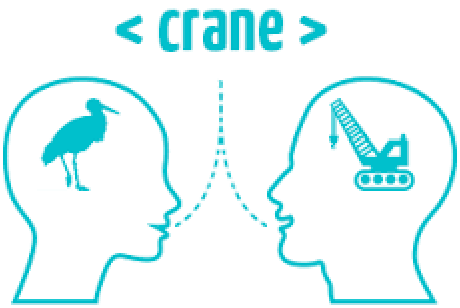
Semantic Analysis



Lexical semantics is concerned with the systematic meaning related connections among lexical items, and the internal meaning-related structure of individual lexical items.

To identify the semantics of lexical items, we need to focus on the notion of lexeme, an individual entry in the lexicon.

What is a *lexeme*?

Lexeme should be thought of as a pairing of a particular orthographic and phonological form with some sort of symbolic meaning representation.

- Orthographic form, and phonological form refer to the appropriate form part of a lexeme
- Sense refers to a lexeme's meaning counterpart.

Example

verge noun
a: BRINK, THRESHOLDa country on the *verge* of destruction— Archibald MacLeish

b: something that borders, limits, or bounds: such as
(1): an outer margin of an object or structural part
(2): the
edge of roof covering (such as tiling) projecting over the gable of a roof
(3)*British* **:** a paved or planted strip of land at the edge of a road **:** SHOULDER

verge
verb (1)

verged; **verging**

intransitive verb
1: to be contiguous

2: to be on the verge or borderthe line where sentiment *verges* on mawkishness— Thomas Hardy

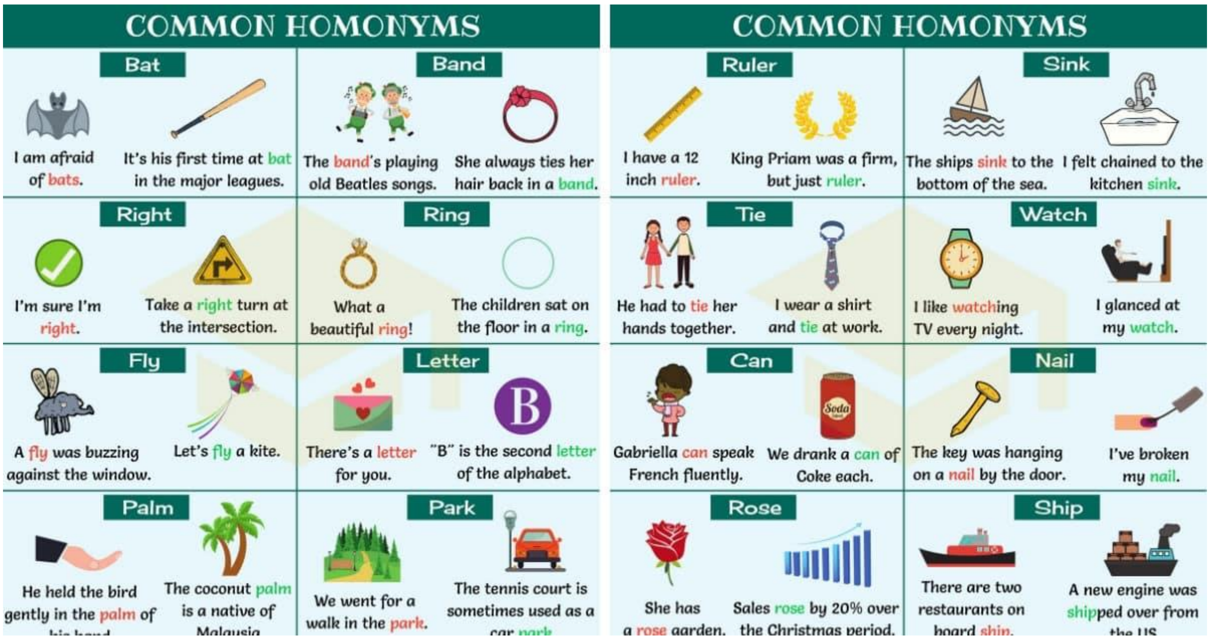
Example: meaning related facts?
Definitions from the American Heritage Dictionary (Morris, 1985)

