Language and Inference

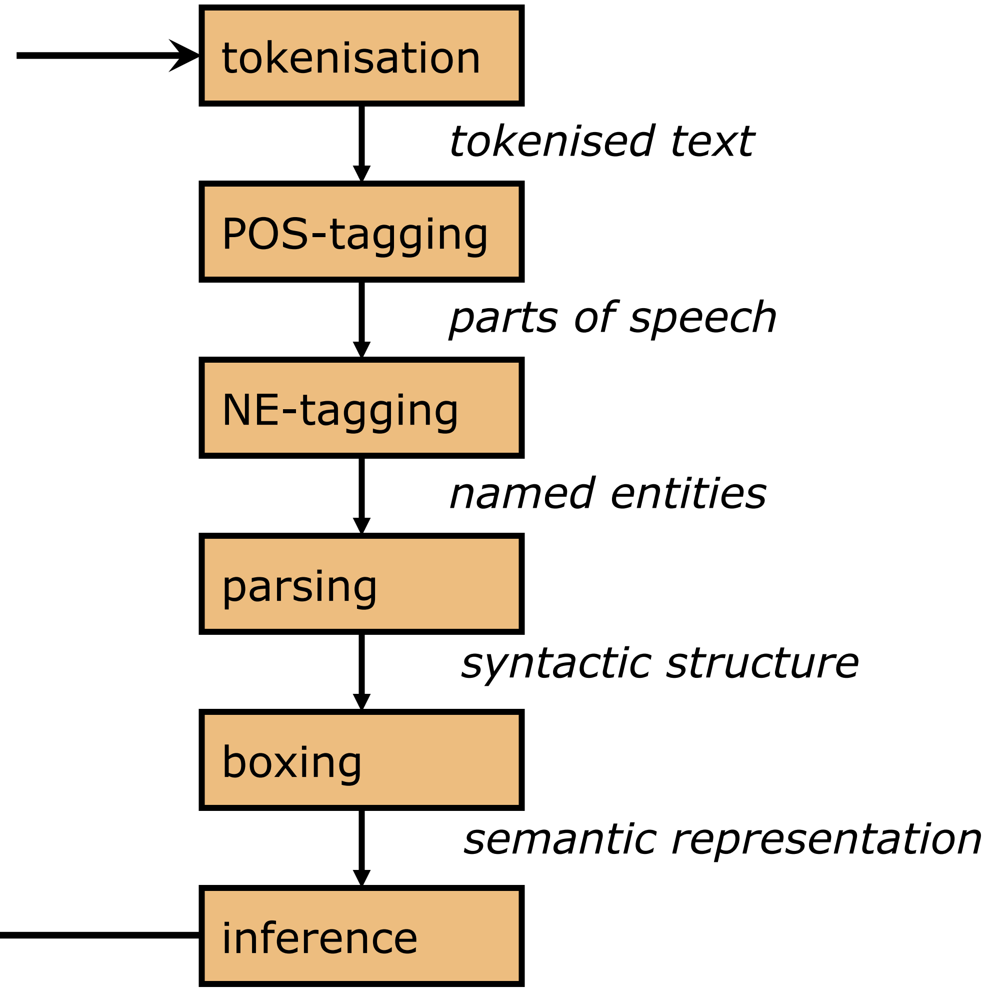
Day 5: Inference in the Real World

Johan Bos
johan.bos@rug.nl

Semantic Analysis Pipeline

WHO COULD BUY CHICAGO

FOUR OF THE RICHEST WOMEN IN THE WORLD.



Low-level formatting issues

- Document headers, tables, diagrams
- Filter required to remove junk
- Errors caused by OCR
  (optical character recognition)

Capitalisation

- Should we treat tokens that are identical disregarding lower- and uppercase as the same?
- Simple heuristics do not exist
  - Change an uppercase word at beginning of sentence into lowercase
  - Assume that all other uppercase words are names

EXAMPLE

the, The, THE
Meg White, a white swan

Segmentation

- Divide an input text into units called <u>tokens</u>
- Distinguish sentence tokens and word tokens
- Usually first step in a NLP pipeline

- Two boundary detection tasks
  - detect boundaries of word tokens
    (separating punctuation symbols from words)
  - detect boundaries of sentence tokens
    (syntactic analysis wants sentences as input)

Punctuation symbols

- Punctuation symbols can be important
- Don’t throw them away!

EXAMPLE

The camel, who crossed Australia, was thirsty.
The camel who crossed Australia was thirsty.

What is a word?

- Even linguists don't have a clear answer!
- An attempt:

A sequence of alphanumeric characters with space on either side, including hyphens and apostrophes

EXAMPLES

$14,00	Micro$oft	:-)	John's	s'pose

Full stops

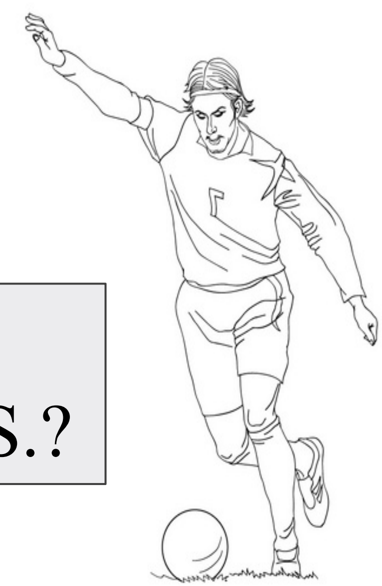
- It looks simple to remove punctuation symbols from word tokens
- But it is problematic for full stops (period)
  - Most full stops indicate the end of a sentence
  - But some full stops mark an <u>**abbreviation**</u>
- Arguably, the full stop of an abbreviation should be part of the word

EXAMPLE

Jack White lives in San Francisco, Calif., where he…

Haplography

- An abbreviation ends with a full stop
- A sentence ends with a full stop
- Therefore, a sentence ending with an abbreviation ends with two full stops (hmm, usually not!)



