Unit 4
Lexical Semantics

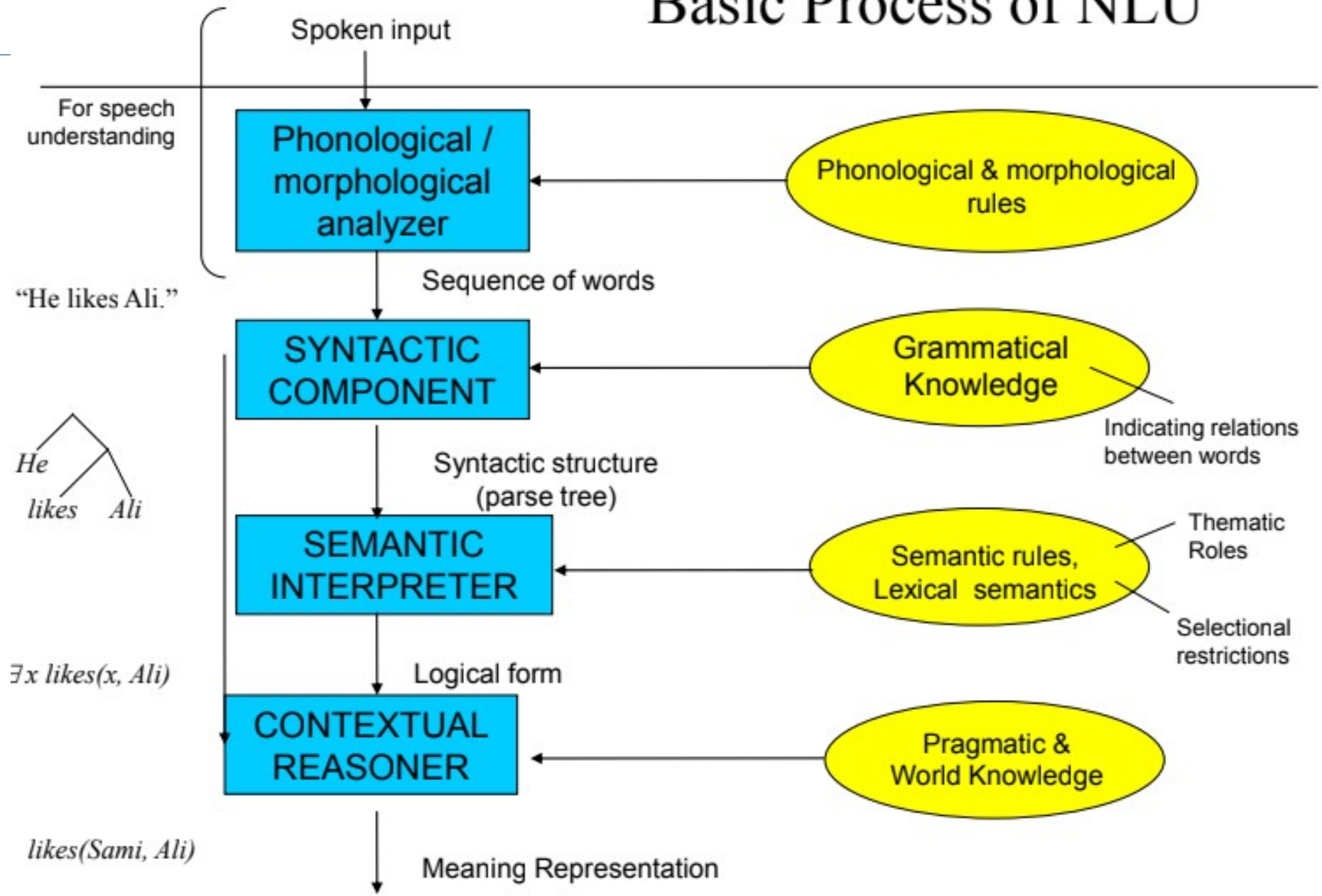
Lexeme, Lexicon, Senses, WordNet, WSD, Word Similarity

Natural Language Processing (NLP)
MDS 555

Objective

- Lexeme
- Lexicon
- Senses
- Lexical relations
- WordNet (Lexical Database)
- Word Sense Disambiguation (WSD)
- Word Similarity

Basic Process of NLU

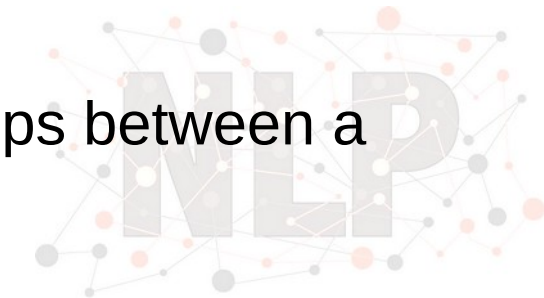


Semantic Analysis

- The purpose of semantic analysis is to draw exact meaning or dictionary meaning from the text.
- The work of semantic analyzer is to check the text for meaningfulness.