- right adj. located near the right hand esp. being on the right when facing the same direction as the observer
- left adj. located near to this side of the body than the right
- red n. the color of blood or a ruby
- blood n. the red liquid that circulates in the heart, arteries and veins of animals

- ✓ The entries are description of lexemes in terms of other lexemes.
- ✓ Definitions make it clear that right and left are similar kind of lexemes that stand in some kind of alternation, or opposition, to one another
- ✓ We can glean that red is a color, it can be applied to both blood and rubies, and that blood is a liquid.

Relations between word meanings

1. **<u>Homonymy</u>**



Homonymy is defined as a relation that holds between words that have the same form with unrelated meanings.

Examples

- ✓ Bat (wooden stick-like thing) vs Bat (flying mammal thing)
- ✓ Bank (financial institution) vs Bank (riverside)

<u>Homophones and Homographs</u>

Homophones are the words with the same pronunciation but different spellings.

- ✓ write vs right
- ✓ piece vs peace

Homographs are the lexemes with the same orthographic form but different meaning. Ex: bass.

<u>Problems for NLP applications</u>

a. **Text-to-Speech**

   Same orthographic form but different phonological form

b. Information Retrieval

   Different meaning but same orthographic form

c. Speech Recognition
   to, two, too
   Perfect homonyms are also problematic

2. <u>Polysemy</u>

Multiple related meanings within a single lexeme.

- ✓ The bank was constructed in 1875 out of local red brick.
- ✓ I withdrew the money from the bank.

Are those the same sense?

Sense 1: "The building belonging to a financial institution"

Sense 2: "A financial institution"

Another example

- ✓ Heavy snow caused the roof of the school to collapse.
- ✓ The school hired more teachers this year than ever before.

Often, the relationships are systematic E.g., building vs. organization school, university, hospital, church, supermarket

More examples:

- ✓ Author (Jane Austen wrote Emma)↔Works of Author (I really love Jane Austen)
- ✓ Animal (The chicken was domesticated in Asia)↔Meat (The chicken was overcooked)
- ✓ Tree (Plums have beautiful blossoms)↔Fruit (I ate a preserved plum yesterday)

Zeugma test

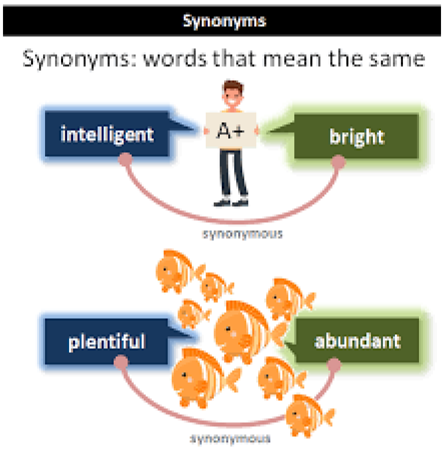
Which of these flights serve breakfast?

Does Midwest Express serve Philadelphia?

*Does Midwest Express serve breakfast and San Jose?

Combine two separate uses of a lexeme into a single example using conjunction
Since it sounds weird, we say that these are two different senses of serve.

3. Synonymy



Words that have the same meaning in some or all contexts.

- ✓ filbert / hazelnut

- ✓ couch / sofa
- ✓ big / large
- ✓ automobile / car
- ✓ vomit / throw up
- ✓ water / H2O

Two lexemes are synonyms if they can be successfully substituted for each other in all situations.

Synonymy: A relation between senses

Consider the words big and large. Are they synonyms?

- ✓ How big is that plane?
- ✓ Would I be flying on a large or small plane?

How about here?

Miss Nelson, for instance, became a kind of big sister to Benjamin.

- ✓ *Miss Nelson, for instance, became a kind of large sister to Benjamin.

Why?

