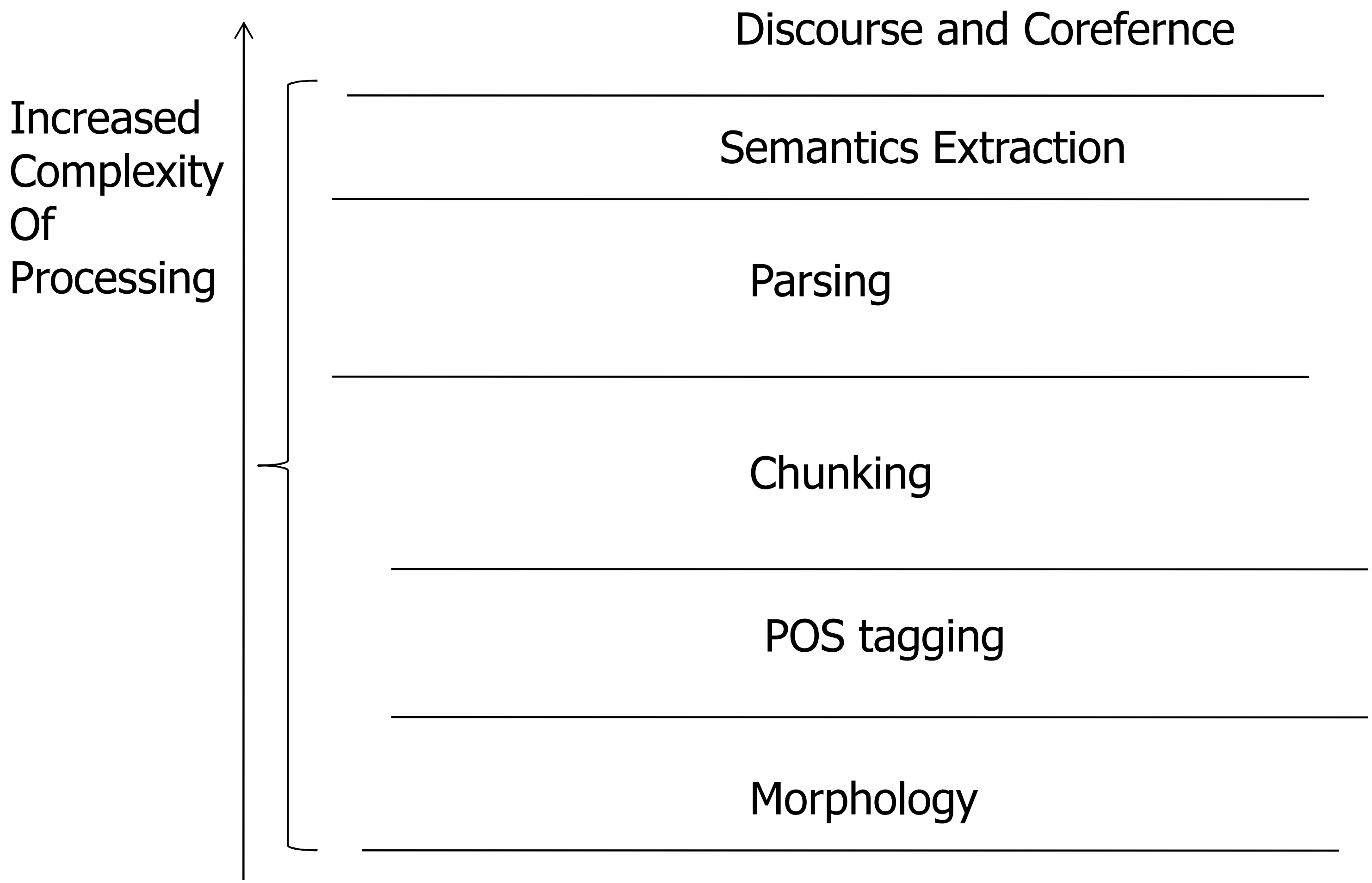
CS460/626 : Natural Language Processing/Speech, NLP and the Web

Lecture 27: Wordnet Relations and Word Sense Disambiguation Approaches; Metonymy

Pushpak Bhattacharyya
CSE Dept.,
IIT Bombay

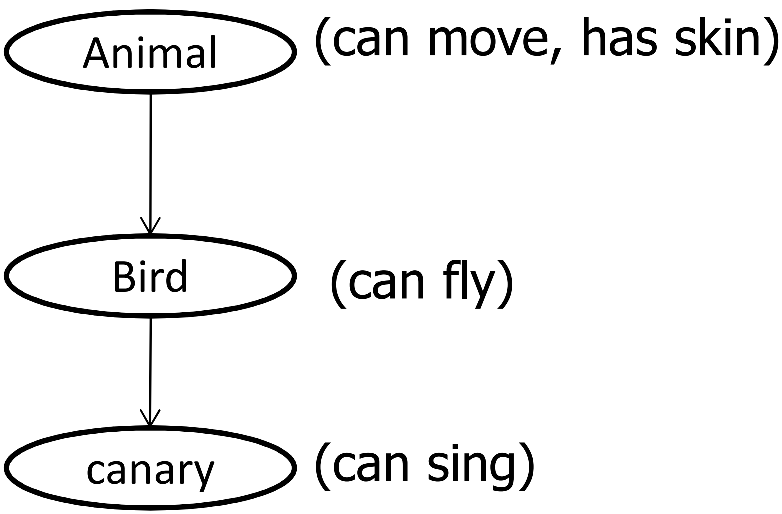
25th Oct, 2012

NLP Layer


Discourse and Corefernce
Semantics Extraction
Parsing
Chunking
POS tagging
Morphology
Increased Complexity Of Processing

Psycholinguistic Theory

- Human lexical memory for nouns as a hierarchy.
- *Can canary sing? - Pretty fast response.*
- *Can canary fly? - Slower response.*
- *Does canary have skin? – Slowest response.*



Wordnet - a lexical reference system based on psycholinguistic theories of human lexical memory.

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					$E_{m,n}$

Wordnet: History

- The first wordnet in the world was for English developed at Princeton over 15 years.
- The Eurowordnet- linked structure of European language wordnets was built in 1998 over 3 years with funding from the EC as a a mission mode project.
- Wordnets for Hindi and Marathi being built at IIT Bombay are amongst the first IL wordnets.
- All these are proposed to be linked into the **IndoWordnet** which eventually will be linked to the English and the Euro wordnets.

