Computational Semantics

Walter Daelemans Walter.Daelemans@uantwerpen.be

Guy De Pauw Guy.Depauw@uantwerpen.be

Computational Linguistics

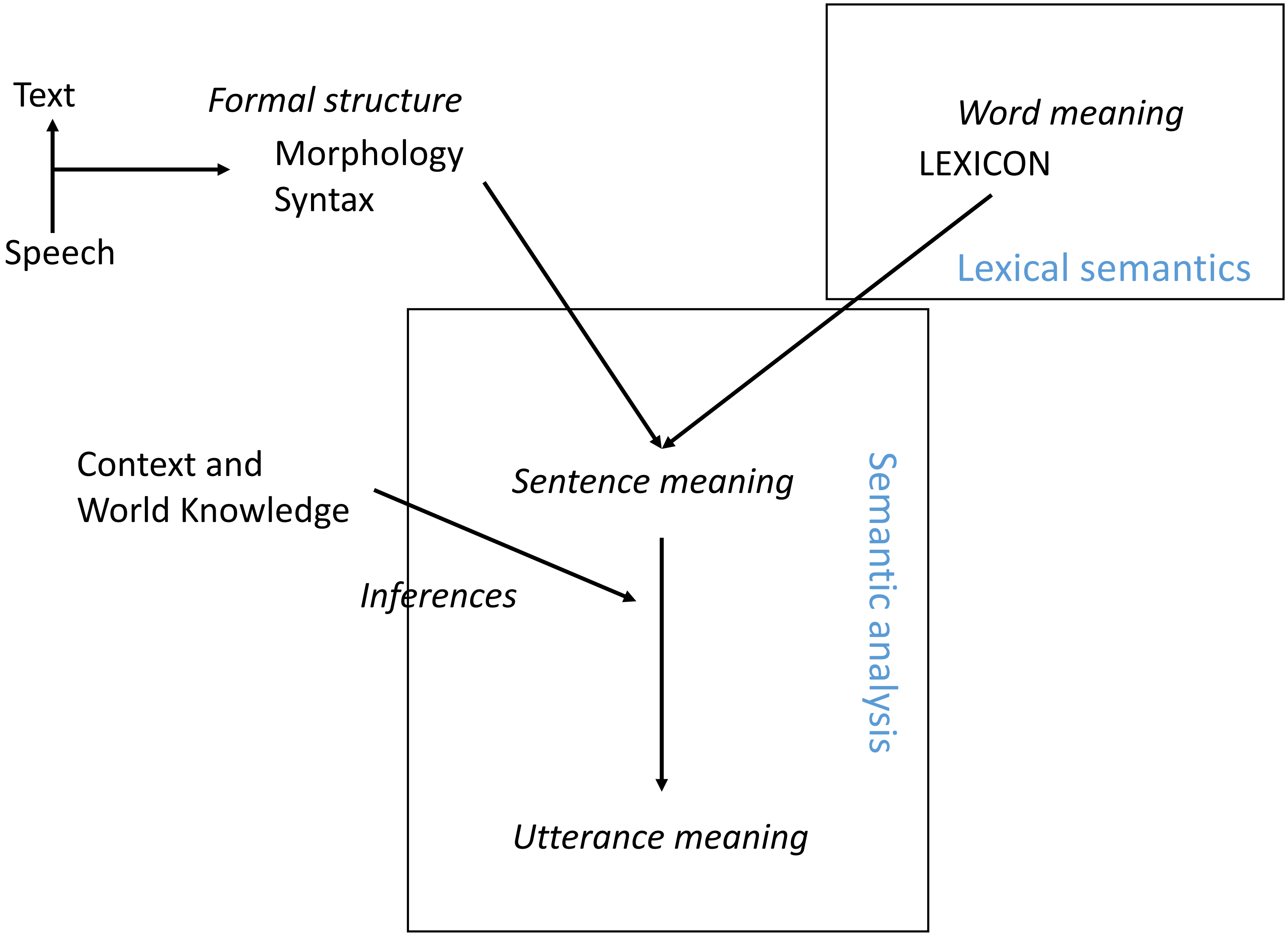
2016-2017

Computational Semantics

- Written exam material:
  - Slides
  - Jurafsky & Martin:
    - Chapters 17 and 18 (optional)
    - Chapter 19.1->19.4; Chapter 20.1->20.7
- Assignment:
  - Deadline = December 21
  - See slides 73-74

Program

- Computational Semantics overview
- Introduction to lexical semantics
  - Senses, relations between senses
- Word Sense Disambiguation (WSD) and word similarity