- big has a sense that means being older, or grown up
- large lacks this sense

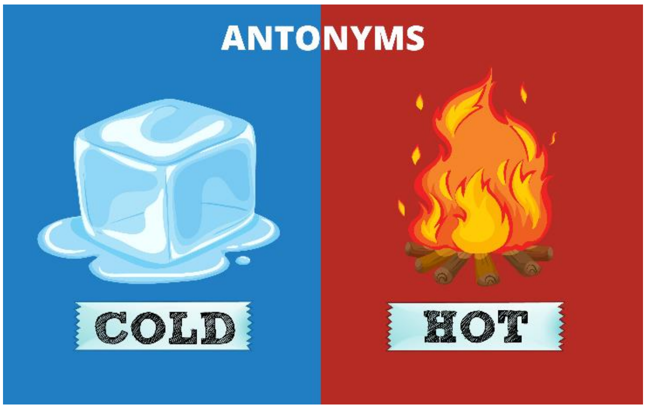
Shades of meaning

- ✓ What is the cheapest first class *fare*?
- ✓ *What is the cheapest first class *price*?

Collocational constraints

- ✓ We frustate 'em and frustate 'em, and pretty soon they make a *big* mistake.
- ✓ *We frustate 'em and frustate 'em, and pretty soon they make a *large* mistake.

4. Antonyms



Senses that are opposites with respect to one feature of their meaning Otherwise, they are similar!

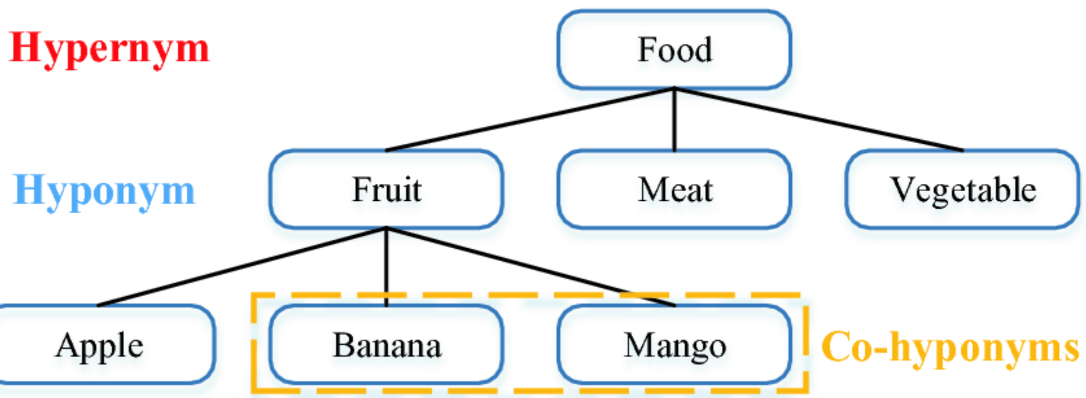
- ✓ dark / light
- ✓ short / long
- ✓ hot / cold
- ✓ up / down
- ✓ in / out

More formally: antonyms can define a binary opposition or at opposite ends of a scale (long/short, fast/slow) Be reversives: rise/fall

5. Hyponymy

One sense is a hyponym of another if the first sense is more specific, denoting a subclass of the other

- ✓ car is a hyponym of vehicle
- ✓ dog is a hyponym of animal
- ✓ mango is a hyponym of fruit



6. Hypernymy

Conversely

- ✓ vehicle is a hypernym/superordinate of car
- ✓ animal is a hypernym of dog
- ✓ fruit is a hypernym of mango

Entailment: Sense A is a hyponym of sense B if being an A entails being a B. Ex: dog, animal

Transitivity: A hypo B and B hypo C entails A hypo C

7. Meronyms and holonyms

Meronymy: an asymmetric, transitive relation between senses. X is a meronym of Y if it denotes a part of Y. The inverse relation is holonymy.

Meronym	**Holonym**
Porch	House
Wheel	Car
Leg	Chair
Nose	Face

WordNet

https://wordnet.princeton.edu/wordnet/

- ✓ A hierarchically organized lexical database
- ✓ A large online machine-readable thesaurus, and aspects of a dictionary
- ✓ Versions for other languages are under development

English WordNet consists of three separate databases, one each for nouns and verbs and a third for adjectives and adverbs; closed class words are not included. Each database contains a set of lemmas, each one annotated with a set of senses. The set of near-synonyms for a WordNet sense is called a synset (for synonym synset set). The entry for bass includes synsets like {bass, deep}, or {bass, bass voice, basso}.

