UNIT 4

SEMANTICS AND PRAGMATICS

SEMANTICS
VERSUS
PRAGMATICS

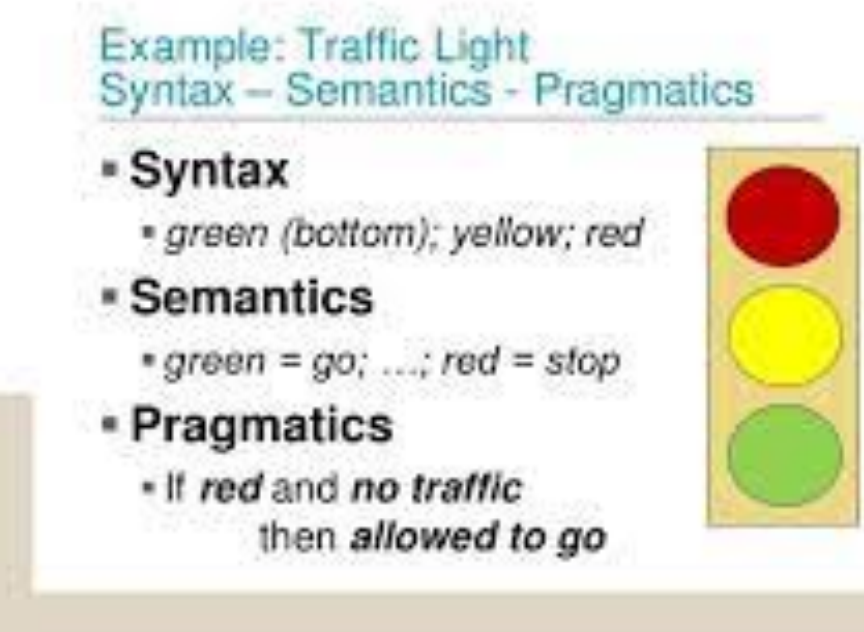
SEMANTICS	PRAGMATICS
Study of words and their meanings in a language	Study of words and their meaning in a language with concern to their context
Focuses mainly on the significance of the meaning of words in a literal sense	Additionally focuses on the meaning of words according to the context and their inferred meanings as well
Studies the literal meaning	Studies the intended or the inferred meaning as well

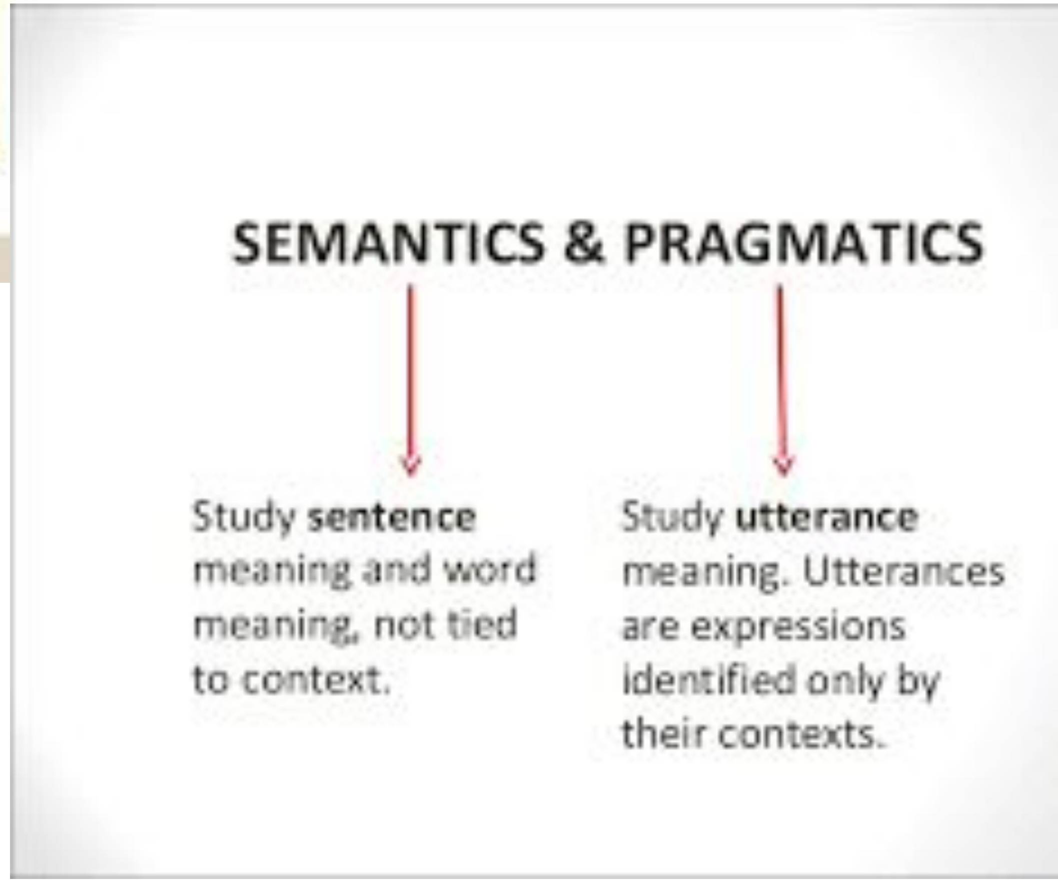
Visit www.PEDIAA.com

For example, this sentence – “**He is so cool.**” **Semantically**, this sentence can be interpreted as – He is very nice, a compliment to the person, which is the literal meaning.