EXAMPLE

David Beckham played soccer in the U.S.
Did he make an impact on soccer in the U.S.?

Contractions

- Are English contractions such as “I’d” and “aren’t” one or two word tokens?
- Not splitting them puts pressure on the grammar
- Splitting them produces funny words: n’t, ’s, ’d, …
- Note: possible difference in meaning
  - I mustn’t grumble.   (negation outscopes modal)
  - I must not grumble.   (modal outscopes negation)

Clitics

- The dog’s walking away.
- The dog’s tail was wagging too much.
- The scary dog’s tail was wagging too much.
- The dog’s owner shouted.
- The dogs’ owner shouted.

Hyphenation

- Do sequence of letters with a hyphen count as one word or two?
- Line-breaks further complicate things

EXAMPLES

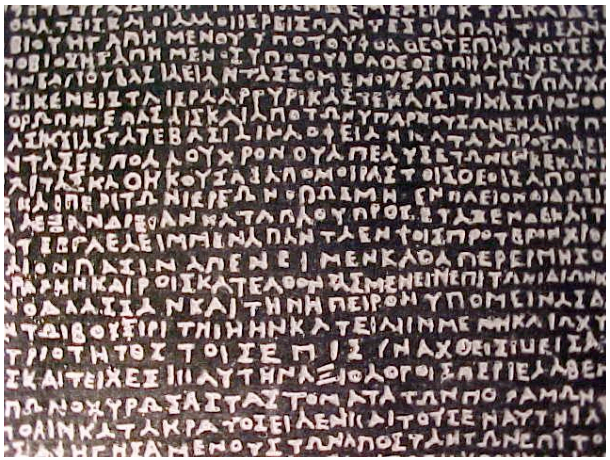
e-mail so-called non-commercial co-operate

the 16-year-old boy—who surprised his friends

the San Francisco-based company

Words in other languages

- Ancient Greek was written without spaces



- Hottentottententententoonstelling (Dutch)
  An exhibition ("tentoonstelling") of tents of the Khoikhoi

What is a sentence?

- Something ending with a ?, !, or .
- Full stops might also indicate abbreviations
- 90% of full stops are sentence boundaries!

EXAMPLE 1

"We are not getting any gas supplies from the gas field. The pipe is blown up," said Imran Khan.

EXAMPLE 2

"Do you mean to say," said Hermione in a hushed voice, "that that little girl dropped the toad-span?"

Semantic Analysis Pipeline

WHO COULD BUY CHICAGO

EMPRESSES of finance are figures of mighty importance in the industrial scheme of the twentieth century. In the United States their hold is strongest. The grip of feminine imperialism has been vastly enhanced by the recent succession of Mrs. Mary M. Harrington, wife of the late railroad Napoleon, Edward H. Harriman, to practically complete control of $100,000,000, representing holdings over almost the complete railroad skeleton of the United States. She is now the world's wealthiest woman, and in the etiquette of the court of dollar marks, she marches to the van of an auriferous galaxy of money queens.

The female sovereigns of history who ruled and practiced despotism over millions of subjects, declared war, surrounded themselves with panoply and display, and through caprice caused the death of thousands, had no such amount of power for good or

York, Chicago and Philadelphia this combination of woman-controlled wealth outweighs the value of the manufacturing product in any city in the country. St. Louis, fourth in the list, measures barely more than half this feminine total in its manufacturing output.

THE queens of old maintained bands of men at arms whom they hurled at rival powers for gain and glory, as well as using them to keep their own subjects in a humble mood. But what an army and navy could be retained by these empresses of to-day! See what Holland accomplishes for scarcely more than one-half the annual income of these women. It maintains an army which has a war footing of 68,000 men, and a navy of 72 ships, manned by 10,750 sailors.

In the cause of mercy and peace the energies of these women mostly are directed, and the above estimate is offered only for purposes of comparison. This court of sovereigns could launch a publicity campaign in the cause of peace and disarmament such as the world has never seen. They could erect temples for the propagation of peace with greater facility than Andrew Carnegie built his widespread libraries. In time of war, combined, they would be the greatest factor in supplying the needs of both land and sea for the hospital corps.

In the name of charity they would be the most powerful combination on earth. It is quite likely that they

visers. A few, including Mrs. Potter Palmer, deem the cares of sordid business a minor feature of life, and turn to the signing of checks and documents only when such matters are pressing. Society represents their world, and the financial tumult beyond its purlieus has no interest for them. Mrs. Russell Sage, perhaps, is better known than the rest for her activities along charitable lines since the death of her husband left at her disposal wealth amounting to many millions. Mrs. Phoebe Hearst also is widely known, particularly in California, for her benefactions in the aid of higher education.

ALL EYES now are turned on Mrs. Harriman, who has mounted the throne left vaccant by her husband and holds the reins of his power almost completely.