Senses / Meaning

- Traditionally, meaning in language has been studied from three perspectives
  - The meaning of individual words
  - The meaning of individual sentences or utterances
  - The meaning of a text or discourse
- Meaning is a notion in semantics classically defined as having two components:
  - **Reference**, anything in the referential realm denoted by a word or expression, and
  - **Sense**, the system of paradigmatic and syntagmatic relationships between a lexical unit and other lexical units in a language.



Semantic Analysis

- It can be done in two parts
  - the study of the meaning of individual words
  - the individual words will be combined to provide the meaning of the sentences

Lexeme

- It is an entry in the lexicon that include
  - An orthographic representation
  - A phonological form
  - A symbolic meaning representation or sense
- Dictionary Entries
  - Red (‘red’) n; the color of blood or a ruby
  - Blood (‘bluhd’) n : the red liquid that circulates in the heart, arteries and veins of animals

Elements of Semantic Analysis

- Lexical Semantics
  - Lexical Semantic Analysis involves understanding the meaning of each word of the text individually.
    - Hyponymy
    - Homonymy
    - Polysemy
    - Synonymy
    - Antonymy

Hypernym and Hyponyms

- Is A relationship
- It may be defined as the relationship between a generic term and instances of that generic term.
- The generic term is called hypernym
- The instances are called hyponyms
  - For example, the word color is hypernym and the color blue, yellow etc. are hyponyms

Homonymy

- Lexemes that share a form
  - Phonological, orthographic or both
- It may be defined as the words having same spelling or same form but having different and unrelated meaning.
  - For example, the word “Bat” is a homonymy word
    - Bat can be an implement to hit a ball
    - Bat is a nocturnal flying mammal

Polysemy

- It is a word or phrase with different but related sense
- We can say that polysemy has the same spelling but different and related meaning
  - For example, the word “bank” is a polysemy word having the following meanings
    - A financial institution
    - The building in which such an institution is located
    - A synonym for “to rely on”

Homonym vs Polysemy

- Homonyms: same word, different meaning
  - bank (river)
  - bank (financial)
- Polysemy: different senses of same word
  - That dog has floppy ears.
  - He has a good ear for jokes.
  - bank(financial) - the building, the institution, the notion of where money is stored

Homonym VS Polysemy

- Both polysemy and homonymy words have the same syntax or spelling.
- The main difference between them is that
  - in polysemy, the meanings of the words are related but
  - in homonymy, the meanings of the words are not related.

    For example, if we talk about the same word “Bank”, we can write the meaning ‘a financial institution’ or ‘a river bank’. In that case it would be the example of homonym because the meanings are unrelated to each other.

Synonymy

- Different ways of expressing related concepts
- It is the relation between two lexical items having different forms but expressing the same or a close meaning.
  - Examples are ‘author/writer’, ‘fate/destiny’
- Synonyms are almost never truly substitutable-
  - Used in different contexts
  - Have different implications

Antonymy

- It is the relation between two lexical items having symmetry between their semantic components relative to an axis. The scope of antonymy is as follows −
  - Application of property or not − Example is ‘life/death’, ‘certitude/incertitude’
  - Application of scalable property − Example is ‘rich/poor’, ‘hot/cold’
  - Application of a usage − Example is ‘father/son’, ‘moon/sun’.

Metonymy

- Use one aspect of something to stand for the whole
  - Metaphore
    - The White House released new figures today
    - Simhadarbar is sent to villages…. => Public Administration
    - As per Officials of Simhadarbar ….. => Government

Lexical Database

WordNet

Lexical Database

- A lexical database is an organized description of the lexemes of a language
- It contains structured information about words
- The main difference between lexical databases and dictionaries is that dictionaries aim to explain or translate words, while lexical databases are primarily developed for research purposes.