But under **pragmatics**, this sentence suggests the context: the positive attitude of the speaker towards the person

Semantics and pragmatics (S & P)





	Focus	Scope / Range	Meaning of an Utterance	Regulated by	Domain
Semantics	Meaning	Narrows as it deals with only meaning	Context independent	General Rules	Grammar
Pragmatics	Language use	Broad as it deals with aspects beyond text	Context dependent	Principles	Rhetoric

Meaning representation

- Requirements that meaning representations should fulfill

- Types of meaning representation:
  - **First order predicate calculus (FOPC)**
  - Frame-based representation
  - Semantic network
  - Conceptual dependency diagram

Requirements

- Verifiability
- Unambiguous representations
- Canonical form
- Inference
- Expressiveness

Verifiability

- A system's ability to compare the state of affairs described by a representation to the state of affairs in some world as modeled in a knowledge base

- Example:
  - Sent: Maharani serves vegetarian dishes.
  - Question: Is the statement true?

Unambiguous representation

- Representations should have a single unambiguous interpretation.

- Example:
  - Mary and John bought a book
  - Two students met three teachers
  - A German teacher
  - A Chinese restaurant
  - A Canadian restaurant

Canonical form

- Sentences with the same thing should have the same meaning representation

- Example:
  - Alternations: active/passive, dative shift
  - Does Maharani have vegetarian dishes?
  - Do they serve vegetarian food at Maharani?

Inference

- a system's ability to draw valid conclusions based on the meaning representation of inputs and its store of background knowledge.

- Example:
  - Sent: Maharani serves vegetarian dishes
  - Question: can vegetarians eat at Maharani?

Expressiveness

- A system should be expressive enough to handle an extremely wide range of subject matter.

- Example:
  – Belief: I think that he is smart.
  – Hypothetical statement: If I were you, I would buy that book.

Meaning representation

- Requirements
  - Verifiability
  - Unambiguous representations
  - Canonical form
  - Inference
  - Expressiveness

- Types of meaning representation:
  - **First order predicate calculus (FOPC)**
  - Frame-based representation
  - Semantic network
  - Conceptual dependency diagram

FOPC

- Elements of FOPC
- Representing
  - Categories
  - Events
  - Time (including tense)
  - Aspect
  - Belief
  - …

Elements of FOPC

- Terms:
  - Constant: specific objects in the world: e.g., Maharani
  - Variable: a particular unknown object or an arbitrary object: e.g., a restaurant
  - Function: concepts: e.g., LocationOf(Maharani)
- Predicates: referring to relations that hold among objects:
  - Ex: Serve(Maharani, food)
  - Arguments of predicates must be terms.

Elements of FOPC (cont)

- Logical connectives: $\wedge, \vee, \Rightarrow$
- Quantifier: $\forall, \exists$
- Example: All restaurants serve food.

$$\forall x \, Restaurant(x) \Rightarrow Serve(x, food)$$

Syntax Driven Semantic Analysis

- Syntax Driven Semantic Analysis
  - How meaning representations are created
  - Syntax driven semantic analysis is a computational approach to semantic analysis that uses static knowledge from the lexicon and the grammar
  - This meaning is literal meaning and is context independent and inference free
  - Based on principle of compositionality – meaning of whole is made up of meaning of its parts
  - Syntax driven semantic analysis uses syntax analysis to guide the process of arranging semantic representations

semantic representations.

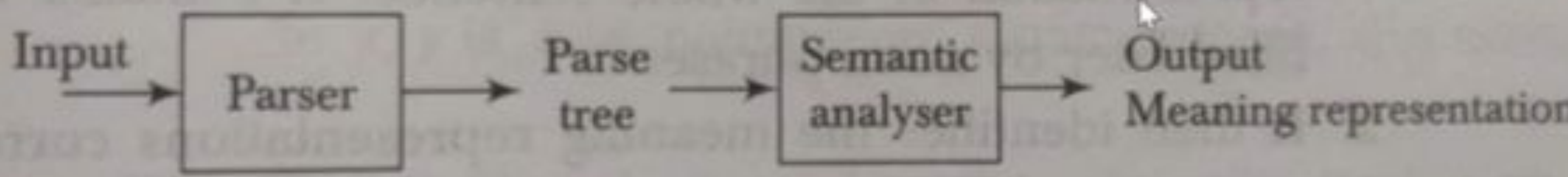


Figure 5.1 A simple approach to syntax-driven semantic analysis

Figure 5.1 shows the schematic diagram of this approach. Th

- Usually, semantic analyser produces multiple ambiguous meaning representations as the output
- This would correspond to ambiguities in syntax and lexicon
- Ambiguity resolution would be done at subsequent stages when domain-specific and contextual information is also considered
- A POS tagger and WSD mechanism can reduce the ambiguity to an extent

President nominates speaker. (5.5)

$\exists e$ is_a (e, nomination) $\wedge$ nominator (e, President) $\wedge$ nominee (e, speaker)

