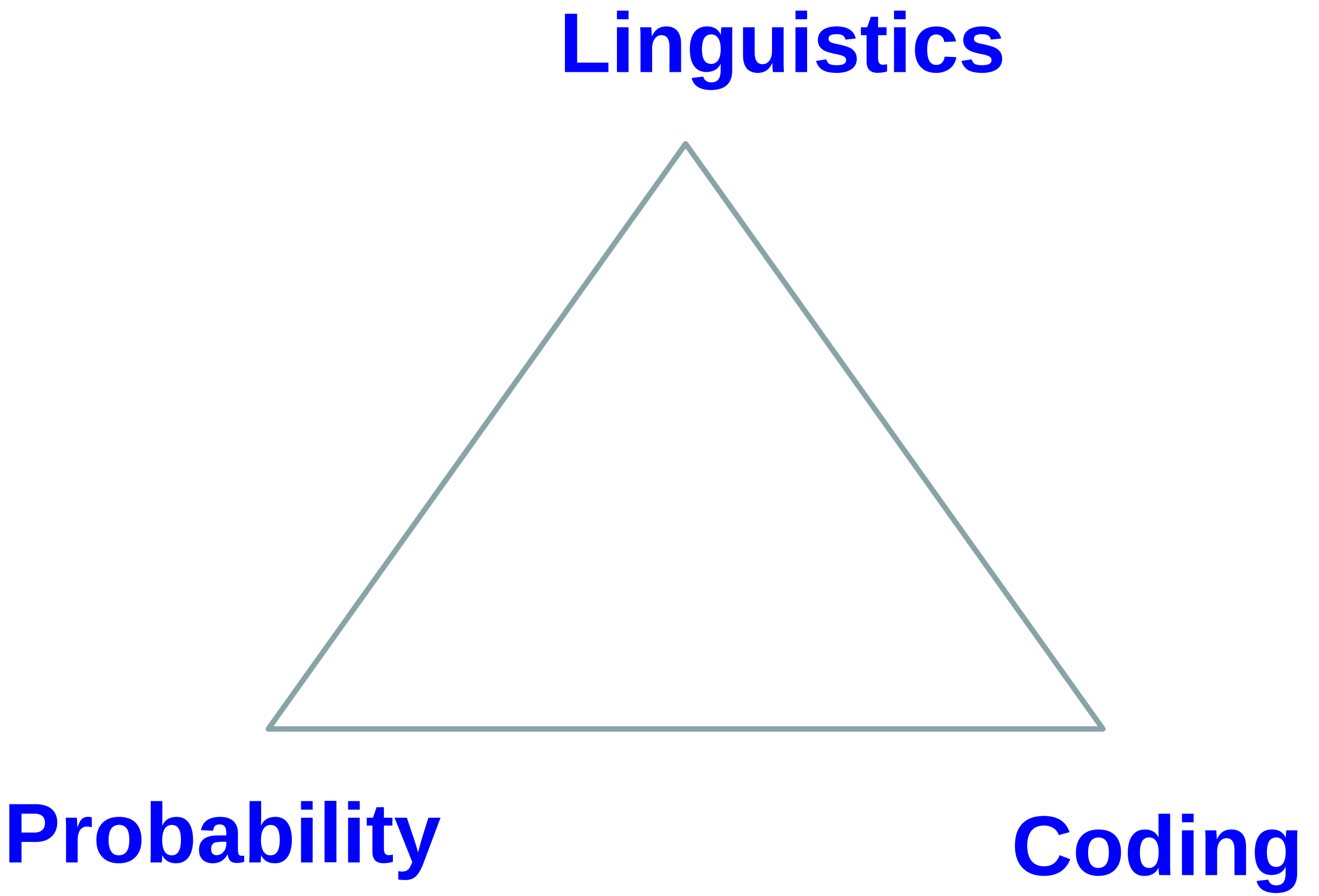
CS626: Speech, Natural Language Processing and the Web

Wordnet and Word Sense Disambiguation

Pushpak Bhattacharyya
Computer Science and Engineering Department
IIT Bombay

Week 3 of 9th August, 2021

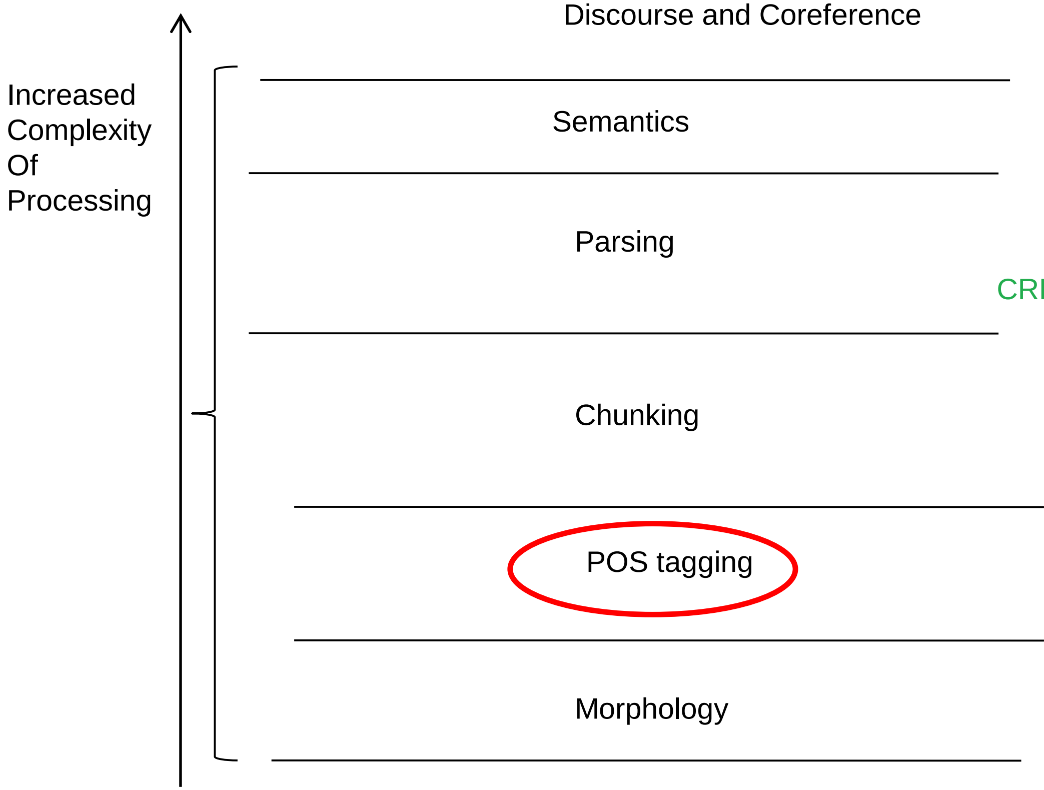
The Trinity

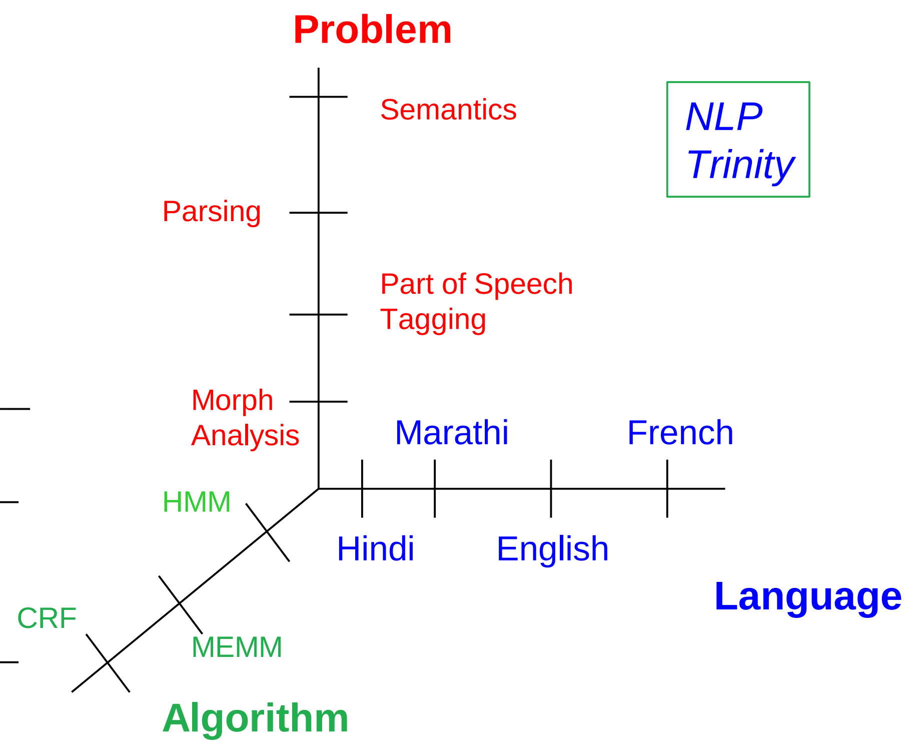


Task *vs.* Technique Matrix

Task (row) vs. Technique (col) Matrix	Rules Based/Knowledge-Based	Classical ML				Deep Learning		
		Perceptron	Logistic Regression	SVM	Graphical Models (HMM, MEMM, CRF)	Dense FF with BP and softmax	RNN-LSTM	CNN
Morphology								
POS								
Chunking								
Parsing								
NER, MWE								
Coref								
WSD								
Machine Translation								
Semantic Role Labeling								
Sentiment								
Question Answering								

NLP: multilayered, multidimensional





Recap on POS tagging

The Task

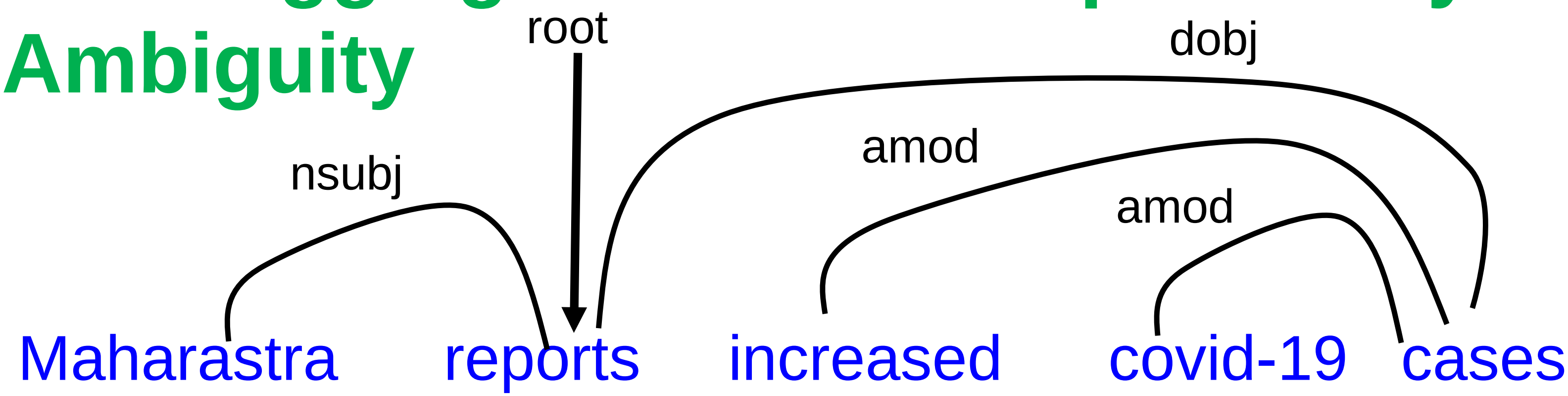
The/at house/nn passed/vbd finally/rb ,/, and/cc sent/vbd to/in the/at Senate/nn-tl ,/, a/at bill/nn extending/vbg the/at State/nn-tl Health/nn-tl Department's/nn$-tl authority/nn to/to give/vb planning/vbg assistance/nn to/in cities/nns ./.

WSD tags can come in addition to POS tags

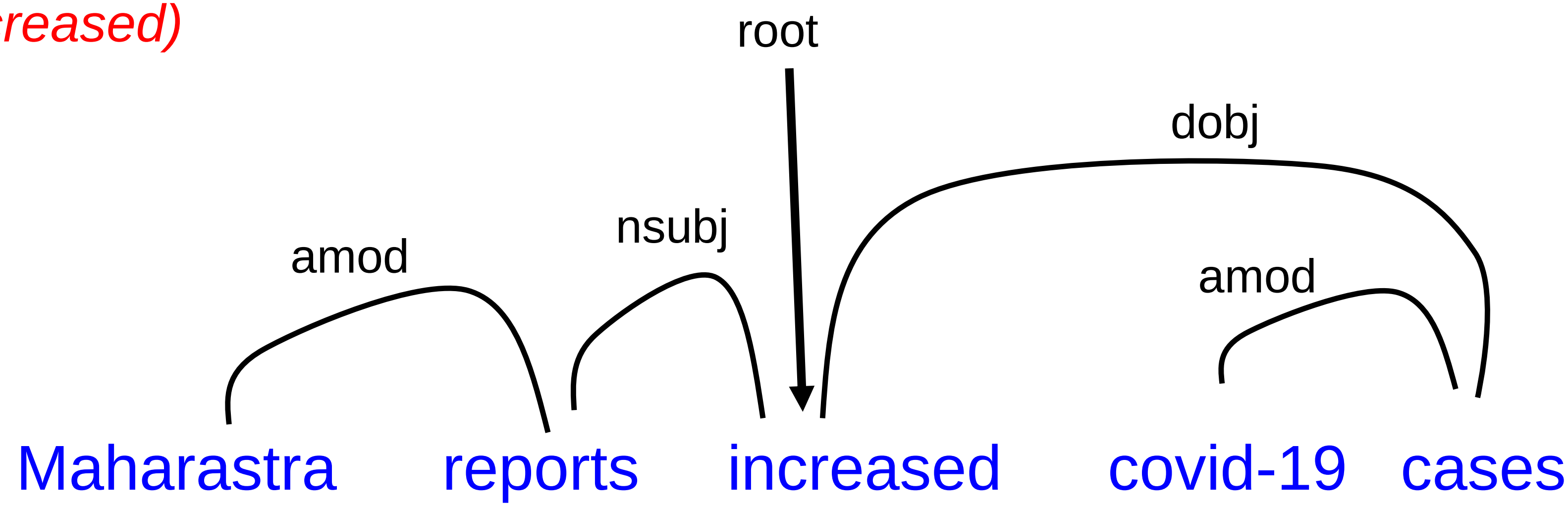
- Similar to POS tags, words can be tagged with senses and discourse
- Example: 'arm' has following senses in OntoNotes (a large scale corpus of semantic annotations)

Sense Tag	Sense Definition	WordNet sense
arm.01	The forelimb of an animal	WN.1
arm.02	A weapon	WN.2
arm.03	A subdivision or branch of an organization	WN.3
arm.04	A projection, a narrow extension of a structure	WN.4 WN.5

POS tagging is needed- Dependency Ambiguity



(it is reported by Maharastra Govt. that covid-19 cases have increased)



(it is the Maharastra reports that have increased covid-19 cases!!!)

