19CSE453 – Natural Language Processing
Semantics

By

Ms. Kavitha C.R.

Dept. of Computer Science and Engineering

Amrita School of Engineering, Bengaluru

Amrita Vishwa Vidyapeetham

AMRITA
VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

School of
Engineering

Semantic - Meaning

- Semantics is the study of the meanings of words, phrases and sentences.

- Linguistics semantics deals with the conventional meaning conveyed by the use of words, phrases and sentences of a language.

- This technical approach is concerned with objective or general meaning and avoids trying to account for subjective or local meaning.

Semantic - Meaning

- To understand language
  - the meaning of words and of the morphemes that compose them
  - Words into phrases and sentences
  - Context which determines the meaning (Pragmatics)

- The study of meaning of words, phrases, and sentences.
  - Lexical semantics (words and meaning relationship among words)
  - Phrasal/ sentential semantics (syntactic units larger than a word)

- What a speaker <u>conventionally</u> means (objective or general meaning)- *not what he is trying to say* (subjective or local meaning)

Semantics Representation

Semantics is a branch of linguistics that looks at the meanings of words and language, including the symbolic use of language. It also refers to the multiple meanings of words as well.

Semantic representation is **an abstract (formal) language** in **which meanings can be represented**.

Opinions differ about whether semantic representation is sufficient or necessary, about its form and about how it relates to syntactic representations.

Semantic representation can take the form of a structure of semantic features or formula of a logical system

Two terms that are related to **semantics** are <u>connotation</u> and <u>denotation</u>.

<u>Connotation</u> refers to the meanings that we associate with the word-beyond the literal dictionary definition. The <u>connotation</u> of a word includes all of the emotions and feelings that go along with the use of the word.

<u>Denotation</u> includes the literal definition of the word.

When we read literature we are looking at language that is used to <u>denote</u> and <u>connote</u> meaning.

Semantics is the study and analysis of how language is used figuratively and literally to produce meaning.

**Semantics** seeks to describe how words are used-not to prescribe how they should be used

Examples of Semantics

1. A toy block could be called a block, a cube, a toy.

2. A child could be called a child, kid, boy, girl, son, daughter.

3. The word "run" has many meanings-physically running, depart or go (I have to run, spent (it has run its course)

Example of Semantic representation:

When a visual scene is interpreted, it is represented semantically in the cognitive system as a network in which objects are identified and represented (as nodes), their properties are represented by links to attributes, and their relationships to each other are represented by particular types of semantic
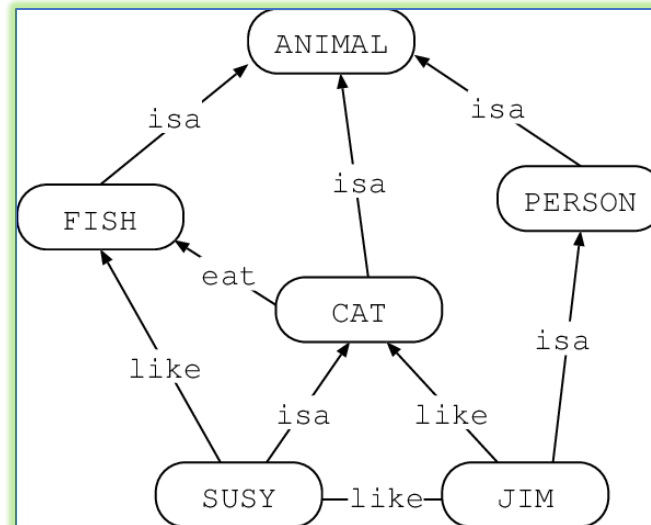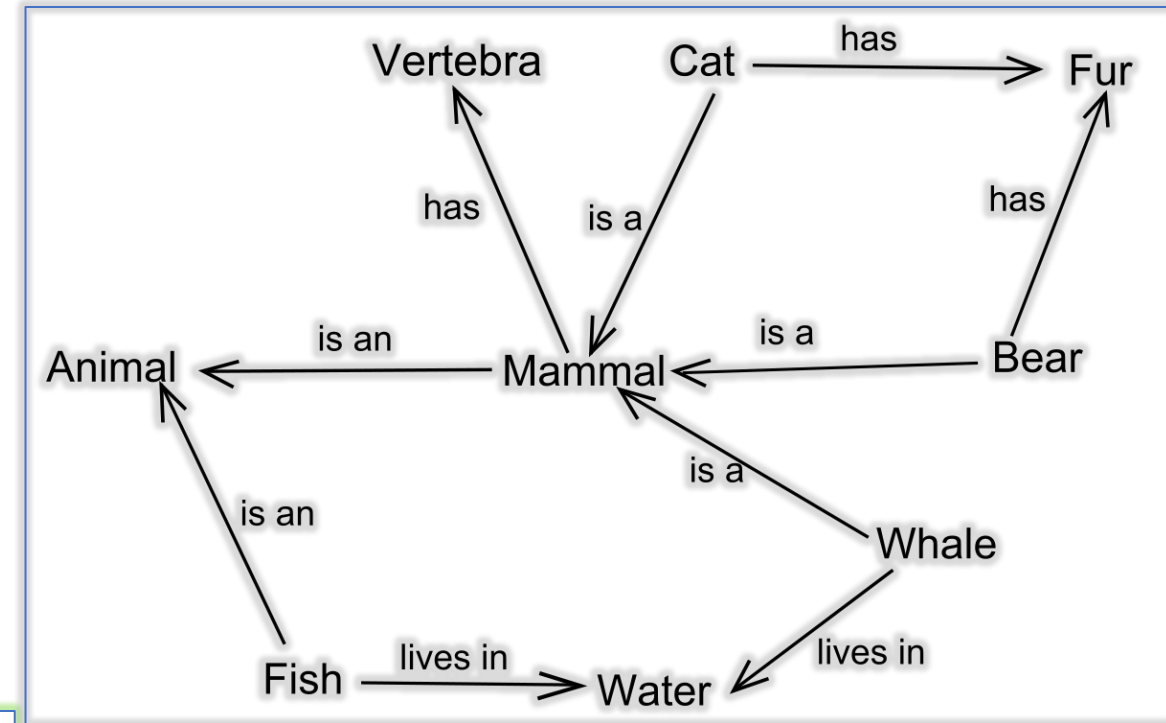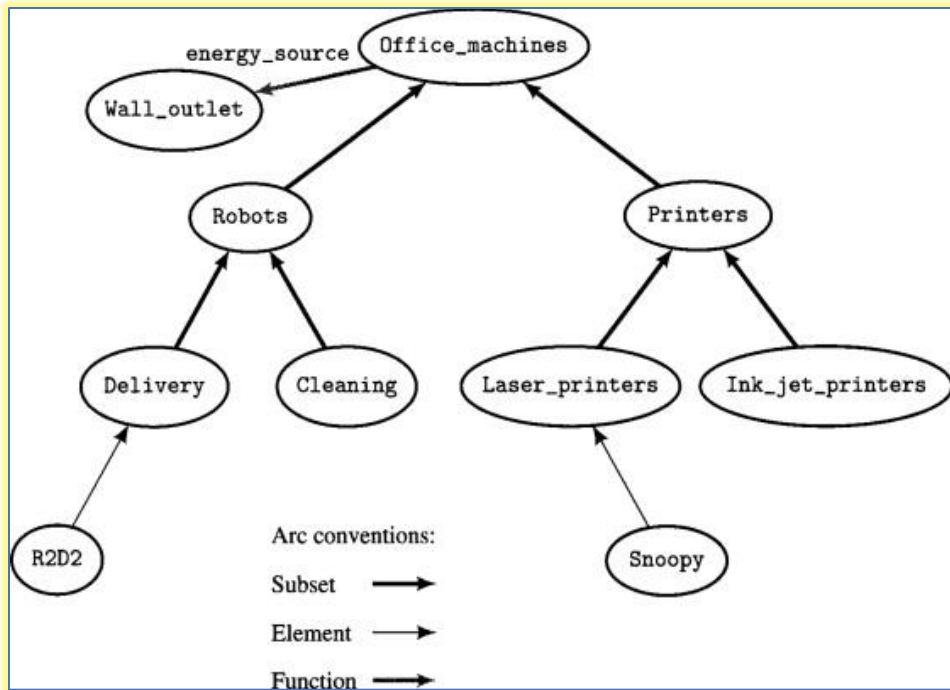
Sentence: While the mother washed the baby cried.