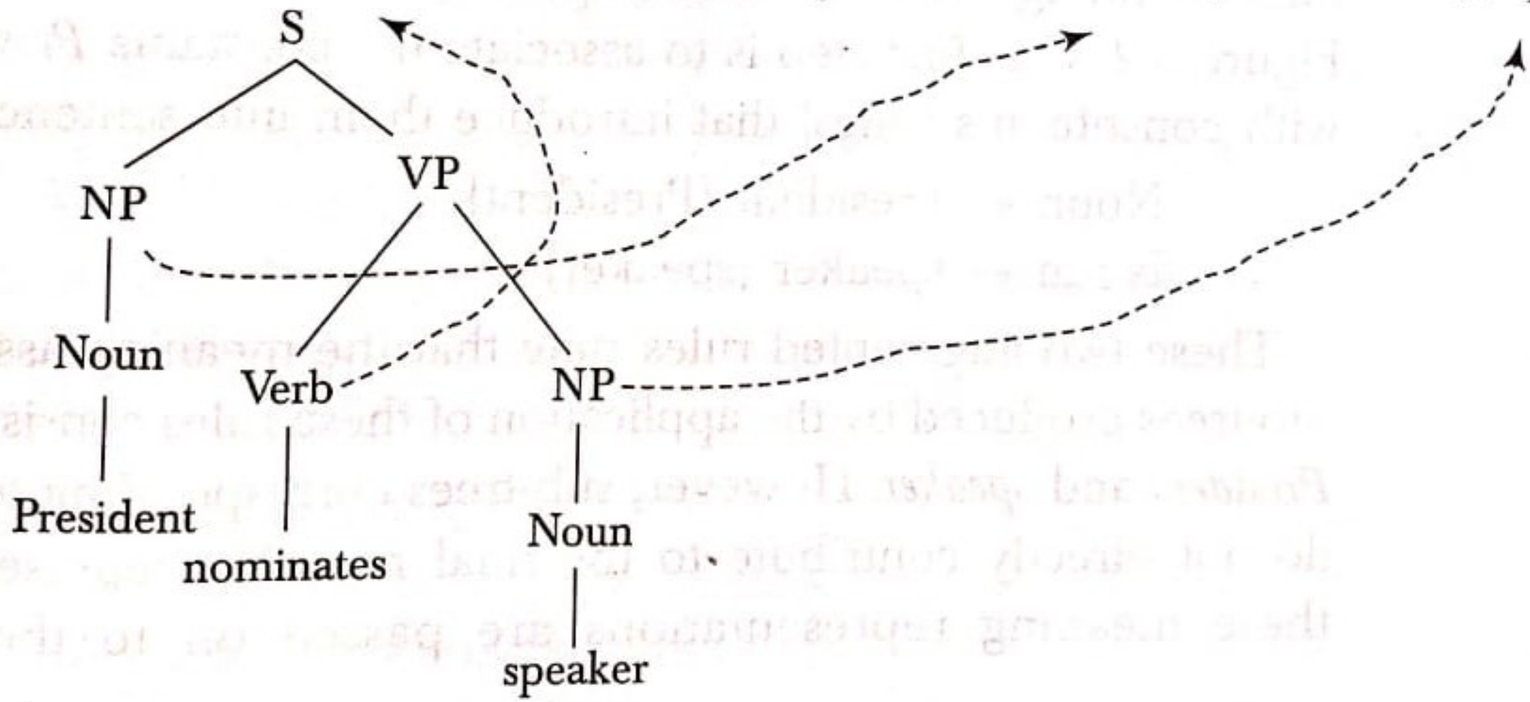


Figure 5.2 Mapping syntactic constituents to meaning representation

- Steps in semantic representation
- 1. Find meaning representation corresponding to verb nominates
  - It is the verb whose meaning defines the meaning of the whole sentence
  - The meaning representation of the verb acts as the template for meaning representation of the whole sentence
  - The Noun Phrases are arguments to the verb and are filled in in the template based on their roles
- 2. Find meaning representation for the two NPs
- 3. Bind the meaning representation of the NPs to the variables in the meaning representation of the verb to get the meaning representation of the whole sentence

How is the mapping from parse tree to meaning representation done?

- Augment the lexicon and grammar rules with semantic attachment – devise a mapping between rules of the grammar and rules of semantic representation (rule to rule hypothesis)
- An augmented rule can take the form

rm:

$$A \rightarrow \alpha_1 \alpha_2 \alpha_3 \ldots \alpha_n \{f(\alpha_i.\mathrm{sem}, \ldots, \alpha_k.\mathrm{sem})\}$$

The text appearing in {...} specifies that the meani

The text appearing within brackets specifies the meaning representation assigned to A as a function of the semantic attachments of A's constituents

- Eg: President nominates speaker
  - Noun -> President {President}
  - Noun -> speaker {speaker}
  - {President} and {speaker} are meanings associated with the augmented rules
  - Np-> Noun {Noun $_{sem}$}

resulting in the following augmented rule:

$$\text{verb} \rightarrow \text{nominates } \{\exists e, x, y \text{ is\_a (nomination)} \wedge \text{nominator}(e, x) \wedge \text{nominee}(e, y)\}$$

$$\text{VP} \rightarrow \text{verb NP } \{\text{verb}_{sem} (\text{NP}_{sem})\}$$

To combine NP_{sem} and verb_{sem} , y has to be replaced with speaker, not specified in verb_{sem}
Need to revise the semantic attachment for verb

- Lambda Calculus used to combine semantic representations systematically
- Lambda Calculus is an extension of FOPC
  - The following three rules define how to build all syntactically valid lambda terms

1. A variable, x, is a valid lambda term
2. If t is a lambda term and x is a variable, then λx.t is a lambda term (abstraction) – an anonymous function that takes a single input term x and substitutes in the the expression t

   Eg: λx. x+2 , : λx. x+y
3. If t and s are lambda terms, then ts is a lambda term (application) – application of a function t to an input s, represents the act of calling a function t on input s to produce t(s)