Basic Principle

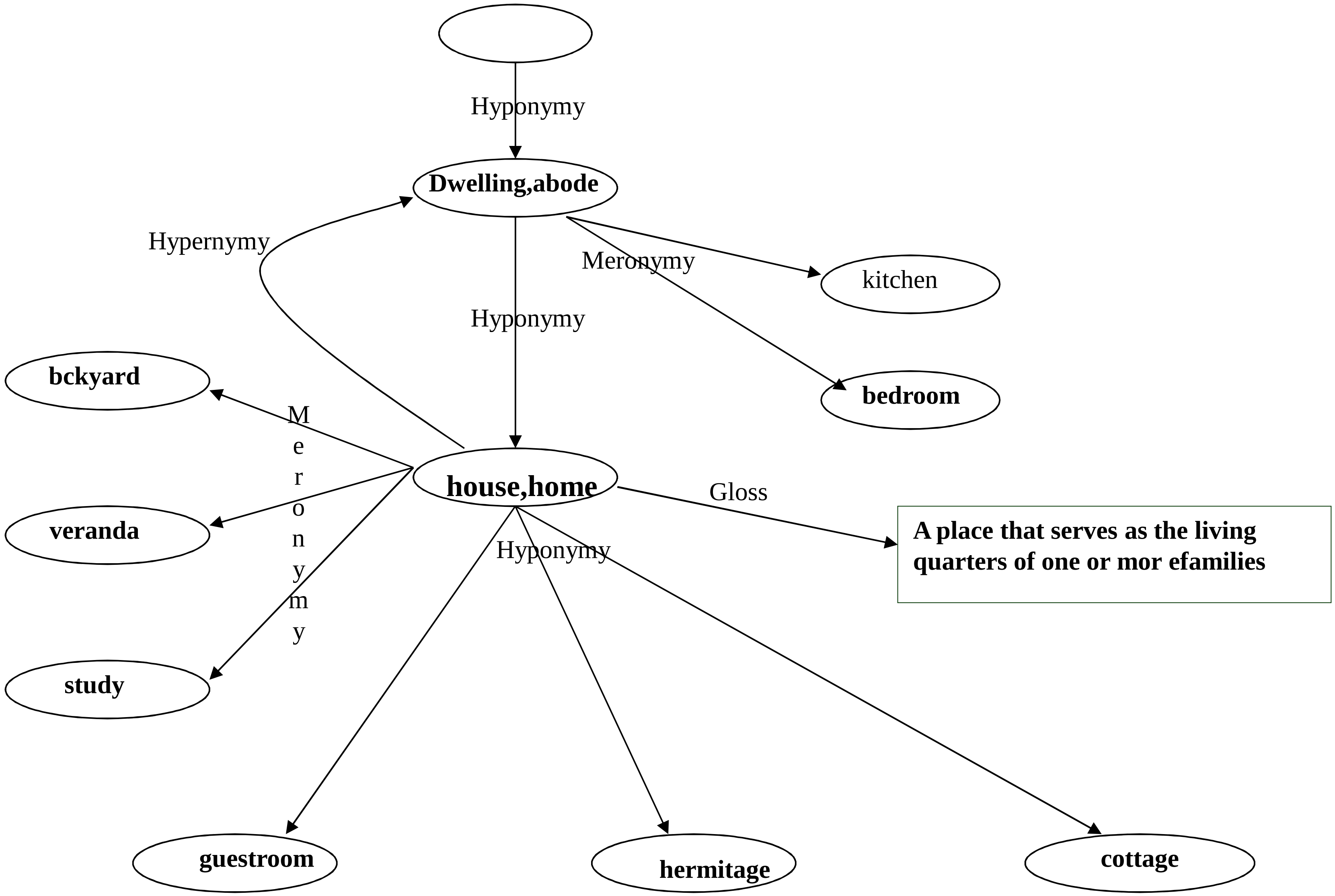
- Words in natural languages are polysemous.
- However, when synonymous words are put together, a unique meaning often emerges.
- Use is made of *Relational Semantics*.

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



Fundamental Design Question

- **Syntagmatic** *vs.* **Paradigmatic relations?**
- Psycholinguistics is the basis of the design.
- When we hear a word, many words come to our mind *by association*.
- For English, about half of the associated words are *syntagmatically related* and half are *paradignatically related.*
- For *cat*
  - *animal, mammal*- paradigmatic
  - *mew, purr, furry*- syntagmatic

Stated Fundamental Application of Wordnet: *Sense Disambiguation*

Determination of the correct sense of the word

The crane ate the fish vs.

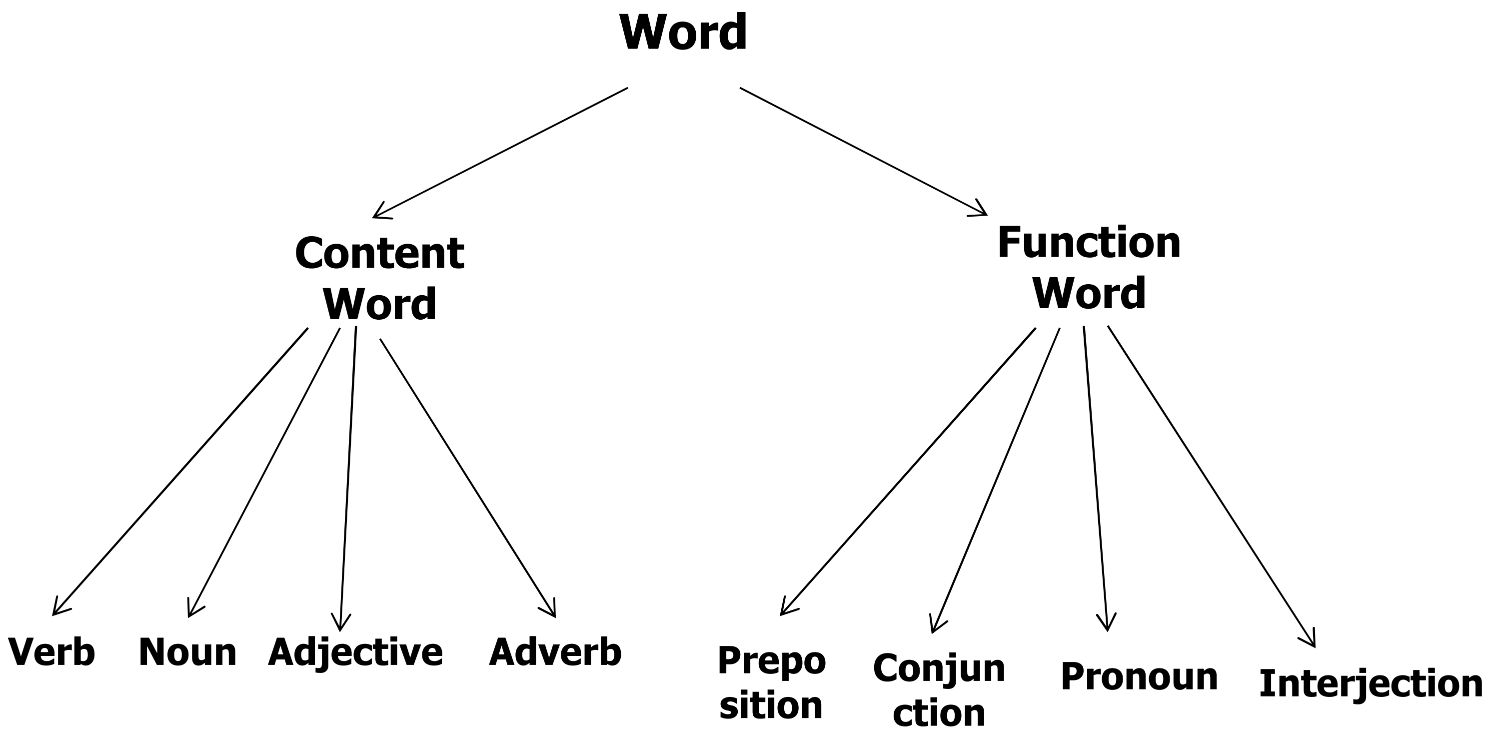
The crane was used to lift the load

bird vs. *machine*

The problem of Sense tagging

- Given a corpora **To Assign correct sense to the words.**
- This is sense tagging. Needs **Word Sense Disambiguation (WSD)**
- **Highly important for Question Answering, Machine Translation, Text Mining tasks.**

Classification of Words



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"***

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.*
- *Componential Semantics* where each word is a **bundle of semantic features** (as in the Schankian *Conceptual Dependency* system or *Lexical Componential Semantics)* is to be examined as a viable alternative.

Componential Semantics

- Consider *cat* and *tiger.* Decide on *componential attributes.*

Furry	Carnivorous	Heavy	Domesticable

- For *cat* (Y, Y, N, Y)
- For *tiger* (Y,Y,Y,N)

  Complete and correct Attributes are difficult to design.

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).*

Synset: the *foundation* (*house*)

1. **house** -- (a dwelling that serves as living quarters for one or more families; "he has a house on Cape Cod"; "she felt she had to get out of the house")
2. **house** -- (an official assembly having legislative powers; "the legislature has two houses")
3. **house** -- (a building in which something is sheltered or located; "they had a large carriage house")
4. family, household, **house**, home, menage -- (a social unit living together; "he moved his family to Virginia"; "It was a good Christian household"; "I waited until the whole house was asleep"; "the teacher asked how many people made up his home")
5. theater, theatre, **house** -- (a building where theatrical performances or motion-picture shows can be presented; "the house was full")
6. firm, **house**, business firm -- (members of a business organization that owns or operates one or more establishments; "he worked for a brokerage house")
7. **house** -- (aristocratic family line; "the House of York")
8. **house** -- (the members of a religious community living together)
9. **house** -- (the audience gathered together in a theatre or cinema; "the house applauded"; "he counted the house")
10. **house** -- (play in which children take the roles of father or mother or children and pretend to interact like adults; "the children were playing house")
11. sign of the zodiac, star sign, sign, mansion, **house**, planetary house -- ((astrology) one of 12 equal areas into which the zodiac is divided)
12. **house** -- (the management of a gambling house or casino; "the house gets a percentage of every bet")

Creation of Synsets

Three principles:

- Minimality
- Coverage
- Replacability

Synset creation (continued)

Home

John's home was decorated with lights on the occasion of Christmas.