Tag Set

- Attach to each word a tag from **Tag-Set**

- Standard Tag-set : Penn Treebank (for English).

Argmax computation (1/2)

Best tag sequence

= T*

= argmax P(T|W)

= argmax P(T)P(W|T) (by Baye's Theorem)

$$\mathbf{P(T) = P(t_0{=}\wedge\ t_1 t_2 \ldots t_{n+1}{=}.)}$$

$$\mathbf{= P(t_0)P(t_1|t_0)P(t_2|t_1t_0)P(t_3|t_2t_1t_0) \ldots P(t_n|t_{n-1}t_{n-2}\ldots t_0)P(t_{n+1}|t_nt_{n-1}\ldots t_0)}$$

$$\mathbf{= P(t_0)P(t_1|t_0)P(t_2|t_1) \ldots P(t_n|t_{n-1})P(t_{n+1}|t_n)}$$

$$= \prod_{i=0}^{N+1} \mathbf{P(t_i|t_{i-1})}$$ Bigram Assumption

Argmax computation (2/2)

$$P(W|T) = P(w_0|t_0\text{-}t_{n+1})P(w_1|w_0t_0\text{-}t_{n+1})P(w_2|w_1w_0t_0\text{-}t_{n+1}) \ldots P(w_n|w_0\text{-}w_{n-1}t_0\text{-}t_{n+1})P(w_{n+1}|w_0\text{-}w_nt_0\text{-}t_{n+1})$$

Assumption: A word is determined completely by its tag. This is inspired by speech recognition

$$= P(w_o|t_o)P(w_1|t_1) \ldots P(w_{n+1}|t_{n+1})$$

$$= \prod_{i=0}^{n+1} P(w_i|t_i)$$

$$= \prod_{i=1}^{n+1} P(w_i|t_i) \quad \text{(Lexical Probability Assumption)}$$

Viterbi for POS Tagging

- E.g.
  - T: Tags
  - W: Words
  - Two special symbol: '^' and '.'

  Find out number of paths in the tree given word sequence.

Exponential *w.r.t.* number of words in the sentence of length *L*

Number of path = Number of leaves in the tree.

$$O(T^L)$$

How to avoid it?

We do not need exponential work!

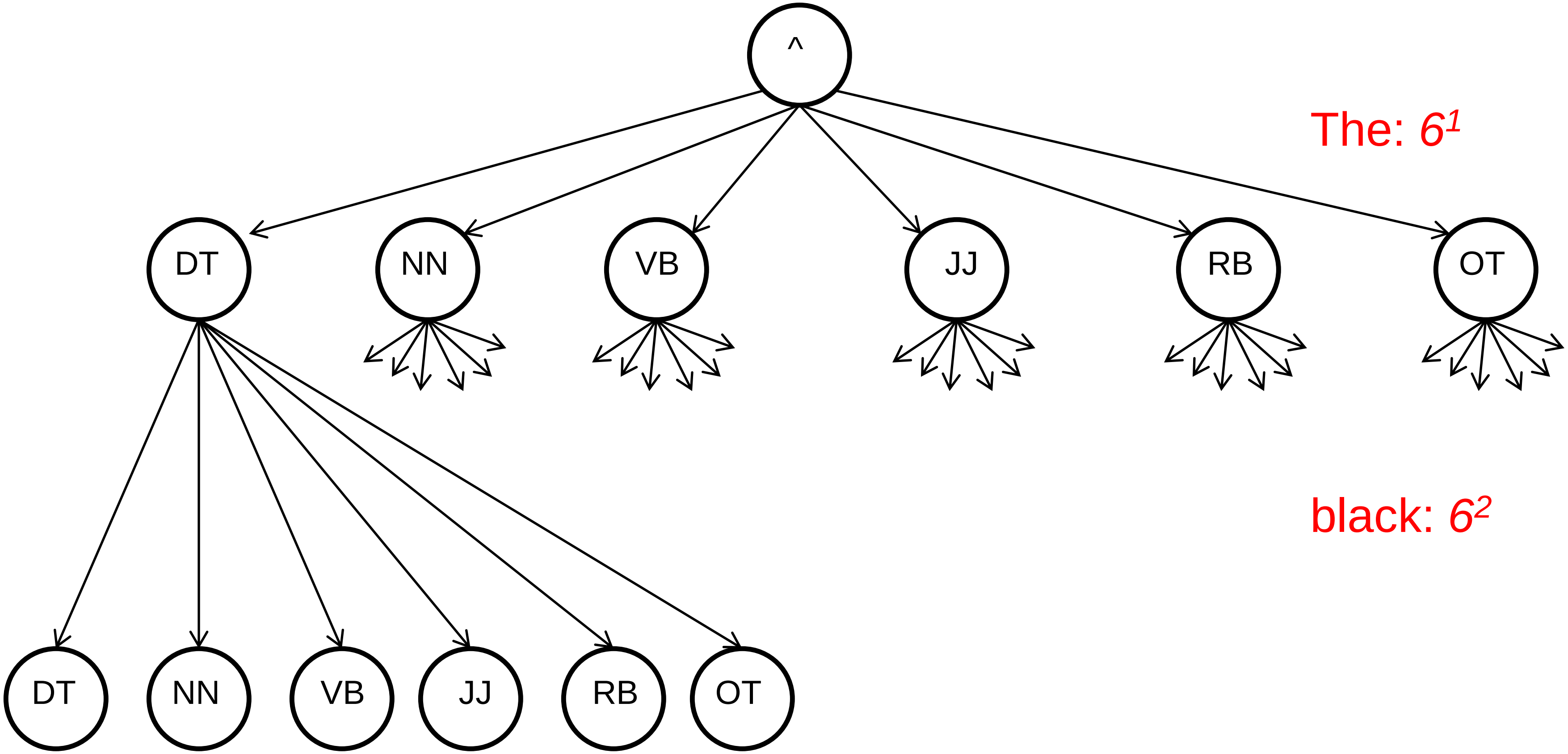
- Suppose our tags are
  - DT, NN, VB, JJ, RB and OT
- E.g.

DT- determiner
NN- Noun
VB- Verb
JJ- Adjective
RB- Adverb
OT- others

^	**The**	**black**	**dog**	**barks**	**.**
^	DT	DT	DT	DT	.
	NN	NN	NN	NN	
	VB	VB	VB	VB	
	JJ	JJ	JJ	JJ	
	RB	RB	RB	RB	
	OT	OT	OT	OT	

Possible tags

So, 6^4 possible path



The: 6^1

black: 6^2

dog: 6^3

barks: 6^4

.: 6^4

Total 6^4 paths

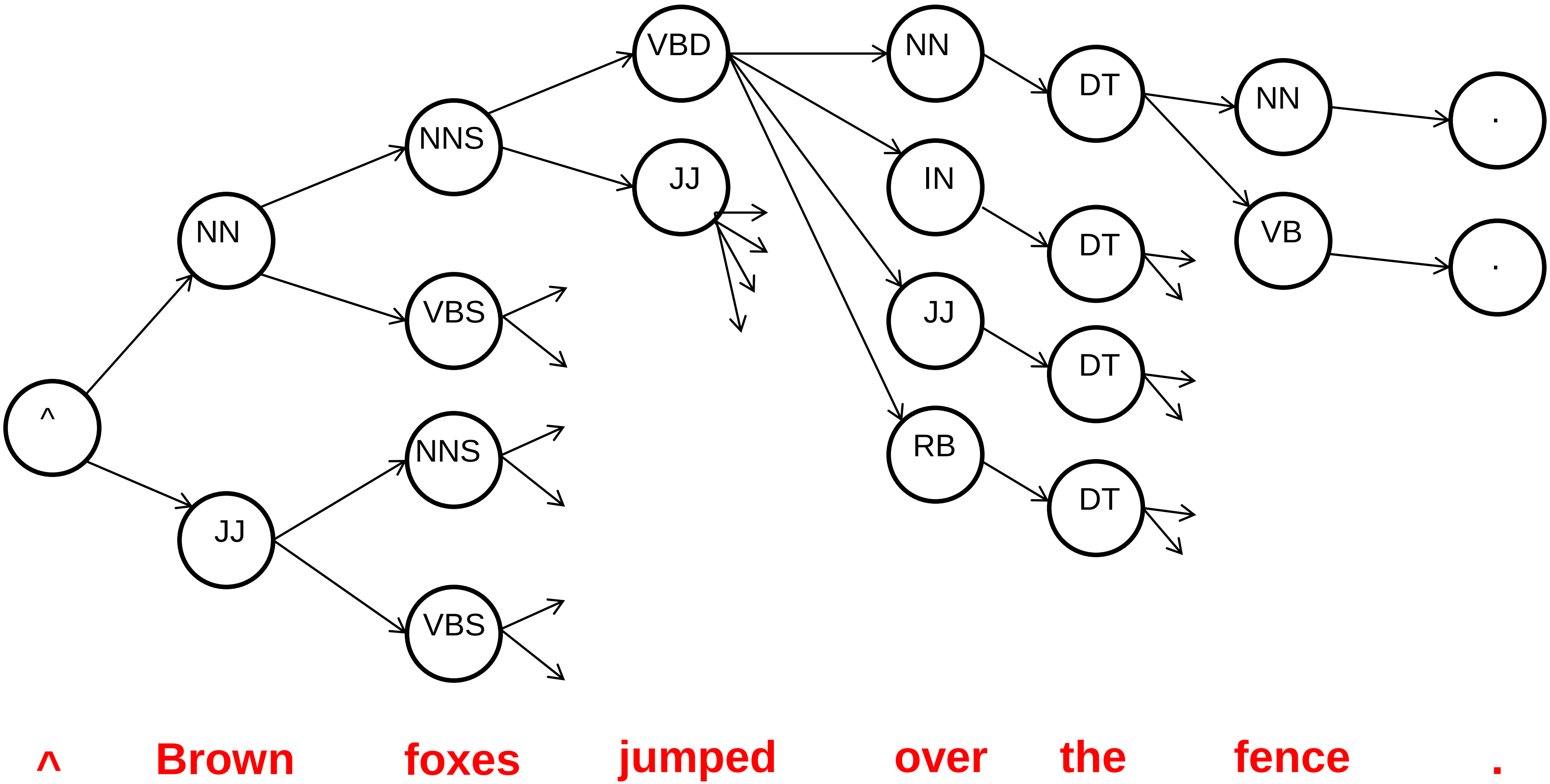
Consider the paths that end in NN after seeing input “The black”

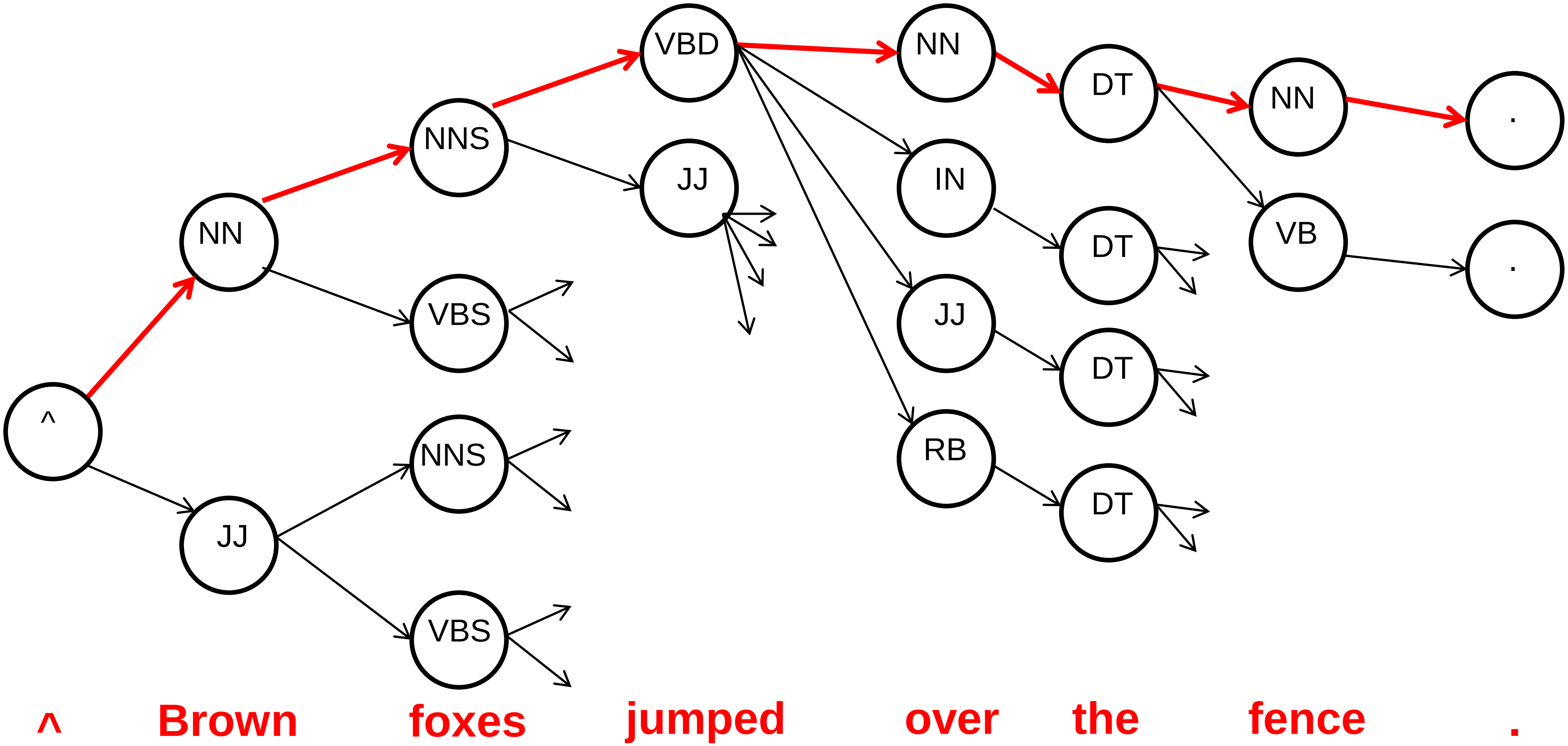
^	The	black	
^	DT	NN	***P*(T).*P*(W|T) =** *P*(DT|^) . *P*(NN|DT) . *P*(The|DT) . *P*(Black|NN)
^	NN	NN	***P*(T).*P*(W|T) =** *P*(NN|^) . *P*(NN|NN) . *P*(The|NN) . *P*(Black|NN)
^	VB	NN	***P*(T).*P*(W|T) =** *P*(VB|^) . *P*(NN|VB) . *P*(The|VB) . *P*(Black|NN)
^	JJ	NN	***P*(T).*P*(W|T) =** *P*(JJ|^) . *P*(NN|JJ) . *P*(The|JJ) . *P*(Black|NN)
^	RB	NN	***P*(T).*P*(W|T) =** *P*(RB|^) . *P*(NN|RB) . *P*(The|RB) . *P*(Black|NN)
^	OT	NN	***P*(T).*P*(W|T) =** *P*(OT|^) . *P*(NN|OT) . *P*(The|OT) . *P*(Black|NN)