WordNet

- Organize lexical information in terms of word meaning, rather than word form
- Lexical knowledge-base
- Large Lexical resource with information about
  - Nouns
  - Verbs
  - Adjectives
  - Adverbs
- https://aclanthology.org/H94-1111.pdf

WordNet

- Each entry is annotated with one or more **senses**
- Each sense provides a variety of information

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: mask Search WordNet

Display Options: (Select option to change) Change

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

Display options for sense: (frequency) {offset} <lexical filename > [lexical file number] (gloss) "an example sentence"

Display options for word: word#sense number (sense key)

Noun

- (1){03730361} <noun.artifact>[06] S: (n) **mask#1 (mask%1:06:00::)** (a covering to disguise or conceal the face)
- (1){01051399} <noun.act>[04] S: (n) **mask#2 (mask%1:04:00::)** (activity that tries to conceal something) *"no mask could conceal his ignorance"; "they moved in under a mask of friendship"*
- {08270371} <noun.group>[14] S: (n) masquerade#1 (masquerade%1:14:00::), masquerade party#1 (masquerade_party%1:14:00::), masque#1 (masque%1:14:00::), **mask#3 (mask%1:14:00::)** (a party of guests wearing costumes and masks)
- {03730526} <noun.artifact>[06] S: (n) **mask#4 (mask%1:06:01::)** (a protective covering worn over the face)

Verb

- (1){02152033} <verb.perception>[39] S: (v) dissemble#2 (dissemble%2:39:00::), cloak#1 (cloak%2:39:00::), **mask#1 (mask%2:39:00::)** (hide under a false appearance) *"He masked his disappointment"*
- (1){01361031} <verb.contact>[35] S: (v) **mask#2 (mask%2:35:00::)** (put a mask on or cover with a mask) *"Mask the children for Halloween"*
- {02163017} <verb.perception>[39] S: (v) disguise#1 (disguise%2:39:00::), **mask#3 (mask%2:39:01::)** (make unrecognizable) *"The herb masks the garlic taste"; "We disguised our faces before robbing the bank"*
- {01361558} <verb.contact>[35] S: (v) **mask#4 (mask%2:35:02::)** (cover with a sauce) *"mask the meat"*

WordNet Entries

- Senses contains
  - **Gloss**
    - A definition of the sense
  - List of Synonyms
    - Commonly referred to as a **synset**
  - Example sentence