Part of Speech	No of Synsets
Noun	82,115
Verb	13,767
Adjective	18,156
Adverb	3,621

WordNet also labels each synset with a lexicographic category drawn from a semantic field for example the 26 categories for nouns as well as 15 for verbs (plus 2 for adjectives and 1 for adverbs). These categories are often called supersenses, because they act as coarse semantic categories or groupings of supersense senses which can be useful when word senses are too fine-grained.

Supersense	**Nouns denoting . . .**
act	acts or actions
animal	animals
artifact	man-made objects
attribute	attributes of people and objects
body	body parts
cognition	cognitive processes and contents
communication	communicative processes and contents
event	natural events
feeling	feelings and emotions
food	foods and drinks
group	groupings of people or objects
location	spatial position
motive	goals
object	natural objects (not man-made)
quantity	quantities and units of measure
phenomenon	natural phenomena
plant	plants
possession	possession and transfer of possession
process	natural processes
person	people
relation	relations between people or things or ideas
shape	two and three dimensional shapes
state	stable states of affairs
substance	substances
time	time and temporal relations
Tops	abstract terms for unique beginners

Wordnet noun and verb relations

Wordnet noun and verb relations

Relation	Also called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$\textit{breakfast}^1 \rightarrow \textit{meal}^1$
Hyponym	Subordinate	From concepts to subtypes	$\textit{meal}^1 \rightarrow \textit{lunch}^1$
Member Meronym	Has-Member	From groups to their members	$\textit{faculty}^2 \rightarrow \textit{professor}^1$
Has-Instance		From concepts to instances of the concept	$\textit{composer}^1 \rightarrow \textit{Bach}^1$
Instance		From instances to their concepts	$\textit{Austen}^1 \rightarrow \textit{author}^1$
Member Holonym	Member-Of	From members to their groups	$\textit{copilot}^1 \rightarrow \textit{crew}^1$
Part Meronym	Has-Part	From wholes to parts	$\textit{table}^2 \rightarrow \textit{leg}^3$
Part Holonym	Part-Of	From parts to wholes	$\textit{course}^7 \rightarrow \textit{meal}^1$
Antonym		Opposites	$\textit{leader}^1 \rightarrow \textit{follower}^1$

Relation	Definition	Example
Hypernym	From events to superordinate events	$\textit{fly}^9 \rightarrow \textit{travel}^5$
Troponym	From a verb (event) to a specific manner elaboration of that verb	$\textit{walk}^1 \rightarrow \textit{stroll}^1$
Entails	From verbs (events) to the verbs (events) they entail	$\textit{snore}^1 \rightarrow \textit{sleep}^1$
Antonym	Opposites	$\textit{increase}^1 \Longleftrightarrow \textit{decrease}^1$

Adjectives are organized in terms of antonymy. Pairs of “direct” antonyms like wet-dry and young-old reflect the strong semantic contract of their members. Each of these polar adjectives in turn is linked to a number of “semantically similar” ones: dry is linked to parched, arid, dessicated and bone-dry and wet to soggy, waterlogged, etc. Semantically similar adjectives are “indirect antonyms” of the contral member of the opposite pole. Relational adjectives ("pertainyms") point to the nouns they are derived from (criminal-crime).

There are only few adverbs in WordNet (hardly, mostly, really, etc.) as the majority of English adverbs are straightforwardly derived from adjectives via morphological affixation (surprisingly, strangely, etc.)

Verb synsets are arranged into hierarchies as well; verbs towards the bottom of the trees (troponyms) express increasingly specific manners characterizing an event, as in {communicate}-{talk}-{whisper}. The specific manner expressed depends on the semantic field; volume (as in the example above) is just one dimension along which verbs can be elaborated. Others are speed (move-jog-run) or intensity of emotion (like-love-idolize). Verbs describing events that necessarily and unidirectionally entail one another are linked: {buy}-{pay}, {succeed}-{try}, {show}-{see}, etc.