Having worked for many years abroad, John Returned home.

House

John's house was decorated with lights on the occasion of Christmas.

Mercury is situated in the eighth house of John's horoscope.

Synsets (continued)

{house} is ambiguous.

{house, home} has the sense of *a social unit living together;*

Is this the minimal unit?

{family, house , home} will make the unit completely unambiguous.

For coverage:

{family, household, house, home} ordered according to frequency.

Replacability of the most frequent words is a requirement.

Synset creation

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 (continued)

From the wordnet of another language 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.
- Translation, Insertion and deletion.

Synset+Gloss+Example

Crucially needed for concept explication, wordnet building using another wordnet and wordnet linking.

English Synset: {earthquake, **quake**, temblor, seism} -- (shaking and vibration at the surface of the earth resulting from underground movement along a fault plane of from volcanic activity)

Hindi Synset: भूकंप, भूचाल, भूडोल, जलजला, भूकम्प, भू-कंप, भू-कम्प, ज़लज़ला, भूमिकंप, भूमिकम्प - प्राकृतिक कारणों से पृथ्वी के भीतरी भाग में कुछ उथल-पुथल होने से ऊपरी भाग के सहसा हिलने की क्रिया *"२००१ में गुज़रात में आये भूकंप में काफ़ी लोग मारे गये थे"*

(shaking of the surface of earth; many were killed in the earthquake in Gujarat)

Marathi Synset: धरणीकंप,भूकंप - पृथ्वीच्या पोटात द्रव्यक्षोभ होऊन पृष्ठभाग हालण्याची क्रिया *"२००१ साली गुजरातमध्ये झालेल्या धरणीकंपात अनेक लोक मृत्युमुखी पडले"*

Semantic Relations

- Hypernymy and Hyponymy
  - Relation between word senses (*synsets*)
  - X is a hyponym of Y *if* X is a kind of Y
  - Hyponymy is transitive and asymmetrical
  - Hypernymy is inverse of Hyponymy

  (lion->animal->animate entity->entity)

Semantic Relations (continued)

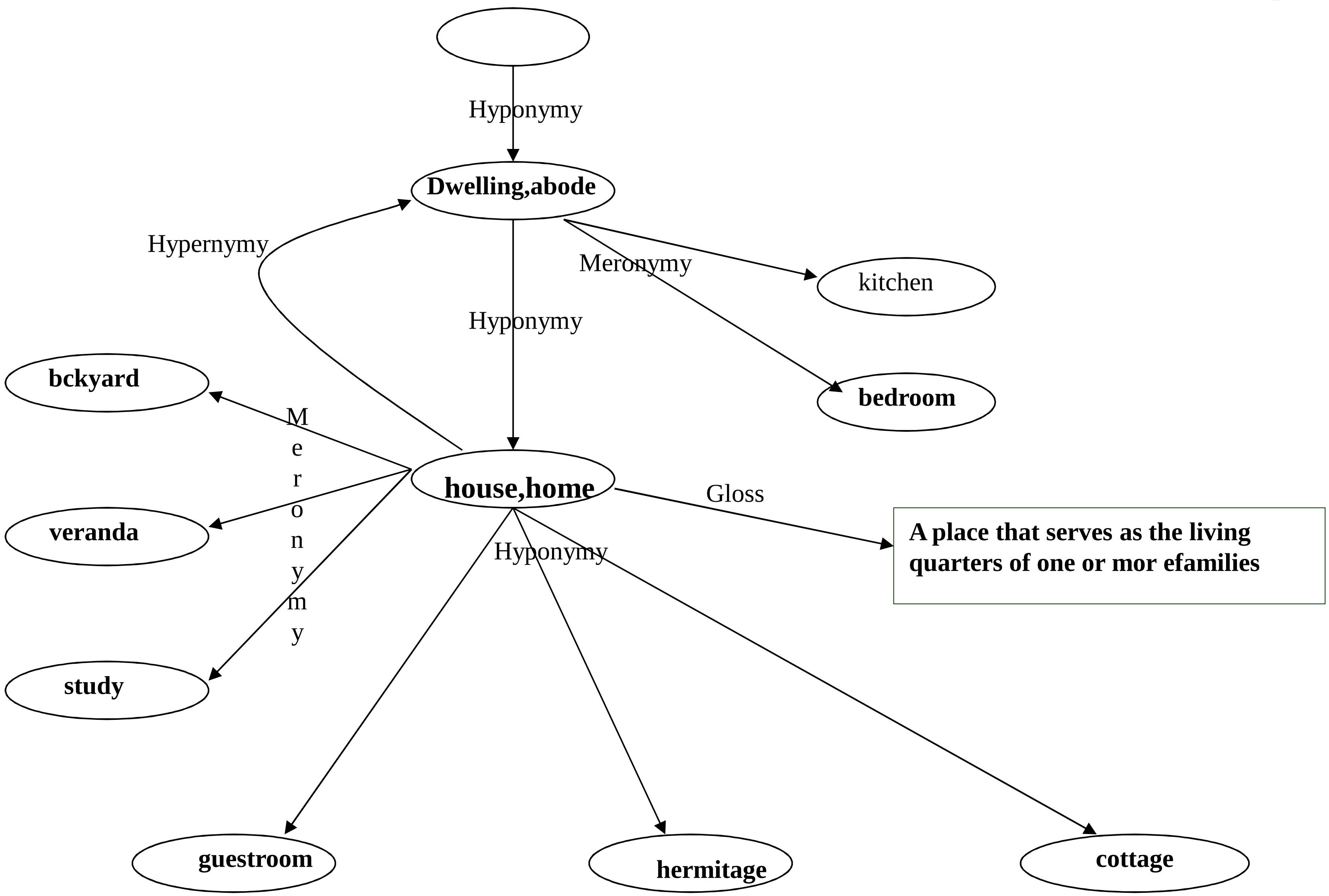
- Meronymy and Holonymy
  - Part-whole relation, branch is a part of tree
  - X is a meronymy of Y *if* X is a part of Y
  - Holonymy is the inverse relation of Meronymy

{kitchen} .... *{house}*

Lexical Relation

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

WordNet Sub-Graph



Troponym and Entailment

- Entailment

  {snoring – sleeping}

- Troponym

  {limp, strut – walk}

  {whisper – talk}

Entailment

Snoring entails sleeping.

Buying entails paying.

- Proper Temporal Inclusion.

  Inclusion can be in any way.

  Sleeping temporally includes snoring.

  Buying temporally includes paying.

- Co-extensiveness. (Troponymy)

  Limping is a manner of walking.

Opposition among verbs.

- {Rise,ascend} {fall,descend}
  Tie-untie (do-undo)
  Walk-run (slow,fast)
  Teach-learn (same activity different perspective)
  Rise-fall (motion upward or downward)

- Opposition and Entailment.
  Hit or miss (entail aim) . Backward presupposition.
  Succeed or fail (entail try.)