Text
Speech
Formal structure
Morphology
Syntax
Word meaning
LEXICON
Lexical semantics
Context and
World Knowledge
Sentence meaning
Inferences
Semantic analysis
Utterance meaning

Representations and Transformations

- Meaning representations
  - Logic (FOPC)
  - Graphs
    - Semantic Networks
    - Conceptual Dependency
    - Frames
- Meaning transformations
  - sentence semantics (from syntax and word meaning to sentence meaning)
  - discourse semantics (from sentence meaning and world knowledge to discourse meaning)

Criteria for semantic representation

- Verifiable
  - linguistic meaning and world knowledge should use compatible representations in order to find matches
- Unambiguous
- Canonical
  - same meaning, same representation (paraphrases)
- Support *inference*
  - from knowledge + sentence meaning, more meaning can be inferred
- Expressive
  - be able to represent any utterance meaning
    - language of thought?

What can a meaning representation look like?

- Must represent objects, properties of objects, relations between objects, … in the world
- Model-theoretic semantics
  - Domain (set of objects in the ‘world’)
  - Property (set of objects sharing that property)
  - Object (element of the domain)
  - relation (set of tuples of elements in the domain)

- Domain = {john, paul, george, ringo, bach, mozart, trump, clinton, bass, guitar, drums, …}
- Properties
  - beatle = {john, paul, george, ringo}
  - musician = {john, paul, george, ringo, johann_sebastian, wolfgang}
- Relations
  - plays = {<john, guitar>, <paul, bass>, <george, guitar>, <ringo, drums>, <wolfgang, harpsichord>, …}
- Interpretation in a model
  - All Beatles are musicians
  - $\forall X\ (beatle(X) \rightarrow musician(X))$
  - beatle $\subset$ musician

Predicate Logic

There is no woman that doesn't like chocolate

$\neg \exists X (\text{woman}(X) \wedge \neg \text{likes}(X, \text{chocolate}))$

- statements
    - predicates (n-ary) and arguments
- properties
    - one place predicates
- or, and, if … then …, not
- for all, there exists, …
- Problems:
    - uncertainty, defaults, beliefs, time, change
    - efficiency

Conceptual Dependency (Schank)

- Interlingua
- Semantic primitives
- All information explicit

John gave the book to Mary.

John loaned the book to Mary.

What is meaning ?

- In AI: symbols all the way down
  - Language
  - Syntactic analysis (nouns, verbs, adjectives, sentences, …)
  - Semantic analysis (predicates, constants, propositions, …)
  - Mathematical basis (set-theoretic operations, functions)
  - ???
- Symbol *grounding* problem
- Meaning is in the mind of the beholder of the representations, not in the representations themselves

from syntax and word meaning to sentence meaning

- Syntax driven
  - Compositionality: the meaning of the whole is a function of the meaning of the parts
  - Problems: idioms
- Semantic grammars

Syntax

S $\rightarrow$ *NP VP*

VP $\rightarrow$ *V NP*

NP $\rightarrow$ *Art N*

Semantics

S: Apply (VP, NP)

VP: Apply (V, NP)

*NP: Apply (lambda (x) (DEF/SING * x), N)*

Lexicon

kissed	*lambda(o)lambda(x)* *(past * kiss [agent x][theme o])*
man	*MALE-HUMAN*
dog	*DOG*

Syntax

S → NP VP

VP → V NP

NP → Art N

Semantics

S: Apply (VP, NP)

VP: Apply (V, NP)

NP: Apply (lambda (x) (DEF/SING * x), N)

Lexicon

kissed	lambda(o)lambda(x) (past * kiss [agent x][theme o])
man	MALE-HUMAN
dog	DOG

The man kissed the dog

S -- NP VP
Apply (VP, NP)

VP -- V NP
Apply (V, NP)

Apply
lambda(o)lambda(x) (past * kiss [agent x][theme o])
Apply lambda(x)(def/sing * x), N

Apply
lambda(o)lambda(x) (past * kiss [agent x][theme o])
(def/sing MALE-HUMAN)

Apply
lambda(o) (past * kiss [agent (def/sing MALE-HUMAN)][theme o])
Apply lambda(x)(def/sing * x), N

Apply
lambda(o) (past * kiss [agent (def/sing MALE-HUMAN)][theme o])
(def/sing DOG)

(past * kiss [agent (def/sing MALE-HUMAN)][theme (def/sing DOG)])