When the above sentence is presented without a comma after the verb 'washed,' readers frequently interpret that the mother washed the baby.

Once the verb 'cried' is encountered, it becomes clear that the mother washed herself and that 'the baby' is the subject of the upcoming clause, not the direct object of the verb 'washed.'

While the processing system eventually comes to the correct reading, it sometimes fails to completely rule out the incorrect interpretation.

Types of Meaning

1. Conceptual meaning
2. Connotative meaning
3. Stylistic meaning
4. Affective meaning
5. Reflective meaning
6. Collocative meaning
7. Thematic meaning

Conceptual Meaning

- Conceptual meaning covers those basic essential components of meaning that are conveyed by the literal use of a word.

- Some basic components of a word like *needle* in English might include 'thin, sharp, steel instrument'. These components would be part of the conceptual meaning of *needle*.

- literally meaning is basically using a word or sentence NOT in the way it sounds, but the way the dictionary would describe it.

  here is an example

  I felt like i was walking on water... (implying that you were not walking on water, it just felt that way)

  Vs the literal

  I was literally walking on water (the device i used made me walk on the water)

08-06-2023

Associative Meaning

- Different people might have different associations or connotations attached to a word like *needle*. They might associate it with 'pain', or 'illness', or 'blood', or 'drugs', or 'thread', or 'knitting', or 'hard to find' etc.

- These associations may differ from one person to the next, and are not treated as part of the word's conceptual meaning.

Note

- Poets, novelists, advertisers and lovers may be very interested in using words in such a way that certain associative meanings are evoked and literary critics often write about this aspect of language use.

- When we investigate the meaning of words in a language, we are normally interested in characterizing the conceptual meaning and less concerned with the associative meaning of the words.

Types of Semantic Analysis

- Words as 'containers' → *Semantic features*
- 'roles' they fulfill → *Semantic roles*
- 'relationship' with other words → *lexical relation*

Semantic Features

- Words as 'containers' of meaning

Semantic Roles

- Words as fulfilling 'roles' within the situation described by a sentence

Lexical Relations

- 'Relationships' between words

Semantic Features

Read the following sentences.

The hamburger ate the boy.

The table listens to the radio.

The horse is reading the newspaper.

The tree ate the elephant

Do you find these sentences odd? Why?

Semantic Features

- We should note that the oddness of the sentences in the previous slide does not derive from their syntactical structure.

    The hamburger ate the boy.

    The boy ate the hamburger.

- The first sentence is syntactically good, but semantically odd.

- The components of the conceptual meaning of the noun *hamburger* must be significantly different from those of the noun *boy*, thereby preventing one, not being the other, from being used as the subject of the verb *ate*.

- The kind of noun that can be the subject of the verb *ate* must denote an entity that is capable of 'eating'. The noun *hamburger* does not have this property and the noun *boy* does.

- We can make this observation more generally applicable by trying to determine the crucial element or feature of meaning that any noun must have in order to be used as the subject of the verb *ate*.

- Such an element may be as general as 'animate being'.

- We can then use this idea to describe part of the meaning of words as having either plus (+) or minus (-) that particular feature.

- The feature that the noun *boy* has is '+ animate' (= denotes an animate being) and the feature that the noun *hamburger* has is '- animate' (= denotes an inanimate being).

Semantic Features

- If we had to provide the crucial distinguishing features of the meaning of a set of English words such as *table*, *horse*, *boy*, *man*, *girl*, *woman*, we could begin with the following diagram.

- Semantic properties: The components of meaning of a word.
- Meaning as collection of properties/features typically with two possible values (+ / -)
- Example of componential analysis: *baby* is [+ young], [+ human], [+animate]

	table	horse	boy	man	girl	woman
animate	-	+	+	+	+	+
human	-	-	+	+	+	+
female	-	-	-	-	+	+
adult	-	+	-	+	-	+

1.   (a) widow, mother, sister, aunt, maid
    (b) widower, father, brother, uncle
    → The (a) and (b) words are [+ human]
    → The (a) words are [+ female]
    → The (b) words are [+ male]
2. (a) bachelor, paperboy, pope, chief
    (b) bull, rooster, drake, ram
    → The (a) and (b) words are [+ male]
    → The (a) words are [+ human]
    → The (b) words are [+ animal]

Semantic Features

- Features such as '+ animate, - animate', '+ human, - human', '+ female, - female' for example, can be treated as the basic elements involved in differentiating the meaning of each word in a language from every other word.

- From a feature analysis like this, we can say that at least part of the meaning of the word *girl* in English involves the elements [+human, +female, -adult].

- We can also characterize the feature that is crucially required in a noun in order for it to appear as the subject of a particular verb, supplementing the syntactic analysis along with semantic features.

The _____is reading the newspaper.

N [+ human]

- This approach gives us the ability to predict which nouns make this sentence semantically odd.

- Some examples would be *table*, *horse* and *hamburger*, because none of them have the required feature [+ human].

- The approach just outlined is a start on analyzing the conceptual components of word meaning, but it is not without problems.

- For many words in a language it may not be easy to come up with neat components of meaning. If we try to think of the components or features, we would use to differentiate the nouns *advice*, *threat* and *warning*, for example, we would not be very successful.