$$Complexity = L * T^2$$

For each tag, only path with ***highest probability*** value are ***retained***, others are ***discarded***.

W:	^	Brown	foxes	jumped	over	the	fence	.
T:	^	JJ	NNS	VBD	NN	DT	NN	.
		NN	VBS	JJ	IN		VB	
					JJ			
					RB			





Probability of a path (e.g. Top most path) = *P*(T) * *P*(W|T)

P(^) . P(NN|^) . P(NNS|NN) . P(VBD|NNS) . P(NN|VBD) . P(DT|NN) . P(NN|DT) . P(.|NN) . P(.)

*

P(^|^) . P(brown|NN) . P(foxes|NNS) . P(jumped|VBD) . P(over|NN) . P(the|DT) . P(fence|NN) . P(.|.)

Decoding Summary

- On every word compute the partial path probability
- Out of all partial paths ending in a particular state, choose the one with highest path probability
- Advance only that leaf
- In case of tie, choose any one arbitrarily

Well known Data: Brown Corpus

- 1,014,312 words of running text of edited English prose printed in the United States
- 500 samples of 2000+ words each
- Facilitate automatic or semi-automatic syntactic analysis

Tag repository and probability

- Where do tags come from?
  - Tag set
- How to get probability values i.e. P(.)?
  - Annotated corpora

After modeling of the problem, emphasis should be on the corpus

Computing P(.) values

Let us suppose annotated corpus has the following sentence

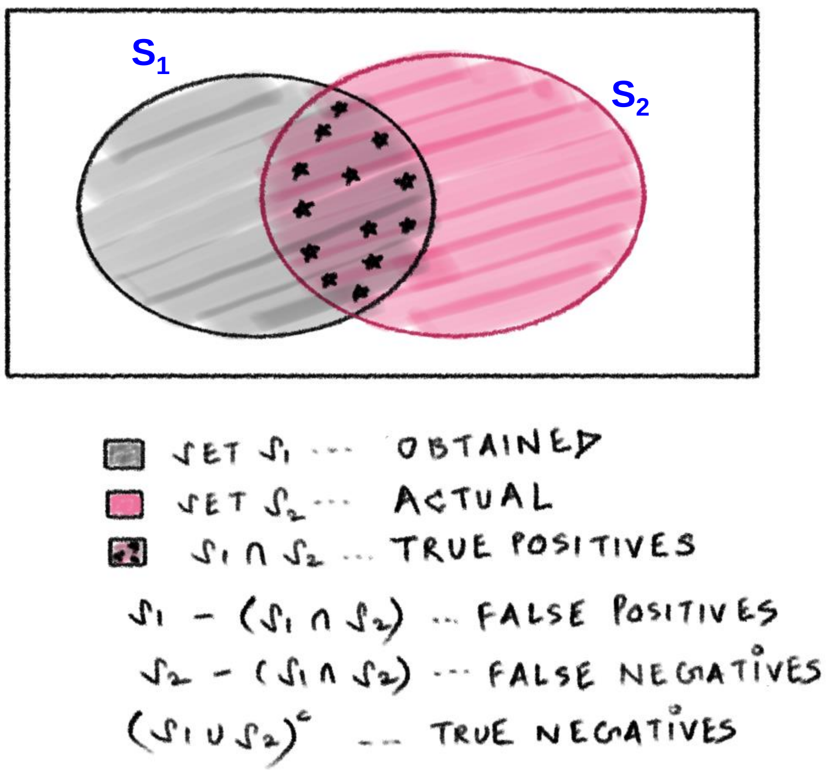
I	**have**	**a**	**brown**	**bag**	**.**
PRN	**VB**	**DT**	**JJ**	**NN**	**.**

$$P(NN \mid JJ) = \frac{Number\_of\_times\_JJ\_followed\_by\_NN}{Number\_of\_times\_JJ\_appeared}$$

$$P(Brown \mid JJ) = \frac{Number\_of\_times\_Brown\_tagged\_as\_JJ}{Number\_of\_times\_JJ\_appeared}$$

Evaluation:
Precision and Recall; Accuracy and F-score

False Positives, False Negatives, Precision, Recall, F-score

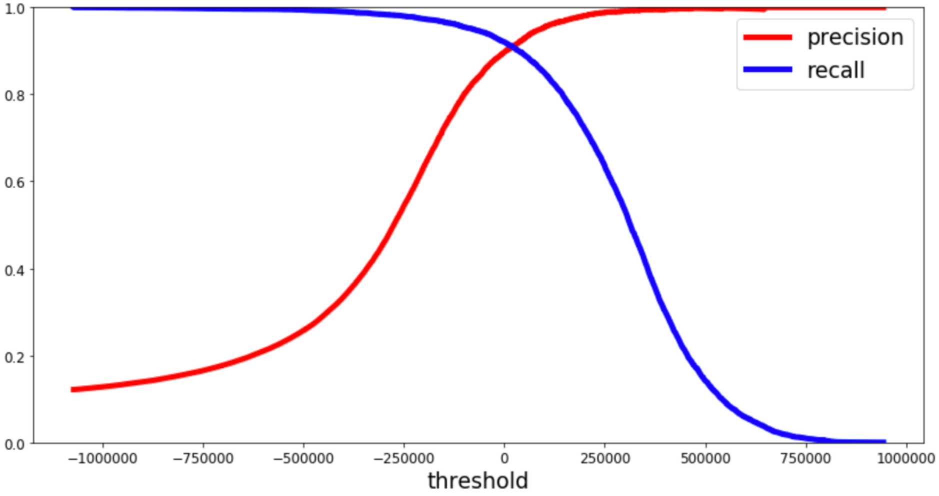


$$Precision = \frac{|S_1 \bigcap S_2|}{|S_1|}$$

$$Recall = \frac{|S_1 \bigcap S_2|}{|S_2|}$$

Precision and Recall Curve

- Precision-Recall Tradeoff : Increasing precision leads to a fall in recall and vice-versa
- Choose a threshold that gives desired precision and recall for the chosen problem



F1-Score

- F1 Score (F-Score) is the harmonic mean of Precision and Recall
- A perfect model has an F-score of 1
- Generalization is F_β (common values of β are 2 and 0.5 respectively, with more wt and less wt to precision)

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{\mathrm{tp}}{\mathrm{tp} + \frac{1}{2}(\mathrm{fp} + \mathrm{fn})}.$$

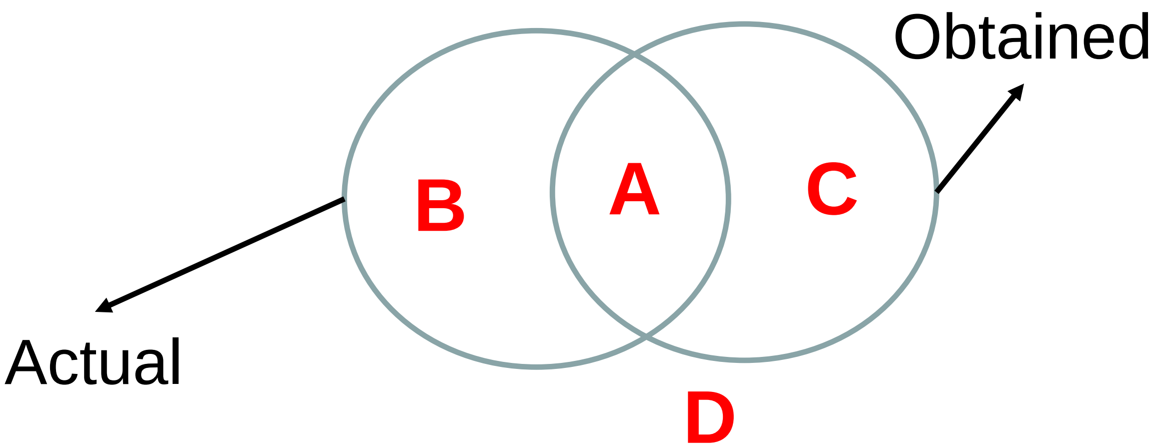
$$F_\beta = (1 + \beta^2)\frac{P.R}{\beta^2.P + R}$$

Exercise

- When will precision be different from recall
  - When will Precision be less than Recall
  - When will Recall be less than Precision

In terms of contingency table

			Obtained	
		Y	**N**	**Row-Total**
	Y	**A**	**B**	**A+B**
Actual	**N**	**C**	**D**	**C+D**
	Col-Total	**A+C**	**B+D**	**A+B+C+D**



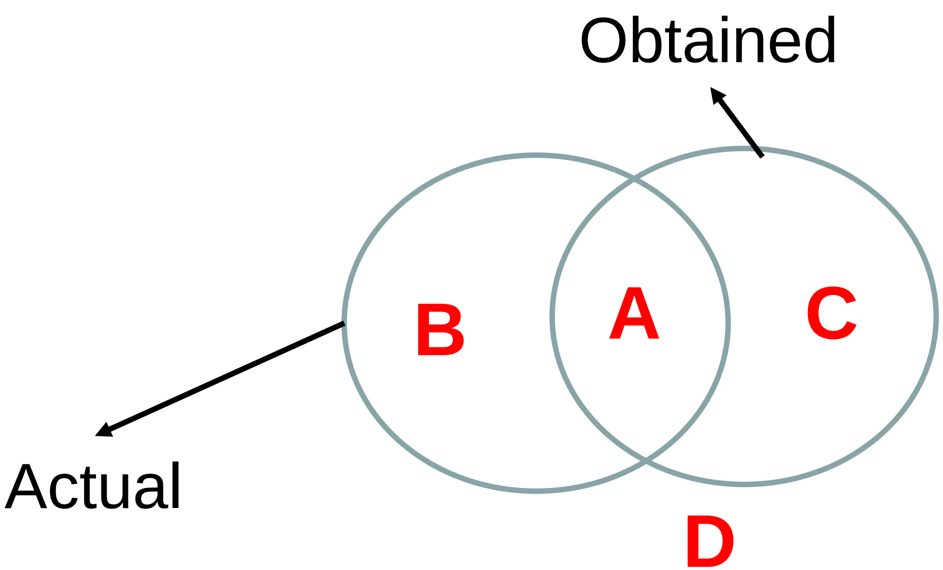
A: TP (true +ve)
B: FN (false -ve)
C: FP
D: TN

In terms of TP, TN, FP, FN

			Obtained	
		Y	**N**	**Row-Total**
	Y	**A**	**B**	**A+B**
Actual	**N**	**C**	**D**	**C+D**
	Col-Total	**A+C**	**B+D**	**A+B+C+D**

