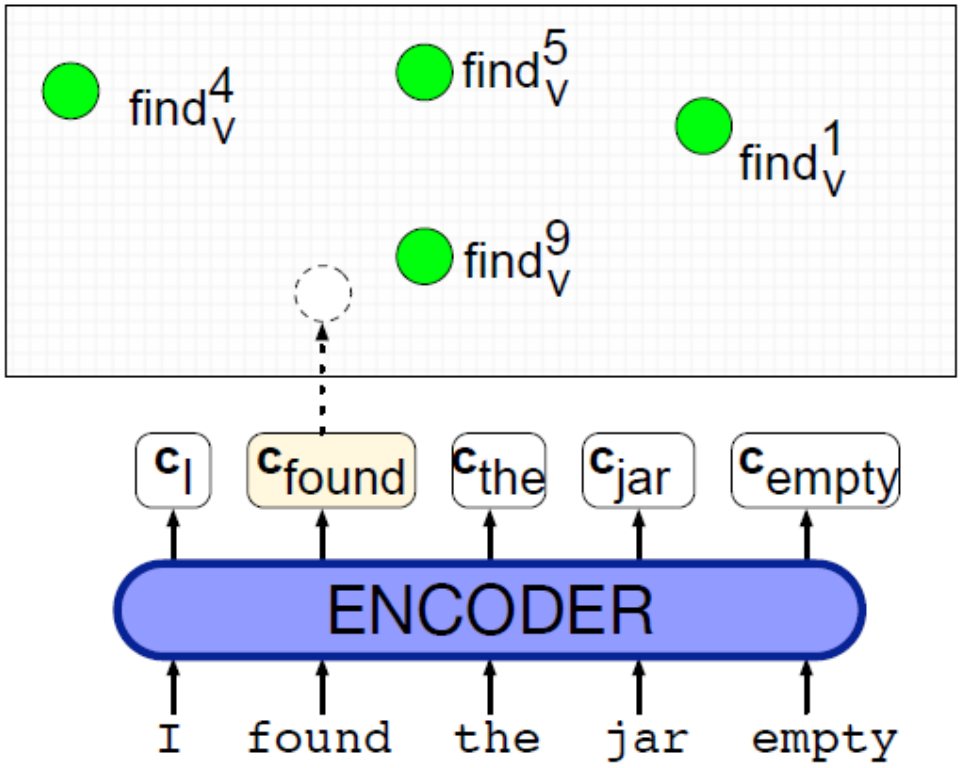
# Word Sense Disambiguation

- Sequence classification problem
  - X = sentence = sequence of words
  - Y = word sense labels = sequence of tags (WordNet sense annotations)



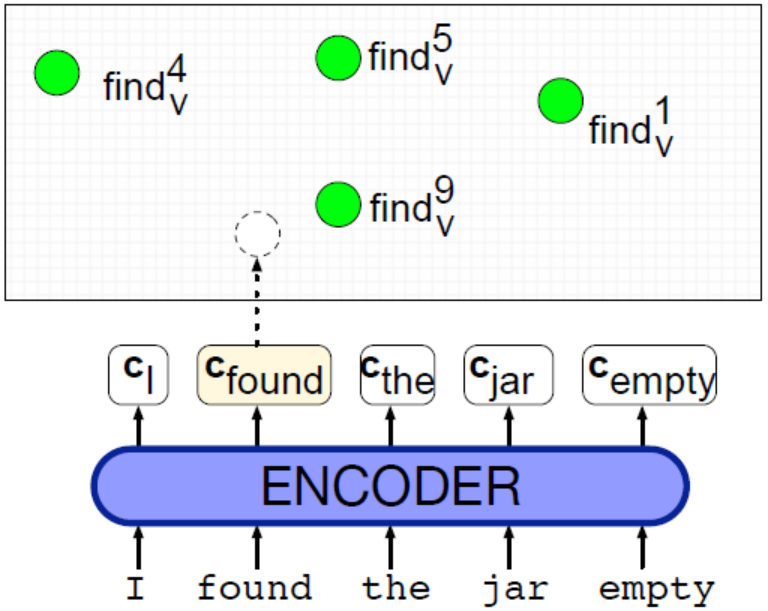
# Word Sense Disambiguation

- Baselines
  - Statistically most frequent sense in corpus
  - One sense per discourse (remember previous sense and use that)
- WSD classifier using contextual embeddings



# Word Sense Disambiguation

- Baselines
  - Statistically most frequent sense in corpus
  - One sense per discourse (remember previous sense and use that)
- WSD classifier using contextual embeddings
  - Encoder = BERT embeddings (pre-trained, embedding weights fixed)
  - Training phase
    >> average (across corpus) each token_embedding for a particular word sense to get a sense_embedding
  - Testing phase
    >> nearest neighbour classifier
    >> cosine distance between input token_embedding and learnt sense_embeddings



# Word Sense Disambiguation

- What about senses not in corpus?
  - There are many more senses in WordNet than is in most training corpora
  - option 1 = statistically most frequent sense in corpus (not good)
  - option 2 = impute sense using WordNet taxonomy (averaging sense_embeddings of children to compute embeddings for hypernyms)
  - Optional further reading
    - Loureiro, D. and Jorge, A. Language modelling makes sense: Propagating representations through WordNet for full-coverage word sense disambiguation. ACL 2019

# Required Reading

- Dependency Parsing
  - Jurafsky and Martin, Speech and Language Processing, 3rd edition (online) >> chapter 18

# Questions

- Panopto Quiz - 1 minute brainstorm for interactive questions

  Please write down in Panopto quiz in **1 minute** two or three questions that you would like to have answered at the next interactive session.

  Do it **right now** while its fresh.

  Take a screen shot of your questions and **bring them with you** at the interactive session so you have something to ask.