Lexicographic Categories

- Super-senses
- Coarse-grained semantic category
-

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

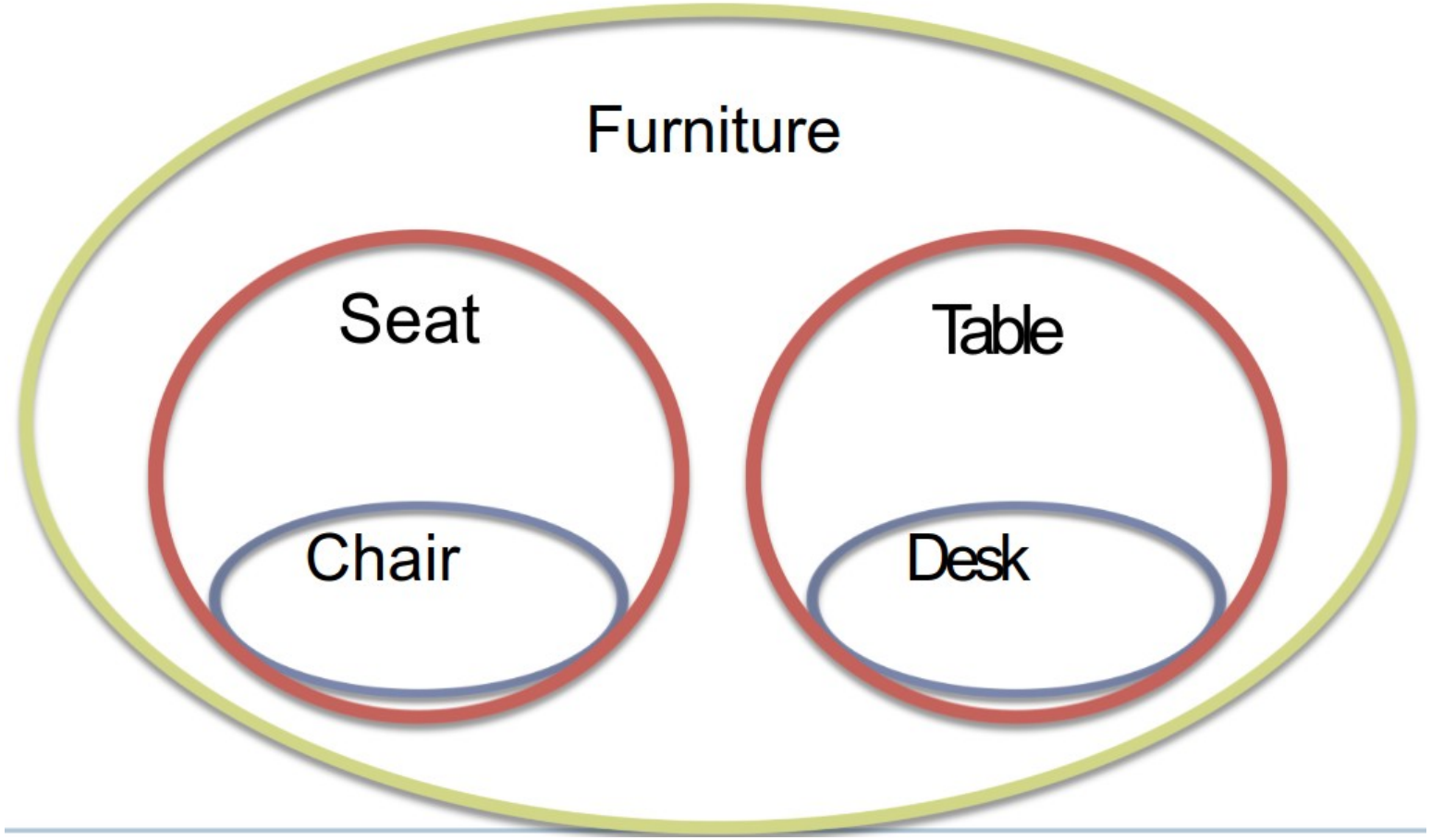
Sense Relations

- Hypernym: Relation between a concept and its superordinate
  - Food is a hypernym of cake
- Hyponym: Relation between a concept and its subordinate
  - Corgi is a hyponym of dog
- Meronym: Relation between a part and its whole
  - Wheel is a meronym of car
- Holonym: Relation between a whole and its parts
  - Car is a holonym of wheel
- Antonym: Relation between two semantically opposite concepts
  - Leader is an antonym of follower