A: TP (true +ve)
B: FN (false -ve)
C: FP
D: TN

$$P = \frac{Actual \cap Obtained}{Obtained} = \frac{A}{A+C} = \frac{TP}{TP+FP}$$



$$R = \frac{Actual \cap Obtained}{Actual} = \frac{A}{A+B} = \frac{TP}{TP+FN}$$

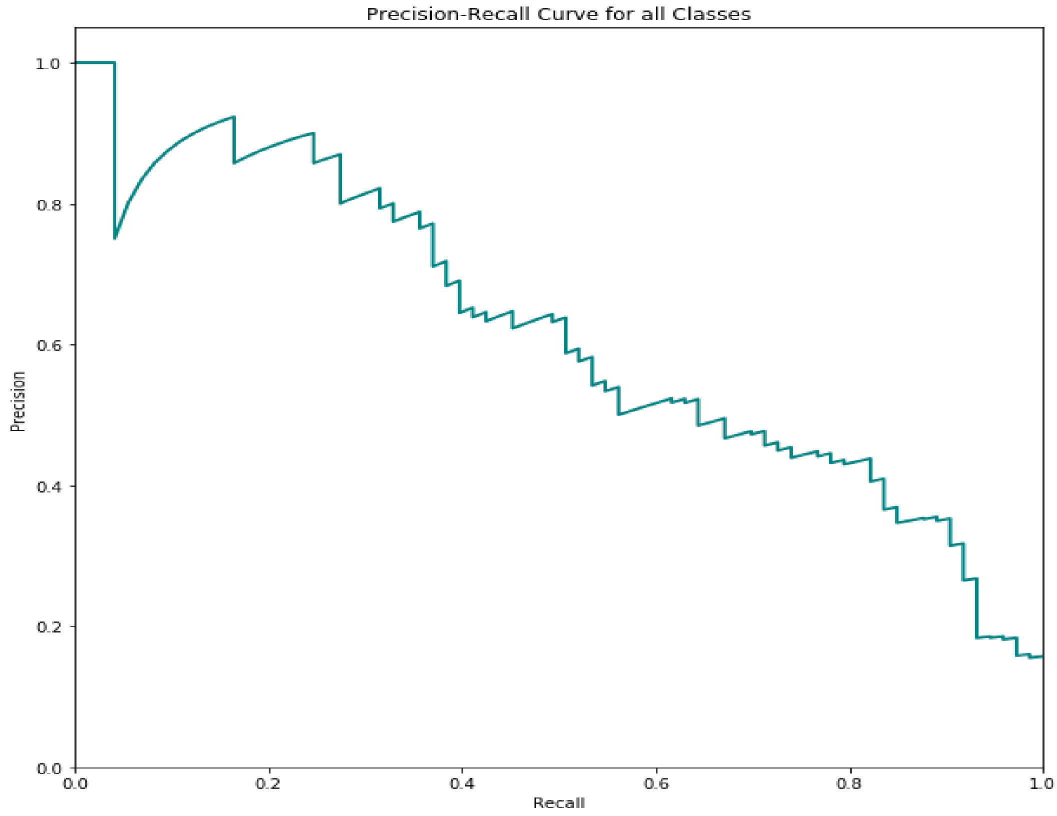
Example: Precision vs Recall

- Sentence - I love cricket
- Ground truth – I_PRP love_VB cricket_NN
- Predicted - I_NN love_VB cricket_NN
- Precision = 2/3
- Recall = 2/3

Precision vs Recall

- If a tag is not produced for a particular position (here for word love) then precision and recall will be different
- Sentence - I love cricket
- Ground truth – I_PRP love_VB cricket_NN
- Predicted - I_NN love cricket_NN
- Precision = 1/2
- Recall = 1/3

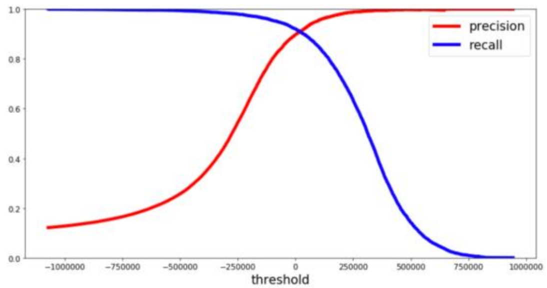
Example of precision-recall curve



Inverse relationship between precision and recall

- If $|S_2|$ decreases value of recall increases (denom influences more than num)
- $|S1 \cap S2|$ also decreases because $|S_2|$ has decreased
- Since there is no change in $|S_1|$, the value of precision decreases
- Hence we can observe that as the value of recall increases, the value of precision decreases and vice-versa

Precision-Recall tradeoff



- Threshold is value based on which we classify whether an instance belongs to a particular class or not .
- Classifier calculate the decision score for each instance and if the decision score is equal or higher than the threshold value, then it predicts positive class
- If decision score is less the threshold then instance belongs to negative class

Precision-Recall tradeoff

- At default threshold value is zero, precision is less than 80% and recall is higher than 80%
- If you increase the threshold value precision increases but recall decreases
- If you decrease the threshold value recall increases but precision increases

Relative weight of Precision-Recall

- Precision is more important than recall when false positives are more costly than false negatives
  - Example : Email spam detection
- Recall is more important than Precision when false negatives are more costly than false positives
  - Example : Detection of Cancer

Exercise

- Why Harmonic mean is chosen for calculating F1 score over other measures? (We know that Harmonic mean(HM) < Geometric mean(GM) < Arithmetic Mean(AM))

Exercise

- Explain the inverse relationship between precision and recall using TP,FP,FN (Make use the notions of contingency table)

Assignment

- Compute confusion matrix as a part of POS Tagging assignment

WSD

What is WSD (1/2)

- The task of Word Sense Disambiguation (WSD) consists of associating words in context with their most suitable entry in a pre-defined sense inventory.
- The de-facto sense inventory for English in WSD is WordNet. For example, given the word “mouse” and the following sentence:

What is WSD (2/2)

- For example, given the word “mouse” and the following sentence:
  - “A mouse consists of an object held in one’s hand, with one or more buttons.”
  - we would assign “mouse” with its electronic device sense (the 4th sense in the WordNet sense inventory).

Training Data for WSD

- The most widely used training corpus used is SemCor, with 226,036 sense annotations from 352 documents manually annotated.
- Some supervised methods, particularly neural architectures, usually employ the SemEval 2007 dataset.
- The most usual baseline is the Most Frequent Sense (MFS) heuristic, which selects for each target word the most frequent sense in the training data.

Example of sense marking: its need

एक_4187 नए शोध_1138 के अनुसार_3123 जिन लोगों_1189 का सामाजिक_43540 जीवन_125623 व्यस्त_48029 होता है उनके दिमाग_16168 के एक_4187
हिस्से_120425 में अधिक_42403 जगह_113368 होती है।

(According to a new research, those people who have a busy social life, have larger space in a part of their brain).

नेचर न्यूरोसाइंस में छपे एक_4187 शोध_1138 के अनुसार_3123 कई_4118 लोगों_1189 के दिमाग_16168 के स्कैन से पता_11431 चला कि दिमाग_16168 का एक_4187 हिस्सा_120425 एमिगडाला सामाजिक_43540 व्यस्तताओं_1438 के साथ_328602 सामंजस्य_166
के लिए थोड़ा_38861 बढ़_25368 जाता है। यह शोध_1138 58 लोगों_1189 पर किया गया जिसमें उनकी उम्र_13159 और दिमाग_16168 की साइज़ के आँकड़े_128065
लिए गए। अमरीकी_413405 टीम_14077 ने पाया_227806 कि जिन लोगों_1189 की सोशल नेटवर्किंग अधिक_42403 है उनके दिमाग_16168 का एमिगडाला
वाला हिस्सा_120425 बाकी_130137 लोगों_1189 की तुलना_में_38220 अधिक_42403 बड़ा_426602 है। दिमाग_16168 का एमिगडाला वाला हिस्सा_120425
भावनाओं_1912 और मानसिक_42151 स्थिति_1652 से जुड़ा हुआ माना_212436 जाता है।

Enrichment of wordnet: SentiWordNet

- SentiWordNet is a lexical resource for opinion mining
- It assigns to each synset of WordNet a sentiment score (positivity, negativity, objectivity)
- The IDs are maintained as the WordNet IDs
- The WordNet ID hence fixes the POS tag, sense, and the pragmatic labels like sentiment.
- The IDs hence act as disambiguating labels.

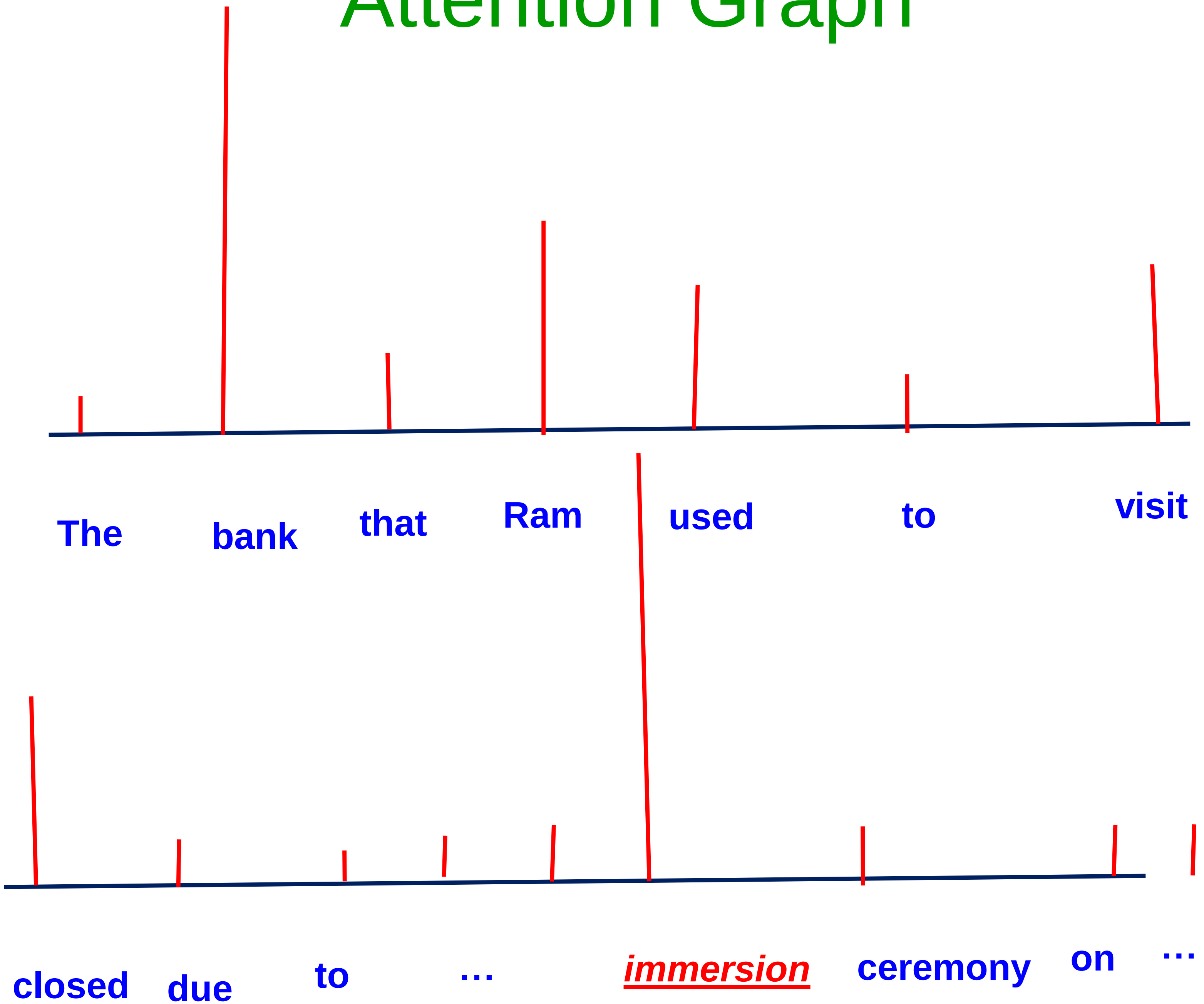
Ambiguity of लोगों (People)

- **लोग**, **जन**, **लोक**, **जनमानस**, **पब्लिक** - एक से अधिक व्यक्ति *"लोगों के हित में काम करना चाहिए"*
  - **(English synset) multitude, masses, mass, hoi_polloi, people, the_great_unwashed - the common people generally *"separate the warriors from the mass" "power to the people"***
- **दुनिया**, **दुनियाँ**, **संसार**, **विश्व**, **जगत**, **जहाँ**, **जहान**, **ज़माना**, **जमाना**, **लोक**, **दुनियावाले**, **दुनियाँवाले**, **लोग** - संसार में रहने वाले लोग *"महात्मा गाँधी का सम्मान पूरी दुनिया करती है / मैं इस दुनिया की परवाह नहीं करता / आज की दुनिया पैसे के पीछे भाग रही है"*
  - **(English synset) populace, public, world - people in general considered as a whole *"he is a hero in the eyes of the public"***

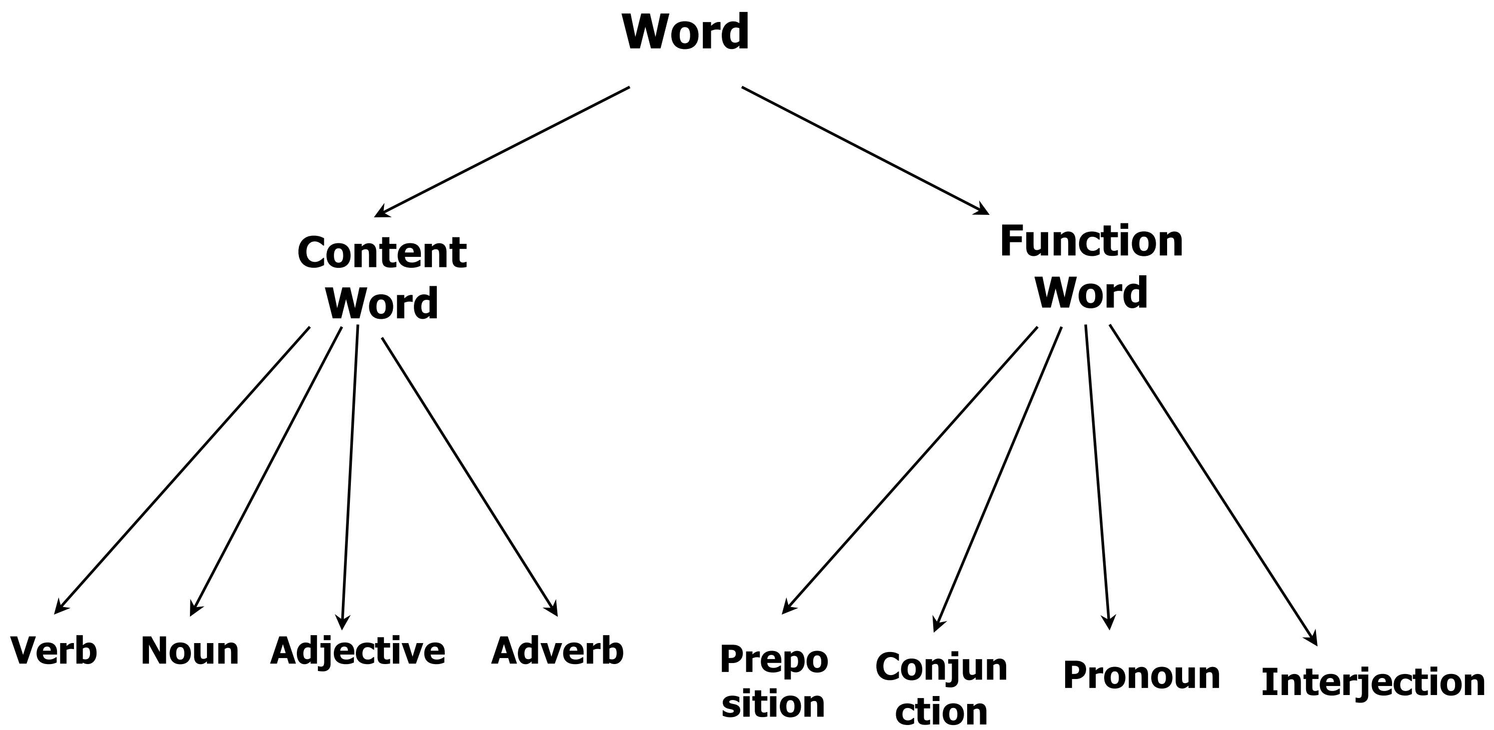
WSD: long distance dependency; contrasted with POS tagging

The *bank* that Ram used to visit 30 years before was closed due to the lockdown with the Govt. getting worried that crowding of people during the immersion ceremony on the river will aggravate the situation.