Semantic Grammars

- RES-VP $\rightarrow$ RESERVING RES-MOD
- RES-VP $\rightarrow$ RESERVING
- DEP-VP $\rightarrow$ DEPARTING DEP-MODS
- RESERVING $\rightarrow$ RESERVE-VERB flight
- RES-MOD $\rightarrow$ for PERSON
- DEPARTING $\rightarrow$ DEPART-VERB
- DEPARTING $\rightarrow$ DEPART-VERB SOURCE-LOCATION
- DEP-MODS $\rightarrow$ DEP-MOD DEP-MODS
- DEP-MODS $\rightarrow$ DEP-MOD
- DEP-MOD $\rightarrow$ to DEST-LOCATION
- DEP-MOD $\rightarrow$ from SOURCE-LOCATION

Book a flight for me from Boston to Chicago

Lexical Semantics

Word senses

- Wordforms and lemmas (citation forms)
  - appeltjes appel
  - lopen lopen (V)
  - lopen loop (N)
- Lemmatization
- Lemmas have lexical meaning
- One lemma can have many different (word) senses
  - Discrete representation of aspects of a lemma's meaning
- Senses, rather than words, are important in NLP systems:
  - Machine translation: bank ➔ "bank" or "oever"
  - Text categorization: python ➔ snake or programming language
  - Text to speech: bass ➔ music or fishing

Distinguishing senses

- Word can have many senses, see WordNet:
  - Bank as noun: 10 senses
  - Bank as verb: 8 senses
- Sometimes subtle differences:
  - Bank: sloping land
  - Bank: a slope in the turn of a road or track
- Rule of thumb:
  - Different truth conditions, syntactic behavior
  - Zeugma
    - ?Lufthansa **serves** breakfast and New York
    - ?John plays and eats **bass**

Relations between senses

- **Homonymy** in case of same form but unrelated meaning
  - bank^1 : financial institution
  - bank^2: sloping mound
- **Polysemy** if there is a semantic relation: bank^1 and bank^3
  - bank^3: biological repository
- **Metonymy**: systematic polysemy
  - E.g. Building – Organization, Author – Work of the author, Animal – Meat
    - *Fortis is around the corner – I will never buy Fortis shares again*
    - *Jane Austen is on the top shelf – I hate Jane Austen*
    - *The chickens was domesticated early – Chicken is considered a vegetable in South-Africa*

Relations between **senses**

- Synonymy (synonyms mean the same in all contexts, same propositional meaning, same truth conditions): *couch/sofa*
- Perfect synonymy is rare
  - big car, large car
  - big sister, large sister ?
- Synonymy is a relation between senses rather than between words

Relations between senses

- **Antonymy**:
  - Different ends of a scale: *long/short*; *dark/light*
  - Reversives: *up/down*
- **Hyponymy**: *car/vehicle* (*x is subordinate, hyponym of y*) (*y is superordinate,* ***hypernym*** *of x*)
  - Hyponymy is mostly associative
    - Grape is hyponym of Fruit, Fruit is hyponym of Edible Things
    - Grape is hyponym of Edible Things
  - Classes and instances
    - Relation between instance and class versus relation between classes
    - ISA-hierarchy, AKO-hierarchy
    - Antwerp ISA city, city AKO location

Relations between senses

- **Meronymy**: *wheel/car* (*x is-part-of y*) (*y is **holonym** of x*)
- Semantic field (domain)
  - *Reservation, flight, travel, buy, price, cost, fare, rates, plane*

Structured lexical resources

- **Dictionaries** available in machine-readable form
  - Contain list of senses, definitions for all senses, typical usage examples for most senses
  - E.g. Oxford English Dictionary, Collins, Longman Dictionary of Ordinary Contemporary English
- **Thesaurus**
  - Contains explicit information about semantic relations between word senses http://www.thesaurus.com/
  - E.g. Roget's Thesaurus
    - https://archive.org/details/rogetsthesauruso10681gut
- **Lexical database**
  - Contains relations between senses, definitions, etc.
  - E.g. WordNet, EuroWordNet
  - Dutch: Open Dutch Wordnet http://wordpress.let.vupr.nl/odwn/
- **Integrated resources** http://babelnet.org/about

WordNet (George A. Miller)

- Combination of dictionary, thesaurus & semantic network
- Database of lexical relations in 3 parts: nouns, verbs, adjectives & adverbs
- Word senses
  - Sense: gloss, synset (= set of near-synonyms)
- Downloadable resource
- Web interface
  - http://wordnetweb.princeton.edu/perl/webwn