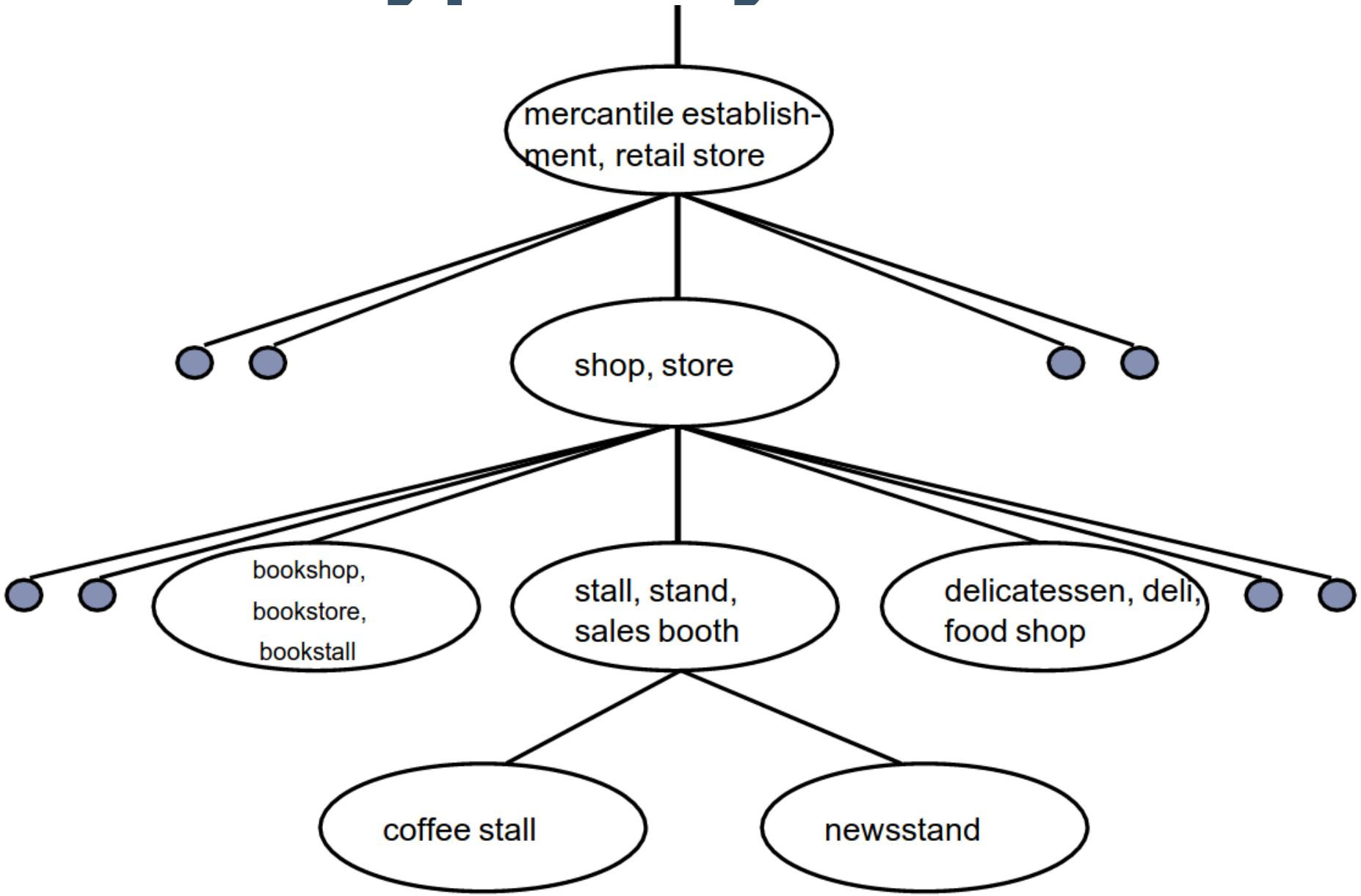
Sense Relations

Category	Relation	*Type*	*Example*
Noun	Hypernym/hypo	Sem	dog IS A KIND OF animal
	Meronym	Sem	arm IS A PART OF body
Verb	Implication: Cause	Sem	to kill CAUSES to die
	Precondition		to succeed ENTAILS DOING to try
	Troponym		to limp IS ONE WAY TO walk
	Inclusion		snore ENTAILS DOING to sleep
	Opposition	Lex	to die ANTONYM to be born
Adj	Antonym	Lex	hot ANTONYM cold
Adv	Derived adj	Lex	quickly DERIVED FROM quick
	Antonym	Lex	quickly ANTONYM slowly