Attention Graph



Classification of Words



Ambiguity in Function Words

- Preposition:
  - I ate some bread with butter
  - I ate some bread with knife
  - I ate some bread with friend
- Conjunction:
  - *I like chocolate, and it is so expensive!*
    - Here ‘and’ acts like ‘but’
  - Different from *I like chocolate and my brother likes icecream*

Ambiguity in Function Words

- Pronouns:
  - With phone in one hand and camera in other, Rahul clicked photos with *it* during the wedding
  - *Pleonastic* *"it"*: I find *it* difficult to sleep
    - The word "It" does not refer to anything in this sentence
- Interjections:
  - *Ohh!* What a surprise (Joy)
  - *Ohh!* That's so unfortunate (Sad)

NLP: Thy Name is Disambiguation

- A word can have multiple meanings

and

- A meaning can have multiple words

Word with multiple meanings

Where there is a will,

Where there is a will,

There are hundreds of relatives

Where there is a will

There is a way

There are hundreds of relatives

WSD should be distinguished from structural ambiguity

- Correct groupings a must
  - …
- *Iran quake kills 87, 400 injured*
- *When it rains cats and dogs run for cover*

Should be distinguished from structural ambiguity

- Correct groupings a must
  - …
- *Iran quake kills 87, 400 injured*
- *When it rains, cats and dogs runs for cover*
- *When it rains cats and dogs, run for cover*

Groups of words (Multiwords) and names can be ambiguous

- *Broken guitar for sale, no strings attached (Pun)*
- *Washington voted Washington to power*
- *pujaa ne pujaa ke liye phul todaa*
- *(Pujaa plucked flowers for worship)*
- *(deep world knowledge) The use of a shin bone is to locate furniture in dark room*

Stages of processing

- Phonetics and phonology
- Morphology
- Lexical Analysis
- Syntactic Analysis
- Semantic Analysis
- Pragmatics
- Discourse

Example of WSD

- **Operation**, surgery, surgical operation, surgical procedure, surgical process -- (a medical procedure involving an incision with instruments; performed to repair damage or arrest disease in a living body; "they will schedule the operation as soon as an operating room is available"; "he died while undergoing surgery") TOPIC->(noun) surgery#1
- **Operation**, military operation -- (activity by a military or naval force (as a maneuver or campaign); "it was a joint operation of the navy and air force") TOPIC->(noun) military#1, armed forces#1, armed services#1, military machine#1, war machine#1
- **Operation** -- ((computer science) data processing in which the result is completely specified by a rule (especially the processing that results from a single instruction); "it can perform millions of operations per second") TOPIC->(noun) computer science#1, computing#1
- mathematical process, mathematical **operation**, **operation** -- ((mathematics) calculation by mathematical methods; "the problems at the end of the chapter demonstrated the mathematical processes involved in the derivation"; "they were learning the basic operations of arithmetic") TOPIC->(noun) mathematics#1, math#1, maths#1

IS WSD NEEDED IN LARGE APPLICATIONS?

Overlap-based WSD

- Find the overlap between words of context and words of definition
- Can we use bi-grams, tri-grams or higher n-grams overlap for WSD instead of Unigrams?
  - If we include higher n-grams in overlap-based WSD, It will incorporate the syntactic and structural information.
  - The structural part and lexicon part are independent of each other. Hence using higher n-grams will not help since WSD requires lexical disambiguation.

Word ambiguity, topic drift in IR

{case, container}

Drifted topic due to inapplicable sense!!!

Query word:
*"Madrid bomb blast **case**"*

{case, suit, lawsuit}

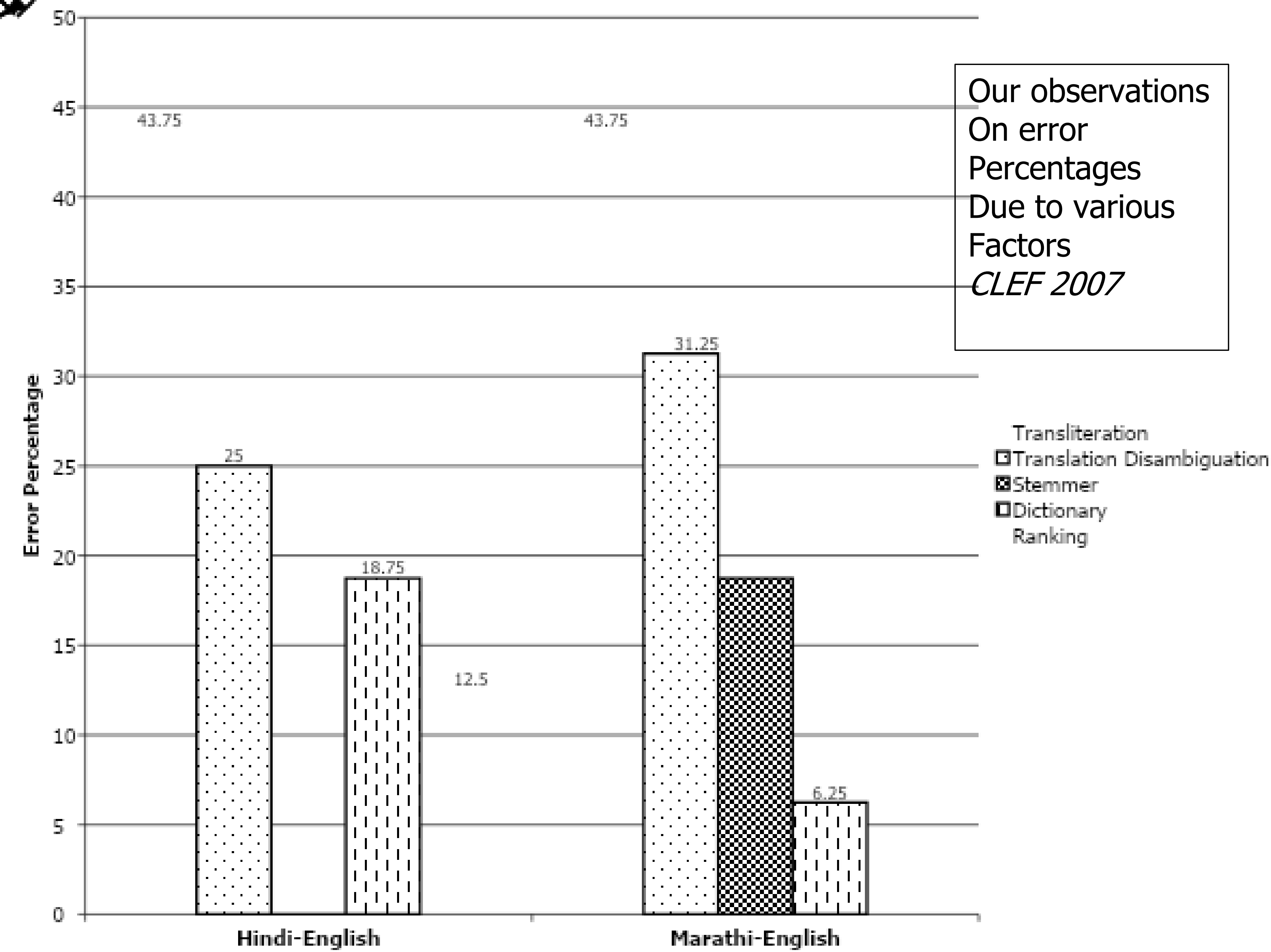
Drifted topic due to expanded term!!!

{suit, apparel}

Word Ambiguity and Topic drift in Information Retrieval

- Query: “Madrid bomb blast *case*”
  - How did the documents related to the clothing sense were added? – Through Query Expansion
  - The word *case* can have two senses: {case, container} and {case, suit, lawsuit}
  - In the first iteration of retrieval, retrieve the documents using initial word senses and then add the keywords from the retrieved documents in the query. This phenomena is called Relevance Feedback
  - “suit” gets added in the query which itself has two meanings – {suit, apparel}
  - In the next iteration of retrieval, documents related to the clothing sense of the word suit are also retrieved.

Our observations
On error
Percentages
Due to various
Factors
CLEF 2007



How about WSD and MT?
(status of google translate 10 years back)

Zaheer Khan, the India fast bowler, has been ruled out of the remainder of the series against England.

He will return to India and will be replaced by left-arm seamer RP Singh.

Zaheer picked up a hamstring injury during the first Test at Lord's.

He had been withdrawn from the squad for India's recent Test series in the West Indies due to a right ankle injury.

भारत के तेज गेंदबाज, जहीर खान, इंग्लैंड के खिलाफ श्रृंखला के शेष के बाहर शासन किया गया है. (ruled in the administrative sense??)

वह भारत लौटने और बाएँ हाथ के तेज गेंदबाज आरपी सिंह द्वारा प्रतिस्थापित किया जाएगा.

जहीर लॉर्ड्स में पहले टेस्ट के दौरान हैमस्ट्रिंग चोट उठाया. (lifted??)

वह भारत की वेस्ट इंडीज में हाल ही में एक सही (correct??) टखने की चोट के कारण टेस्ट श्रृंखला के लिए टीम से वापस ले लिया गया था.

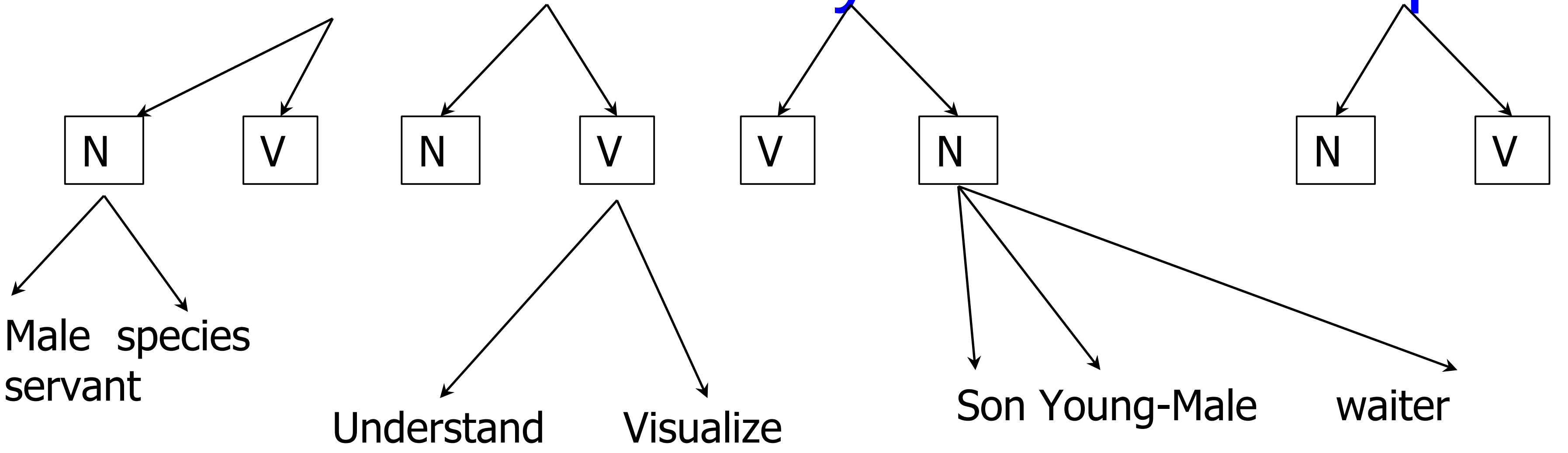
Current status of these translations

Zaheer Khan, the Indian fast bowler, has been ruled of the remainder of the series against England.	भारतीय तेज गेंदबाज जहीर खान इंग्लैंड के खिलाफ सीरीज के बाकी बचे मैचों से बाहर हो गए हैं
He will return to India and will be replaced by left-arm seamer RP Singh.	वह भारत लौटेंगे और उनकी जगह बाएं हाथ के तेज गेंदबाज आरपी सिंह लेंगे।
Zaheer picked up a hamstring injury during the first Test at Lord's.	लॉर्ड्स में पहले टेस्ट के दौरान जहीर को हैमस्ट्रिंग में चोट लग गई थी।
He had been withdrawn from the squad for India's recent Test series in the West Indies due to right ankle injury	उन्हें दाहिने टखने की चोट के कारण वेस्टइंडीज में भारत की हालिया टेस्ट श्रृंखला के लिए टीम से हटा दिया गया था