- Part of the problem seems to be that the approach involves a view of words in a language as some sort of 'containers' that carry meaning components.

- There is clearly more to the meaning of words than these basic types of features.

- Instead of thinking of words as 'containers' of meaning, we can look at the 'roles' they fulfil within the situation described by a sentence.

- If the situation is a simple event, as in *The boy kicked the ball*, then the verb describes an action (*kick*). The noun phrases in the sentence describe the roles of entities, such as people and things, involved in the action.

Semantic Roles

We can identify a small number of semantic roles (also called 'thematic roles') for these noun phrases.

agent and theme	instrument and experiencer	location, source and goal
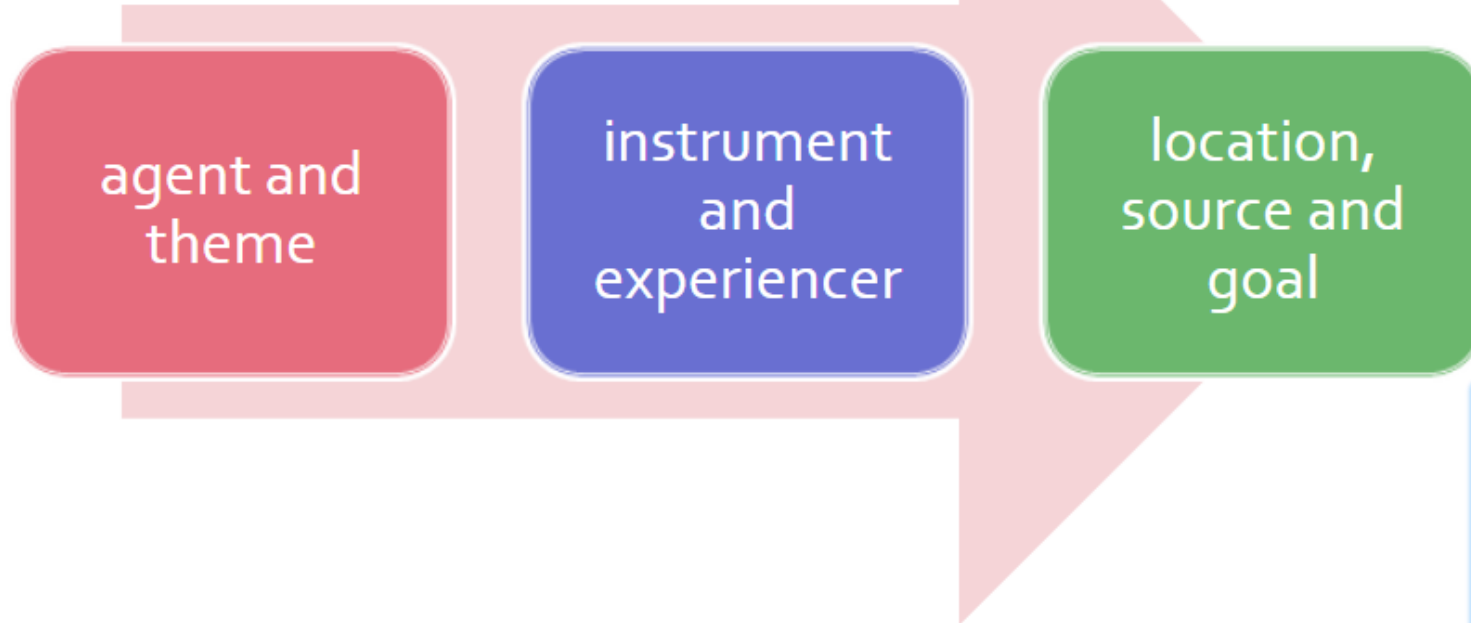
- Words are described according to the roles they fulfill with the situation described in a sentence.
  - The boy *kicked* the ball
    - verb → indicates action
    - Boy → performs the action= **agent**
    - Ball → undergoes the action= **theme**
- The NPs describe the role of entities (people or things) involved in the action, i.e. they have certain semantic (or thematic) roles.

Semantic Roles

- Linguistic semantics deals with the meaning conveyed by words, phrases and sentences that are used in a language.

- These words, phrases and sentences tell us about an event that occurs in the real world.

- e.g. Ali cut the cake

  tells us about something that happened in reality.
  A real boy Ali, did an action [cut]
  to a real thing [cake]

- **Agent** = the entity that performs the action
- **Theme** = the entity that undergoes the action
- **Experiencer** = one who perceives something
- **Instrument** = an entity used to perform an action
- **Location** = the place where the action happens
- **Source** = the place from which an action originates
- **Goal** = the place where the action is directed

The boy kicked the ball.

- **Agent** : the entity that performs the action, e.g. *the boy*

- **Theme** (or the 'patient'): the entity that is involved in or affected by the action, e.g. *the ball*

- The theme can also be an entity (*The ball*) that is simply being described (i.e. not performing an action), as in *The ball was red.*

- John is writing with a pen
  - agent   instrument
- Mary saw a mosquito on the wall
  - experiencer   theme location
- The children ran from the playground to the pool
  - agent   source   goal
- The boy opened the door with a key
- The dog bit the stick
- With a stick, the man hit the dog.

- Agents and themes are the most common semantic roles.

- Although agents are typically human (*The boy*), they can also be non-human entities that cause actions, as in noun phrases denoting a natural force (*The wind*), a machine (*A car*), or a creature (*The dog*), all of which affect *the ball* as theme.

The boy kicked the ball.
The wind blew the ball away.

A car ran over the ball.

The dog caught the ball.

- The theme is typically non-human, but can be human (*the boy*), as in *The dog chased the boy.*

- The same physical entity can appear in two different semantic roles in a sentence, as in *The boy cut himself.* Here *The boy* is agent and *himself* is theme.

Instrument and Experiencer

- If an agent uses another entity in order to perform an action, that other entity fills the role of **instrument** .

  The boy cut the rope with an old razor.

  He drew the picture with a crayon.

- In the above sentences, the noun phrases *an old razor* and *a crayon* are being used in the semantic role of instrument.

- When a noun phrase is used to designate an entity as the person who has a feeling, perception or state, it fills the semantic role of **experiencer** .

- If we *see*, *know*, or *enjoy* something, we are not really performing an action (hence we are not agents). We are in the role of an experiencer.

- In the sentence *The boy feels sad*, the experiencer (*The boy*) is the only semantic role.

- In the question, *Did you hear that noise?*, the experiencer is *you* and the theme is *that noise*.

  - **Location** : where the entity is (*on the table*, *in the room*)

  - **Source** : where the entity moves from (*from Chicago*)

  - **Goal** : where the entity moves to (*to New Orleans*)

    We drove from Chicago to New Orleans.