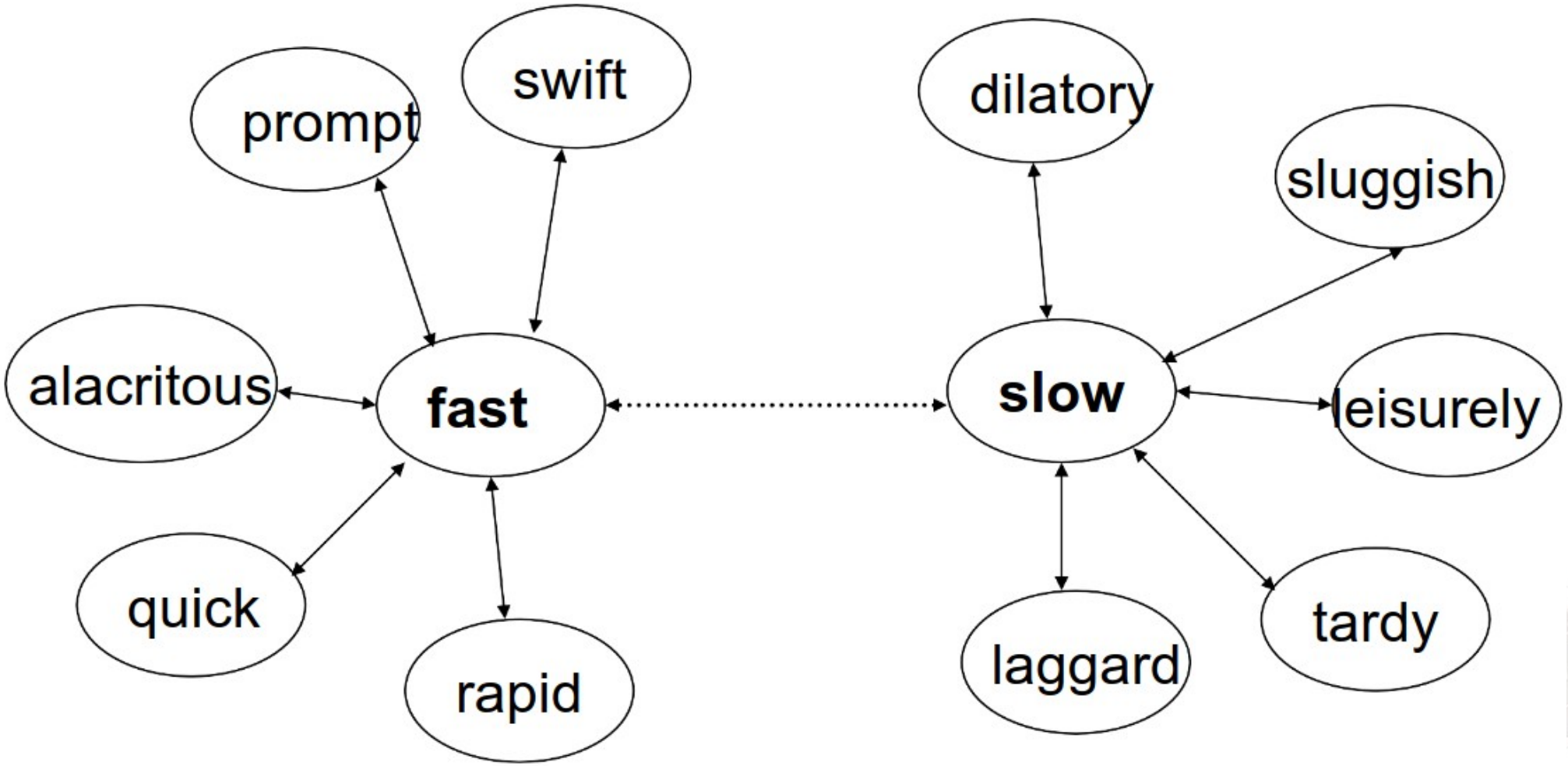
Nouns : Hyperonym



Nouns : Hyperonym



Adjectives (Antonymy, Similarity)

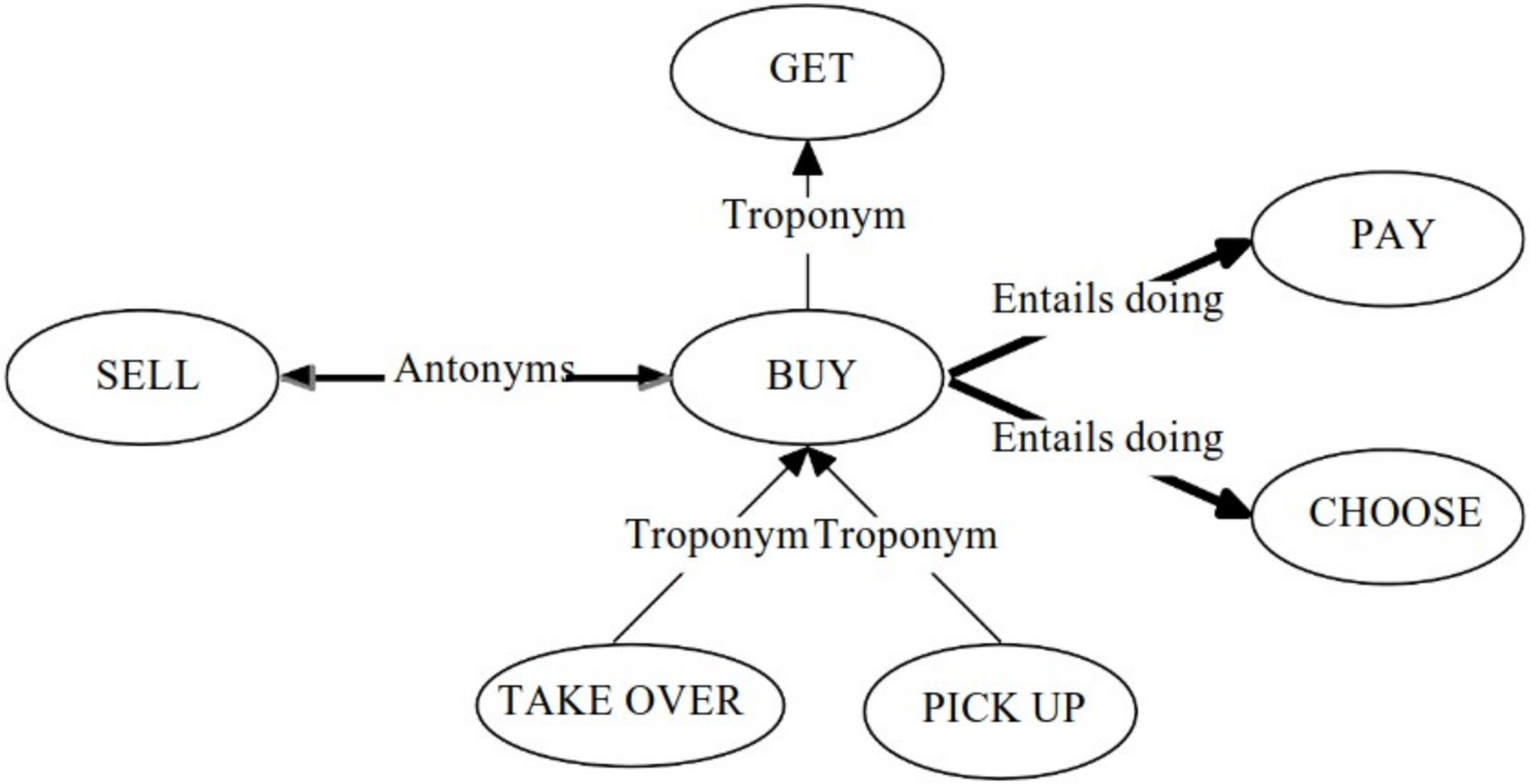


Synset

- it does not provide a full specification of the word meaning
- it points to a lexical concept and represent its (partial) meaning by means of its lexical and semantics relations with other lexical concepts