Applications of WordNet

Machine translation is the original and most obvious application for WSD but WSD has actually been considered in almost every application of language technology, including information retrieval, lexicography, knowledge mining/acquisition and semantic interpretation, and is becoming increasingly important in new research areas such as bioinformatics and the Semantic Web.

Machine translation

WSD is required for lexical choice in MT for words that have different translations for different senses. For example, in an English-French financial news translator, the English noun *change* could translate to either *changement* ('transformation') or *monnaie* ('pocket money'). However, most translation systems do not use a separate WSD module. The lexicon is often pre-disambiguated for a given domain, or hand-crafted rules are devised, or WSD is folded into a statistical translation model, where words are translated within phrases which thereby provide context.

Information retrieval

Ambiguity has to be resolved in some queries. For instance, given the query "depression" should the system return documents about illness, weather systems, or economics? Current IR systems (such as Web search engines), like MT, do not use a WSD module; they rely on the user typing enough context in the query to only retrieve documents relevant to the intended sense (e.g., "tropical depression"). In a process called mutual disambiguation, reminiscent of the Lesk method (below), all the ambiguous words are disambiguated by virtue of the intended senses co-occurring in the same document.

Information extraction and knowledge acquisition

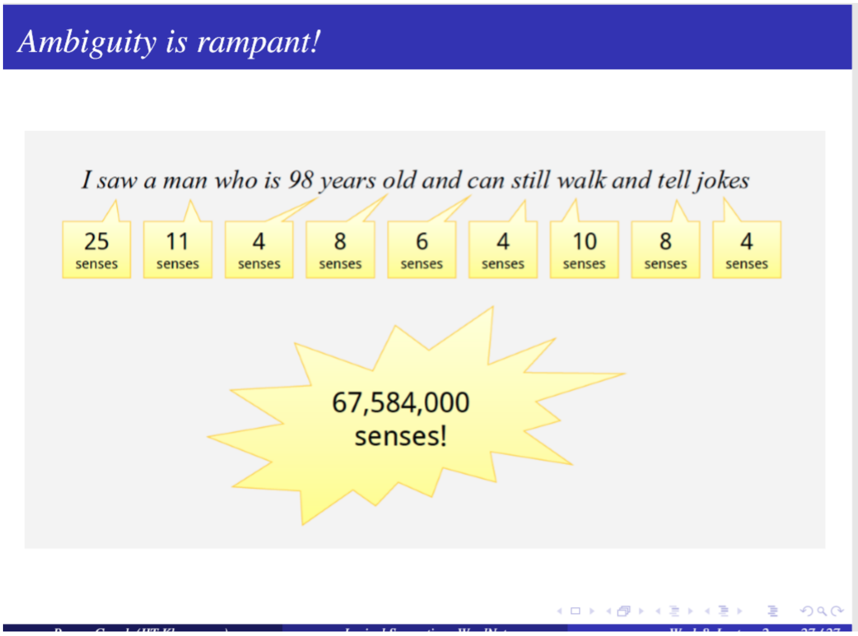
In information extraction and text mining, WSD is required for the accurate analysis of text in many applications. For instance, an intelligence gathering system might need to flag up references to, say, illegal *drugs*, rather than medical *drugs*. Bioinformatics research requires the relationships between genes and gene products to be catalogued from the vast scientific literature; however, genes and their proteins often have the same name. More generally, the Semantic Web requires automatic annotation of documents according to a reference ontology.

Word Sense Disambiguation

ENGLISH	
Bark	The bark of the tree is bown
	I was awakened by the bark of the dog.
Fan	The fan stopped worked.
	I am a die hard fan of Tom Cruise.
HINDI	
मत	मत बोलो । (Don’t Talk)
	मत के आधार पर फैसला हुआ। The decision was based on votes.
गुलाब	गुलाब जामुन अच्छा है। Gulab Jamun (Indian Sweet dish) is sweet.
	गुलाब लाल है। The rose is red.