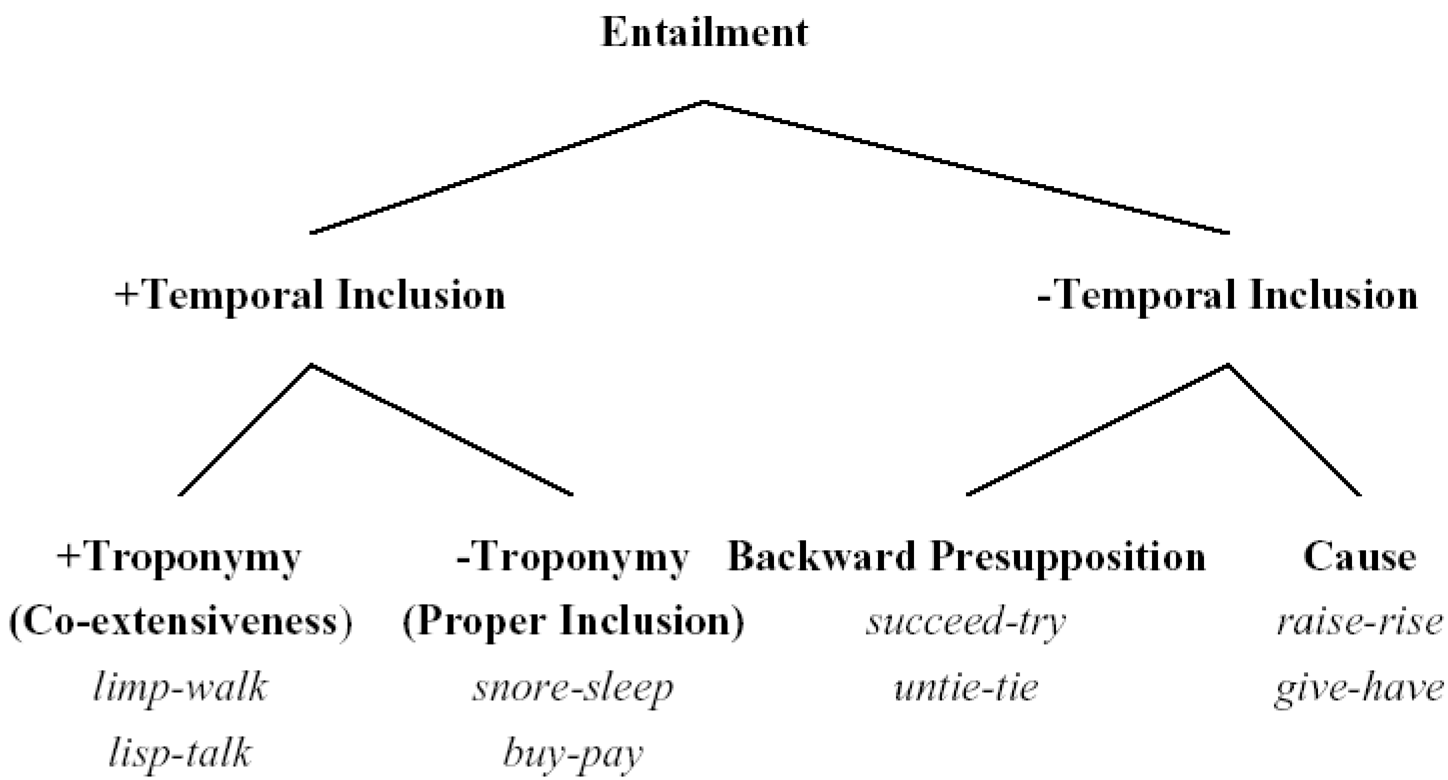
The causal relationship.

Show- see.

Give- have.

Causation and Entailment.

- Giving entails having.
- Feeding entails eating.


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

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

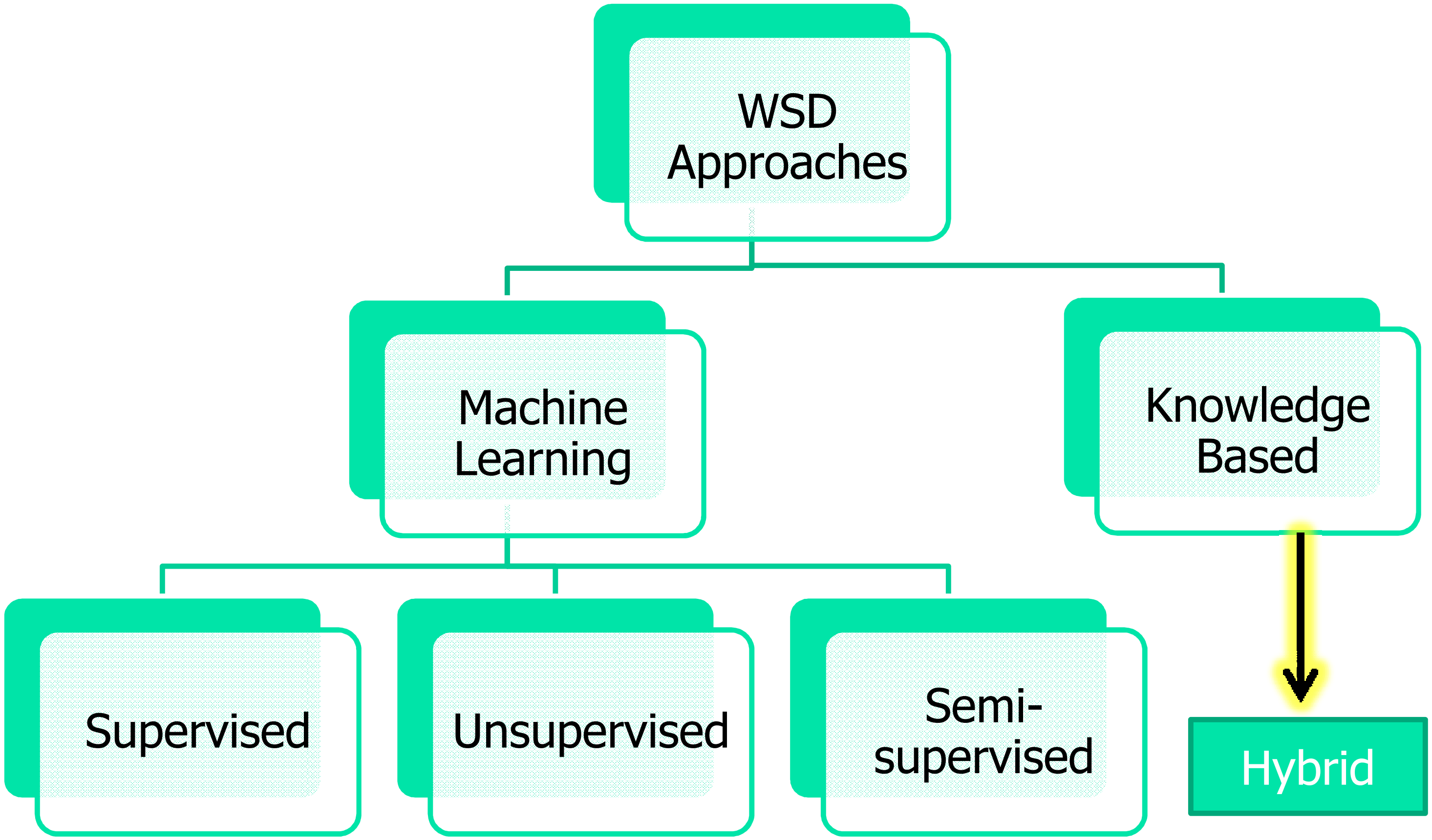
Component-object	**Head - Body**
Staff-object	**Wood - Table**
Member-collection	**Tree - Forest**
Feature-Activity	**Speech - Conference**
Place-Area	**Palo Alto - California**
Phase-State	**Youth - Life**
Resource-process	**Pen - Writing**
Actor-Act	**Physician - Treatment**

Gradation

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

Overview of WSD techniques

Bird's eye view



OVERLAP BASED APPROACHES

- Require a ***Machine Readable Dictionary*** *(MRD).*
- Find the overlap between the features of different senses of an ambiguous word (sense bag) and the features of the words in its context (context bag).
- These features could be sense definitions, example sentences, hypernyms etc.
- The features could also be given weights.
- The sense which has the maximum overlap is selected as the contextually appropriate sense.

LESK'S ALGORITHM

<u>Sense Bag</u>: *contains the words in the definition of a candidate sense of the ambiguous word.*

<u>Context Bag</u>: *contains the words in the definition of each sense of each context word.*

E.g. "On burning ***coal*** we get ***ash***."

From Wordnet

- The noun ash has 3 senses (first 2 from tagged texts)
- 1. (2) ash -- (the residue that remains when something is burned)
- 2. (1) ash, ash tree -- (any of various deciduous pinnate-leaved ornamental or timber trees of the genus Fraxinus)
- 3. ash -- (strong elastic wood of any of various ash trees; used for furniture and tool handles and sporting goods such as baseball bats)
- The verb ash has 1 sense (no senses from tagged texts)
- 1. ash -- (convert into ashes)

CRITIQUE

- Proper nouns in the context of an ambiguous word can act as strong disambiguators.

  E.g. **"Sachin Tendulkar"** will be a strong indicator of the category **"sports".**

  Sachin Tendulkar plays **cricket.**

- Proper nouns are not present in the thesaurus. Hence this approach fails to capture the strong clues provided by proper nouns.
- <u>Accuracy</u>
  - 50% when tested on 10 highly polysemous English words.