# Semantic Roles : Illustrations

Mary	saw	a fly	on the wall.
EXPERIENCER		THEME	LOCATION
She	borrowed	a magazine	from George.
AGENT		THEME	SOURCE
She	squashed	the bug	with the book.
AGENT		THEME	INSTRUMENT
She	handed	the magazine	back to George.
AGENT		THEME	GOAL
"Gee thanks,"	said	George.	
		AGENT	

Lexical Relations

- What is the meaning of 'big'?
  - 'Large' or the opposite of 'small'

- What is the meaning of 'daffodil'?
  - A kind of flower

- Analysis in terms of lexical relations- *explain the meaning in terms of the relationship with other words*
    - Synonymy
    - Antonymy
    - Hyponymy
    - Prototype
    - Homophones and Homonyms
    - Polysemy

Lexical Relations

Synonymy → Antonymy → Hyponymy → Prototypes

Homophones and Homonyms → Polysemy → Metonymy → Collocation
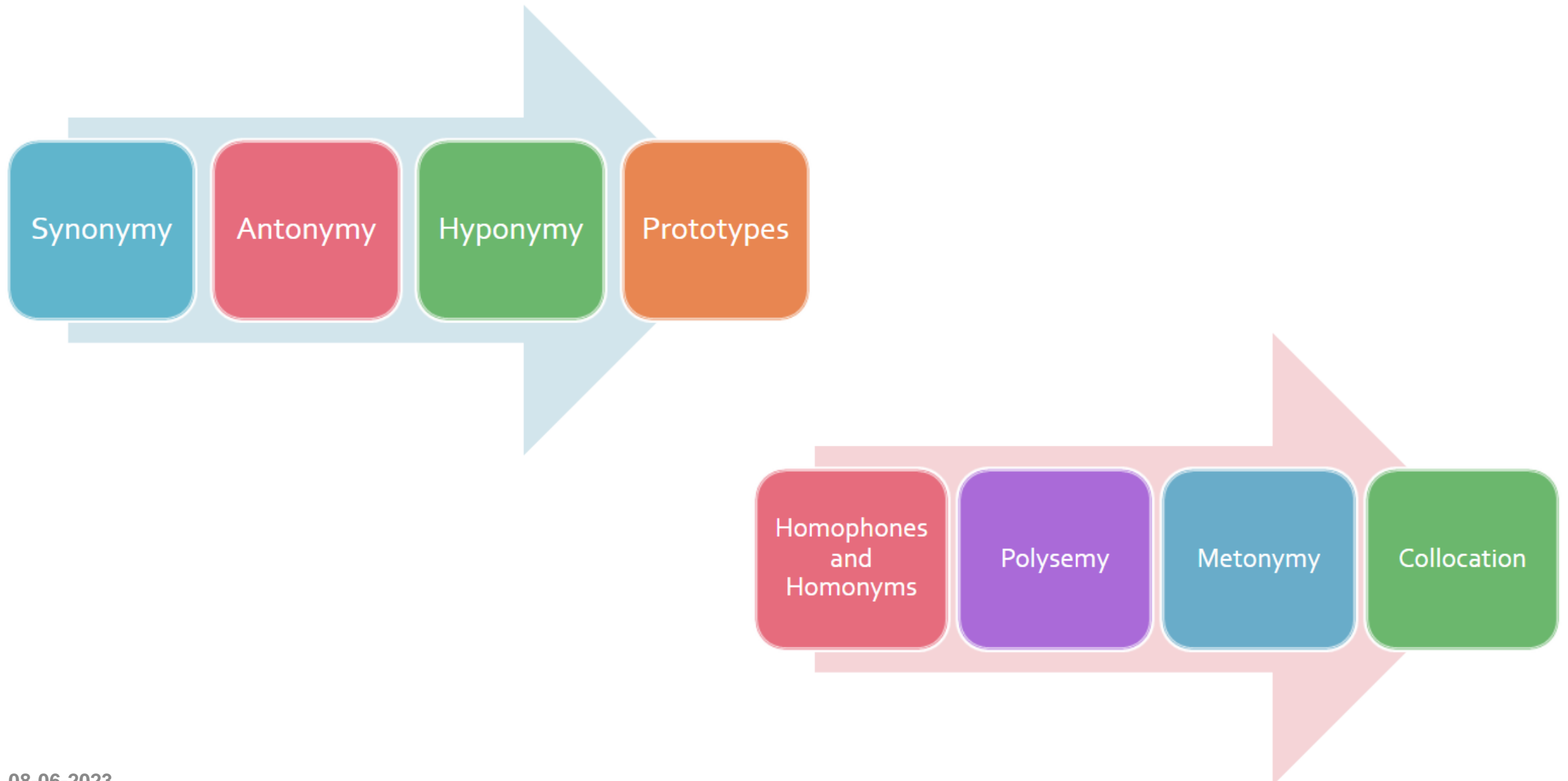
Synonyms

- Two or more words with very closely related meanings are called **synonyms** .

- Synonyms can often, though not always, be substituted for each other in sentences.

- <u>In appropriate circumstances</u>, we can say *What was his answer?* or *What was his reply?* With much the same meaning.

> - **<u>Synonymy</u>**: words that have the same meanings or that are closely related in meaning
>
> - E.g. answer/reply – almost/nearly – broad/wide – buy/purchase – freedom/ liberty
>
> - 'sameness' is not 'total sameness'- only one word would be appropriate in a sentence.
>   - E.g. *Sandy only had one <u>answer</u> correct on the test.* (but NOT reply)
>
> - Synonyms differ in formality
>   - E.g buy/purchase – automobile/car

Synonyms

- The idea of 'sameness' of meaning used in discussing synonymy is not necessarily 'total sameness'. There are many occasions when <u>one word is appropriate in a sentence, but its synonym would be odd</u>.

- For example, *reply* would be odd in this instance.
  - *Sandy had only one* **answer** *correct on the test*
  - *Sandy had only one* **reply** *correct on the test.*

- Synonymous forms may also differ in terms of formal versus informal uses.

- For example, the second version, with four synonymous replacements, sounds much more casual or informal than the first.
  - *My father purchased a large automobile.*
  - *My dad bought a big car.*

Antonyms

- Two forms with opposite meanings are called **antonyms** .

- Some common examples are the pairs: *alive / dead, big / small,* fast/ *slow, happy / sad, hot / cold, long / short, male / female, married/ single, old / new, rich / poor, true / false*

- Antonyms are usually divided into two main types:
  - gradable – opposites along a scale
  - non-gradable – direct opposites

- **Antonymy**: words that are opposites in meaning, e.g. hot & cold.

- Types
- *Gradable*= not absolute, question of degree
  - Hot & cold – small & big
- *Non-gradable*:
  - Dead & alive – asleep & awake

E.g.   happy/sad   married/single
           present/absent   fast/slow

- **Gradable antonyms**, such as *big / small*, can be used in comparative constructions like *I'm bigger than you* and *A pony is smaller than a horse*.