Table I: Examples of Ambiguity

The task of selecting the correct sense for a word is called word sense disambiguation, or WSD. WSD algorithms take as input a word in context and a fixed inventory word sense disambiguation WSD of potential word senses and outputs the correct word sense in context.



Sense ambiguity

Many words have several meanings or senses

The meaning of bass depends on the context

Are we talking about music, or fish? I An electric guitar and bass player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps. I And it all started when fishermen decided the striped bass in Lake Mead were too skinny.

Disambiguation: The task of disambiguation is to determine which of the senses of an ambiguous word is invoked in a particular use of the word. This is done by looking at the context of the word's use.

Approaches:

- **Dictionary- and knowledge-based methods:** These rely primarily on dictionaries, thesauri, and lexical knowledge bases, without using any corpus evidence.
- **Supervised methods:** These make use of sense-annotated corpora to train from.
- **Semi-supervised or minimally-supervised methods:** These make use of a secondary source of knowledge such as a small annotated corpus as seed data in a bootstrapping process, or a word-aligned bilingual corpus.

- **Unsupervised methods:** These eschew (almost) completely external information and work directly from raw unannotated corpora. These methods are also known under the name of *word sense discrimination*.

Dictionary- and knowledge-based methods

The Lesk method (Lesk 1986) is the seminal dictionary-based method. It is based on the hypothesis that words used together in text are related to each other and that the relation can be observed in the definitions of the words and their senses. Two (or more) words are disambiguated by finding the pair of dictionary senses with the greatest word overlap in their dictionary definitions. For example, when disambiguating the words in *pine cone*, the definitions of the appropriate senses both include the words *evergreen* and *tree* (at least in one dictionary).

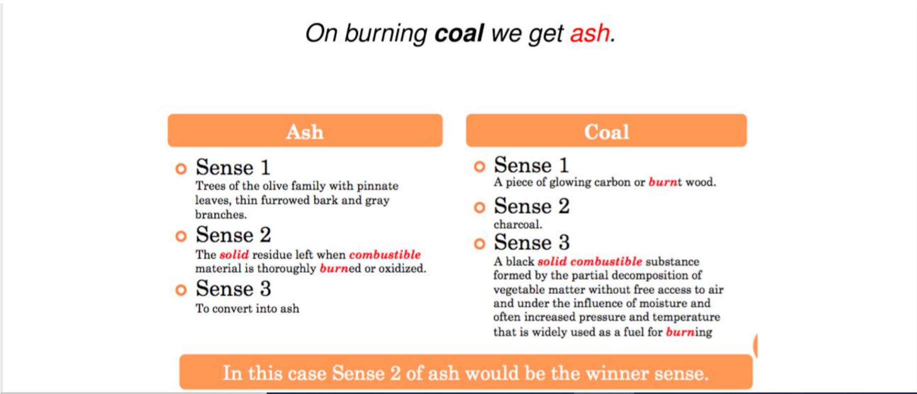
An alternative to the use of the definitions is to consider general word-sense relatedness and to compute the semantic similarity of each pair of word senses based on a given lexical knowledge-base such as WordNet. Graph-based methods reminiscent of spreading-activation research of the early days of AI research have been applied with some success.

The use of selectional preferences (or selectional restrictions) are also useful. For example, knowing that one typically cooks food, one can disambiguate the word *bass* in *I am cooking bass* (i.e., it's not a musical instrument).

Lesk's Algorithm

Sense Bag: contains the words in the definition of a candidate sense of the ambiguous word.

Context Bag: contains the words in the definition of each sense of each context word.