Word to search for: be [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) beryllium, **Be**, glucinium, atomic number 4 (a light strong brittle grey toxic bivalent metallic element)

Verb

- S: (v) **be** (have the quality of being; (copula, used with an adjective or a predicate noun)) *"John is rich"; "This is not a good answer"*
- S: (v) **be** (be identical to; be someone or something) *"The president of the company is John Smith"; "This is my house"*
- S: (v) **be** (occupy a certain position or area; be somewhere) *"Where is my umbrella?"; "The toolshed is in the back"; "What is behind this behavior?"*
- S: (v) exist, **be** (have an existence, be extant) *"Is there a God?"*
- S: (v) **be** (happen, occur, take place) *"I lost my wallet; this was during the visit to my parents' house"; "There were two hundred people at his funeral"; "There was a lot of noise in the kitchen"*
- S: (v) equal, **be** (be identical or equivalent to) *"One dollar equals 1,000 rubles these days!"*
- S: (v) constitute, represent, make up, comprise, **be** (form or compose) *"This*

NLTK Wordnet

http://www.nltk.org/howto/wordnet.html

NLTK book, Chapter 2

```
>>> from nltk.corpus import wordnet as wn
```

Which synsets does a word have (of particular POS: VERB, NOUN, ADJ, ADV)

```
>>> wn.synsets('dog')
[Synset('dog.n.01'), Synset('frump.n.01'),
Synset('dog.n.03'), Synset('cad.n.01'),
Synset('frank.n.02'), Synset('pawl.n.01'),
Synset('andiron.n.01'), Synset('chase.v.01')]
>>> wn.synsets('dog', pos=wn.VERB)
[Synset('chase.v.01')]
```

Properties of Synsets

```
>>> dog1 = wn.synset('dog.n.01')

>>> dog1.definition()

'a member of the genus Canis (probably descended from
the common wolf) that has been domesticated by man since
prehistoric times; occurs in many breeds'

>>> dog1.hypernyms()

[Synset('domestic_animal.n.01'), Synset('canine.n.02')]

>>> dog1.hyponyms()

[Synset('puppy.n.01'), Synset('great_pyrenees.n.01'),
Synset('basenji.n.01'), Synset('newfoundland.n.01'),
Synset('lapdog.n.01'), Synset('poodle.n.01'),
Synset('leonberg.n.01'), Synset('toy_dog.n.01'),
Synset('spitz.n.01'), Synset('pooch.n.01'),
Synset('cur.n.01'), Synset('mexican_hairless.n.01'),
Synset('hunting_dog.n.01'), Synset('working_dog.n.01'),
Synset('dalmatian.n.02'), Synset('pug.n.01'),
Synset('corgi.n.01'), Synset('griffon.n.02')]
```

```
>>> dog1.lemmas()
[Lemma('dog.n.01.dog'), Lemma('dog.n.01.domestic_dog'),
Lemma('dog.n.01.Canis_familiaris')]
>>> dog1.examples()
['the dog barked all night']
>>> for synset in wn.synsets('dog'):
        print(synset.lemma_names())
```

Comparisons

```
>>> cat1 = wn.synset('cat.n.01')
>>> dog1.lowest_common_hypernyms(cat1)
[Synset('carnivore.n.01')]
>>> dog1.wup_similarity(cat1)
0.8571428571428571
# This is Wu and Palmer similarity (look it up!)
```

Semantics of events (event participants)

- John breaks the window

  BREAKER BROKEN THING

- Semantic roles: deep roles and thematic roles
  - Deep: breakers, eaters, openers, givers, …
  - Thematic: agent, instrument, beneficiary, theme, …
- Selectional restrictions: only things that can be broken can be a BROKEN THING
  - Use wordnet hierarchies

Thematic roles

- Deep roles:

$\exists$e,x,y Breaking(e) ∧ Breaker(e,John) ∧ BrokenThing(e,y) ∧ Window(y)

- Thematic roles:

AGENT:John, THEME:Window

- Syntax helps in finding thematic roles
- Problems: Fragmentation, hard to define
  - PropBank: proto-roles & verb-specific semantic roles
  - FrameNet: frame-specific semantic roles