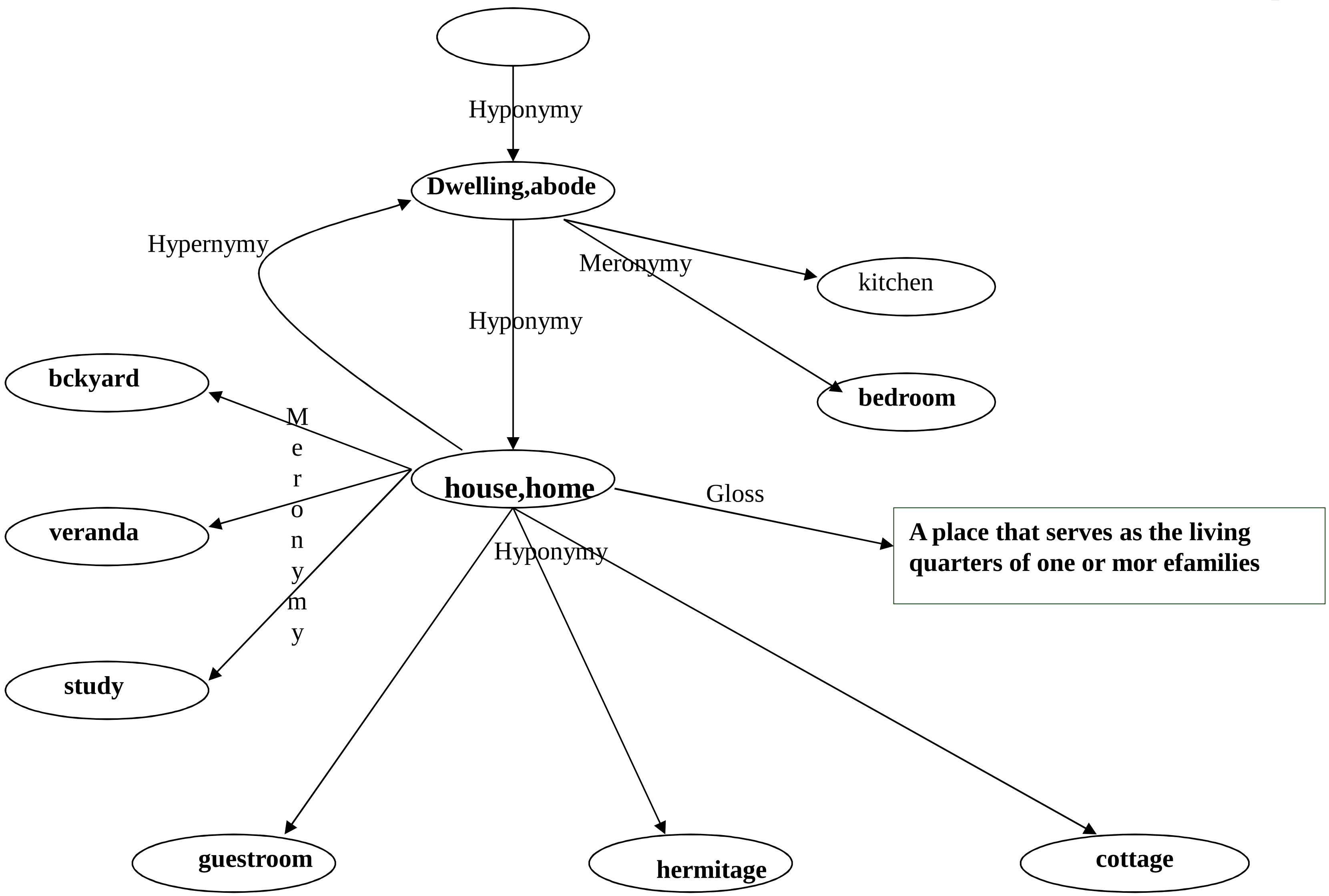
Extended Lesk's algorithm

- Original algorithm is sensitive towards exact words in the definition.
- Extension includes glosses of semantically related senses from WordNet (e.g. *hypernyms*, *hyponyms*, etc.).
- The scoring function becomes:

$$score_{ext}(S) = \sum_{s' \in rel(s) \text{ or } s \equiv s'} |context(w) \cap gloss(s')|$$

  - where,
    - *gloss(S) is the gloss of sense S from the lexical resource.*
    - *Context(W) is the gloss of each sense of each context word.*
    - *rel(s)* gives the senses related to s in WordNet under some relations.

WordNet Sub-Graph



Example: Extended Lesk

- *"On combustion of coal we get ash"*

From Wordnet

- The noun ash has 3 senses (first 2 from tagged texts)
- 1. (2) ash -- (the residue that remains when something is burned)
- 2. (1) ash, ash tree -- (any of various deciduous pinnate-leaved ornamental or timber trees of the genus Fraxinus)
- 3. ash -- (strong elastic wood of any of various ash trees; used for furniture and tool handles and sporting goods such as baseball bats)
- The verb ash has 1 sense (no senses from tagged texts)
- 1. ash -- (convert into ashes)

Example: Extended Lesk (cntd)

- *"On combustion of coal we get ash"*

From Wordnet (through hyponymy)

- ash -- (the residue that remains when something is burned)

  => fly ash -- (fine solid particles of ash that are carried into the air when fuel is combusted)

  => bone ash -- (ash left when bones burn; high in calcium phosphate; used as fertilizer and in bone china)

Critique of Extended Lesk

- Larger region of matching in WordNet
  - Increased chance of Matching

    BUT
  - Increased chance of Topic Drift

WALKER'S ALGORITHM

- A Thesaurus Based approach.
- ***Step 1:*** *For each sense of the target word find the thesaurus category to which that sense belongs.*
- ***Step 2:*** *Calculate the score for each sense by using the context words. A context word will add 1 to the score of the sense if the thesaurus category of the word matches that of the sense.*
  - *E.g. The money in this **bank** fetches an interest of 8% per annum*
  - Target word: ***bank***
  - Clue words from the context: ***money, interest, annum, fetch***

	Sense1: Finance	**Sense2: Location**
Money	+1	0
Interest	+1	0
Fetch	0	0
Annum	+1	0
Total	3	0

Context words add 1 to the sense when the topic of the word matches that of the sense

CFILT - IITB

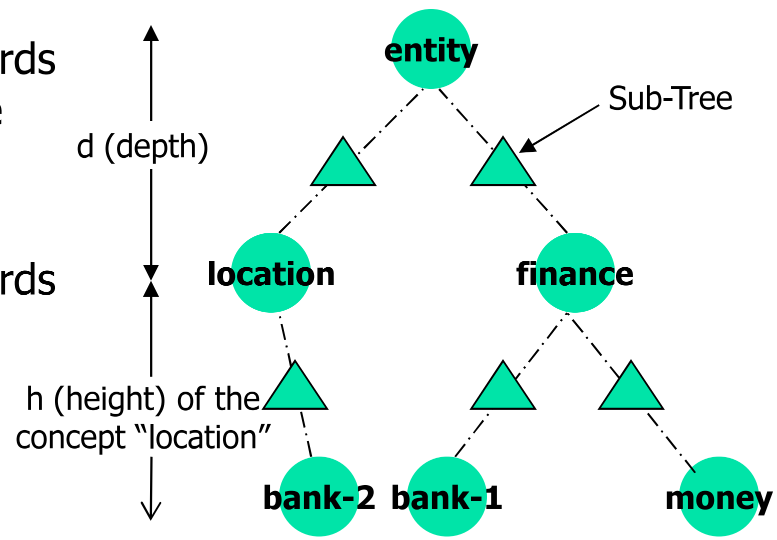
WSD USING CONCEPTUAL DENSITY *(Agirre and Rigau, 1996)*

- Select a sense based on the *relatedness* of that word-sense to the context.
- Relatedness is measured in terms of conceptual distance
  - (i.e. how close the concept represented by the ***word*** and the concept represented by its ***context words*** are)
- This approach uses a structured hierarchical semantic net (*WordNet*) for finding the conceptual distance.
- Smaller the conceptual distance higher will be the conceptual density.
  - (i.e. if all words in the context are strong indicators of a particular concept then that concept will have a higher density.)