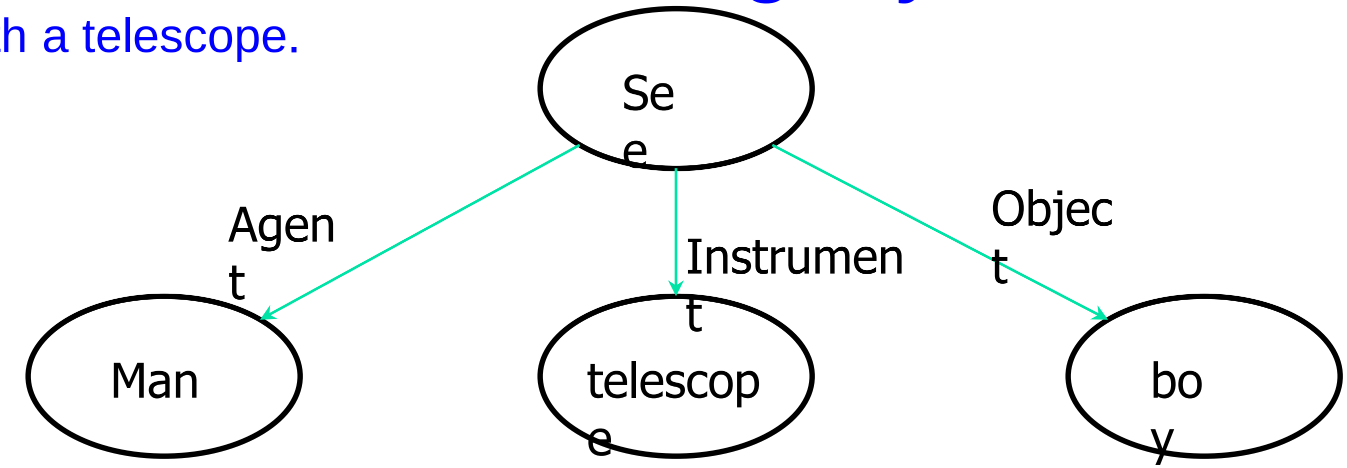
Word Sense Disambiguation-perspective

WSD (multiple options for word meaning)

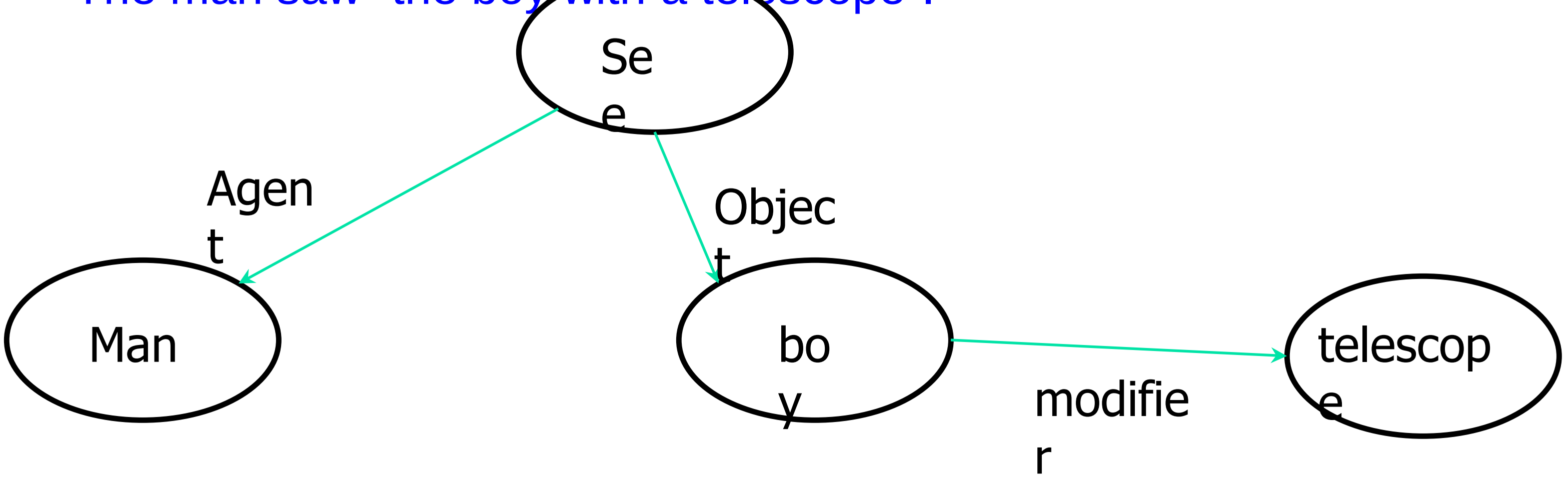
The man saw the boy with a telescope.



Semantic Role Ambiguity: The man saw “the boy” with a telescope.



The man saw “the boy with a telescope”.



Ambiguity

- Referential ambiguity
  - The dog chased a cat and it bit it.
- Preposition ambiguity
  - Ram ate some rice **with** vegetables.
  - Ram ate some rice **with** soon.
  - Ram ate some rice **with** Shayam.

Origin of a polysemous word

- How a word gets multiple sense in a language?
  - Metaphoric use of a word ('fruit fell' vs. 'kingdom fell')
  - Language contact ('bank' from German 'Banque')

Modeling of WSD- sense *S* given word *W* and context *C*

$$S^* = \arg\max_S P(S \mid w, C) \qquad w \in C$$

$$P(S \mid w, C) = \frac{\#(w\_tagged\_as\_S\_in\_context\ C)}{\#(w\_in\_context\ C)}$$

Isolate "*prior*" probablity

$$
\begin{aligned}
P(S \mid w, C) &= P(S, w, C) \\
&= \frac{P(w).P(S, C \mid w)}{P(w).P(C \mid w)} \\
&= \frac{P(S, C \mid w)}{P(C \mid w)} \\
&= \frac{P(S \mid w).P(C \mid S, w)}{P(C \mid w)}
\end{aligned}
$$

Constant in *argmax* calculation

$$S^* = \arg\max_S P(S \mid w, C) = \arg\max_S P(S \mid w).P(C \mid S, w)$$

Prio r

$$P(S \mid w) = \frac{\#(w\_tagged\_as\_S)}{\# w}$$

Likeliho od

$$P(C \mid S, w) = P(w_{c1} w_{c2} w_{c3} \cdots w_{cn} \mid S, w)$$

Let ***W^S^* = *W*** **in sense *S***

Apply chain rule

$$= P(w_{c1} \mid W^S).P(w_{c2} \mid w_{c1}, W^S) \cdots P(w_{cn} \mid w_{cn-1} .. w_{c1}, W^S)$$

Make Markov assumption (*Bigram*)

$$= P(w_{c1} \mid W^S).P(w_{c2} \mid w_{c1}, W^S) \cdots P(w_{cn} \mid w_{cn-1} W^S)$$

Two sentences

- "The navy performed a successful operation, the doctor said" (A)
- "The doctor performed a successful operation, the navy officer said" (B)
- Without a scoring mechanism, the system will not know what sense to output for operation, *medical/military/mathematics/cse*
- The scoring mechanism is probability
- Probability makes use of proportion of association counts (how many times doctor/navy is associated with operation)
- Even with scoring mechanism, system will go wrong in one of the examples; *A correct* □ *B wrong* and *vice versa*

Long-distance dependency

- I went to the *bank*^{Finance} to draw money.
- I went to the *bank*^{River} to draw water.

$$P(I \mid bank^{Finance}).P(went \mid I, bank^{Finance}) \square\ P(money \mid draw, bank^{Finance})$$

$$P(I \mid bank^{River}).P(went \mid I, bank^{River}) \square\ P(water \mid draw, bank^{River})$$

$$P(I, bank^{Finance}) = \frac{\# < I, bank^{Finance} >}{\# < bank^{Finance} >} \qquad P(went \mid I, bank^{Finance}) = \frac{\# < I, went, bank^{Finance} >}{\# < T, bank^{Finance} >}$$

$$P(I, bank^{River}) = \frac{\# < I, bank^{River} >}{\# < bank^{River} >} \qquad P(went \mid I, bank^{River}) = \frac{\# < I, went, bank^{River} >}{\# < T, bank^{River} >}$$

Annotation

- We need sense-tagged corpus
- WordNET [https://wordnet.princeton.edu/]
  - Bank
    - 10 Noun senses
      - Bank_883: Financial institution
      - Bank_99: Sloppy land
      - …
    - 8 Verb senses

Rule based approach

- If “water” appears in the context of “bank” then
  - “bank” most likely has the sense “River bank”

- If “money” appears in the context of “bank” then
  - “bank” most likely has the sense “Financial bank”

Decision List

- Decision list are clusters of rules

```
ContextOf(Bank)
    case "water" : □  River bank
    case "sand"  : □  River bank
    case "money" : □  Financial
  bank
```

...

How many such rules ?

Assume English has 100K words and on an average there are 3 senses per word.

#Rules > 300K

Robustness of rule-based systems

- Decision lists to be “complete”- Is that enough for WSD?

- No!
  - Need MA- If morphological form of “water” appears e.g. “watering”
  - Need POS tagging- POS tag: E.g. “He waters plants in front of a bank every week.”
  - Unrelated reference- “I went to bank to draw money and heard someone shouting ‘water’.”

Machine Learning for WSD

- HMM– Captures local context only (not adequate for WSD)
- Condition Random Field (CRF)– Captures long distance dependency
- Neural Network- Captures long distance dependency

Feature Abstraction in Deep learning

What is the intuition behind deep learning?

Deep Learning **discovers** “water” as important features for disambiguation, instead of being ***given*** that feature

1. Closeness between “water” and “river” in word embeddings
2. Using clustering formation:
   - Some neurons will produce a cluster whose members share a common feature which represents ‘water’
   - However, the neurons will not be able to put a name to the feature

Essential Resource for WSD: ***Wordnet***

Word Meanings	Word Forms				
	F_1	**F_2**	**F_3**	**…**	**F_n**
M_1	(*depend*) $E_{1,1}$	(*bank*) $E_{1,2}$	(rely) $E_{1,3}$		
M_2		(*bank*) $E_{2,2}$		(*embankment*) $E_{2,\ldots}$	
M_3		(*bank*) $E_{3,2}$	$E_{3,3}$		
…				…	
M_m					$E_{m,n}$

Example of sense marking: its need

एक_4187 नए शोध_1138 के अनुसार_3123 जिन लोगों_1189 का सामाजिक_43540 जीवन_125623 व्यस्त_48029 होता है उनके दिमाग_16168 के एक_4187
हिस्से_120425 में अधिक_42403 जगह_113368 होती है।

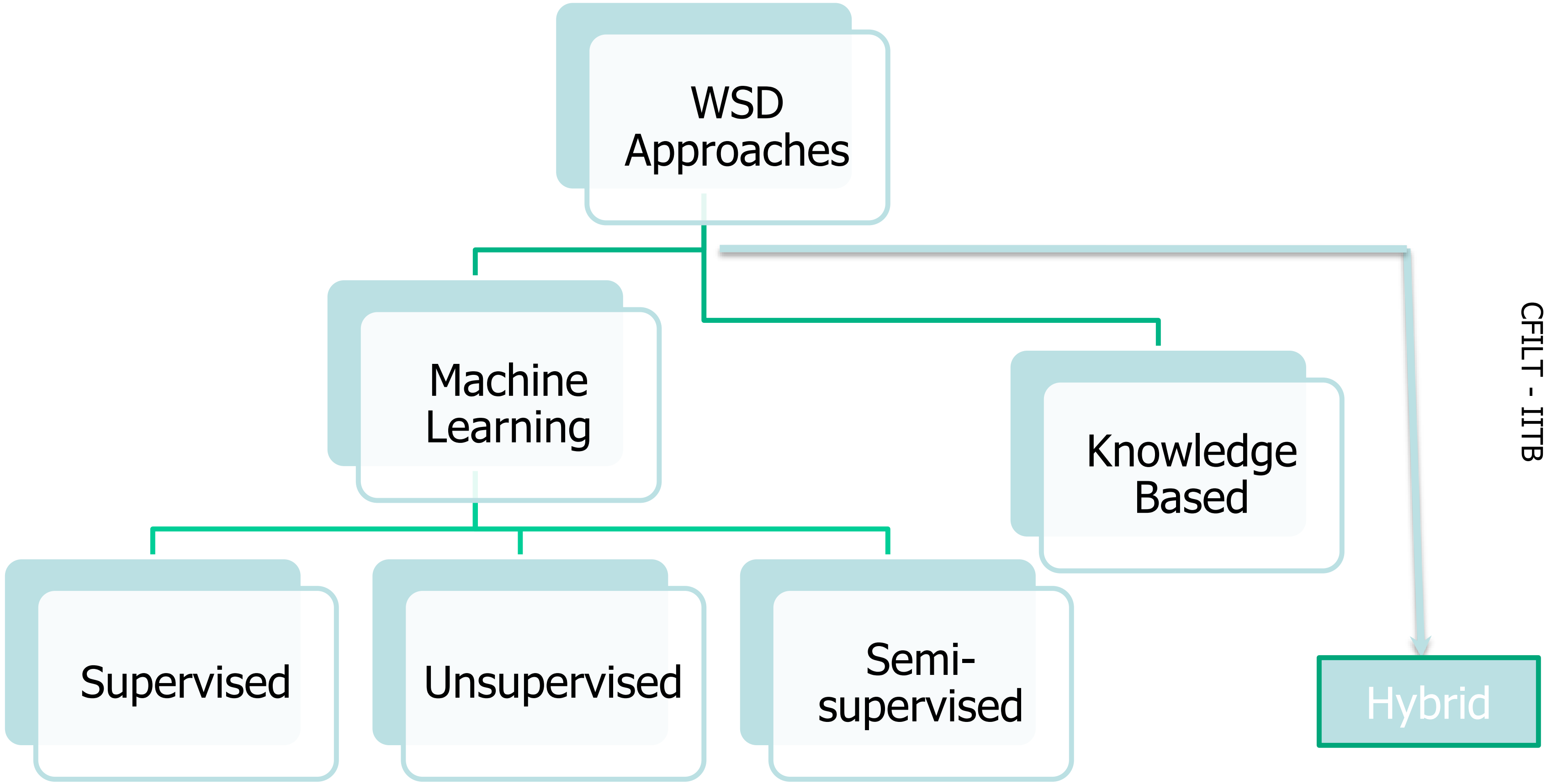
(According to a new research, those people who have a busy social life, have larger space in a part of their brain).