Word Sense Disambiguation

Extreme cases of ambiguity

Drunk Gets Nine Years In Violin *Case*

Farmer *Bill* Dies In House

Prostitutes *Appeal To* Pope

Stolen Painting Found *By* Tree

Red Tape *Holds Up* New Bridge

Include Children When Baking Cookies

Miners Refuse To Work *After Death*

Problems & solutions

Drunk Gets Nine Years In Violin Case

Farmer Bill Dies In House

Prostitutes Appeal To Pope

Stolen Painting Found By Tree

Red Tape Holds Up New Bridge

Include Children When Baking Cookies

Miners Refuse To Work After Death

Lexical, syntactic, referential ambiguity

World Knowledge
Fixed Expressions

Word Sense Ambiguity

- Most of the time no problem for humans, except in some extreme cases
- Computers need help to disambiguate even the ‘simplest’ of cases

Computationally explosive problem

I saw a man who is 98 years old and can still walk and tell jokes

26 senses

11 senses

4 senses

8 senses

5 senses

4 senses

10 senses

8 senses

3 senses

43,929,600 senses

How big is the problem?

- Most words in English have only one sense
  - 62% in Longmans Dictionary of Contemporary English (LDOCE)
  - 79% in WordNet
- Average number of senses per word
  - 3.83 in LDOCE vs. 2.96 in WordNet
- But … ambiguous words are used more frequently!
  - BNC (British National Corpus): 84% of words (tokens) have more than one sense
  - Some senses are more frequent than others

Word Sense Disambiguation (WSD)

- = automatically identify the intended sense of a word *in context*
- Assumes a fixed inventory of senses that you can select the right one from
- Can be seen as a categorization task (cf. POS-tagging)
  - Senses = classes
  - Context = features

Relevance

- Important aspect of many NLP applications
- Relevant for all languages
- Needed in
  - *Machine translation*: select the right sense to translate
  - *Information retrieval*: resolve ambiguity in query
  - *Information extraction*: accurate analysis of text

Upper bound and baseline

- Human performance as an upper bound
  - Fine-grained sense inventories: 75-80% human agreement
  - Coarser-grained inventories: 90% human agreement possible
- Predict the most frequent sense in a given lexical resource (‘MFS baseline’)
  - *bank* 97.20%
  - *bar* 47.38%

Evaluation of WSD

- Internal: measure accuracy of sense selection compared to gold standard
- External: integrate WSD in MT or IR system and evaluate
- Test data
  - Lexical sample: the occurrences of a small sample of target words need to be disambiguated
  - All-words: all words in running text need to be disambiguated
  - Example: Senseval en Semeval competitions
    http://www.senseval.org
    https://en.wikipedia.org/wiki/SemEval

Development of research in WSD

- Noted as problem for Machine Translation (Weaver, 1949)
- Bar-Hillel (1960) declared it unsolvable, left the field of MT
  - “The box is in the pen.” vs. “The pen is in the box.”
- 1970s-80s
  - Rule-based approaches
- 1990s
  - Corpus-based approaches
  - Dependence on sense-tagged training texts
- From 2000s
  - Hybrid Systems
  - Unsupervised learning -> deep neural nets
  - Taking advantage of the Web

Approaches to WSD

- Knowledge-based : External lexical resources
- Supervised : Labeled training data
  - Semi-supervised
- Unsupervised : Large collections of raw text

Knowledge-based approaches

WSD from sense definitions

- LESK algorithm (Lesk, 1986)
  - Retrieve from dictionary all sense definitions of the word to be disambiguated
  - Determine the overlap (in content words) between each sense definition and definitions of words in the current context
  - Choose the sense that leads to highest overlap

LESK algorithm example

e.g. Pine cones hanging in a tree

$Pine^1$	kind of evergreen tree with needle-shaped leaves
$Pine^2$	waste away through sorrow or illness
$Cone^1$	solid body which narrows to a point
$Cone^2$	something of this shape whether solid or hollow
$Cone^3$	fruit of certain evergreen trees

Pine 1 $\cap$ Cone 1 = 0
Pine 2 $\cap$ Cone 1 = 0
Pine 1 $\cap$ Cone 2 = 0
Pine 2 $\cap$ Cone 2 = 0
Pine 1 $\cap$ Cone 3 = 2
Pine 2 $\cap$ Cone 3 = 0

Problems with LESK algorithm

- Problems
  - Very sensitive to the exact wording of definitions: absence of a particular word can radically change the results
  - Dictionary glosses tend to be fairly short; often not sufficient vocabulary to relate fine-grained sense distinctions

LESK variants

- Simplified LESK
  - Retrieve all sense definitions of target word
  - Compare with *context* instead of sense definitions of the context
  - e.g. <u>Pine</u> cones hanging in a tree