Mrs. Harriman was Miss Mary Averell, daughter of the late W. J. Averell, a banker of Rochester, N. Y. She has railroad antecedents, for her father built up his tidy fortune in the Rome, Watertown & Ogdensburg Railroad Company. The Harrimans' 36 years of married life was ideally happy. Mrs. Harriman brought her husband financial aid in his early struggles in the market. It was assistance which came when it was most valuable to him. He never needed it again, for his administrative ability soon won him the support of Kuhn, Loeb & Co., the National City bank and the powerful Standard Oil clique. Thereafter the help of his wife did not lie

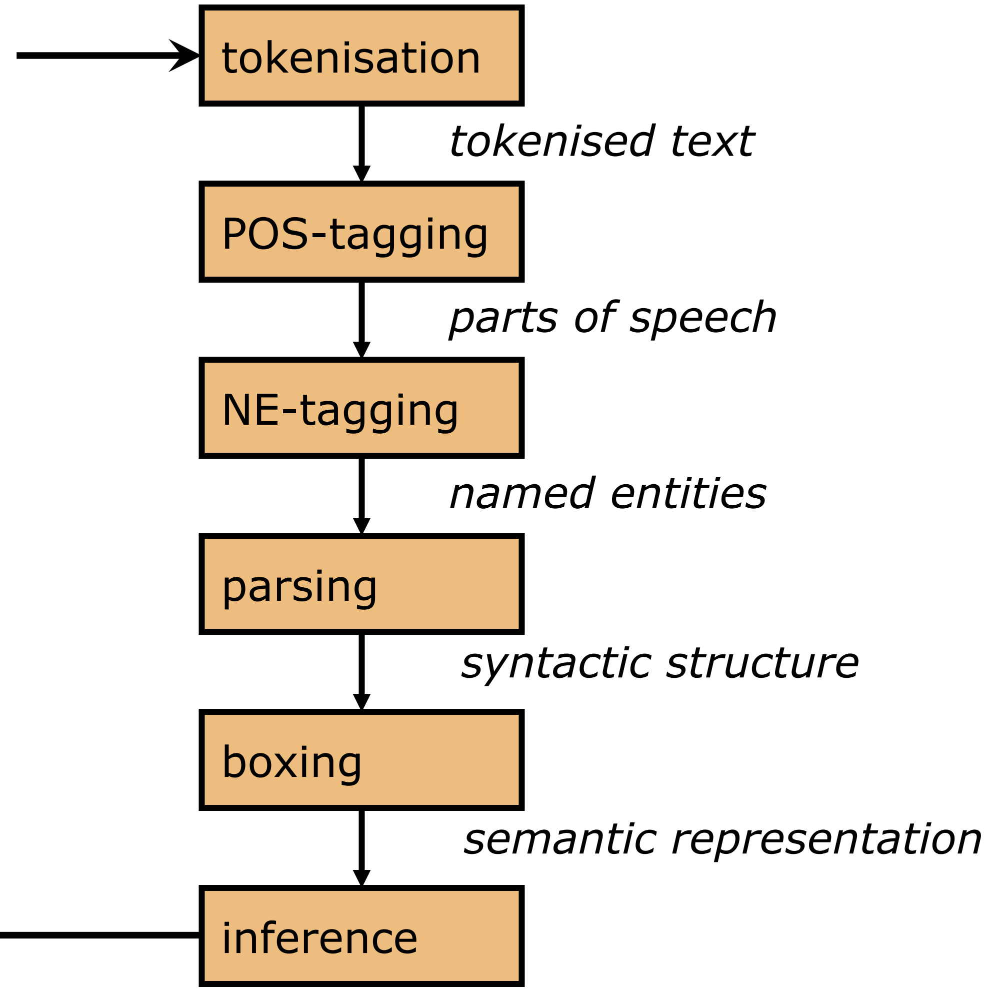
Mrs. Wm. K. Thaw	
Mrs. Potter Palmer	7,500,000
Mrs. H. H. Rogers	5,000,000
	$368,000,000
Assessed Valuation of all Chicago real estate for 1908	$344,399,927

which, under the laws of New York, otherwise would be imposed. Mrs. Harriman's dower rights are subject to no tax whatever, and the remaining two-thirds of the estate to a tax of only one per cent., as against a graded classification which in some instances runs to five per cent.

Besides Mrs. Harriman, those who survive the railway king are Mrs. Cornelia Gerry of Newport, R. I. and Misses Mary and Carol Harriman, daughters; and William Averell and Edward Roland Harriman, sons.

IT IS interesting to view the extent of Mrs. Harriman's railway holdings which place her well at the head of the world's rail magnates. In the absolute control of the Harriman estate are the Southern Pacific system, the Union Pacific system, the Southern Pacific of Mexico, the San Pedro Los Angeles & Salt Lake, the St. Joseph & Grand Island, the Illinois Central and the Central of Georgia. Dominant interest is held in the Baltimore & Ohio; Cincinnati, Hamilton & Dayton, and the Delaware & Hudson. Interest of importance is maintained in the Erie, the New York Central lines, Wheeling & Lake Erie and the Wabash Pittsburg terminal. Minor holdings are retained in the Atchison, Topeka & Santa Fe, the Chicago & Northwestern, and the Chicago, Mil

FOUR OF THE RICHEST WOMEN IN THE WORLD.