- The negative of one member of a gradable pair does not necessarily imply the other. For example, the sentence *My car isn't old* does not necessarily mean *My car is new*.

- With **non-gradable antonyms** (also called 'complementary pairs'), comparative constructions are not normally used. We don't typically describe someone as *deader* or *more dead* than another.

- The negative of one member of a non-gradable pair does imply the other member. That is, *My grandparents aren't alive* does indeed mean *My grandparents are dead*.

- Although we can use the 'negative test' to identify non-gradable antonyms in a language, we usually avoid describing one member of an antonymous pair as the negative of the other.

- For example, while *undress* can be treated as the opposite of *dress*, it does not mean 'not dress'. It actually means 'do the reverse of dress'. Antonyms of this type are called **reversives** .

- Other common examples of **reversives** are: *enter / exit, pack / unpack, lengthen / shorten, raise / lower, tie / untie*

Synonymy or Antonymy

- Flourish – thrive
- Intelligent – stupid
- Casual – informal
- deep-profound
- Drunk – sober
- Sofa – couch
- Hide – conceal
- cheap – expensive
- Rich - wealthy

a. synonym
b. antonym
c. synonym
d. synonym
e. antonym
f. Synonym
g. Synonym
h. Antonym
i. synonym

Hyponymy

- When the meaning of one form is included in the meaning of another, the relationship is described as **hyponymy** .

- Examples are the pairs: *animal / dog, dog / poodle, vegetable / carrot, flower / rose, tree / banyan.*

- The concept of 'inclusion' involved in this relationship is the idea that if an object is a *rose*, then it is necessarily a *flower*, so the meaning of *flower* is included in the meaning of *rose*. Or *rose* is a hyponym of *flower*.

- **Hyponymy**: Words whose meanings are specific instances of a more general word, i.e. *one thing is included (kind of) in another thing.*
  - e.g. cats and dogs are hyponyms of the word animal.
- In this case cats and dogs are co-hyponyms share the same 'superordinate'
- Other e.g. daffodil & flower / carrot & vegetable / ant & insect

Hyponymy

- When we consider hyponymous connections, we are essentially looking at the meaning of words in some type of hierarchical relationship.

- We can represent the relationships between a set of words as a hierarchical diagram.
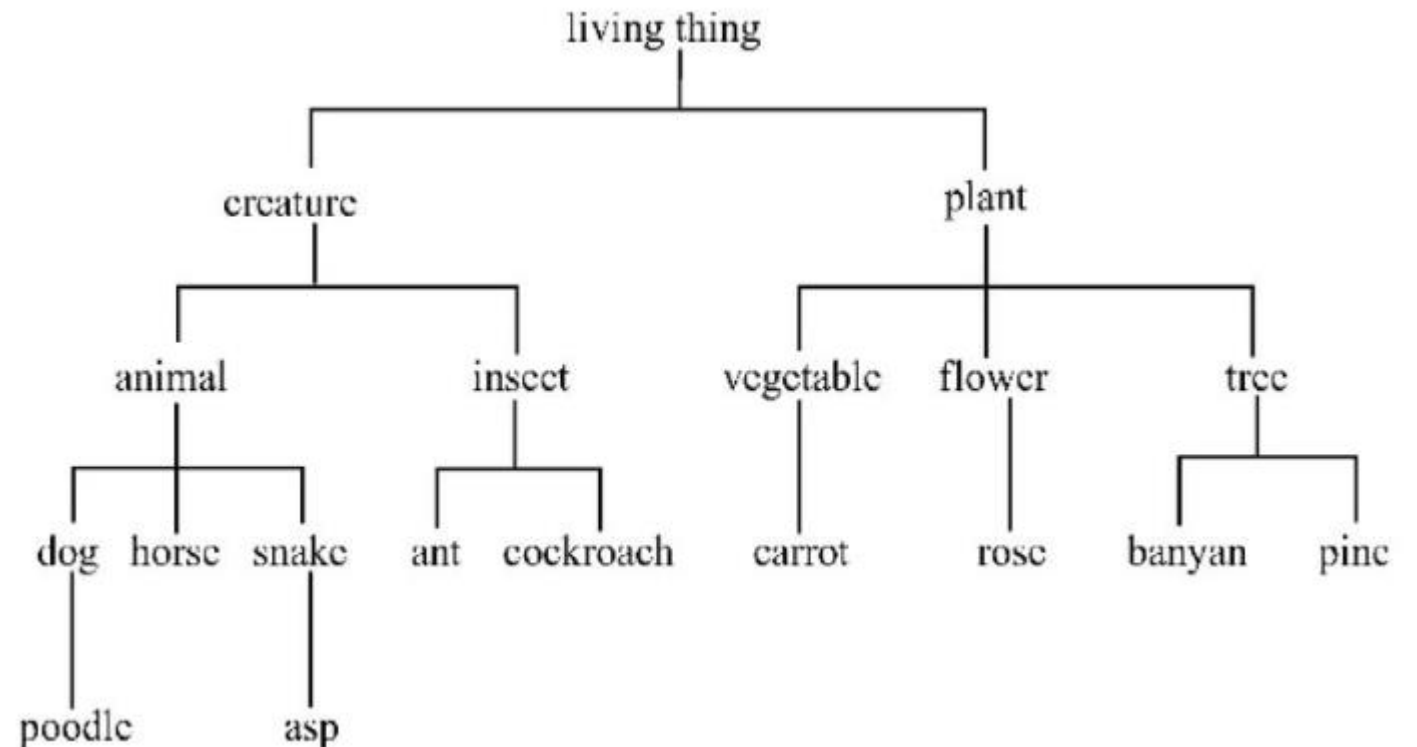
Prototypes

- The concept of a **prototype** helps explain the meaning of certain words like *bird*, not in terms of component features (e.g. 'has feathers', 'has wings'), but in terms of resemblance to the clearest example.

  - While words like *canary, cormorant, dove, duck, flamingo, parrot, pelican* and *robin* are all equally co-hyponyms of the super-ordinate *bird*, they are not all considered to be equally good examples of the category 'bird'.

  - According to some researchers, the most characteristic instance of the category 'bird' is *robin*.

    - Thus, even native speakers of English might wonder if *ostrich* or *penguin* should be hyponyms of *bird* (technically they are), but have no trouble deciding about *sparrow* or *pigeon*. These last two are much closer to the prototype.

Homophones and Homonyms

- When two or more different (written) forms have the same pronunciation, they are described as **homophones** .

- Common examples are *bare / bear, meat / meet, flour / flower, pail / pale, right / write, sew / so* and *to / too / two.*

  - We use the term **homonyms** when one form (written or spoken) has two or more unrelated meanings, as in these examples:

    bank (of a river) – *bank* (financial institution)

    Bat (flying creature) – *bat* (used in sports)

    mole (on skin) – *mole* (small animal)

    pupil (at school) – *pupil* (in the eye)

    race (contest of speed) – *race* (ethnic group)

bank (of a river) – *bank* (financial institution)

- The temptation is to think that the two types of *bank* must be related in meaning. They are not.