Pine 1 $\cap$ Sentence = 1
Pine 2 $\cap$ Sentence = 0

- Corpus-based LESK
  - Add context words from sense tagged corpus to definitions
  - Weight words by inverse document frequency (IDF)
  - Gloss is the document
  - $\mathrm{IDF}(w) = \log(|D|/|D_w|)$ (function words have low IDF)
  - Best-performing LESK variant, was used as a baseline in Senseval competitions

Supervised approaches

Supervised learning

- Last 20 years: shift from manually crafted systems to automated classification methods
- Basic steps
  - Collect a set of examples that illustrate the various possible senses of a word (annotated corpus)
  - Extract predictive features (context words, collocations, pos tags, …); senses are output classes
  - Machine Learning method identifies patterns in the examples (between input features and output classes) and creates a model representing these patterns
  - The resulting model can be applied to new data

Supervised WSD

- Resources
  - Sense-tagged text (*unstructured*)
  - Dictionaries, thesauri, semantic networks (*structured*)
  - Syntactic Analysis (POS tagger, chunker, parser, etc.)
- WSD as a classification problem
  - target word is assigned the most appropriate sense
  - from a given set of possibilities
  - based on the context in which it occurs
  - = word expert approach

Sense-tagged corpora

- Word sense disambiguation resources page of ACL
  - https://goo.gl/N352ps

Sense-tagged corpora

e.g.

Bonnie and Clyde are two really famous criminals, I think they were bank/1 robbers.

My bank/1 charges too much for an overdraft.

I went to the bank/1 to deposit my check and get a new ATM card.

The University of Minnesota has an East and a West Bank/2 campus right on the Mississippi River.

My grandfather planted his pole in the bank/2 and got a great big catfish!

The bank/2 is pretty muddy, I can’t walk there.

Simple supervised system

- Extract bags of words from sense-tagged text
  - #1 financial-bank

    a an and are ATM Bonnie card charges check Clyde criminals deposit famous for get I much My new overdraft really robbers the they think to too two went were
  - #2 river-bank

    a an and big campus cant catfish East got grandfather great has his I in is Minnesota Mississippi muddy My of on planted pole pretty right River The the there University walk West

Simple supervised system

Given a sentence S containing <u>bank</u>

For each word W_i in S

If W_i is in *financial-bank* then Sense#1 = Sense#1 + 1

If W_i is in *river-bank* then Sense#2 = Sense#2 + 1

If Sense 2 > Sense 1 then decide "River"

else decide "Financial" (majority sense)

Supervised methodology

- Features extracted for WSD
  - local features: local context of word usage (e.g. POS, lemma, etc.)
  - topical features: general topic of a text or discourse
  - syntactic features: syntactic cues and argument-head relations between the target word and other words
  - semantic features: representing semantic information, e.g. previously established senses of words in context, domain indicators, etc.

Representing context

- Using these features, convert each word occurrence into a feature vector

My father used to fish along the banks/SHORE of the river

The bank/FINANCE issued a check for the amount of interest

P-1	P+1	P+2	Fish	Check	River	Interest	SENSE TAG
det	prep	det	1	0	1	0	SHORE
det	verb	det	0	1	0	1	FINANCE

Representing context

- Which context words are taken into account?
  - No function words
  - Only words that are in a specific grammatical relation
  - Size of the window
- How are they represented in the vector?
  - Binary: present/not present
  - Continuous: Relative frequency, mutual information
- How is the similarity between vectors used in Machine Learning?

Supervised methodology

- Use any supervised learning algorithm
  - Lazy learners
    - e.g. k-Nearest Neighbor Classifiers
  - Eager learners
    - e.g. Support Vector Machines, Decision Trees, naïve Bayes, neural nets
- Training data to train and validate the machine learner
- Procedure: *n*-fold cross-validation
- Hold-out test data to test the resulting classifier

Shortcomings

- Supervised approaches to WSD achieve best results, but
  - heavily rely on large sense-tagged corpora
  - fixed sense inventory: often arbitrary divisions of word meanings into dictionary senses
  - low inter-annotator agreements on sense tagging
- WSD should be integrated in real applications such as MT or multilingual IR (extrinsic evaluation)

Unsupervised and semi-supervised approaches

Motivation

- Supervised yields highest performance, but…
    - Limited to words whose senses are tagged
    - *Corpus Annotation Bottleneck*
- Solutions:
    - pseudo-ambiguity
        - take any two words, e.g. guitar and mouse
        - replace each instance in a text of guitar or mouse with guitar-mouse
        - train and test on this new corpus
        - (does not really work like real ambiguity)
    - unannotated corpora instead of sense-tagged text or lexical resources