Tagging

- Assigning a label to each word (token) in a sentence (text)
- The label indicates to what class the token belongs
- Examples:
  part of speech, named entities, chunks

- How can we feed a machine some new, unseen linguistic data (a text) and expect it to come back with certain predictions?
- Basic idea: **learn from examples**

Machine Learning

POS tagging

- POS tagging is the task of labelling each token with a part of speech
- Most current approaches use statistical techniques
- There are two main issues
  - Dealing with ambiguity
  - Choice of tagset

Tag	Description
CC	coordinating conjunction
CD	cardinal number
DT	determiner
EX	existential there
FW	foreign word
IN	preposition/subordinating conjunction
JJ	adjective
JJR	adjective, comparative
JJS	adjective, superlative
LS	list marker
MD	modal
NN	noun, singular or mass
NNS	noun plural
NNP	proper noun, singular
NNPS	proper noun, plural
PDT	predeterminer
POS	possessive ending

Tag	Description
PRP	personal pronoun
PRP$	possessive pronoun
RB	adverb
RBR	adverb, comparative
RBS	adverb, superlative
RP	particle
TO	to
UH	interjection
VB	verb, base form
VBD	verb, past tense
VBG	verb, gerund/present participle
VBN	verb, past participle
VBP	verb, sing. present, non-3d
VBZ	verb, 3rd person sing. present
WDT	wh-determiner
WP	wh-pronoun
WP$	possessive wh-pronoun
WRB	wh-abverb

POS tagset (Penn)

Named Entity Recognition

- The task of finding domain-relevant names in texts
- Most common types of named entity are:
  - Person
  - Organisation
  - Location
- Two phases
  - Detect proper names (or entities)
  - Classify detected phrases

Tag	Description
B-PER	Person (first word)
I-PER	Person (subsequent words)
B-LOC	Location (first word)
I-LOC	Location (subsequent words)
B-ORG	Organisation (first word)
I-ORG	Organisation (subsequent words)
B-NAM	Miscellaneous (first word)
I-NAM	Miscellaneous (subsequent words)
O	not a named entity

NE tagset (IOB-2 format)

Data selection

- Select data (a corpus)
- Enrich it with the information you want a machine to predict for you

EXAMPLE

I will use the back door.
He promised to back my proposal.

Annotation (POS)

- Select data (a corpus)
- Label each word correctly

NN

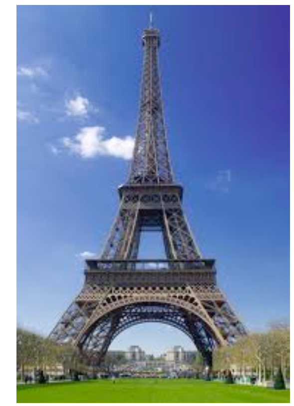
EXAMPLE

I will use the back door.
He promised to back my proposal.

VB

Annotation (NE)

- Select data (a corpus)
- Label each word correctly



B-LOC

EXAMPLE

Discover what's on and things to do in Paris.
The footwear collection from celebrity Paris Hilton
will be launched next month.

B-PER

I-PER

Preparation

- Enrich it with the information you want a machine to predict for you
- Put in the correct format

EXAMPLE

Michael|NNP J.|NNP Fox|NNP replaced|VBD Bruce|NNP Willis|NNP in|IN third|JJ place|NN

EXAMPLE

I will use the <lex pos="NN">back</lex> door.
He promised to <lex pos="VB">back<lex> my proposal.

Feature selection (POS)

- Prefixes of current word (up to 4 characters)
- Suffixes of current word (up to 4 characters)
- Word contains a number (yes/no)
- Word contains uppercase character (yes/no)
- Word contains hyphen (yes/no)
- Values of previous words and tags

Feature selection (NE)

- Word contains period
- Word contains punctuation
- Word is only digits
- Word is a number
- Word is upper/lower/title/mixed case
- Word is alphanumeric
- Length of word
- Word has only Roman numerals
- Word is an initial
- Word is an acronym
- Word is in a gazetteer (geographical dictionary)
- POS tag
- NE memory tag (most recently assigned tag to Word)
- is Word seen more frequently with uppercase or lowercase?

Feature extraction

EXAMPLE

The	stories	about	well-heeled	communities	and	developers
DT	NNS	IN	JJ	NNS	CC	NNS

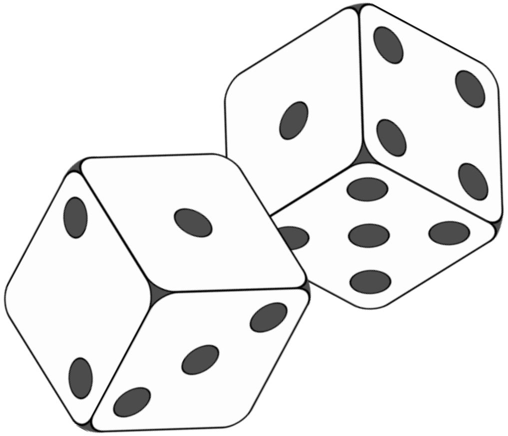
FEATURES

well-heeled

Feature	Value	Feature	Value
current word	well-heeled	contains uppercase	no
previous word	about	contains number	no
next word	communities	prefix-2	we
previous tag	IN	preffix-3	wel
next tag	NNS	suffix-2	ed
contains hyphen	yes	suffix-3	led

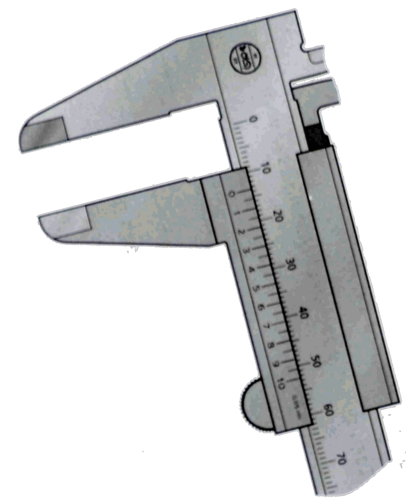
Statistical modelling

- Now we are ready to pick a learning algorithm and make a model
- We can use this model on new, unseen data
- The performance on the unseen data will show us how good this model is



The performance of a tagger depends mainly on three factors:

- Amount of training data
- Feature sets
- Machine learning method



Tagging performance

- Most words in natural languages have multiple possible meanings

"pen" (noun)

- The dog is in the pen.
- The ink is in the pen.