- Homonyms are words that have separate histories and meanings, but have accidentally come to have exactly the same form.

- **Homonymy**: A word which has two or more *entirely distinct* (unrelated) meanings,
  - e.g. **bank**: 'financial institution' ; 'of a river'.
  - **Bat**: 'flying creature' or 'used in sports'
  - **Race**: 'contest of speed' or 'ethnic group'
- **Homophony**: Different words pronounced the same but spelled differently,
  - e.g. *two, to* and *too*.
  - *Flour* and *flower*
  - *Meat* and *meet*
  - *Right* and *write*

Polysemy

- When we encounter two or more words with the same form and related meanings, we have what is technically known as **polysemy** .

- Polysemy can be defined as one form (written or spoken) having multiple meanings that are all related by extension.
- Examples are the word *head*, used to refer to the object on top of your body, on top of a glass of beer, person at the top of a company or department and many other things.

- Other examples of polysemy are *foot* (of person, of bed, of mountain) or *run* (person does, water does, colours do).

- **Polysemy**: A word which has *multiple meanings related by extension*,
  - e.g. *bright*: 'shining' ; 'intelligent'
  - '*Head*' of the body and the person at the top of a company.
  - 'Foot' of a body and of a mountain and of the bed or chair.
  - 'Run' a person runs, the water runs

- If we are not sure whether different uses of a single word are examples of homonymy or polysemy, we can check in a dictionary.
- If the word has multiple meanings (i.e. it is polysemous), then there will be a single entry, with a numbered list of the different meanings of the word.

- If the two words are treated as homonyms, they will typically have two separate entries.

- It is possible for two forms to be distinguished via homonymy and for one of the forms also to have uses via polysemy.

- *The words date* (= a thing we eat) and *date* (= a point in time) are homonyms.

- *Date* (= a point in time) is polysemous in terms of a particular day and month (= on a letter), an arranged meeting time (= an appointment), a social meeting (= with someone we like), and even a person (= that person we like).

Metonymy

- A type of relationship based on a close connection in everyday experience, which can be based on:

  - container-contents relation (*bottle / water, can / juice*)

  - whole-part relation (*car / wheels, house / roof*)

  - representative-symbol relationship (*king / crown, the President / White House*)

- Using one of these words to refer to the other is an example of **metonymy** .

  - It is our familiarity with metonymy that makes it possible for us to understand *He drank the whole bottle* although it sounds absurd literally (i.e. he drank the liquid, not the glass object).

  - We also accept *The White House has announced...* or *Downing Street protested...* without being puzzled that buildings appear to be talking.

- Many examples of metonymy are highly conventionalized and easy to interpret. However, other examples depend on an ability to infer what the speaker has in mind. Examples include:

   Get your butt over here.

   The strings are too quiet.

   I prefer cable.

- Making sense of such expressions often depends on context, background knowledge and inference.

- *What do you think about these sentence?*
  - He drank the whole bottle. *(container-content)*
  - The White House announced. *(king-crown)*
  - I gave her a hand. *(whole-part)*

- A word substituted for another word with which it is closely associated e.g. bottle is used for water

- **Metonymy** is "a figure of speech in which an attribute or commonly associated feature is used to name or designate something." A short definition is "part for whole."

Collocation

- We know that some words tend to occur with other words. If you ask a thousand people what they think of when you say *hammer*, more than half will say *nail*. If you say *table*, they will mostly say *chair*, and *butter* elicits *bread*, *needle* elicits *thread* and *salt* elicits *pepper*.

- One way we seem to organize our knowledge of words is simply on the basis of **collocation**, or frequently occurring together.

- Words tend to occur with other words.

- E.g. table/chair
- Butter/bread
- Salt/pepper
- Hammer/ nail

Kinds of Semantics

There are three main kinds of semantics:
- Formal semantics
- Lexical semantics
- Conceptual semantics

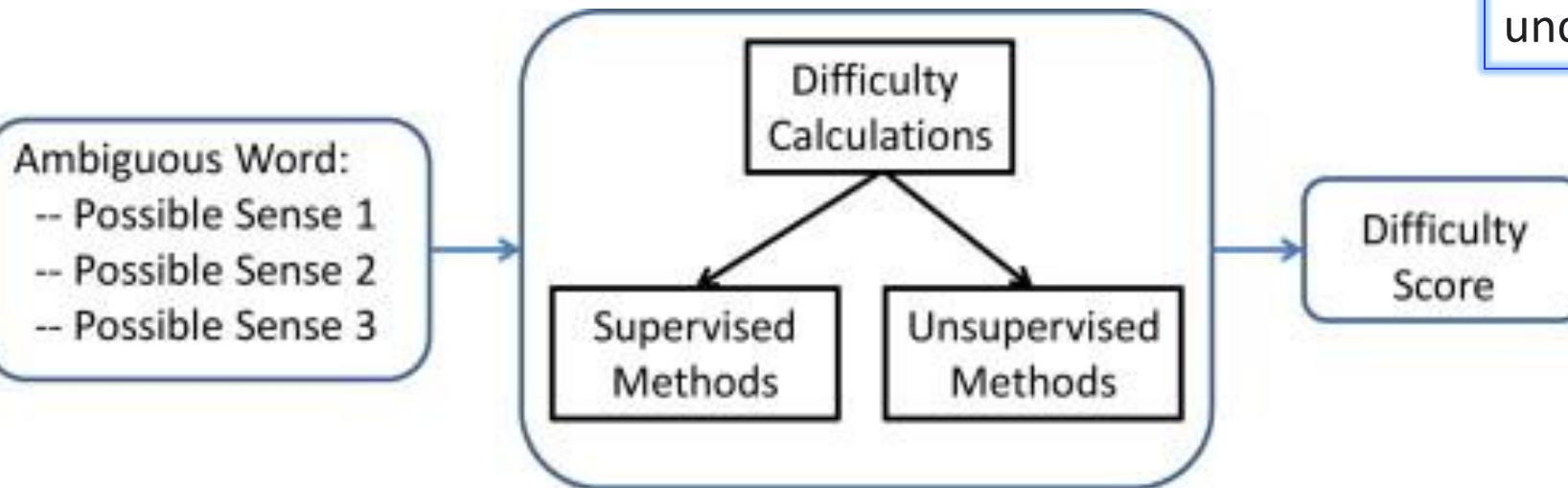# Word Sense Disambiguation (WSD)