Unsupervised / minimally supervised

- = learning sense classifiers from unannotated data, with minimal or no human supervision
- Examples
  - Sense clustering
  - Automatically bootstrap a corpus starting with *a few human annotated examples*
  - Cross-lingual evidence
  - Use Wikipedia as sense-tagged text

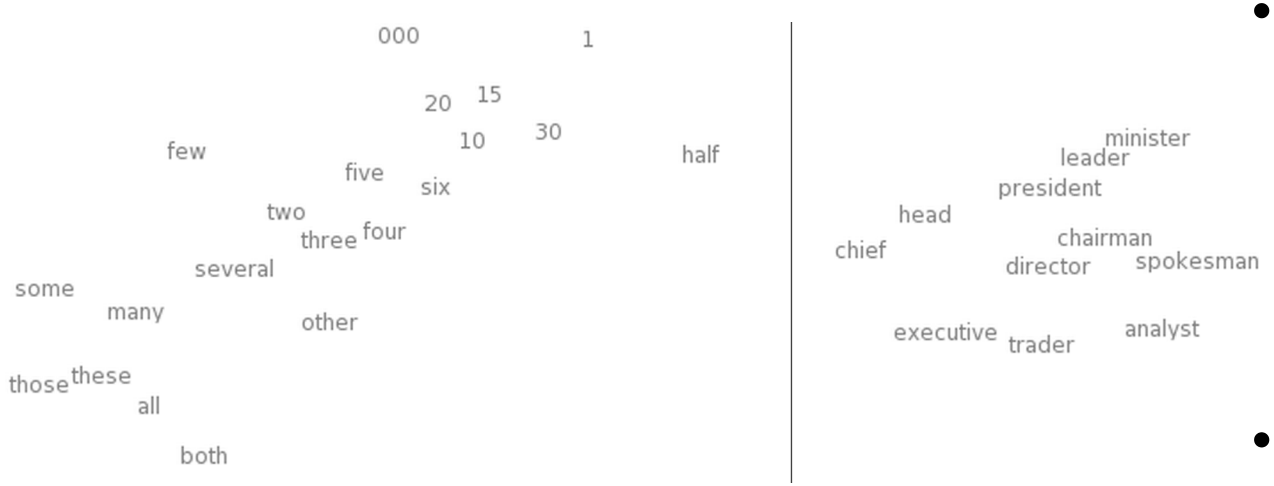
Sense clustering

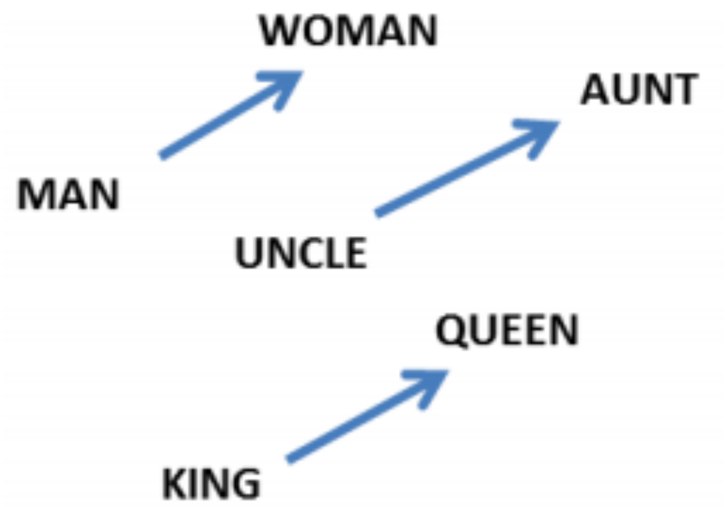
- Also *word sense induction/discrimination*
- Cluster words on similarity of context (using distributions and similarity metrics)
- Hypothesis:
  - Words with similar meanings tend to occur in similar contexts (Miller and Charles, 1991)
  - Cf. ‘You shall know a word by the company it keeps’ (Firth, 1957)

Distributional similarity

- Given a large corpus
- Construct for each word a vector of occurrences, in its immediate context, of other words
- Two words are semantically similar if they have similar vectors
- Parameters
  - definition of relevant immediate context
    - only content words, only words in a syntactic relation, …
  - weighting of the context terms (binary, frequency, …)
  - distance metric (cosine, Euclidean, …)

Word Embeddings





- Training a backprop neural network to predict target words given contexts or contexts given target words
- Weights to the hidden layer are the target word’s embedding (vector)
- Fast approximations:
  - word2vec
  - glove