नेचर न्यूरोसाइंस में छपे एक_4187 शोध_1138 के अनुसार_3123 कई_4118 लोगों_1189 के दिमाग_16168 के स्कैन से पता_11431 चला कि दिमाग_16168 का एक_4187 हिस्सा_120425 एमिगडाला सामाजिक_43540 व्यस्तताओं_1438 के साथ_328602 सामंजस्य_166
के लिए थोड़ा_38861 बढ़_25368 जाता है। यह शोध_1138 58 लोगों_1189 पर किया गया जिसमें उनकी उम्र_13159 और दिमाग_16168 की साइज़ के आँकड़े_128065
लिए गए। अमरीकी_413405 टीम_14077 ने पाया_227806 कि जिन लोगों_1189 की सोशल नेटवर्किंग अधिक_42403 है उनके दिमाग_16168 का एमिगडाला
वाला हिस्सा_120425 बाकी_130137 लोगों_1189 की तुलना_में_38220 अधिक_42403 बड़ा_426602 है। दिमाग_16168 का एमिगडाला वाला हिस्सा_120425
भावनाओं_1912 और मानसिक_42151 स्थिति_1652 से जुड़ा हुआ माना_212436 जाता है।

Ambiguity of लोगों (People)

- **लोग**, **जन**, **लोक**, **जनमानस**, **पब्लिक** - एक से अधिक व्यक्ति *"लोगों के हित में काम करना चाहिए"*
  - **(English synset) multitude, masses, mass, hoi_polloi, people, the_great_unwashed - the common people generally *"separate the warriors from the mass" "power to the people"***
- **दुनिया**, **दुनियाँ**, **संसार**, **विश्व**, **जगत**, **जहाँ**, **जहान**, **ज़माना**, **जमाना**, **लोक**, **दुनियावाले**, **दुनियाँवाले**, **लोग** - संसार में रहने वाले लोग *"महात्मा गाँधी का सम्मान पूरी दुनिया करती है / मैं इस दुनिया की परवाह नहीं करता / आज की दुनिया पैसे के पीछे भाग रही है"*
  - **(English synset) populace, public, world - people in general considered as a whole *"he is a hero in the eyes of the public”***

Bird’s eye view of WSD



What is expected of *Course Project*

Activities in the paper

- You will take a problem
- Implement ONE paper (may be allied papers will support)
- Prepare the demo well
- DO result presentation and error analysis

Example

- Say, POS tagging for Malayalam
- Understand complexities
- Take a paper that gives the best results for Malayalam POS tagger
- Implement and see results

Wordnet

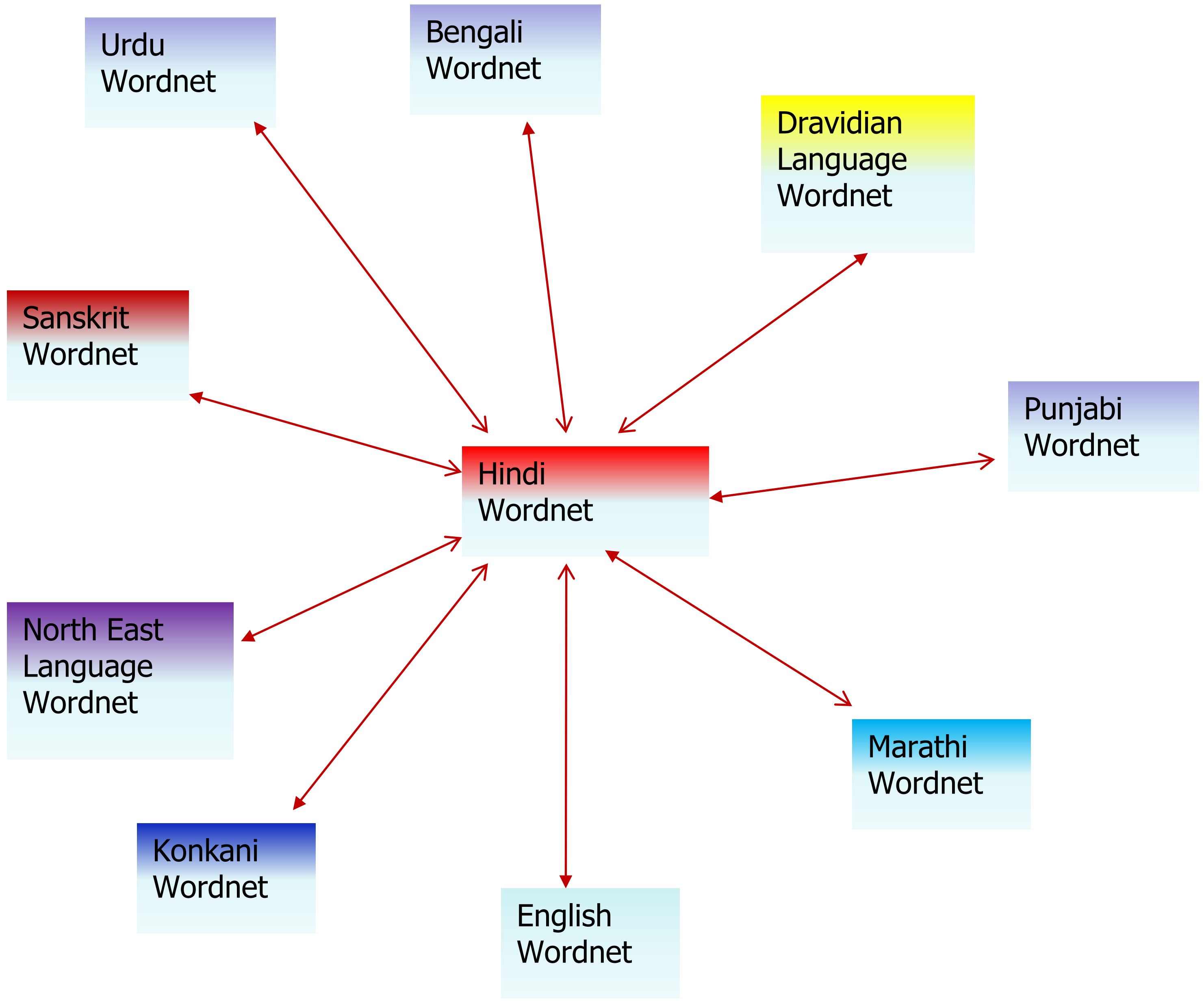
Wordnet: main purpose

- Disambiguation: **Sense Disambiguation**
- Main instrument: Relational Semantics
- Disambiguate a *by other words*
  - *{house}: "house" as a kind of "physical structure"*
  - *{family, house}: "family" as an abstract concept*
  - *{house, astrological position}: "astrological place" as a concept*

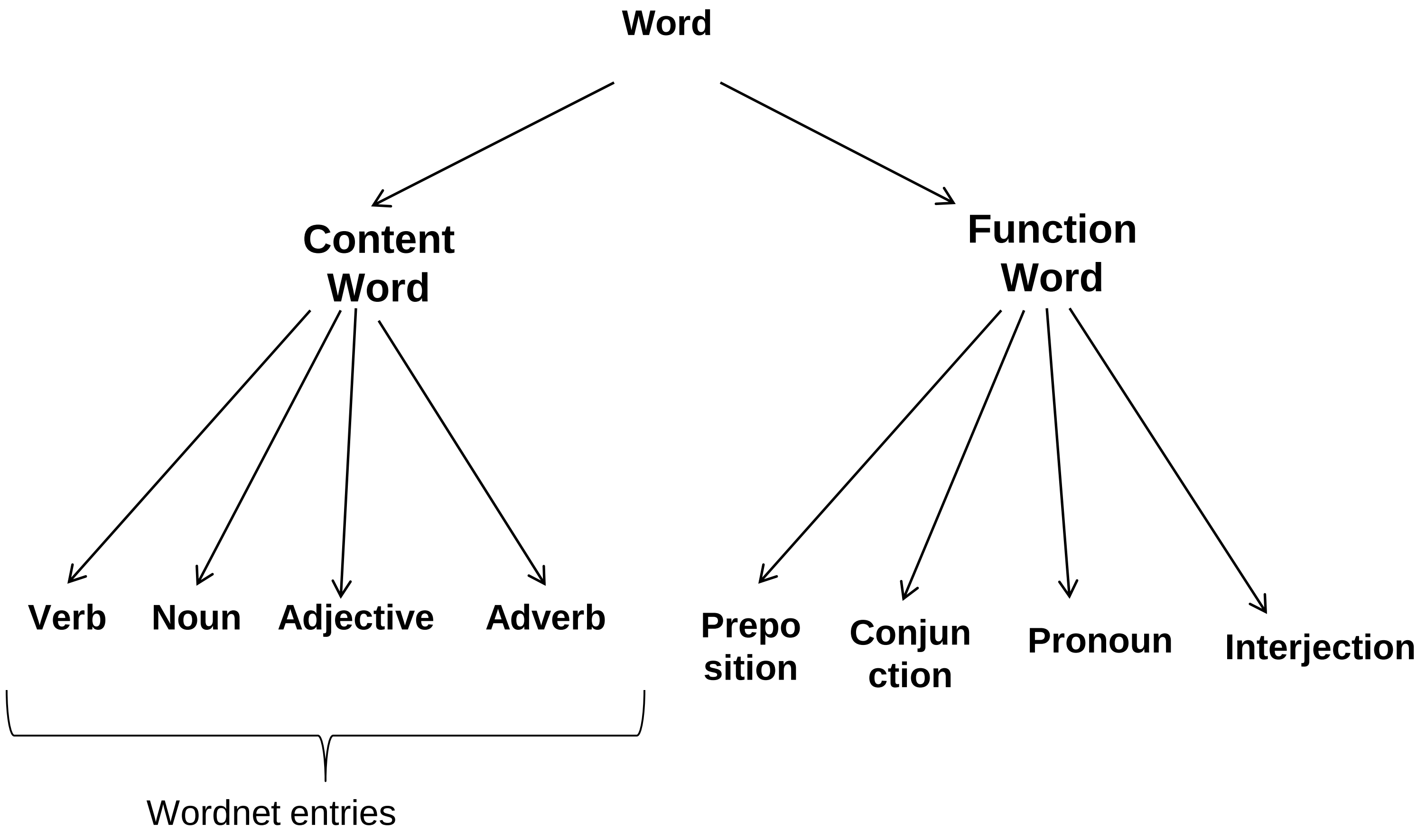
Wordnet - Lexical Matrix (with examples)

Word Meanings	Word Forms				
	F_1	**F_2**	**F_3**	**…**	**F_n**
M_1	(*depend*) $E_{1,1}$	(*bank*) $E_{1,2}$	(rely) $E_{1,3}$		
M_2		*(bank)* $E_{2,2}$		(*embankment*) $E_{2,\ldots}$	
M_3		*(bank)* $E_{3,2}$	$E_{3,3}$		
…				…	
M_m					$E_{m,n}$

INDOWORDNET



Classification of Words



Sense tagged corpora (task: sentiment analysis)

- I have enjoyed_21803158 #LA#_18933620 every_42346474 time_17209466 I have been_22629830 there_3110157 , regardless_3119663 if it was for work_1578942 or pleasure_11057430.
- I usually_3107782 fly_21922384 into #LA#_18933620, but this time_17209466 we decided_2689493 to drive_21912201 .
- Interesting_41394947, to say_2999158 the least_3112746 .

Senses of “pleasure”

The noun pleasure has 5 senses (first 2 from tagged texts)

1. (21) pleasure, pleasance -- (a fundamental feeling that is hard to define but that people desire to experience; "he was tingling with pleasure")
2. (4) joy, delight, pleasure -- (something or someone that provides pleasure; a source of happiness; "a joy to behold"; "the pleasure of his company"; "the new car is a delight")
3. pleasure -- (a formal expression; "he serves at the pleasure of the President")
4. pleasure -- (an activity that affords enjoyment; "he puts duty before pleasure")

Basic Principle

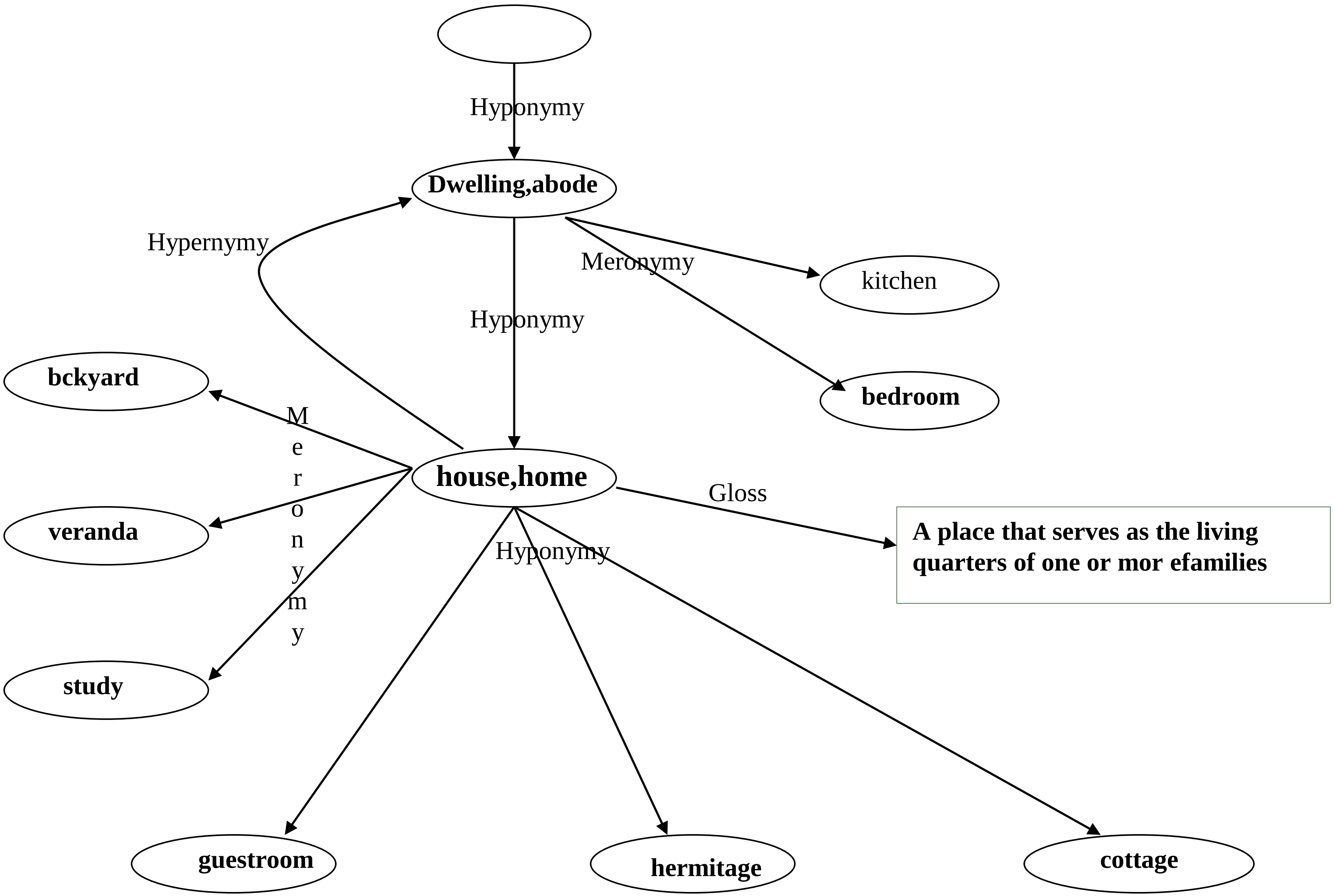
- Words in natural languages are polysemous.
- However, when synonymous words are put together, a unique meaning often emerges.
- Use is made of *Relational Semantics.*