CONCEPTUAL DENSITY FORMULA

Wish list

- The conceptual distance between two words should be proportional to the length of the path between the two words in the hierarchical tree (WordNet).
- The conceptual distance between two words should be proportional to the depth of the concepts in the hierarchy.



$$CD(c,m) = \frac{\sum_{i=0}^{m-1} nhyp^{i^{0.20}}}{descendants_c}$$

where,

c= concept

nhyp = mean number of hyponyms

h= height of the sub-hierarchy

m= no. of senses of the word and senses of context words contained in the sub-hierarchy

CD= Conceptual Density

and 0.2 is the smoothing factor

CONCEPTUAL DENSITY (cntd)

- The dots in the figure represent the senses of the word to be disambiguated or the senses of the words in context.
- The CD formula will yield highest density for the sub-hierarchy containing more senses.
- The sense of W contained in the sub-hierarchy with the highest CD will be chosen.

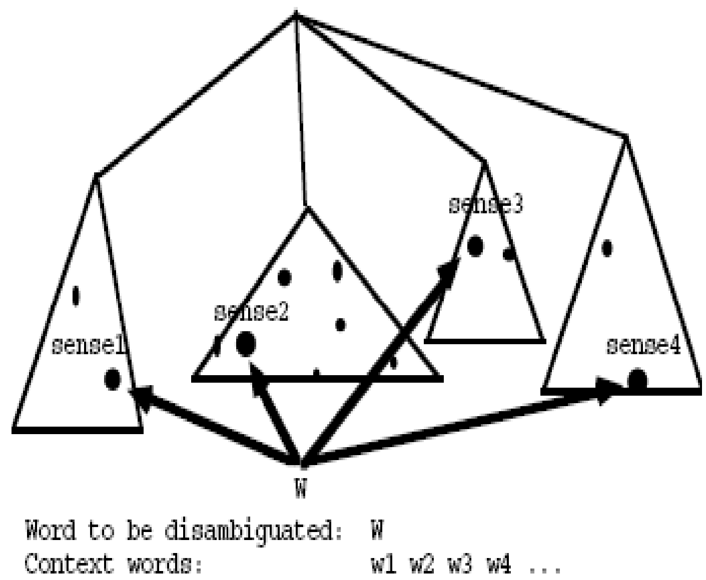
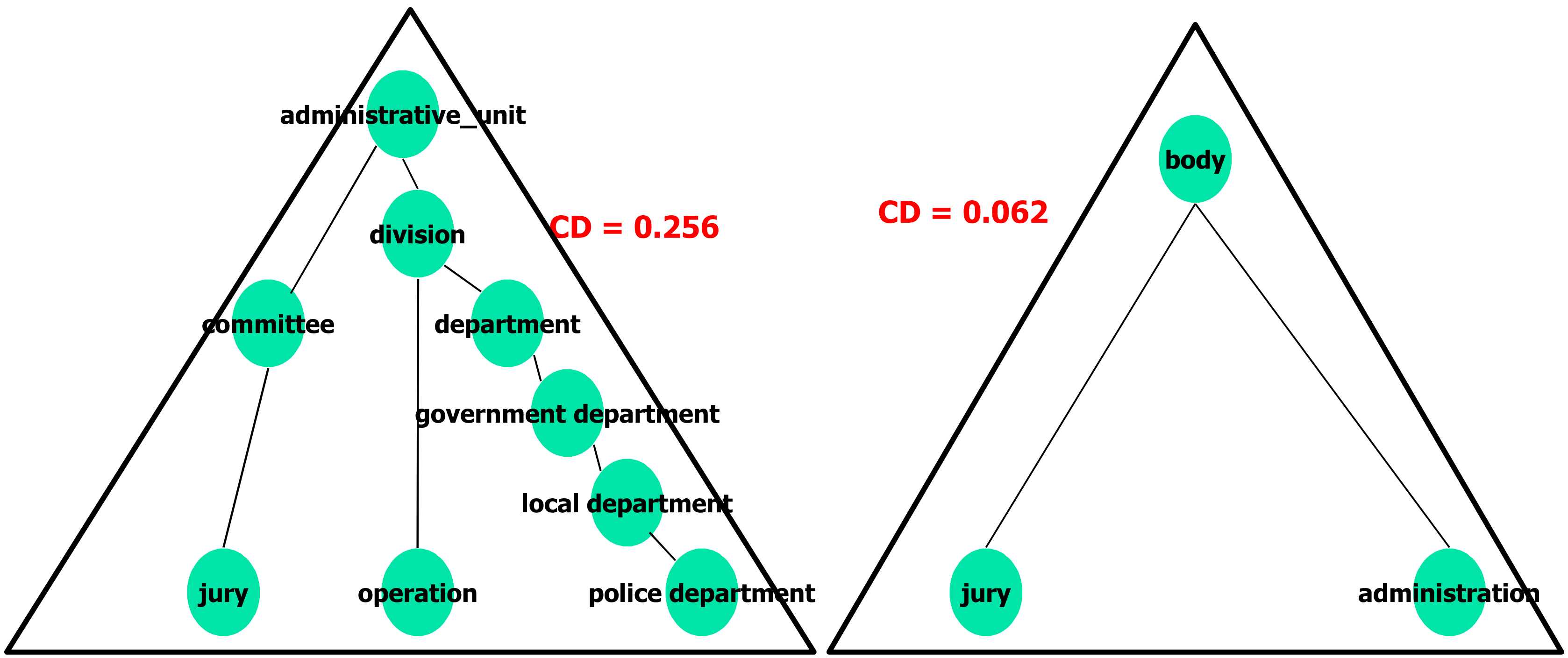


Figure 1: senses of a word in WordNet

CONCEPTUAL DENSITY (EXAMPLE)



The jury(2) praised the administration(3) and **operation** (8) of Atlanta Police Department(1)

Step 1: Make a lattice of the nouns in the context, their senses and hypernyms.

Step 2: Compute the conceptual density of resultant concepts (sub-hierarchies).

Step 3: The concept with the highest CD is selected.

Step 4: Select the senses below the selected concept as the correct sense for the respective words.

CRITIQUE

- Resolves lexical ambiguity of ***nouns*** by finding a combination of senses that maximizes the total Conceptual Density among senses.
- The Good
  - Does not require a tagged corpus.
- The Bad
  - Fails to capture the strong clues provided by proper nouns in the context.
- Accuracy
  - 54% on Brown corpus.

WSD USING RANDOM WALK ALGORITHM (Page Rank) *(sinha and Mihalcea, 2007)*

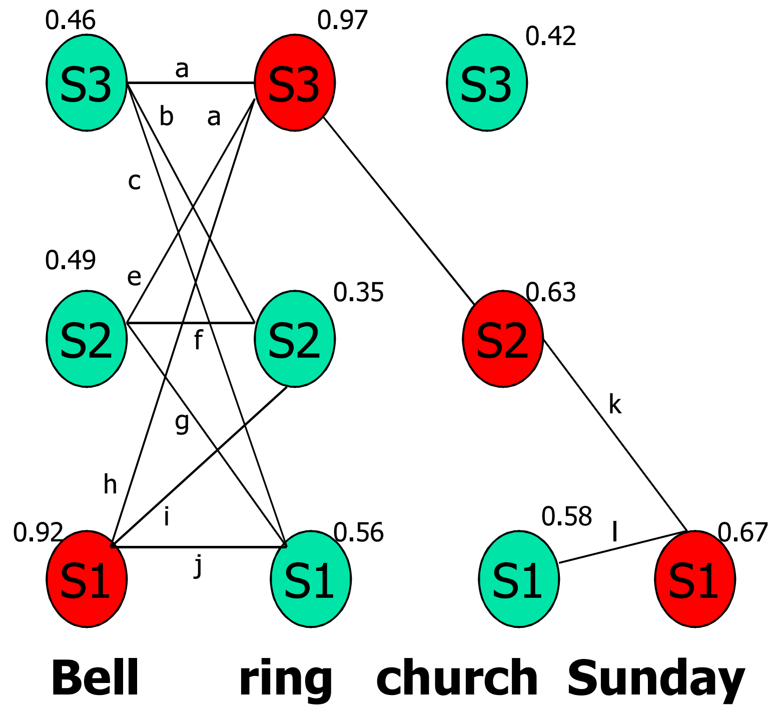
The **church bells** no longer **rung** on **Sundays**.

church
1: one of the groups of Christians who have their own beliefs and forms of worship
2: a place for public (especially Christian) worship
3: a service conducted in a church

bell
1: a hollow device made of metal that makes a ringing sound when struck
2: a push button at an outer door that gives a ringing or buzzing signal when pushed
3: the sound of a bell

ring
1: make a ringing sound
2: ring or echo with sound
3: make (bells) ring, often for the purposes of musical edification

Sunday
1: first day of the week; observed as a day of rest and worship by most Christians



Step 1: Add a vertex for each possible sense of each word in the text.