WSD is basically **solution to the ambiguity which arises due to different meaning of words in different context**.
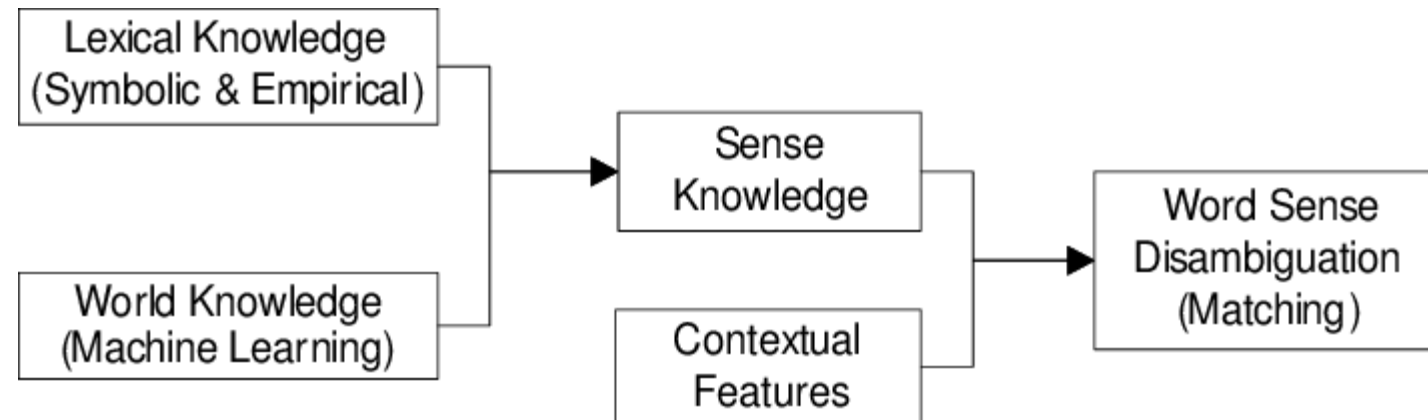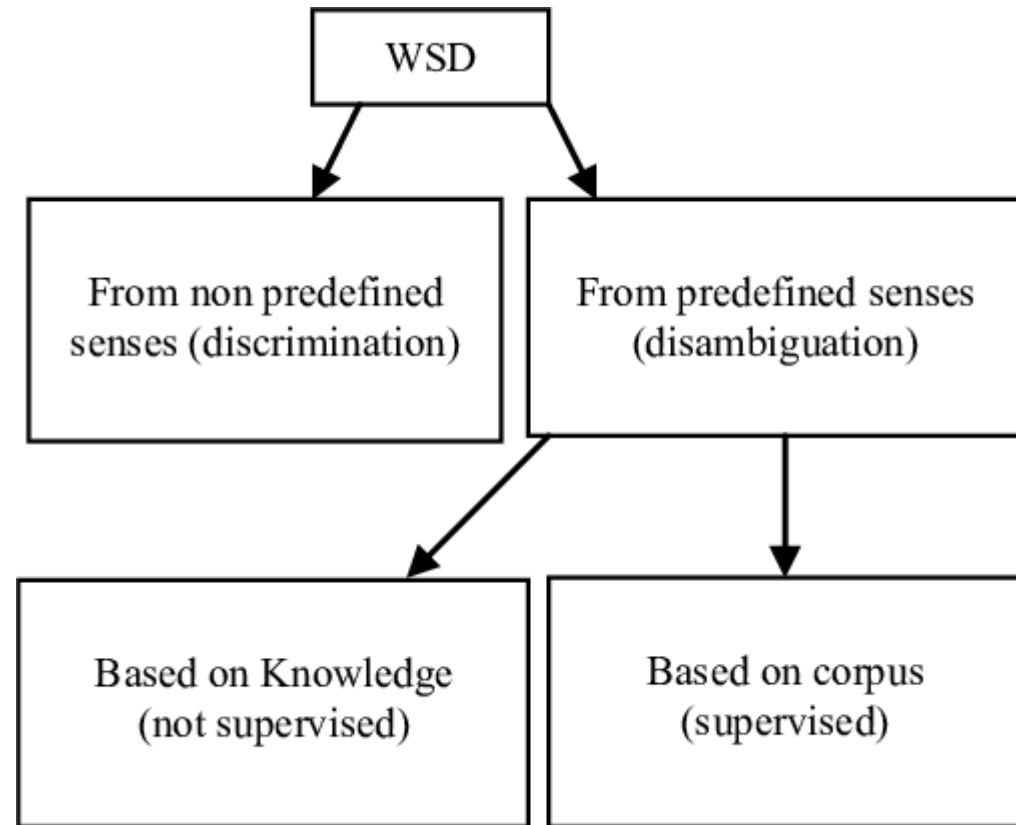For example, consider the two sentences.
"The bank will not be accepting cash on Saturdays".
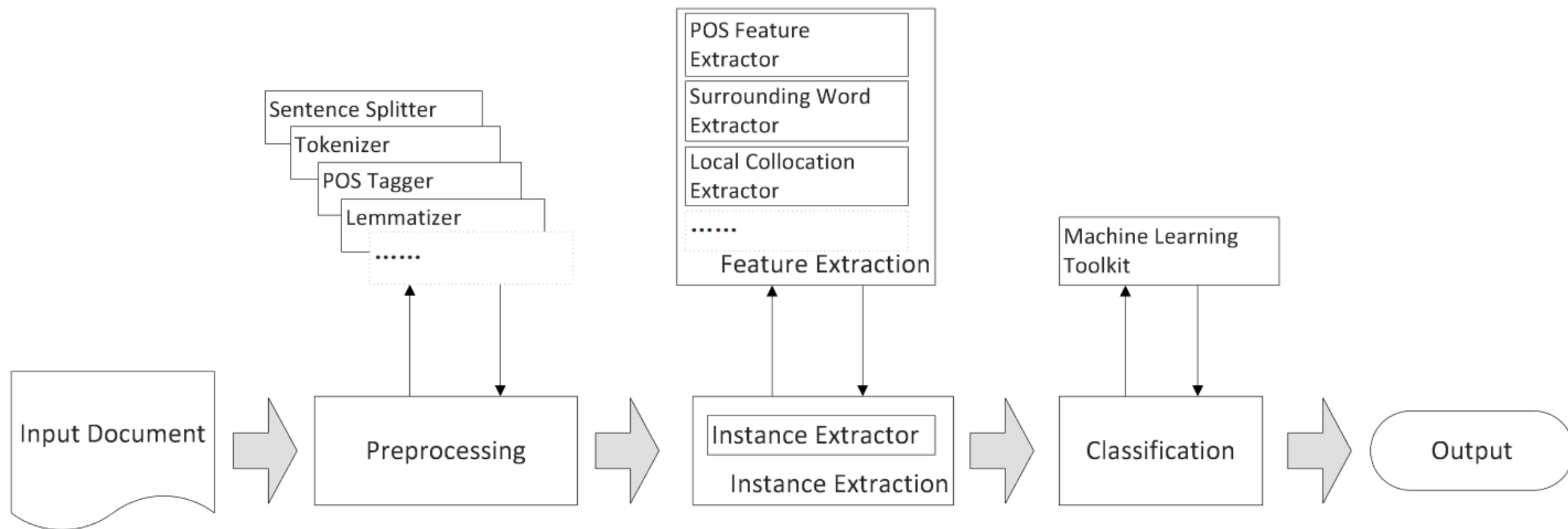"The river overflowed the bank".