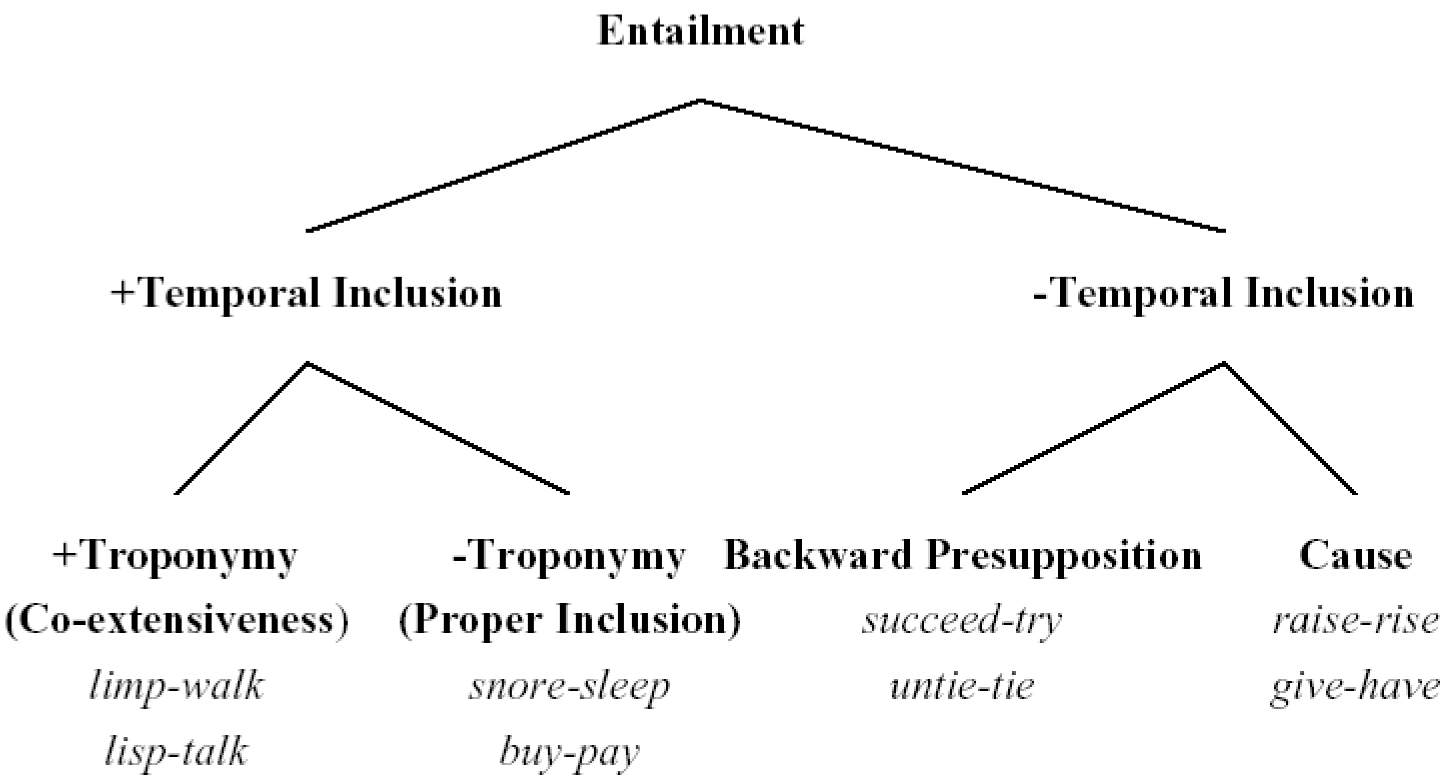
Lexical and Semantic relations in wordnet

1. Synonymy
2. Hypernymy / Hyponymy
3. Antonymy
4. Meronymy / Holonymy
5. Gradation
6. Entailment
7. Troponymy

1, 3 and 5 are lexical (*word to word)*, rest are semantic (*synset to synset).*

WordNet Sub-Graph




Entailment
+Temporal Inclusion
-Temporal Inclusion
+Troponymy
(Co-extensiveness)
limp-walk
lisp-talk
-Troponymy
(Proper Inclusion)
snore-sleep
buy-pay
Backward Presupposition
succeed-try
untie-tie
Cause
raise-rise
give-have

Principles behind creation of Synsets

Three principles:

- Minimality
- Coverage
- Replacability

Synset creation: from first principles

From first principles

- Pick all the senses from good standard dictionaries.
- Obtain synonyms for each sense.
- Needs hard and long hours of work.

Synset creation: Expansion approach

From the wordnet of another language preferably in the **same family**

- Pick the synset and obtain the sense from the gloss.
- Get the words of the target language.
- Often same words can be used- especially for words with the same etymology borrowed from the parent language in the typology.
- Translation, Insertion and deletion.

Illustration of expansion approach with noun[1]

English

- bank (sloping land (especially the slope beside a body of water)) "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"

French (*wrong!*)

- ~~banque~~ (les terrains en pente (en particulier la pente à côté d'un plan d'eau)) "ils ont tiré le canot sur la rive», «il était assis sur le bord de la rivière et j'ai vu les courants"

Illustration of expansion approach with noun²

No hypernymy in the synset

English

- bank (sloping land (especially the slope beside a body of water)) "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the '

French

- {rive, rivage, ~~bord~~} (les terrains en pente (en particulier la pente à côté d'un plan d'eau)) "ils ont tiré le canot sur la rive», «il était assis sur le bord de la rivière et j'ai vu les courants"

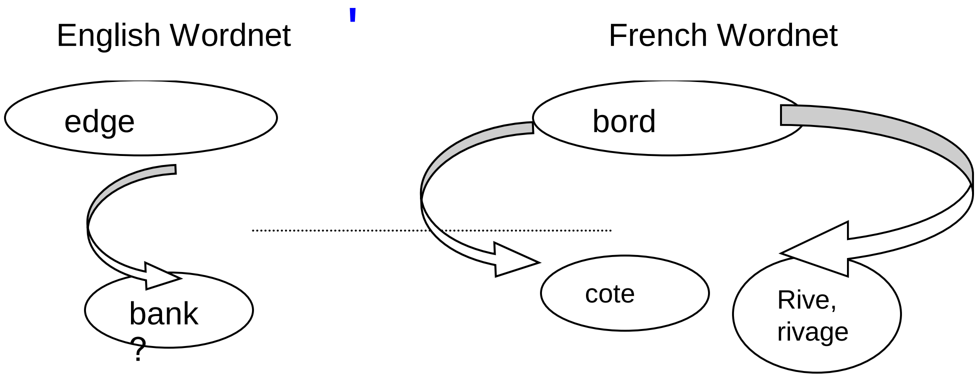


Illustration of expansion approach with verb[3]

English

- trust, swear, rely, bank (have confidence or faith in) "We can trust in God"; "Rely on your friends"; "bank on your good education"

French

- compter_sur, avoir_confiance_en, se_fier_a ’, faire_confiance_a’ (avoir confiance ou foi en) "Nous pouvons faire confiance en Dieu»,«Fiez-vous à vos amis",

Ordered by frequency

Lexical Relation

- Antonymy
  - Oppositeness in meaning
  - Relation between word forms
  - Often determined by phonetics, word length etc. ({rise, ascend} *vs.* {fall, descend})

Kinds of Antonymy

Size	**Small - Big**
Quality	**Good – Bad**
State	**Warm – Cool**
Personality	**Dr. Jekyl- Mr. Hyde**
Direction	**East- West**
Action	**Buy – Sell**
Amount	**Little – A lot**
Place	**Far – Near**
Time	**Day - Night**
Gender	**Boy - Girl**

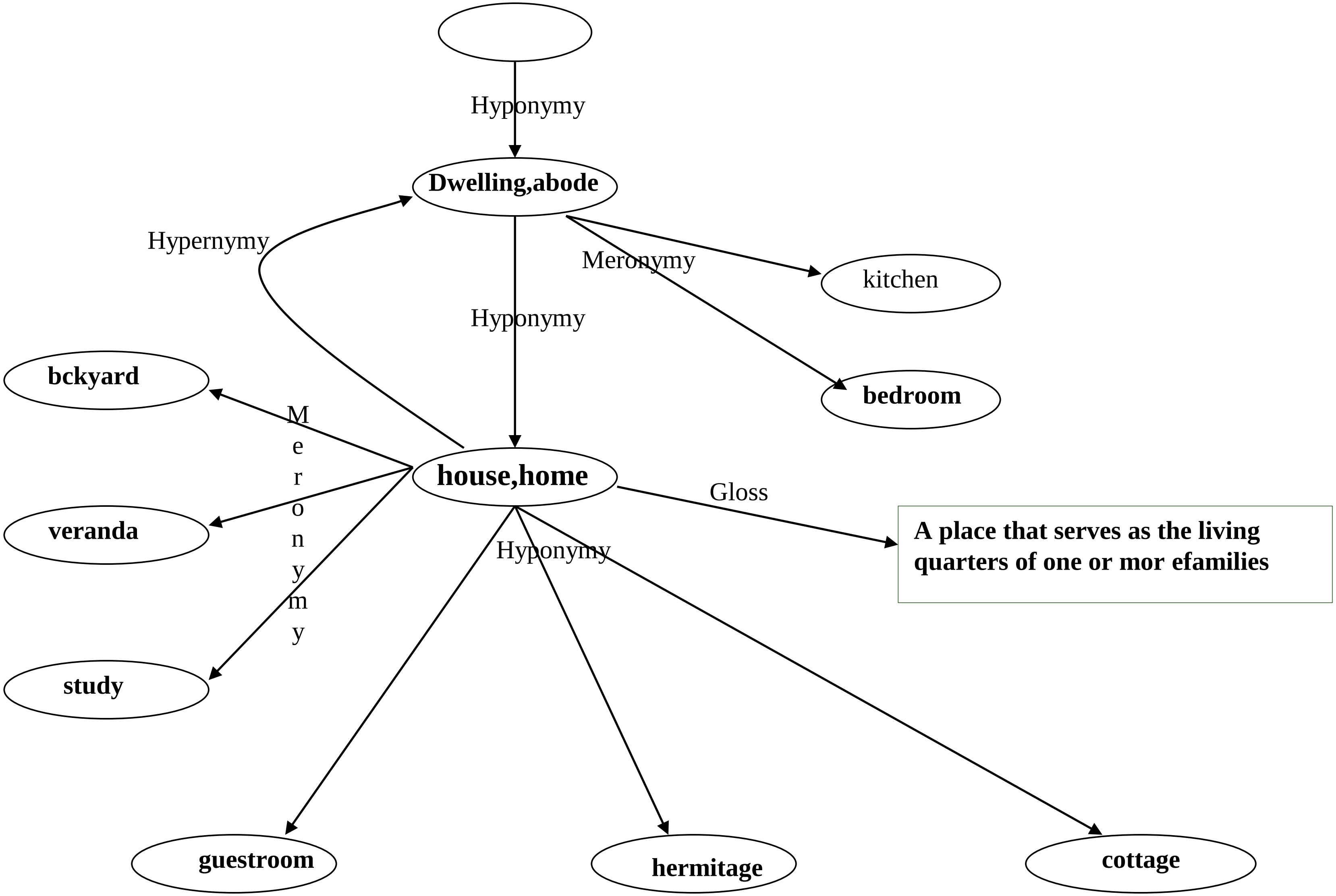
Kinds of Meronymy

Component-object	**Head - Body**
Staff-object	**Wood - Table**
Member-collection	**Tree - Forest**
Feature-Activity	**Speech - Conference**
Place-Area	**Palo Alto - California**
Phase-State	**Youth - Life**
Resource-process	**Pen - Writing**
Actor-Act	**Physician - Treatment**

Gradation

State	Childhood, Youth, Old age
Temperature	Hot, Warm, Cold
Action	Sleep, Doze, Wake

WordNet Sub-Graph



Metonymy

- Associated with *Metaphors* which are epitomes of semantics
- Oxford Advanced Learners Dictionary definition: “The use of a word or phrase to mean something different from the literal meaning”
- *Does it mean Careless Usage?!*

WordNet: limitations

- Contains little syntactic information
- No explicit predicate argument structures
- No systematic extension of basic senses
- Sense distinctions are very fine-grained, **IAA 73%**
- No hierarchical entries