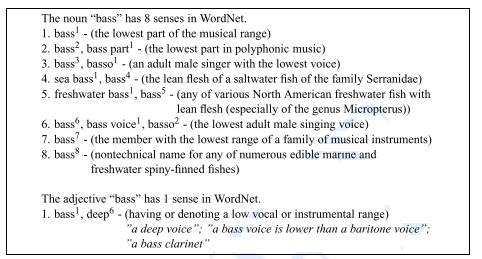
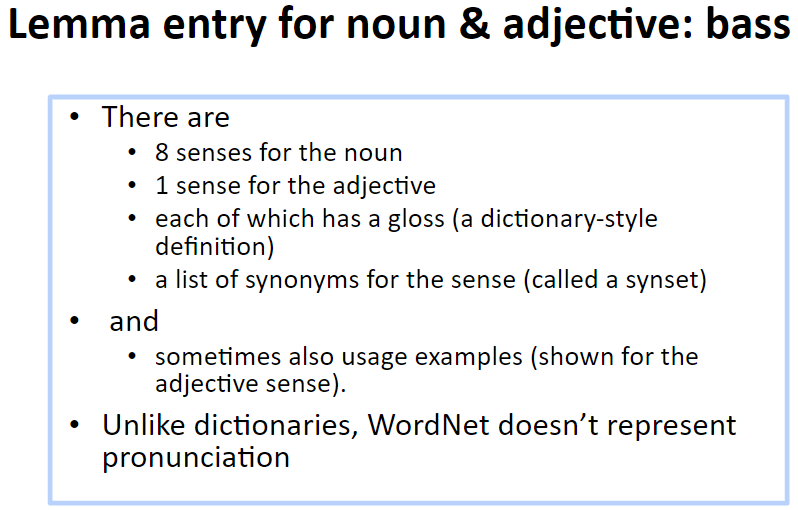
mod 4

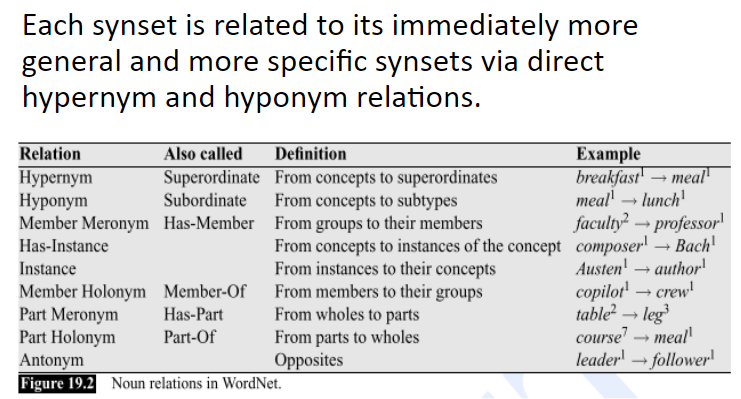
## What is Word net? Explain with an example

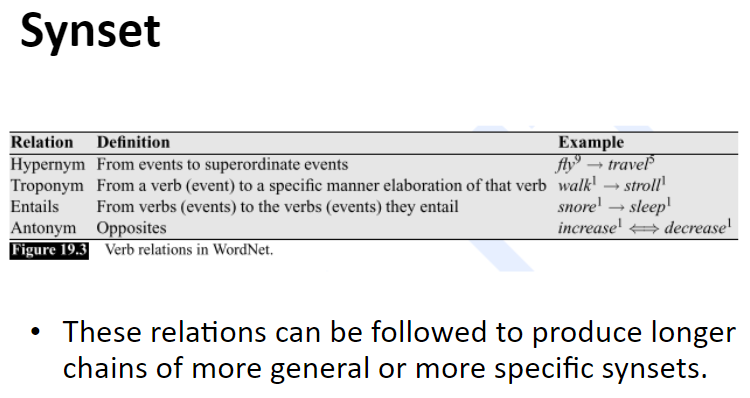
* It is a database of Lexical Relations.
* The most commonly used resource for English sense relations is the WordNet lexical database.
* WordNet consists of three separate databases:
  1. For nouns
  2. For verbs
  3. For adjectives and adverbs
* NB: closed class words are not included.
* Each database consists of a set of lemmas, each one annotated with a set of senses.
* The WordNet 3.0 release has
  1. 117,097 nouns
  2. 11,488 verbs
  3. 22,141 adjectives
  4. 4,601 adverbs.
* WordNet can be accessed via the web or downloaded and accessed locally.





1. What is Synset? Explain with an example.

* The set of near-synonyms for a WordNet sense is called a synset (for synonym set)
* Synsets are an important primitive in WordNet.
* Synsets actually constitute the senses associated with WordNet entries
* It is synsets, not wordforms, lemmas or individual senses, that participate in most of the lexical sense relations in WordNet



1. Explain Robust Word Sense Disambiguation with different approaches in detail.
2. Explain Lesk Algorithm.

* In this approach, all the senses definitions of the word to be disambiguated are retrieved from the dictionary
* Each of these senses is then compared to the dictionary definitions of all the remaining words in the context
* The sense with the highest overlap with these context words is chosen as the correct sense.
* E.g. Finding the appropriate sense of cone in the phrase pine cone given the following definitions:
* Pine:
  1. Kind of evergreen tree
  2. Waste away through sorrow or illness
* Cone:
  1. Solid body which narrows to a point
  2. Something of this shape whether solid or hollow
  3. Fruit of certain evergreen trees
* The lesk algorithm will select cone(c) as the correct sense since two of the words in its entry: evergreen and tree, overlaps with the words in the entry for pine.
* Neither of the other entries have any overlap with words in the definition of pine.
* **Disadvantage of Lesk Algorithm** : The dictionary entries for the target words are relatively short and may not provide sufficient material to create adequate classifiers.
* Remedy: Expand the list of words used in the classifier to include words related to but not contained in their individual sense definitions.
* This can be accomplished by including words whose definitions make use of the target word.
* E.g. the word deposit does not occur in the definition of bank in the American Heritage Dictionary. However, banks do occur in the definition of deposit. Therefore the classifier for bank must include deposit as a relevant feature.
* **Subject Codes :** The word deposit can be related to financial deposit or mud deposit. So there is still ambiguity with respect to its usage. This problem can be solved with the help of Subject Codes.
* Many dictionaries include tags known as subject codes in their entries that correspond roughly to broad conceptual categories.
* E.g.the entry for bank in Longman’s dictionary includes subject code EC(Economics) for the financial sense of the bank.
* Given such subject codes,the expanded terms in subject code can be related to this sense of bank