"take" (verb)

- Take one pill every morning.
- Take the first right past the stoplight.

Lexical Ambiguity

Lexical Ambiguity

- Sometimes syntax helps distinguish meanings for different parts of speech of an ambiguous word

"conduct" (noun or verb)

- John's **conduct** in class is unacceptable.
- John will **conduct** the orchestra on Thursday.

How many different senses for table are used in these five sentences?

① “See table 4.”

② “It was a sturdy table.”

③ "I reserved a table at my favorite restaurant.”

④ “She sets a fine table.”

⑤ “He entertained the whole table with his witty remarks.”

How many different senses for see are used in these 14 sentences?

1) "Can you see the bird in that tree?"
2) "I just can't see your point."
3) "You'll see a lot of cheating in this school."
4) "I can see what will happen."
5) "I don't see the situation quite as negatively as you do."
6) "I see that you have been promoted."
7) "This program will be seen all over the world."
8) "I'll probably see you at the meeting."
9) "See whether it works."
10) "See that the curtains are closed."
11) "You should see a lawyer."
12) "We went to see the Eiffel Tower in the morning."
13) "The doctor will see you now."
14) "Did you know that she is seeing an older man?"

What is a “sense” of a word?

- Homonyms
  (same words, disconnected meanings)
- Polysemes
  (same words, connected meanings)
- Metonyms
  (systematically related meanings)

bank

- financial institute

bank

- sloping land next to river

Homonyms: disconnected meanings

fan

- device used to induce an airflow for the purpose of cooling or refreshing oneself

fan

- a person with a liking and enthusiasm for something

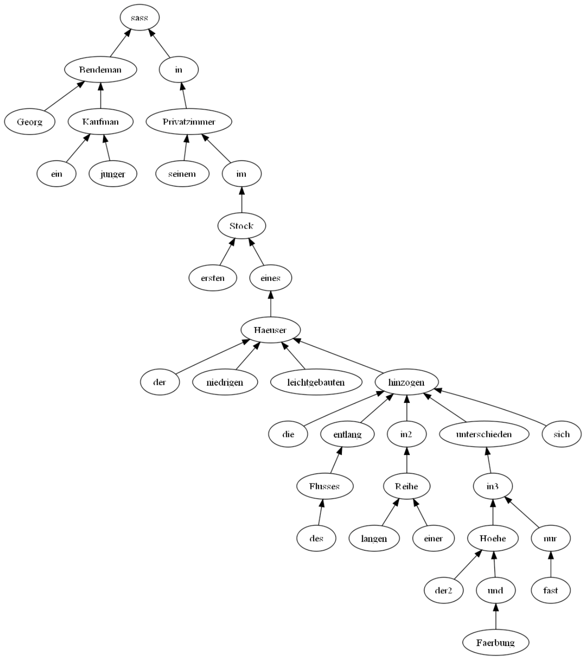
Homonyms: disconnected meanings

tree

- a woody plant

tree

- a data structure



Polysemy: connected meanings

fiat fired 100 employees

- the company

I bought a **fiat**

- a product

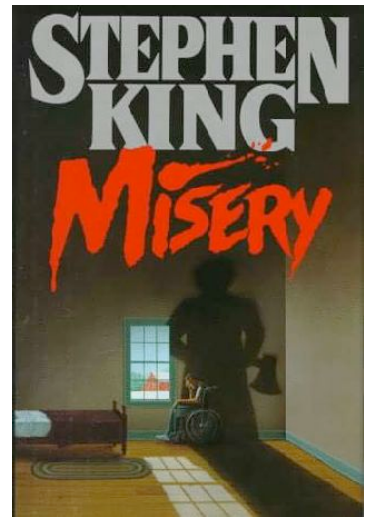
Metonomy: systematically connected meanings

Stephen King is an author.

- the author

I am reading a **Stephen King**

- the book



Metonomy:
systematically connected meanings

Don’t get confused...

- <u>homonyms</u>
  - senses that share pronunciation and orthography
  - example: *bank* vs *bank*
- <u>homophones</u>
  - words that share pronunciations but are spelled differently
  - example: would/wood, to/two/too
- <u>homographs</u>
  - words with distinct senses pronounced differently
  - example: conduct (noun) vs conduct (verb)
bass (animal) vs bass (music)

Relations between senses

- Synonymy / Antonomy
  (same / different)
- Hyponomy / Hyperonomy
  (subclass / generalisation)
- Meronomy / Holonomy
  (part-whole / whole-part)

Synonymy

- When two senses of two different words are (nearly) identical, they are <u>synonyms</u>

  couch — sofa
  vomit — throw up
  water — H_2O
  car — automobile

- Note:
  - relation between senses, not between words
  - probably no two words are true synonyms