In natural language processing, word sense disambiguation (WSD) is **the problem of determining which "sense" (meaning) of a word is activated by the use of the word in a particular context**, a process which appears to be largely unconscious in people.

Ambiguous Word:
-- Possible Sense 1
-- Possible Sense 2
-- Possible Sense 3

Difficulty Calculations

Supervised Methods

Unsupervised Methods

Difficulty Score

Classification of Word Sense Disambiguation

Conceptual Model for word sense disambiguation

Word Sense Disambiguation Applications

WSD has many applications in various text processing and NLP fields.

• Lexicography - Much of the modern Lexicography is corpus-based. WSD, used in Lexicography can provide significant textual indicators.

• Text Mining and Information Extraction tasks. As the major purpose of WSD is to accurately understand the meaning of a word in particular usage or sentence, it can be used for the correct labeling of words.

For example: From a security point of view, a text system should be able to understand the difference between a **coal "mine"** and a **land "mine"**.

While the former serves industrial purposes, the latter is a security threat. So a text mining application must be able to determine the difference between the two.

•Information Retrieval - Information Retrieval systems work through text data primarily based on textual information. Knowing the relevance of using a word in any sentence will surely help.

Challenges in Word Sense Disambiguation

WSD faces a lot of challenges and problems.

•The most common problem is the difference between various dictionaries or text corpus.

Different dictionaries have different meanings for words, which makes the sense of the words to be perceived as different. A lot of text information is out there and often it is not possible to process everything properly.

•Different applications need different algorithms and that is often a challenge for WSD.

•A problem also arises is that words cannot be divided into discrete meanings. Words often have related meanings and this causes a lot of problems.

Selectional restriction

Selectional restrictions are **the semantic restrictions that a word imposes on the environment in which it occurs**.

Selectional restrictions **place semantic constraints on arguments and account for the implausibility of sentences** such as "Colorless green ideas slept furiously"

They have been used in natural language understanding for disambiguation and pronoun resolution.

Machine learning approaches

Machine Learning Based Approach: **A classifier is used to learn features and assigns senses to unseen examples**.

In these approaches, the initial input is the word to be disambiguated called target word, and the text in which it is embedded, called as context.
In this approach features are themselves served by the words.

1. Naïve Bayes Classifier
2. Decision Tree
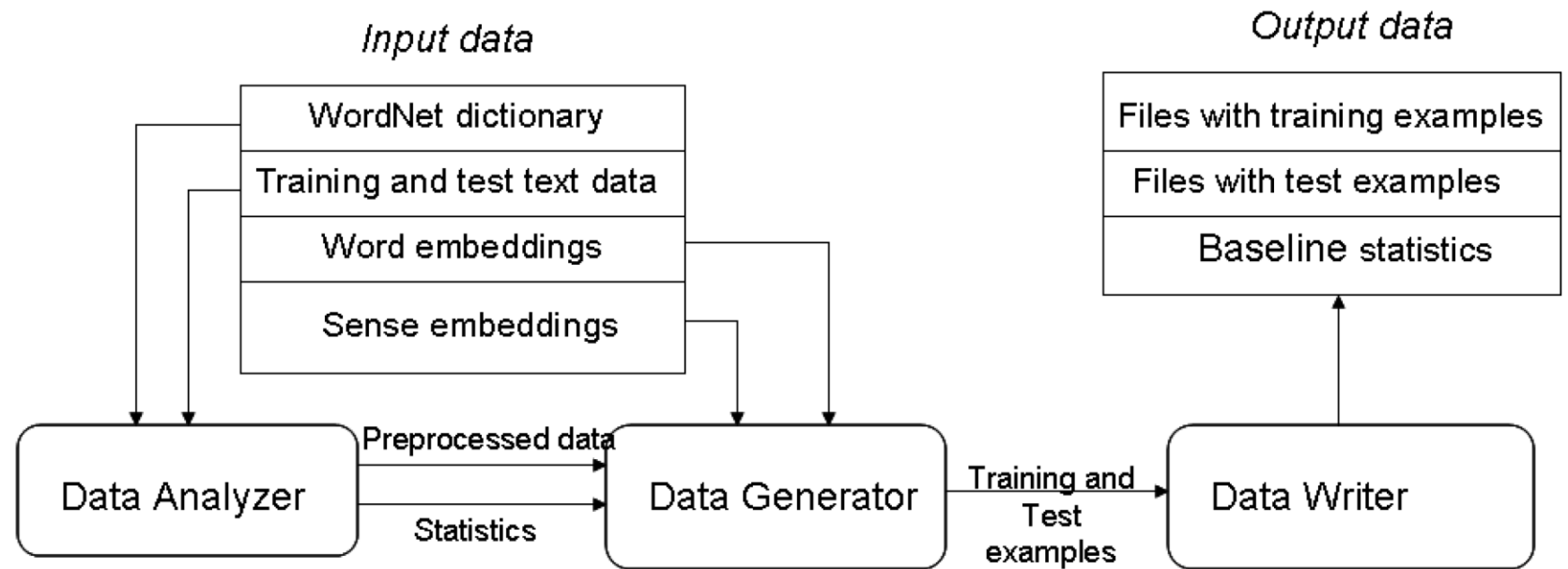
Dictionary based or Knowledge-based approaches

Dictionary Based Approach: Knowledge based approach based on knowledge resources of machine readable dictionaries in form of corpus, WordNet etc. they may use grammar rules for disambiguation.

The aim of Knowledge based approach (Dictionary based approach) WSD is to exploit knowledge resources to infer the senses of words in context.

The knowledge resources are **dictionaries, thesauri, ontology's, collocations** etc.

The above methods have lower performance than their supervised alternative methods but they have an advantage of a wider range.

Knowledge-based Approach

- Knowledge based algorithms use various lexical resources such as Machine Readable Dictionaries (MRDs) or WordNet to identify the correct sense of words.
- A knowledge based system only needs access to commercial dictionary resources to start process of disambiguation.

52
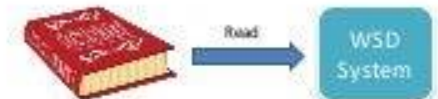
Dictionary Based Approaches

- Large-scale disambiguation possible
- Sense definitions retrieved from the dictionary
  - The sense with highest overlap within context words
- Dictionary entries relative short
  - Not enough overlap
  - expand word list, subject codes

Thank You