Bootstrapping

- Build sense classifiers with little training data
- Components
  - (Some) labelled data
  - (Large amounts of) unlabelled data
  - (One or more) basic supervised learning classifiers
- Output
  - Classifier that improves over the basic classifiers

Bootstrapping algorithm

(Yarowsky, 1995)

- Bootstrapping algorithm
  - Start from small seed set of hand-labeled data Λ_0
  - Learn classifier from Λ_0
  - Use learned classifier to label unlabeled data V_0
  - Move high-confidence examples in V_0 to Λ_1
  - Repeat until classifier no longer confident about output
- 2 heuristics to automatically select Λ_0
  - One sense per collocation: bass/fish & bass/play
  - One sense per discourse: within a text or discourse, you will find either bass/fish or bass/play, not both

WSD using cross-lingual evidence

- Corpus-based approach: using translations from a parallel corpus instead of human-defined sense labels
- Advantages
  - easier to integrate in real applications
  - implicitly deals with granularity problem
  - language-independent approach
- Hypothesis: different sense distinctions are often lexicalized across languages

Cross-lingual WSD

- living on the bank of the river

Dutch	oever/dijk
French	rives/rivage/bord/bords
German	Ufer
Italian	riva
Spanish	orilla

- money supply is of direct interest to any bank

Dutch	bank/kredietinstelling
French	banque/établissement de crédit
German	Bank/Kreditinstitut
Italian	banca
Spanish	banco

Wikipedia as sense-tagged corpus

(Mihalcea, 2007)

In 1834, Sumner was admitted to the [[bar (law)|bar]] at the age of twenty three, and entered private practice in Boston.

It is danced in 3/4 time (like most waltzes), with the couple turning approx. 180 degrees every [[bar (music)|bar]].

- For most investigated words, performance using Wikipedia improves over MFS and LESK baselines
- Advantages
  - Size of Wikipedia is growing
  - Wikipedia is available for about 200 different languages

Word sense	Labels in Wikipedia	Wikipedia de□nition	WordNet de□nition
bar (establishment)	bar_(establishment), nightclub gay_club, pub	a retail establishment which serves alcoholic beverages	a room or establishment where alcoholic drinks are served over a counter
bar (counter)	bar_(counter)	the counter from which drinks are dispensed	a counter where you can obtain food or drink
bar (unit)	bar_(unit)	a scienti□c unit of pressure	a unit of pressure equal to a million dynes per square centimeter
bar (music)	bar_(music), measure_music musical_notation	a period of music	musical notation for a repeating pattern of musical beats
bar (law)	bar_association, bar_law law_society_of_upper_canada state_bar_of_california	the community of persons engaged in the practice of law	the body of individuals quali□ed to practice law in a particular jurisdiction
bar (landform)	bar_(landform)	a type of beach behind which lies a lagoon	a submerged (or partly submerged) ridge in a river or along a shore
bar (metal)	bar_metal, pole_(object)	-	a rigid piece of metal or wood
bar (sports)	gymnastics_uneven_bars, handle_bar	-	a horizontal rod that serves as a support for gymnasts as they perform exercises
bar (solid)	candy_bar, chocolate_bar	-	a block of solid substance

Table 1: Word senses for the word *bar*, based on annotation labels used in Wikipedia

Taken from Mihalcea 2007 'Using Wikipedia for Automatic Word Sense Disambiguation'

Conclusions

- Introduction to lexical semantics & WSD
  - How to measure semantic similarity of words
  - How to disambiguate semantically ambiguous words (find the contextually correct sense)
    - Knowledge-based approaches
    - Supervised approaches
    - Semi-supervised/unsupervised approaches
      - clustering
      - bootstrapping
      - using multilingual data
      - using encyclopedic data (wikipedia)

Assignment 1

- Do research on different methods and approaches for computing word similarity using taxonomies such as Wordnet
  - NLTK book chapter 2 is a good place to start for methods implemented in NLTK
  - Chapter 20.6 in J&M is also a good source
- For each method:
  - Describe concisely how it works
  - If not yet present in NLTK, implement it in Python (1 extra method only)
  - Compare them by computing similarity using all available methods between the following word pairs, and try to explain the differences:
    - analogy - simulation
    - dog - cat
    - good - conscious
    - drink - kiss
    - book - bible

Assignment 2

- Implement *simplified Lesk* WSD method in Python, using the NLTK Wordnet module
- Test the code for at least 5 ambiguous words (nouns or verbs), with for each word 10 real sentences (i.e. not invented by you), selected from the web or from existing corpora