Walker's Algorithm

A Thesaurus Based approach

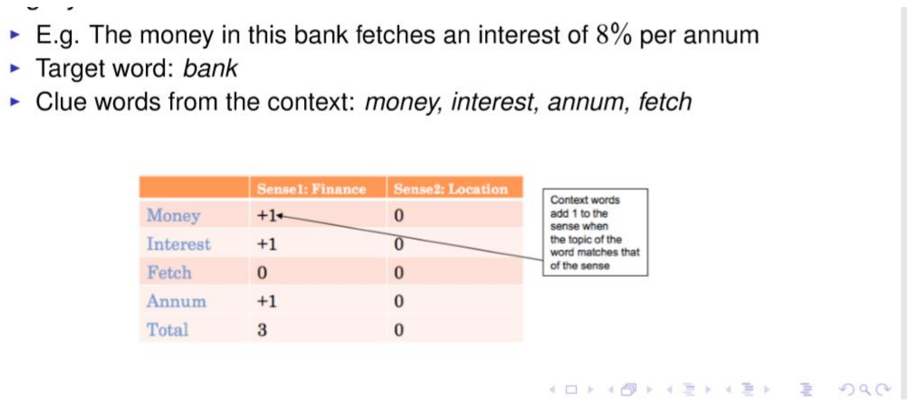
Step 1: For each sense of the target word find the thesaurus category to which that sense belongs

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



We can think of WSD as a kind of contextualized similarity task, since our goal is to be able to distinguish the meaning of a word like bass in one context (playingmusic) from another context (fishing). Somewhere in between lies the word-in-context task. Here the system is given word-in-context two sentences, each with the same target word but in a different sentential context. The system must decide whether the target words are used in the same sense in the two sentences or in a different sense.

---There's a lot of trash on the ***bed*** of the river

— I keep a glass of water next to my ***bed*** when I sleep

The WiC sentences are mainly taken from the example usages for senses in WordNet. But WordNet senses are very fine-grained. For this reason tasks like word-in-context first cluster the word senses into coarser clusters, so that the two sentential contexts for the target word are marked as T if the two senses are in the same cluster. The baseline algorithm to solve the WIC task uses contextual embeddings like BERT with a simple thresholded cosine. We first compute the contextual embeddings for the target word in each of the two sentences, and then compute the cosine between them. If it's above a threshold tuned on a devset we respond true (the two senses are the same) else we respond false.

Supervised methods

Supervised methods are based on the assumption that the context can provide enough evidence on its own to disambiguate words (hence, world knowledge and reasoning are deemed unnecessary). Probably every machine learning algorithm going has been applied to WSD, including associated techniques such as feature selection, parameter optimization, and ensemble learning. Support vector machines and memory-based learning have been shown to be the most successful approaches, to date, probably because they can cope with the high-dimensionality of the feature space. However, these supervised methods are subject to a new knowledge acquisition bottleneck since they rely on substantial amounts of manually sense-tagged corpora for training, which are laborious and expensive to create.

Naïve Bayes for WSD (Classification)

Definition of Classification

Input:

• a word w and some features f

• a fixed set of classes C = {c1, c2,…, cJ}

• Output: a predicted class c∈C

Supervised Machine Learning

Input:

• a word w in a text window d (which we'll call a "document")

• a fixed set of classes C = {c1, c2,…, cJ }

• A training set of m hand-labeled text windows again called "documents"

(d1,c1),....,(dm,cm) •

Output: • a learned classifier γ:d -> c

Applying Naive Bayes to WSD

P(c) is the prior probability of that sense

• Counting in a labeled training set.

• P(w|c) conditional probability of a word given a particular sense

• P(w|c) = count(w,c)/count(c)

• We get both of these from a tagged corpus like SemCor

Can also generalize to look at other features besides words.

• Then it would be P(f|c)

• Conditional probability of a feature given a sense

P ˆ (c) = Nc/ N

P ˆ (w | c) = count(w, c)+1/ count(c)+ |V |

	Doc	Words	Class
Training	1	Fish smoked fish	F
	2	Fish line	F
	3	Fish haul smoked	F
	4	Guitar jazz line	G
Test	5	Line guitar jazz jazz	?