Semantic Network

- Meaning of a word = relations with other words
  - e.g.: to buy



Relations

A closer look on the word “get”

- 17. {catch, get}
- 18. {catch, arrest, get}
- 19. {get, catch}
- 20. {get}
- 21. {get}
- 22. {get}
- 23. {catch, get}
- 24. {catch, get}

Senses

“get” semses

- 17. {catch, get} => {understand}
- 18. {catch, arrest, get} => {attract, pull, pull in, draw, draw in}
- 19. {get, catch} => {hit}
- 20. {get}
- 21. {get} => {get, acquire}
- 22. {get} => {buy, purchase}
- 23. {catch, get} => {hear}
- 24. {catch, get} => {hurt, ache, suffer}

Glosses

- “get” glosses
  - 17. {catch, get} -- (grasp with the mind or develop an understanding of) "did you catch that allusion?"; "We caught something of his theory in the lecture"; "don't catch your meaning"; "did you get it?"; "She didn't get the joke"; "I just don't get him“
  - 18. {catch, arrest, get} -- (attract and fix) "His look caught her"; "She caught his eye"; "Catch the attention of the waiter“
  - 19. {get, catch} -- (reach with a blow or hit in a particular spot) "the rock caught her in the back of the head"; "The blow got him in the back"; "The punch caught him in the stomach“
  - 20. {get} -- (reach by calculation) "What do you get when you add up these numbers?“
  - 21. {get} -- (acquire as a result of some effort or action) "You cannot get water out of a stone"; "Where did she get these news?“
  - 22. {get} -- (purchase) "What did you get at the toy store?“
  - 23. {catch, get} -- (perceive by hearing) "I didn't catch your name"; "She didn't get his name when they met the first time“
  - 24. {catch, get} -- (suffer from the receipt of) "She will catch hell for this behavior!"

WordNet for multiple languages

- EuroWordNet
  - Create synsets, create relations for every language
  - Then map sysnets
- MultiWordNet
  - Create synsets for a new WordNet mapped to the English wordnet synsets (Princeton WordNet, PWN)
  - Importing the semantic relations the new wordnet

Word Sense Disambiguation (WSD)

Word sense disambiguation

- *Word sense disambiguation* is the problem of selecting a sense for a word from a set of predefined possibilities.
  - Sense Inventory usually comes from a dictionary or thesaurus.
  - Knowledge intensive methods, supervised learning, and (sometimes) bootstrapping approaches
- *Word sense discrimination* is the problem of dividing the usages of a word into different meanings, without regard to any particular existing sense inventory.
  - Unsupervised techniques

Computer vs Humans

- *Polysemy* – most words have many possible meanings.
- A computer program has no basis for knowing which one is appropriate, even if it is obvious to a human…
- Ambiguity is rarely a problem for humans in their day to day communication, except in extreme cases…

Ambiguity for Humans

- DRUNK GETS NINE YEARS IN VIOLIN CASE
- FARMER BILL DIES IN HOUSE
- PROSTITUTES APPEAL TO POPE
- STOLEN PAINTING FOUND BY TREE
- RED TAPE HOLDS UP NEW BRIDGE
- DEER KILL 300,000
- RESIDENTS CAN DROP OFF TREES
- INCLUDE CHILDREN WHEN BAKING COOKIES
- MINERS REFUSE TO WORK AFTER DEATH

Ambiguity for Computer

- The fisherman jumped off the **bank** and into the water.
- The **bank** down the street was robbed!
- Back in the day, we had an entire **bank** of computers devoted to this problem.
- The **bank** in that road is entirely too steep and is really dangerous.
- The plane took a **bank** to the left, and then headed off towards the mountains.