Step 2: Add weighted edges using definition based semantic similarity (Lesk's method).

Step 3: Apply graph based ranking algorithm to find score of each vertex (i.e. for each word sense).

Step 4: Select the vertex (sense) which has the highest score.

A look at Page Rank (from Wikipedia)

Developed at Stanford University by Larry Page (hence the name *Page*-Rank) and Sergey Brin as part of a research project about a new kind of search engine

The first paper about the project, describing PageRank and the initial prototype of the Google search engine, was published in 1998

Shortly after, Page and Brin founded Google Inc., the company behind the Google search engine

While just one of many factors that determine the ranking of Google search results, PageRank continues to provide the basis for all of Google's web search tools

A look at Page Rank (cntd)

PageRank is a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page.

Assume a small universe of four web pages: **A**, **B**, **C** and **D**.

The initial approximation of PageRank would be evenly divided between these four documents. Hence, each document would begin with an estimated PageRank of 0.25.

If pages **B**, **C**, and **D** each only link to **A**, they would each confer 0.25 PageRank to **A**. All PageRank **PR( )** in this simplistic system would thus gather to **A** because all links would be pointing to **A**.

PR(A)=PR(B)+PR(C)+PR(D)

This is 0.75.

A look at Page Rank (cntd)

Suppose that page **B** has a link to page **C** as well as to page **A**, while page **D** has links to all three pages

The *value of the link-votes is divided among all the outbound links on a page*.

Thus, page **B** gives a vote worth 0.125 to page **A** and a vote worth 0.125 to page **C**.

Only one third of **D**'s PageRank is counted for A's PageRank (approximately 0.083).

PR(A)=PR(B)/2+PR(C)/1+PR(D)/3

In general,

$$PR(U) = \sum_{V \in B(U)} PR(V)/L(V),$$

where B(u) is the set of pages u is linked to, and L(V) is the number of links from V

A look at Page Rank (damping factor)

The PageRank theory holds that even an imaginary surfer who is randomly clicking on links will eventually stop clicking.

The probability, at any step, that the person will continue is a damping factor *d*.

$$PR(U) = (1-d)/N + d.\sum_{V \epsilon B(U)} PR(V)/L(V),$$

N=size of document collection

For WSD: Page Rank

- Given a graph G = (V,E)
  - $In(V_i)$ = predecessors of V_i
  - $Out(V_i)$ = successors of V_i

$$S(V_i) = \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j)$$

- In a weighted graph, the walker randomly selects an outgoing edge with higher probability of selecting edges with higher weight.

$$WS(V_i) = \sum_{j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j)$$

Other Link Based Algorithms

- *HITS* algorithm invented by Jon Kleinberg (used by Teoma and now Ask.com)
- IBM *CLEVER project*
- *TrustRank* algorithm.

CRITIQUE

- Relies on random walks on graphs encoding label dependencies.
- The Good
  - Does not require any tagged data (a WordNet is sufficient).
  - The weights on the edges capture the definition based semantic similarities.
  - Takes into account global data recursively drawn from the entire graph.
- The Bad
  - Poor accuracy
- Accuracy
  - 54% accuracy on SEMCOR corpus which has a baseline accuracy of 37%.

KB Approaches – Comparisons

Algorithm	Accuracy
WSD using Selectional Restrictions	44% on Brown Corpus
Lesk's algorithm	50-60% on short samples of "*Pride and Prejudice*" and some "*news stories*".
Extended Lesk's algorithm	32% on Lexical samples from Senseval 2 (Wider coverage).
WSD using conceptual density	54% on Brown corpus.
WSD using Random Walk Algorithms	54% accuracy on SEMCOR corpus which has a baseline accuracy of 37%.
Walker's algorithm	50% when tested on 10 highly polysemous English words.

KB Approaches – Conclusions

- Drawbacks of WSD using Selectional Restrictions
  - Needs exhaustive Knowledge Base.
- Drawbacks of Overlap based approaches
  - Dictionary definitions are generally very small.
  - Dictionary entries rarely take into account the distributional constraints of different word senses (e.g. selectional preferences, kinds of prepositions, etc. → ***cigarette*** and ***ash*** never co-occur in a dictionary).
  - Suffer from the problem of sparse match.
  - Proper nouns are not present in a MRD. Hence these approaches fail to capture the strong clues provided by proper nouns.

SUPERVISED APPROACHES

NAÏVE BAYES

- The Algorithm find the winner sense using

$$\hat{s} = \operatorname{argmax}_{s \in senses} Pr(s \mid V_w)$$

- 'V_w' is a feature vector consisting of:
  - POS of w
  - Semantic & Syntactic features of w
  - Collocation vector (set of words around it) → typically consists of next word(+1), next-to-next word(+2), -2, -1 & their POS's
  - Co-occurrence vector (number of times w occurs in bag of words around it)

- Applying Bayes rule and naive independence assumption

$$\hat{s} = \operatorname{argmax}_{s \in senses} Pr(s).\Pi_{i=1}^{n} Pr(V_w^i \mid s)$$

BAYES RULE AND INDEPENDENCE ASSUMPTION

$$\mathbf{s\hat{}= argmax_{\ s\ \varepsilon\ senses}\ Pr(s|V_w)}$$

where V_w is the feature vector.

- Apply Bayes rule:

$$\mathbf{Pr(s|V_w)=Pr(s).Pr(V_w|s)/Pr(V_w)}$$

$$\mathbf{s\hat{}= argmax_{\ s\ \varepsilon\ senses}\ Pr(s|V_w)}$$

- $Pr(V_w|s)$ can be approximated by independence assumption:

$$\mathbf{Pr(V_w|s) = Pr(V_w^1|s).Pr(V_w^2|s,V_w^1)...Pr(V_w^n|s,V_w^1,..,V_w^{n-1})}$$

$$\mathbf{= \Pi_{i=1}^{n}Pr(V_w^i|s)}$$

Thus,

$$\mathbf{s\hat{}= argmax_{\ s\hat{I}senses}\ Pr(s).\Pi_{i=1}^{n}Pr(V_w^i|s)}$$

ESTIMATING PARAMETERS

- Parameters in the probabilistic WSD are:
  - **$Pr(s)$**
  - **$Pr(V_w^i|s)$**
- Senses are marked with respect to sense repository (WORDNET)

$$Pr(s) = count(s,w) / count(w)$$

$$Pr(V_w^i|s) = Pr(V_w^i,s)/Pr(s)$$

$$= \frac{count(V_w^i, s, w)/count(w)}{count(s, w)/count(w)}$$

$$= c(V_w^i,s,w)/c(s,w)$$

DECISION LIST ALGORITHM

- Based on 'One sense per collocation' property.
  - Nearby words provide strong and consistent clues as to the sense of a target word.
- Collect a large set of collocations for the ambiguous word.
- Calculate word-sense probability distributions for all such collocations.
- Calculate the log-likelihood ratio

$$\mathrm{Log}\left(\frac{\Pr(\text{Sense-A} \mid \text{Collocation}_i)}{\Pr(\text{Sense-B} \mid \text{Collocation}_i)}\right)$$

Assuming there are only two senses for the word. Of course, this can easily be extended to 'k' senses.

- Higher log-likelihood = more predictive evidence
- Collocations are ordered in a ***decision list***, with most predictive collocations ranked highest.

DECISION LIST ALGORITHM (CONTD.)

Training Data

Sense	Training Examples (Keyword in Context)
A	used to strain microscopic *plant* life from the ...
A	... zonal distribution of *plant* life . ...
A	close-up studies of *plant* life and natural ...
A	too rapid growth of aquatic *plant* life in water ...
A	... the proliferation of *plant* and animal life ...
A	establishment phase of the *plant* virus life cycle ...
A	... ...
B	... ...
B	computer **manufacturing** *plant* and adjacent ...
B	discovered at a St. Louis *plant* **manufacturing**
B	... copper **manufacturing** *plant* found that they
B	copper wire **manufacturing** *plant* , for example ...
B	's cement **manufacturing** *plant* in Alpena ...
B	polystyrene **manufacturing** *plant* at its Dow ...
B	company **manufacturing** *plant* is in Orlando ...