Antonymy

- Words with opposite meanings are called <u>antonyms</u>

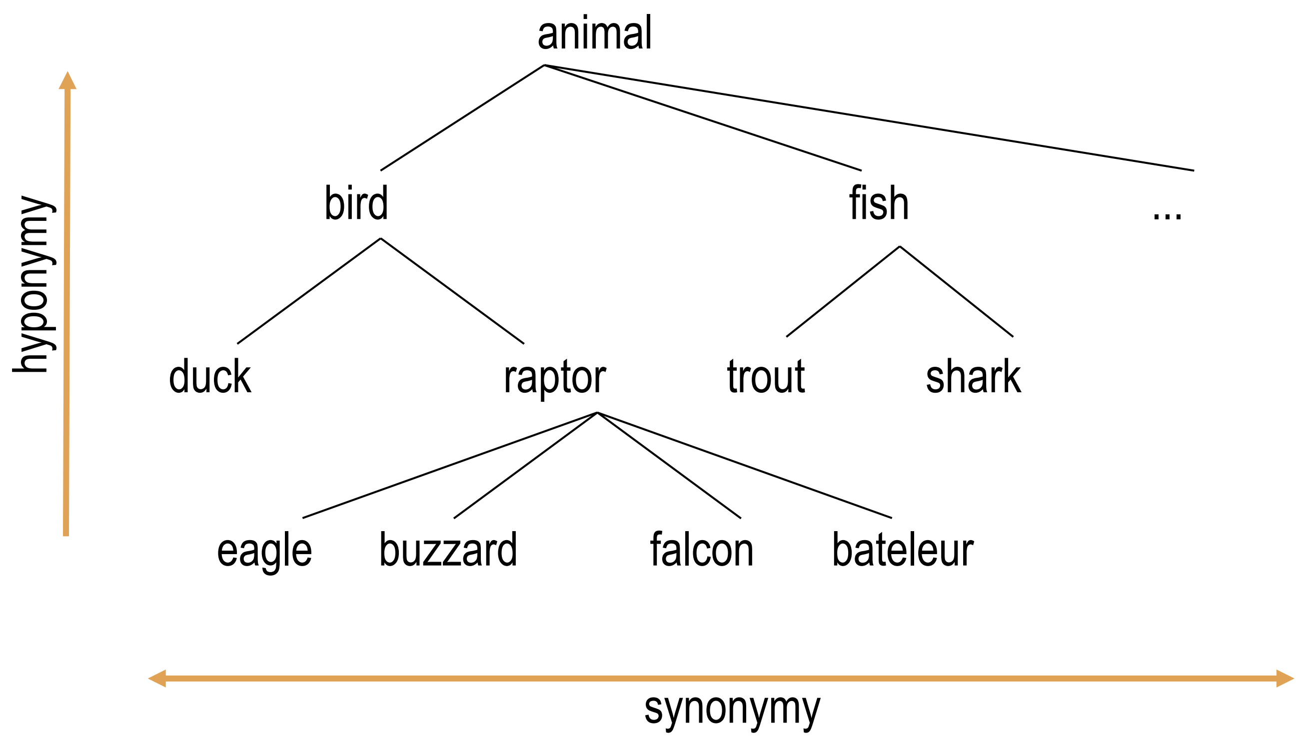
long — short
cold — hot
in — out
boring — interesting

Hyponymy

- A sense is a hyponym of another sense if the first sense is more specific than the other (i.e., forms a subclass)

  dog — pet
  falcon — bird
  house — building
  company — organisation

- Note: similar to ISA links in a knowledge base



ISA-hierarchy

Hyperonymy

- A sense is a hyperonym of another sense if the first sense is more general than the other (i.e., forms a superclass)

  dog — boxer
  falcon — kestrel
  house — villa
  company — agency

- Note: inverse of hyponomy

Meronomy (part-whole)

- A sense is a meronym of another sense if the first is a part of the second

leg — chair
door — house
wheel — car
leaf — tree

Holonomy (whole-part)

- A sense is a holonym of another sense if the first contains the second
(i.e., the opposite of meronym)

table — leg
door — keyhole
wheel — spoke
tree — branch

- A detailed database of semantic relationships between English words
- Developed by famous cognitive psychologist George Miller and team at Princeton University.
- Comprises about 155K English words.
- Nouns, adjectives, verbs, and adverbs grouped into about 117K synonym sets called *synsets*.

WordNet

WordNet is Big!

DESCRIPTION

Number of words, synsets, and senses

POS	Unique Strings	Synsets	Total Word-Sense Pairs
Noun	117798	82115	146312
Verb	11529	13767	25047
Adjective	21479	18156	30002
Adverb	4481	3621	5580
Totals	155287	117659	206941

Polysemy information

POS	Monosemous Words and Senses	Polysemous Words	Polysemous Senses
Noun	101863	15935	44449
Verb	6277	5252	18770
Adjective	16503	4976	14399
Adverb	3748	733	1832
Totals	128391	26896	79450

WordNet synsets

- How are word meanings represented in WordNet?
  - By **synsets** (synonym sets) as basic units
  - A concept (word meaning) is represented by listing the word forms that can be used to express it

Example: two senses of *board*

- Sense 1: a piece of lumber:
  {board, plank, ...}

- Sense 2: a group of people assembled
  for some purpose
  {board, committee, ...}

Example of WordNet synset

Division of the lexicon into four main categories:

- Nouns
- Verbs
- Adjectives
- Adverbs

WordNet: global organisation

WordNet: nouns

Noun

- hyponym
- hypernym
- holonym
- meronym

Noun

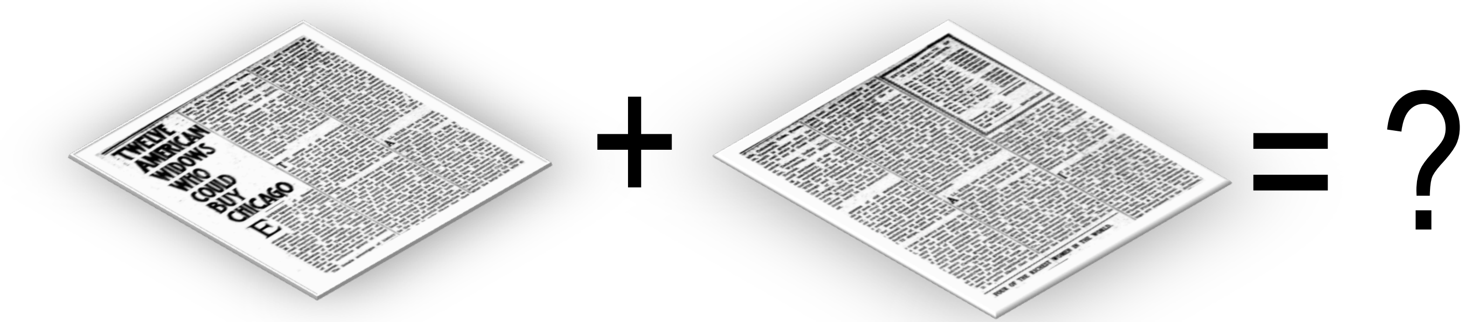
- S: (n) **orchestra** (a musical organization consisting of a group of instrumentalists including string players)
  - *direct hyponym* / *full hyponym*
    - S: (n) chamber orchestra (small orchestra; usually plays classical music)
    - S: (n) string orchestra (an orchestra playing only stringed instruments)
    - S: (n) symphony orchestra, symphony, philharmonic (a large orchestra; can perform symphonies) *"we heard the Vienna symphony"*
  - *part meronym*
    - S: (n) section (a division of an orchestra containing all instruments of the same class)
  - *direct hypernym* / *inherited hypernym* / ***sister term***
    - S: (n) musical organization, musical organisation, musical group (an organization of musicians who perform together)
      - S: (n) chorus (a group of people assembled to sing together)
      - S: (n) ensemble (a group of musicians playing or singing together) *"a string ensemble"*
      - S: (n) section (a division of an orchestra containing all instruments of the same class)
      - S: (n) duet, duette, duo (two performers or singers who perform together)
      - S: (n) trio (three performers or singers who perform together)
      - S: (n) quartet, quartette (four performers or singers who perform together)
      - S: (n) quintet, quintette (five performers or singers who perform together)
      - S: (n) sextet, sextette, sestet (six performers or singers who perform together)
      - S: (n) septet, septette (seven performers or singers who perform together)
      - S: (n) octet, octette (eight performers or singers who perform together)
      - S: (n) **orchestra** (a musical organization consisting of a group of instrumentalists including string players)
      - S: (n) band (instrumentalists not including string players)
      - S: (n) dance band, band, dance orchestra (a group of musicians playing popular music for dancing)
  - *derivationally related form*
- S: (n) **orchestra** (seating on the main floor in a theater)

Textual Entailment

Text:	**Mary bought a bottle of red wine.**	
Hypothesis:	**Someone bought a bottle of wine.**	YES!
Text:	**Mary bought a bottle of red wine.**	
Hypothesis:	**Someone bought a bottle of dry red wine.**	NO!
Text:	**Mary bought a bottle of red wine.**	
Hypothesis:	**John bought a pack of crisps.**	NO!

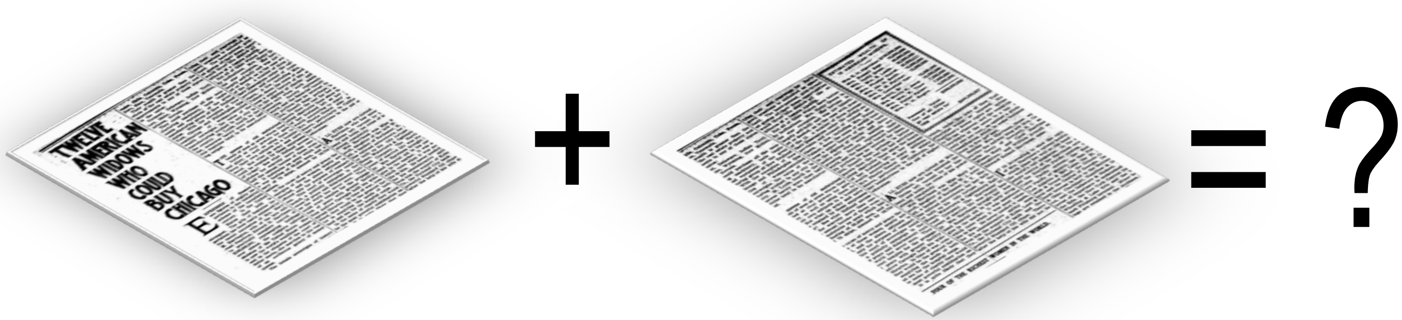
Recognising Textual Entailment

- Two-way classification:
  - **T entails H** if H contains no new information
  - **T does not entail H** if H contains new information



Recognising Textual Entailment

- Three-way classification:
  - Are the texts (taken together) contradictory?
  - If not, does one text contain information that the other doesn't?