Two variants of WSD task

- Lexical Sample task
  - Small pre-selected set of target words
  - And inventory of senses for each word
- All-words task
  - Every word in an entire text
  - A lexicon with senses for each word
  - Sort of like part-of-speech tagging
  - Except each lemma has its own tagset

WSD - Applications

- Machine Translation
- Information Retrieval
- Text Summarization
- Question Answering
- Sentiment Analysis
- Text Classification
- Grammar and Style checking of writing
- Speech Context Identification

Approaches to WSD

- Dictionary-Based Approaches
- Supervised Machine Learning
- Unsupervised Machine Learning
- Knowledge-Based Approaches

WSD: Dictionary-Based Approaches

- These approaches use dictionaries or lexical resources like WordNet to look up word senses.
- WordNet, for example, provides a structured database of words and their senses, along with relationships between these senses (e.g., hypernyms and hyponyms).
- Dictionary-based methods match the word in context to its sense in the dictionary.

Supervised Machine Learning

- n supervised approaches, a WSD model is trained on labeled data where words are tagged with their correct senses.
- Features derived from the context (surrounding words) are used to train classifiers like
  - decision trees,
  - support vector machines
  - neural networks to predict senses.

Unsupervised Machine Learning

- Unsupervised methods do not rely on labeled data.
- Instead, they use clustering or similarity-based techniques to group instances of a word into clusters representing different senses.
- Common approaches include clustering algorithms like
  - k-means
  - distributional similarity based on word co-occurrence.

Knowledge-Based Approaches

- These approaches incorporate external knowledge sources, such as ontologies, semantic networks, or domain-specific databases, to disambiguate word senses.
- They rely on semantic relationships and knowledge about the entities mentioned in the text.

Hybrid Approaches

- Some WSD systems combine multiple techniques, such as using a dictionary-based method as a fallback when supervised machine learning models are uncertain.

Lesk Algorithm

- Intuition: word overlap between context and dictionary entries
  - Unsupervised, but knowledge rich

The **bank** can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.

WordNet

bank[1]	Gloss:	a financial institution that accepts deposits and channels the money into lending activities
	Examples:	"he cashed a check at the bank", "that bank holds the mortgage on my home"
bank[2]	Gloss:	sloping land (especially the slope beside a body of water)
	Examples:	"they pulled the canoe up on the bank", "he sat on the bank of the river and watched the currents"

Lesk Algorithm

- Simplest implementation:
  - Count overlapping content words between glosses and context
- Lots of variants:
  - Include the examples in dictionary definitions
  - Include hypernyms and hyponyms
  - Give more weight to larger overlaps (e.g., bigrams)
  - Give extra weight to infrequent words (e.g., idf weighting)
  - …
- Works reasonably well!

Lesk Algorithm

```
function SIMPLIFIED LESK(word, sentence) returns best sense of word

  best-sense ← most frequent sense for word
  max-overlap ← 0
  context ← set of words in sentence
  for each sense in senses of word do
    signature ← set of words in the gloss and examples of sense
    overlap ← COMPUTEOVERLAP(signature, context)
    if overlap > max-overlap then
        max-overlap ← overlap
        best-sense ← sense
  end
  return(best-sense)
```

Figure 20.3 The Simplified Lesk Algorithm. The COMPUTEOVERLAP function returns the number of words in common between two sets, ignoring function words or other words on a stop list. The original Lesk algorithm defines the *context* in a more complex way. The *Corpus Lesk* algorithm weights each overlapping word w by its $-\log P(w)$, and includes labeled training corpus data in the *signature*.

Word Similarity

Word similarity

- Word similarity refers to the degree of resemblance or likeness between two words based on their meaning or semantic content.
- It quantifies how close or related two words are in terms of their semantic interpretation.
- Word similarity is a fundamental concept in natural language processing (NLP) and computational linguistics,
- It plays a crucial role in various NLP tasks
  - including information retrieval,
  - machine translation,
  - document clustering, and more.

Lexical Similarity

- Lexical similarity measures focus on the surface form of words and consider factors such as spelling and character overlap. Common lexical similarity measures include:
  - **Jaccard Similarity**: It measures the size of the intersection of characters or n-grams divided by the size of the union of characters or n-grams between two words.

Lexical Similarity

- **Edit Distance** (Levenshtein Distance): It quantifies the minimum number of single-character edits (insertions, deletions, substitutions) required to transform one word into another.
- **Cosine Similarity**: It computes the cosine of the angle between two vectors representing the word frequencies in a document-term matrix. While it's often used for documents, it can also be applied to words.

Study Materials

- Word Net: https://wordnetcode.princeton.edu/5papers.pdf
- http://disi.unitn.it/~ldkr/ldkr2017/slides/11.KDI.WordNet.A.Closer.Look.pdf
- http://lintool.github.io/UMD-courses/CMSC723-2009-Fall/session11-slides.pdf
- https://web.stanford.edu/~jurafsky/slp3/slides/Chapter18.wsd.pdf
-

Thank you