V = {fish, smoked, line, haul, guitar, jazz}

Priors: P(f)= 3/4

P(g)= ¼

Conditional Probabilities:

P(line|f) = (1+1) / (8+6) = 2/14

P(guitar|f) = (0+1) / (8+6) = 1/14

P(jazz|f) = (0+1) / (8+6) = 1/14

P(line|g) = (1+1) / (3+6) = 2/9

P(guitar|g) = (1+1) / (3+6) = 2/9

P(jazz|g) = (1+1) / (3+6) = 2/9

Choosing a class:

P(f|d5) ∝ 3/4 * 2/14 * (1/14)2 * 1/14 ≈ 0.00003

P(g|d5) ∝ 1/4 * 2/9 * (2/9)2 * 2/9 ≈ 0.0006

Semi-supervised methods

The bootstrapping approach starts from a small amount of seed data for each word: either manually-tagged training examples or a small number of surefire decision rules (e.g., *play* in the context of *bass* almost always indicates the musical instrument). The seeds are used to train an initial classifier, using any supervised method. This classifier is then used on the untagged portion of the corpus to extract a larger training set, in which only the most confident classifications are included. The process repeats, each new classifier being trained on a successively larger training corpus, until the whole corpus is consumed, or until a given maximum number of iterations is reached.

Other semi-supervised techniques use large quantities of untagged corpora to provide co-occurrence information that supplements the tagged corpora. These techniques have the potential to help in the adaptation of supervised models to different domains.

Also, an ambiguous word in one language is often translated into different words in a second language depending on the sense of the word. Word-aligned bilingual corpora have been used to infer cross-lingual sense distinctions, a kind of semi-supervised system.

Unsupervised methods

Unsupervised learning is the greatest challenge for WSD researchers. The underlying assumption is that similar senses occur in similar contexts, and thus senses can be induced from text by clustering word occurrences using some measure of similarity of context. Then, new occurrences of the word can be classified into the closest induced clusters/senses. Performance has been lower than other methods, above, but comparisons are difficult since senses induced must be mapped to a known dictionary of word senses. Alternatively, if a mapping to a set of dictionary senses is not desired, cluster-based evaluations (including measures of entropy and purity) can be performed. It is hoped that unsupervised learning will overcome the knowledge acquisition bottleneck because they are not dependent on manual effort.

There are two kinds of test corpora:

- **Lexical sample:** the occurrences of a small sample of target words need to be disambiguated, and
- **All-words:** all the words in a piece of running text need to be disambiguated.

Why is WSD hard?

WSD is hard for many reasons, three of which are discussed here.

A sense inventory cannot be task-independent

A task-independent sense inventory is not a coherent concept: each task requires its own division of word meaning into senses relevant to the task. For example, the ambiguity of *mouse* (animal or device) is not relevant in English-French machine translation, but is relevant in information retrieval. The opposite is true of *river*, which requires a choice in French (*fleuve* 'flows into the sea', or *rivière* 'flows into a river').

Different algorithms for different applications

Completely different algorithms might be required by different applications. In machine translation, the problem takes the form of target word selection. Here the "senses" are words in the target language, which often correspond to significant meaning distinctions in the source language (*bank* could translate to French *banque* 'financial bank' or *rive* 'edge of river'). In information retrieval, a sense inventory is not necessarily required, because it is enough to know that a word is used in the same sense in the query and a retrieved document; what sense that is, is unimportant.

Word meaning does not divide up into discrete senses

Word meaning is in principle infinitely variable and context sensitive. It does not divide up easily into distinct or discrete sub-meanings. Lexicographers frequently discover in corpora loose and overlapping word meanings, and standard or conventional meanings extended, modulated, and exploited in a bewildering variety of ways.

Expected Questions:

1. What is semantic analysis? Why is it difficult? Explain various approaches to semantic analysis.
2. Explain with suitable examples following relationships between word meanings Homonymy, Polysemy, Synonymy, Antonymy, Hpernomy, Hyponomy, Meronymy.
3. What is semantic analysis? Discuss different semantic relationships between the words.
4. What is WordNet? How is sense defined in WordNet? Explain with example.
5. What do you mean by word sense disambiguation? Discuss dictionary based approach for WSD.
6. What do you mean by word sense disambiguation? Discuss knowledge based approach for WSD.
7. What do you mean by word sense disambiguation? Discuss machine learning based(Naïve Bayes) approach for WSD.
8. Explain how a supervised learning algorithm can be applied for word sense disambiguation.