T: **Johan has a beautiful black bicycle.**

H: **Johan has a beautiful bicycle.**

Entailment

T: **Bologna is the cultural capital of Italy.**

H: **Bologna is the capital of Italy.**

No entailment

RTE Examples

RTE baseline algorithms

- *Flipping a coin*
  accuracy: 50%

- *Lexical overlap*
  accuracy: ≈58%

1. Translate text and hypothesis into logic
2. Check if text entails hypothesis (not informative)
3. If it does, then hypothesis contains no novel information

Method: basic idea

- Entailment Engine
  - Input: an RTE problem
  - Output: prediction (yes, no)

- Includes:
  - The CCG parser and Boxer
  - WordNet

- Interface to external inference engines
  - Theorem provers
  - Model builders

Nutcracker

- Construct with Boxer
  - a DRS for Text (Box 1)
  - a DRS for Text+Hypothesis (Box 2)
- Translate Box 1 and Box 2 into first-order logic with the standard translation function FO(…)
- Generate the following formulas for the theorem prover:
  1. ~ [FO(Box 1) & FO(Box 2)] (proof => inconsistent)
  2. ~ [FO(Box 1) & ~ FO(Box 2)] (proof => entailed)

Looking under the hood

- Compile **WordNet** relations into FOL
  - Hyponyms, synonyms
  - “if X is a <u>poodle </u>then X a <u>dog</u>”

- Compile **NomLex** rules into FOL
  - Nominalisations
  - “<u>destruction </u>of X” implies that “X was <u>destructed</u>”

  (not part of Nutcracker yet)

Background Knowledge

Inference Engines (FOL)

Theorem Provers

- Vampire
- Spass
- Otter
- Bliksem

Model Builders

- Mace
- Paradox

THF /300	Satallax 2.1	LEO-II 1.2.8	*LEO-II 1.2*	Isabelle 2011	TPS 3.110228S1a
Solved	246/300	208/300	204/300	201/300	190/300
Av. CPU Time	12.04	8.97	4.95	36.55	18.69

TNT /100	Nitpick 2011	Refute 2011	Satallax 2.1
Solved	71/100	69/100	29/100
Av. CPU Time	6.75	4.75	0.45

TFA /100	SPASS+T 2.2.14	Z3 2.20	CVC3 2.4	H2WO4 11.07	MetiTarsk 1.8	SPASS-X 0.8	MELIA 0.1.2
Solved	96/100	92/100	77/100	60/100	54/100	48/100	12/100
Av. CPU Time	2.74	2.15	0.26	0.12	19.85	0.31	1.10

FOF /300	*Vampire 0.6*	Vampire 1.8	E-MaLeS 1.0	EP 1.4pre	iProver 0.9	leanCoP 2.2	iProver-E 0.7	E-KRHyp 1.2	E-Darwin 1.4	Metis 2.3	LEO-II 1.2.8	Otter 3.3	Muscadet 4.1
Solved	269/300	263/300	233/300	232/300	192/300	136/300	135/300	109/300	103/300	101/300	97/300	62/300	42/300
Av. CPU Time	12.95	13.62	18.85	22.55	9.22	46.80	8.68	8.93	6.97	24.75	25.18	5.84	8.99
Solutions	269/300	263/300	233/300	232/300	0/300	136/300	0/300	0/300	0/300	101/300	94/300	62/300	40/300

FNT /200	*Paradox 3.0*	FIMO 0.2	iProver-S 0.9	Nitrox 0.2	iProver-E 0.7	E-KRHyp 1.2	EP 1.4pre	E-Darwin 1.4
Solved	169/200	162/200	159/200	140/200	86/200	85/200	78/200	57/200
Av. CPU Time	3.33	14.43	34.93	17.42	7.52	15.92	2.40	7.23
Solutions	169/200	162/200	133/200	140/200	0/200	0/200	78/200	57/200

CNF /200	E 1.4pre	*Vampire 0.6*	Vampire 1.8	iProver 0.9	E-Darwin 1.4	iProver-E 0.7	E-KRHyp 1.2	LEO-II 1.2.8	Metis 2.3	Otter 3.3
Solved	178/200	177/200	173/200	116/200	72/200	64/200	63/200	57/200	56/200	55/200
Av. CPU Time	13.06	14.62	12.95	24.16	14.42	10.93	24.60	26.62	20.41	10.50

EPR /150	iProver 0.9	*iProver 0.8*	Vampire 1.8	iProver-E 0.7	E 1.4pre	Metis 2.3	E-Darwin 1.4	FIMO 0.2	E-KRHyp 1.2
Solved	145/150	145/150	127/150	121/150	91/150	78/150	70/150	62/150	60/150
Av. CPU Time	12.70	14.95	15.79	24.78	7.90	20.85	12.64	1.81	10.02

2011 World Cup Theorem Proving (CASC-23)

- RTE system for English
- Based on DRT and theorem proving
- Distributed with the C&C tools

Demo of Nutcracker

Inference

- check:
  - bin/nc
  - make bin/nc
- try the following t/h pairs:
  - T: Bill Gates has a blue cat.
    H: He has no animal.
  - T: John has a dog.
    H: John has an animal.
  - T: John likes no animal.
    H: John likes a dog.
  - T: Mr. Jones likes a dog.
    H. A dog is liked by Mr. Jones.

Method	Accuracy	Coverage
Flip a coin	50.0%	100%
Token overlap	57.6%	100%
Wordnet overlap	58.6%	98%
Model overlap	61.4%	88%
Proof	81.0%	4%

Performance on RTE-3 (800 pairs)