Resultant Decision List

Final decision list for *plant* (abbreviated)

LogL	Collocation	Sense
10.12	*plant* growth	$\Rightarrow$ A
9.68	car (within $\pm k$ words)	$\Rightarrow$ B
9.64	*plant* height	$\Rightarrow$ A
9.61	union (within $\pm k$ words)	$\Rightarrow$ B
9.54	equipment (within $\pm k$ words)	$\Rightarrow$ B
9.51	assembly *plant*	$\Rightarrow$ B
9.50	nuclear *plant*	$\Rightarrow$ B
9.31	flower (within $\pm k$ words)	$\Rightarrow$ A
9.24	job (within $\pm k$ words)	$\Rightarrow$ B
9.03	fruit (within $\pm k$ words)	$\Rightarrow$ A
9.02	*plant* species	$\Rightarrow$ A
...	...	

Classification of a test sentence is based on the highest ranking collocation found in the test sentence.

E.g.

…plucking **flowers** affects ***plant growth***…

CRITIQUE

- Harnesses powerful, empirically-observed properties of language.
- The Good
  - Does not require large tagged corpus. Simple implementation.
  - Simple semi-supervised algorithm which builds on an existing supervised algorithm.
  - Easy understandability of resulting decision list.
  - Is able to capture the clues provided by Proper nouns from the corpus.
- The Bad
  - The classifier is word-specific.
  - A new classifier needs to be trained for every word that you want to disambiguate.
- Accuracy
  - Average accuracy of **96%** when tested on a set of 12 highly polysemous words.

Exemplar Based WSD (k-nn)

- An exemplar based classifier is constructed for each word to be disambiguated.
- **<u>Step1</u>:** From each ***sense marked sentence*** containing the ambiguous word , a training example is constructed using:
  - POS of ***w*** as well as POS of neighboring words.
  - Local collocations
  - Co-occurrence vector
  - Morphological features
  - Subject-verb syntactic dependencies
- **<u>Step2</u>:** Given a test sentence containing the ambiguous word, a test example is similarly constructed.
- **<u>Step3</u>:** The test example is then compared to all training examples and the k-closest training examples are selected.
- **<u>Step4</u>:** The sense which is most prevalent amongst these "k" examples is then selected as the correct sense.

WSD Using SVMs

- SVM is a binary classifier which finds a hyperplane with the largest margin that separates training examples into 2 classes.
- As SVMs are binary classifiers, a separate classifier is built for each sense of the word
- **Training Phase:** Using a tagged corpus, f or every sense of the word a SVM is trained using the following features:
  - POS of ***w*** as well as POS of neighboring words.
  - Local collocations
  - Co-occurrence vector
  - Features based on syntactic relations (e.g. headword, POS of headword, voice of head word etc.)
- **Testing Phase:** Given a test sentence, a test example is constructed using the above features and fed as input to each binary classifier.
- The correct sense is selected based on the label returned by each classifier.

WSD Using Perceptron Trained HMM

- WSD is treated as a sequence labeling task.
- The class space is reduced by using WordNet's super senses instead of actual senses.
- A discriminative HMM is trained using the following features:
  - POS of ***w*** as well as POS of neighboring words.
  - Local collocations
  - Shape of the word and neighboring words

    E.g. for s = "Merrill Lynch & Co shape(s) =Xx*Xx*&Xx
- Lends itself well to NER as labels like "person", location", "time" etc are included in the super sense tag set.

Supervised Approaches – Comparisons

Approach	Average Precision	Average Recall	Corpus	Average Baseline Accuracy
Naïve Bayes	64.13%	Not reported	Senseval3 – All Words Task	60.90%
Decision Lists	96%	Not applicable	Tested on a set of 12 highly polysemous English words	63.9%
Exemplar Based disambiguation (k-NN)	68.6%	Not reported	WSJ6 containing 191 content words	63.7%
SVM	72.4%	72.4%	Senseval 3 – Lexical sample task (Used for disambiguation of 57 words)	55.2%
Perceptron trained HMM	67.60	73.74%	Senseval3 – All Words Task	60.90%

Supervised Approaches – Conclusions

- <u>General Comments</u>
  - Use corpus evidence instead of relying of dictionary defined senses.
  - Can capture important clues provided by proper nouns because proper nouns do appear in a corpus.
- <u>Naïve Bayes</u>
  - Suffers from data sparseness.
  - Since the scores are a product of probabilities, some weak features might pull down the overall score for a sense.
  - A large number of parameters need to be trained.
- <u>Decision Lists</u>
  - A word-specific classifier. A separate classifier needs to be trained for each word.
  - Uses the single most predictive feature which eliminates the drawback of Naïve Bayes.

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?!*

Insight from Sanskritic Tradition

- Power of a word
  - Abhidha, Lakshana, Vyanjana
- Meaning of **Hall:**
  - *The hall is packed (avidha)*
  - *The hall burst into laughing (lakshana)*
  - *The Hall is full (*unsaid*: and so we cannot enter) (vyanjana)*

Metaphors in Indian Tradition

- *upamana* and *upameya*
  - Former: object being compared
  - Latter: object being compared with
  - *Puru was like a lion in the battle with Alexander* (Puru: *upameya;* Lion: *upamana)*

Upamana, rupak, atishayokti

- *upamana*: Explicit comparison
  - *Puru was like a lion in the battle with Alexander*
- *rupak*: Implicit comparison
  - *Puru was a lion in the battle with Alexander*
- *Atishayokti (exaggeration)*: upamana and upameya dropped
  - *Puru's army fled. But the lion fought on.*

Modern study (1956 onwards, Richards et. al.)

- Three constituents of metaphor
  - *Vehicle* (items used metaphorically)
  - *Tenor* (the metaphorical meaning of the former)
  - *Ground* (the basis for metaphorical extension)
- "*The foot of the mountain*"
  - Vehicle: :foot"
  - Tenor: "lower portion"
  - Ground: "spatial parallel between the relationship between the foot to the human body and the lower portion of the mountain with the rest of the mountain"

Interaction of semantic fields (*Haas)*

- Core vs. peripheral semantic fields
- Interaction of two words in metonymic relation brings in new semantic fields with selective inclusion of features
- *Leg of a table*
  - Does not *stretch* or *move*
  - Does *stand* and *support*

Lakoff's (1987) contribution

- Source Domain
- Target Domain
- Mapping Relations

Mapping Relations: ontological correspondences

- *Anger is heat of fluid in container*

Heat	Anger
(i) Container (ii) Agitation of fluid (iii) Limit of resistence (iv) Explosion	Body Agitation of mind Limit of ability to suppress Loss of control

Image Schemas

- Categories: Container Contained
- Quantity
  - More is up, less is down: *Outputs rose dramatically; accidents rates were lower*
  - Linear scales and paths: *Ram is by far the best performer*
- Time
  - Stationary event: *we are coming to exam time*
  - Stationary observer: *weeks rush by*
- Causation: *desperation drove her to extreme steps*

Patterns of Metonymy

- Container for contained
  - *The kettle boiled* (water)
- Possessor for possessed/attribute
  - *Where are you parked?* (car)
- Represented entity for representative
  - The government will announce new targets
- Whole for part
  - *I am going to fill up the car with petrol*

Patterns of Metonymy *(contd)*

- Part for whole
  - *I noticed several new faces in the class*
- Place for institution
  - *Lalbaug witnessed the largest Ganapati*

Question: Can you have part-part metonymy

Purpose of Metonymy

- More idiomatic/natural way of expression
  - More natural to say *the kettle is boiling* as opposed to *the water in the kettle is boiling*
- Economy
  - *Room 23 is answering* (but not **is asleep*)
- Ease of access to referent
  - *He is in the phone book* (but not *on the back of my hand)
- Highlighting of associated relation
  - *The car in the front decided to turn right* (but not **to smoke a cigarette)*

Feature sharing not necessary

- In a restaurant:
  - *Jalebii ko abhi dudh chaiye* (no feature sharing)
  - *The elephant now wants some coffee* (feature sharing)

Proverbs

- Describes a specific event or state of affairs which is applicable metaphorically to a range of events or states of affairs provided they have the same or sufficiently similar image-schematic structure