- Eg: λx.P(x) (Taj) -> P(Taj)
  - Replaces the variable x with Taj and removes λ
- With λ calculus, the VP semantics problem can be solved

verb as follows:

$$\text{Verb} \rightarrow \text{nominates } \{\lambda y\, \lambda x\, \exists\, e \text{ is\_a } (e, \text{nomination}) \wedge \text{nominator } (e, x) \wedge \text{nominee } (e, y)\}$$

The attachment is now a nested λ-expression. In a nested λ-expression

$$\exists\, e \text{ is\_a } (e, \text{nomination}) \wedge \text{nominator } (e, \text{President}) \wedge \text{nominee } (e, \text{speaker}))\}$$

Lexical Semantic

- How should we represent the meaning of a word?
  - Mouse (multiple meaning)
    - any of numerous small rodents...
    - a hand-operated device that controls a cursor...
  - form mouse is the **lemma**, also called the **citation form**
  - The form mouse would also be the lemma for the word mice
  - dictionaries don't have separate definitions for inflected forms like mice.
  - Similarly sing is the lemma for sing, sang, sung (**wordforms**).

Lexical Semantic

- Each of these aspects of the meaning of mouse is called **word sense**.
- The fact that lemmas can be polysemous (have multiple senses) can make interpretation difficult (is someone who types "mouse info" into a search engine looking for a pet or a tool?).

Lexical semantics

- Meanings of individual words
  - Sense and Reference
  - What do we understand by the word *lion* ?
  - Is a toy lion a lion? Is a toy gun a gun? Is a fake gun a gun?
- Grammatical meaning
  - What do we understand by *the lion*, *lions*, *the lions*, ... as in

    The lion is a dangerous animal

    The lion was about to attack

Word Senses,

Relations between Senses

Lemmas have senses

- One lemma “bank” can have many meanings:
  - …a **bank** can hold the investments in a custodial account…
  - “…as agriculture burgeons on the east **bank** the river will shrink even more”
- **Sense** (or **word sense**)
  - A discrete representation of an aspect of a word’s meaning.

Homonymy

Spelling/pronunciation

Homonyms: words that share a ***form*** but have unrelated, distinct meanings:

1. Homographs (bank/bank, bat/bat): same spelling , same/different pronunciation , different meaning
    1. $bank_1$: financial institution, $bank_2$: sloping land
    2. bat_1: club for hitting a ball, bat_2: nocturnal flying mammal
    3. $bass_1$ (بيس): stringed instrument, $bass_2$ (باس): fish
2. Homophones: different spelling, same pronunciation, different meaning
    1. Write and right
    2. Piece and peace

Homonymy causes problems for NLP applications

1. Information retrieval
   - If user search for “bat care”
   - Problem: you do not know whether user looking for bat(mammal) or bat(for basball)
2. Machine Translation (MT)
   - bat: translate to (animal) or translate to (baseball)
3. Text-to-Speech
   - bass (stringed instrument) vs. bass (fish)

Polysemy

- 1. I withdrew the money from the **bank**
- 2. The **bank** was constructed in 1875 out of local red brick.
- Are those the same sense?
  - Sense 2: “A financial institution”
  - Sense 1: “The building belonging to a financial institution”
- A **polysemous** word has **related** meanings
  - **bank** is a polysemous word
  - Most non-rare words have multiple meanings

Metonymy or Systematic Polysemy: A systematic relationship between senses

- Lots of types of polysemy are systematic
  - **School , university , hospital**
  - All can mean the institution or the building.
- A systematic relationship:
  - Building ⟺ Institution

Synonyms

- Word that have the same meaning in some or all contexts.
  - couch / sofa
  - big / large
  - automobile / car
  - vomit / throw up
- Two lexemes are perfect synonyms
  - if they can be substituted for each other in all situations

Synonymy is a relation between senses rather than words

- Consider the words *big* and *large*
- Are they synonyms? **YES**
  - How **big** is that plane?

    ↓

  - How **large** is that plane?

Antonyms

- Senses that are opposites with respect to one feature of meaning

  dark/light short/long fast/slow rise/fall

  hot/cold up/down

- More formally: antonyms can
  - define different scales
    - long/short, fast/slow
  - Define different directions:
    - rise/fall, up/down

Hyponymy and Hypernymy

One sense is a **hyponym** ("hypo is sub")of another if the first sense is more specific, denoting a subclass of the other.

Examples:

- car is a **hyponym** of *vehicle*
- *mango* is a **hyponym** of *fruit*

- Conversely **hypernym /superordinate** ("hyper is super")

Examples:

- *vehicle* is a **hypernym** of *car*
- *fruit* is a **hypernym** of *mango*

Superordinate/hypernym	vehicle	fruit	furniture
Subordinate/hyponym	car	mango	chair

Semantic roles

- Seeing words according to the meaningful role they play in language instead of seeing them as the container of meaning is called semantic roles.
- It is also called thematic roles, or thematic relation.

Agent

- Doer of an action
- Someone who initiates an act.
- We usually have agent as the grammatical subject of the sentence.

Theme/ Patient

- Receiver of the action
- Undergoes the action, change, event expressed by verb.

Instrument and experiencer

- Action performed through something is called instrument.
- Experiences a perception, feeling or state.

Examples of agent

- Boy plays cricket.
- John opened the door of the car, and got into it.
- Cat was drinking milk.
- You put the keys on the desk.

Examples of theme/ patient

- The boy kicked the ball.
- The wind blew the ball away.
- The dog caught the ball.
- A car ran over the ball.
- Cat drank the milk.

Examples of instrument and experiencer

- The cook cut the cake with *a* *knife.*
- She used *a* *crayon* to scribble a note.
- That window was broken by *a* *hammer.*

Table 5.1 Commonly used thematic roles

Thematic role	Definition
Agent	Deliberate performer of the event or action, e.g., *I* opened the lock with a key.
Theme	The arguments that are most directly affected by an event, e.g., I opened the *lock* with a key.
Instrument	The instrument used to carry out the action, e.g., I opened the lock with a *key*.
Experiencer	Arguments that undergo a sensory, cognitive, or emotional experience, e.g., *Shweta* hates cricket.
Force	Unconscious performer of the event, e.g., The *storm* blown over many trees.
Location	Places where the event occurs, e.g., Children are playing in the *garden*.
Source	Origin of the object of a transfer event, e.g., Flight will take off in a short while from *Chennai*.
Goal	The destination of a transfer event.

Thematic Role	**Example**
AGENT	*the waiter* spilled the soup.
EXPERIENCER	*john* has a headache.
FORCE	*the wind* blows debris from the mall into our yards
THEME	only after benjamin franklin broke *the ice.*
INSTRUMENT	he poached catfish, stunning them *with a shocking device*
BENEFICIARY	whenever ann callahan makes hotel reservation *for her boss*
SOURCE	I flew in *from boston.*
GOAL	I drove to *portland.*

SRL and Syntactic Cues

- Frequently semantic role is indicated by a particular syntactic position (e.g. object of a particular preposition).
  - Agent: subject
  - Patient: direct object
  - Instrument: object of "with" PP
  - Beneficiary: object of "for" PP
  - Source: object of "from" PP
  - Destination: object of "to" PP
- However, these are preferences at best:
  - The hammer hit the window.
  - The book was given to Mary by John.
  - John went to the movie with Mary.
  - John bought the car for $21K.
  - John went to work by bus.

Selectional Restrictions

- Associated with senses, not words themselves
- Vary in their specificity
  - To eat: THEME should be edible
  - To sip: THEME should be edible and liquid

Use of Sectional Restrictions

- Selectional restrictions can help rule in or out certain semantic role assignments.
  - "John bought the car for $21K"
    - Beneficiaries should be Animate
    - Instrument of a "buy" should be Money
  - "John went to the movie with Mary"
    - Instrument should be Inanimate
  - "John drove Mary to school in the van"
    "John drove the van to work with Mary."
    - Instrument of a "drive" should be a Vehicle

Representing Selectional Restrictions

$\exists e, x, y\ \text{Eating}(e) \wedge \text{Agent}(e, x) \wedge \text{Theme}(e, y)$

$\exists e, x, y\ \text{Eating}(e) \wedge \text{Agent}(e, x) \wedge \text{Theme}(e, y) \wedge \text{EdibleThing}(y)$

$\exists e, x, y\ \text{Eating}(e) \wedge \text{Eater}(e, x) \wedge \text{Theme}(e, y) \wedge \text{EdibleThing}(y) \wedge \text{Pizza}(y)$

What other ways can we represent selectional restrictions?

- WordNet synsets!
  - Selectional restriction for semantic role = one or more synsets
  - Input is considered reasonable if the word filling that semantic role is a member or hyponym of the specified synset

Sense 1

apple—(fruit with red, yellow, or green skin, sweet to tart, and crisp whitish flesh)

⇒ edible fruit—(edible reproductive body of a seed plant especially one having sweet flesh)

⇒ produce, green goods, green groceries, garden truck—(fresh fruits and vegetables grown for the market)

⇒ food—(any solid substance that is used as a source of nourishment; "food and drink")

⇒ solid—(a substance that is solid at room temperature and pressure)

⇒ substance, matter—(that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")

⇒ entity—(that which is perceived, known, or inferred to have its own distinct existence (living or nonliving))

⇒ fruit—(the ripened reproductive body of a seed plant)

Figure 5.4 Part of WordNet hypernym hierarchy of word *'apple'*

Selectional Preferences

- **Selectional restrictions** → hard constraints
- **Selectional preferences** → soft constraints

She was way faster than everyone else …the other runners **ate her dust**.

Spit that out, you **can't eat plastic!**

Selectional Preference

- Selectional preferences, *$S_R(v)$*, are defined as the difference between two distributions:
  - Distribution of the expected semantic classes, *$P(c)$*
  - Distribution of the expected semantic classes for a specific verb, *$P(c|v)$*
- This difference can be quantified using **Kullback-Leibler (KL) divergence**, *$D(P||Q)$*:
  - $D(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$
  - $S_R(v) = D(P(c|v)||P(c)) = \sum_c P(c|v) \log \frac{P(c|v)}{P(c)}$

Selectional Association

- **Selectional association** then indicates how much a given class contributes to a verb's overall selectional preference
  - $A_R(v,c) = \frac{1}{S_R(v)} P(c|v) \log \frac{P(c|v)}{P(c)}$
- When using very large corpora, you can also estimate selectional association based on conditional probabilities or log co-occurrence frequencies of predicates with arguments

Video on detailed programming explanation on this
https://www.youtube.com/watch?v=byy19WPLPBQ

Ambiguity

- **Defination:**

Ambiguities are words or phrases that have more than one meaning. This distracts the conscious mind because it then tries to figure out which meaning is appropriate.

- **Examples:**

"I bought herbs from the apothecary."

actually spoke to the apothecary (pharmacist)

or

apothecary(drug store).

Types of Ambiguity

- Lexical Ambiguity
- Syntactic Ambiguity
- Semantic Ambiguity
- Pragmatic Ambiguity

Lexical ambiguity

Defination:

The lexical ambiguity of a word or phrase consists in its having more than one meaning in the language to which the word belongs. “Meaning” hereby refers to whatever should be captured by a good dictionary.

Examples:

Bank refers to:

- Finincial bank
- edge of river.

Syntactic ambiguity

Defination:

Syntactic ambiguity is a property of sentences, which may be reasonably interpreted in more than one way, or reasonably interpreted to mean more than one thing. Ambiguity may or may not involve one word having two parts of speech or homonyms.

- Syntactic ambiguity arises from relationship between the words and clauses of a sentence.
- understand the sentense by its syntax.

Syntactic ambiguity

Examples:

- The dog bit the man.
- man is bit by dog.
- The Gov,t ask us to save soap and waste paper

- "they were milking cows"

don't know whether the person is referring to milking a cow or the type of cow.

Lexical Ambiguity

The presence of two or more possible meanings within a single word.

"I saw her **duck**."

Syntactic Ambiguity

The presence of two or more possible meanings within a single sentence or sequence of words.

"The chicken is **ready to eat**."

Semantic ambiguity

Defination:

"when a particular word has to diifferent interpretation"

Semantic ambiguity arises when a word or concept has an inherently diffuse meaning based on widespread or informal usage.

Examples:

- Iraqi head allow his arms.
- He allows to opperate.
- Gandhi stoned in rally in india.
- drugs user are not qualified to lead.

The Problem of Semantic Ambiguity

context=food

context=hardware

Did you say you were looking for **mixed nuts**?

People use **context** to derive the correct meaning.

Pragmatic ambiguity

- **Defination:**
- All languages depend on words and sentences in constructing meaning. However, one of the fundamental facts about words and sentences is that many of them in our languages have more one meaning. Therefore, ambiguity may occur when an utterance can be understood in two or more distinct senses.
- **Examples:**

Cake is on the table, I prepared snacks, give it to kids

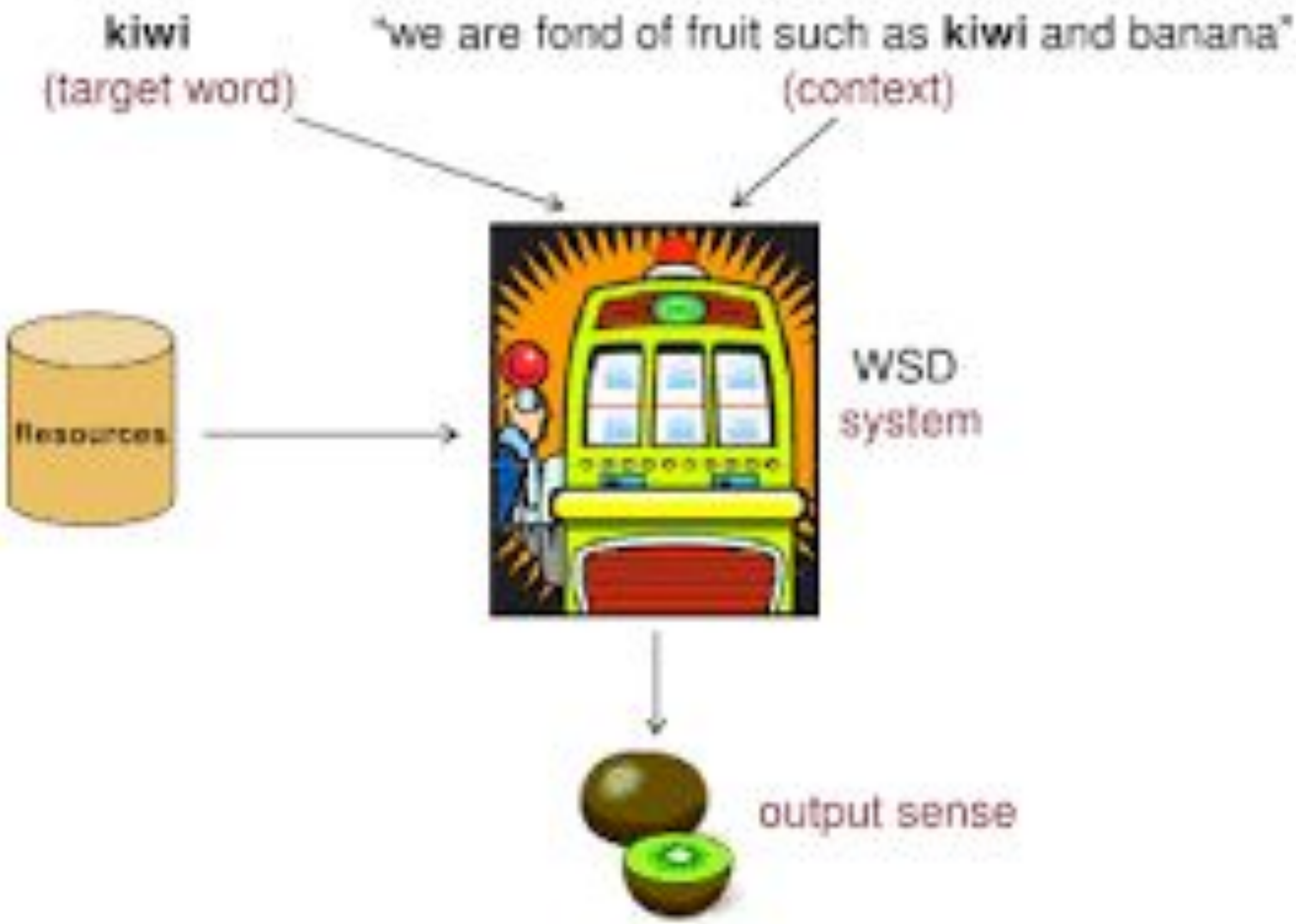
Do you know what time it is?

Asking for time
angry about time

Video on detailed programming explanation on Ambiguity and its types
https://www.youtube.com/watch?v=cYYUteTBtxo

Word-Sense Disambiguation

- Word sense disambiguation refers to the process of selecting the right sense for a word from among the senses that the word is known to have
- Semantic selection restrictions can be used to disambiguate
  - Ambiguous arguments to unambiguous predicates
  - Ambiguous predicates with unambiguous arguments
  - Ambiguity all around


kiwi
(target word)
"we are fond of fruit such as kiwi and banana"
(context)
Resources
WSD
system
output sense

Word Sense Disambiguation

- Many words have multiple meanings
  - E.g, river *bank*, financial *bank*
- **Problem:** Assign proper sense to each ambiguous word in text

- **Applications:**
  - Machine translation
  - Information retrieval
  - Semantic interpretation of text

Tagging?

- **Idea:** Treat sense disambiguation like POS tagging, just with "semantic tags"
- The problems differ:
  - POS tags depend on specific structural cues (mostly neighboring tags)
  - Senses depend on semantic context – less structured, longer distance dependency

Approaches

- Supervised learning:
  Learn from a pretagged corpus
- Dictionary-Based Learning
  Learn to distinguish senses from dictionary entries
- Unsupervised Learning
  Automatically cluster word occurrences into different senses

Supervised Learning

- Each ambiguous word token $w_i = w^k$ in the training is tagged with a sense from $s^k{}_1, \ldots, s^k{}_{n_k}$
- Each word token occurs in a context c_i
  (usually defined as a window around the word occurrence – up to ~100 words long)
- Each context contains a set of words used as features $v^i{}_j$

Bayesian Classification

- Bayes decision rule:

  Classify $s(w_i) = \arg\max_s \mathrm{P}(s \mid c_i)$
- Minimizes probability of error
- How to compute? Use Bayes' Theorem:

$$P(s_k \mid c) = \frac{P(c \mid s_k)P(s_k)}{P(c)}$$

Bayes' Classifier (cont.)

- Note that P(c) is constant for all senses, therefore:

$$
\begin{aligned}
s &= \arg\max_{s_k} P(s_k \mid c) \\
&= \arg\max_{s_k} \frac{P(c \mid s_k)}{P(c)} P(s_k) \\
&= \arg\max_{s_k} P(c \mid s_k) P(s_k) \\
&= \arg\max_{s_k} [\log P(c \mid s_k) + \log P(s_k)]
\end{aligned}
$$

Naïve Bayes

- Assume:
  - Features are conditionally independent, given the example class,
  - Feature order doesn't matter

(*bag of words* model – repetition counts)

$$P(c \mid s_k) = P(\{v_j : v_j \text{ in } c\} \mid s_k)$$
$$= \prod_{v_j \text{ in } c} P(v_j \mid s_k)$$
$$\log P(c \mid s_k) = \sum_{v_j \text{ in } c} \log P(v_j \mid s_k)$$

Do

Naïve Bayes Training

- For all senses s_k of w, do:
  - For all words v_j in the vocabulary, do:

$$P(v_j \mid s_k) = \frac{C(v_j, s_k)}{C(v_j)}$$

- For all senses s_k of w, do:

$$P(s_k) = \frac{C(s_k)}{C(w)}$$

Naïve Bayes Classification

- For all senses s_k of w_i, do:
  - score(s_k) = log P(s_k)
  - For all words *vj* in the context window c_i, do:
    - score(s_k) += log P(v_j | s_k)
- Choose $s(w_i) = \arg\max_{s_k} \text{score}(s_k)$

Significant Features

- Senses of *drug* (Gale et al. 1992):

  'medication' *prices*, *prescription*, *patent*, *increase*, *consumer*, *pharmaceutical*

  'illegal substance' *abuse*, *paraphernalia*, *illicit*, *alcohol*, *cocaine*, *traffickers*

Dictionary-Based Disambiguation

Idea: Choose between senses of a word given in a dictionary based on the words in the definitions

Cone:

1. A mass of ovule-bearing or pollen-bearing scales in trees of the pine family or in cycads that are arranged usually on a somewhat elongated axis
2. Something that resembles a cone in shape: as a crisp cone-shaped wafer for holding ice cream

Algorithm (Lesk 1986)

Define $D_i(w)$ as the bag of words in the ith definition for w

Define $E(w)$ as $\bigcup_i D_i(w)$

- For all senses s_k of w, do:
  - $\text{score}(s_k) = \text{similarity}(D_k(w), [\bigcup_{v_j \text{ in } c} E(v_j)])$
- Choose $s = \arg\max_{s_k} \text{score}(s_k)$

```
function Simplified_Lesk(word, sentence)
  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_examples of sense
      overlap←Compute_overlap(signature, context)
      if overlap > max-overlap then
                 max-overlap←overlap
                 best-sense←sense
   end
   return(best-sense){ returns best sense of word}
```

for $i = 1$ to k do

  score(s_i) = overlap $(d_{sk}, \cup_{w_j \in c}\, dw_j)$

end

choose s' s.t. $s' = \text{argmax}_{sk}$ score(s_k)

Figure 5.8 Lesk's algorithm

Similarity Metrics

$$\text{similarity}(X, Y) = \begin{cases} \text{matching coefficient} & |X \cap Y| \\ \text{Dice coefficient} & \dfrac{2|X \cap Y|}{|X| + |Y|} \\ \text{Jaccard coefficient} & \dfrac{|X \cap Y|}{|X \cup Y|} \\ \text{overlap coefficient} & \dfrac{|X \cap Y|}{\min(|X|, |Y|)} \end{cases}$$

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

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”

Simple Example

ash:

s_1: a tree of the olive family

s_2: the solid residue left when combustible material is burned

This cigar <u>burns</u> slowly and creates a stiff *ash*$_2$

The *ash*$_1$ is one of the last <u>trees</u> to come into <u>leaf</u>

After being struck by lightning the <u>olive</u> <u>tree</u> was reduced to *ash*$_2$

Some Improvements

- Lesk obtained results of 50-70% accuracy

Possible improvements:

- Run iteratively, each time only using definitions of "appropriate" senses for context words
- Expand each word to a set of synonyms, using also a thesaurus

Thesaurus-Based Disambiguation

- ***Idea:*** the semantic categories of the words in a context determine the semantic category of the context as a whole. This category, in turn, determines which word senses are used.
- ***(Walker, 87):*** each word is assigned one or more subject codes which corresponds to its different meanings. For each subject code, we count the number of words (from the context) having the same subject code. We select the subject code corresponding to the highest count.
- ***(Yarowski, 92):*** adapted the algorithm for words that do not occur in the thesaurus but that are very . Informative. E.g., Navratilova --> Sports

for $i = 1$ to k do

$$\text{score}(s_k) = \sum_{w_j \text{ in } c} d(\text{subj}(s_k), w_j)$$

end

choose s' s.t. $s' = \arg\max_{s_k} \text{score}(s_k)$

$d(\text{subj}(s_k), w_j) = 1$ iff $\text{subj}(s_k)$ is one of the subject code of w_j and 0 otherwise. The score is number of words in context whose subject code matches with the subject code of sense s_k.

Figure 5.10 Walker's algorithm for sense disambiguation

Thesaurus-Based Disambiguation

- **Thesaurus** assigns subject codes to different words, assigning multiple codes to ambiguous words
- $t(s_k)$ = subject code of sense s_k for word w in the thesaurus
- $\delta(t,v) = 1$ iff t is a subject code for word v

Simple Algorithm

- Count up number of context words with same subject code:

for each sense s_k of w_i, do:

$$\text{score}(s_k) = \sum_{v_j \text{ in } c_i} \delta(t(s_k), v_j)$$

$$s(w_i) = \arg\max_{s_k} \text{score}(s_k)$$

Thesaurus-Based Disambiguation

- (Walker, 87): each word is assigned one or more subject codes in a dictionary corresponding to its different meanings.
  - If more than one subject code is found, then assume that each code corresponds to a different word sense.
  - Let $t(s_k)$ be the subject code for sense s_k of word w in context c.
  - Then w can be disambiguated by counting the number of words from the context c for which the thesaurus lists $t(s_k)$ as a possible subject code. We select the sense that has the subject code with the highest count.

Thesaurus-Based Disambiguation

Walker's Algorithm

comment: Given context c
for all senses s_k of w **do**
  $\text{score}(s_k) = \Sigma_{v_j \, in \, c} \, \delta(\text{t}(s_k), v_j)$
end
choose $s' = \text{argmax}_{s_k} \text{ score } (s_k)$

- Note that $\delta(\text{t}(s_k), v_j) = 1$ iff $\text{t}(s_k)$ is one of the subject codes for v_j and 0 otherwise. The score is the number of words compatible with the subject code of s_k.
- One problem with this algorithm is that a general categorization of words into topics may be inappropriate in a particular domain (e.g., *mouse* as a mammal or electronic device in the context of computer manual).
- Another problem is coverage, e.g., names like *Navratilova* suggests the topic of sports and yet appear in no lexicon.

Yarowsky, 1995

- Yarowsky uses an approach that is similar to Brown et al.'s information theoretic method in that it selects the strongest collocational feature for a particular context and disambiguates using this feature alone.
- The features are ranked using the ratio: $P(s_{k_1}|f)/P(s_{k_2}|f)$, the ratio of the number occurrences with sense s_{k_1} with collocation f divided by the number occurrences with sense s_{k_2} with collocation f (with the possibility of smoothing in the case of sparse data).
- Selecting the strongest feature removes the need to combine different sources of evidence (given that independence rarely holds, it may be better to avoid the combination).
- Achieves accuracies between 90.6% and 96.5%, with a 27% improvement from adding the discourse constraint.

Unsupervised Disambiguation

- It may be useful to disambiguate among different word senses in cases where there are no available lexical resources.
  - in a specialized domain (e.g., linguistics)
  - could be quite important for information retrieval in a domain
- Of course, it is impossible to do sense tagging in a situation where there is no labeled data; however, it is possible to carry out sense discrimination in a completely unsupervised manner.

Unsupervised Disambiguation

- Without supporting tools such as dictionaries and thesauri and in the absence of labeled text, we can simply cluster the contexts of an ambiguous word into a number of groups and discriminate between these groups without labeling them.
- Context-group discrimination (Schutze, 1998):
  - Clusters uses of an ambiguous word with no additional knowledge.
  - For an ambiguous word w with senses $s_1, \ldots, s_k, \ldots, s_K$, estimate the conditional probability of each word v_j occurring in w's context being used with sense s_k, $P(v_j|s_k)$.