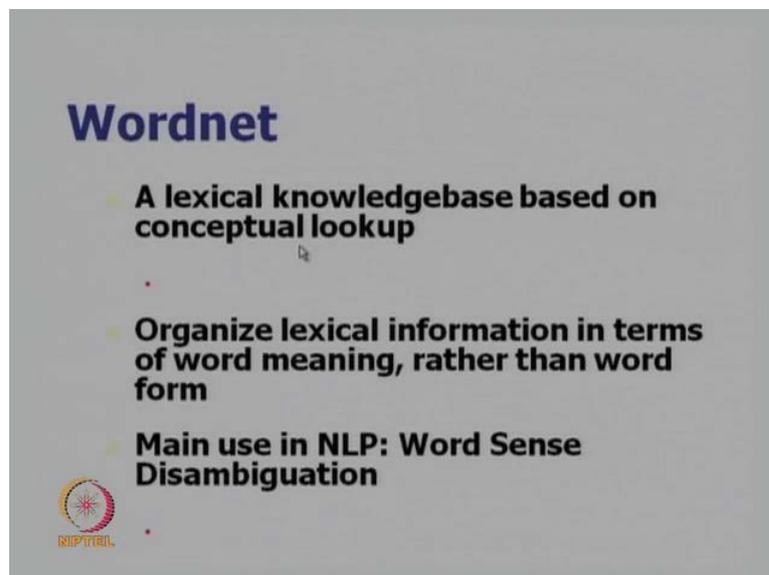


Natural Language Processing
Prof. Pushpak Bhattacharya
Department of Computer Science and Engineering
Indian Institute of Technology, Bombay

Lecture - 29
Wordnet and Word Sense Disambiguation

Today, we are going to discuss Wordnet and Word Sense Disambiguation. Word sense disambiguation is known to be an extremely important problem in natural language processing. This is almost the holy grail of natural language processing, because most words in natural language are ambiguous and it is amazing that human beings can comprehend sentences and text. In spite of the multiple meaning of word and they can identify the correct meanings of the words from the context and from the way the text is used. So, we will discuss wordnet, a very important lexical resource for the purpose of word sense disambiguation and the techniques for word sense disambiguation.

(Refer Slide Time: 01:17)

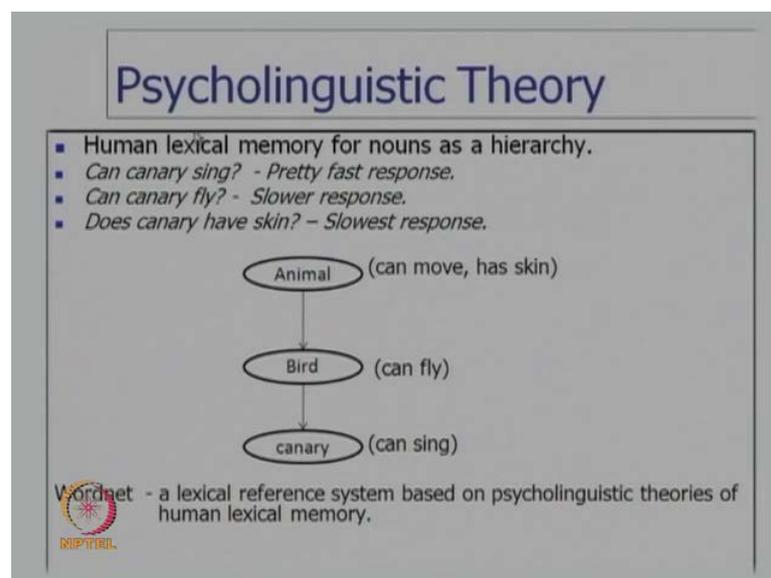


So, first of all what is a wordnet? Wordnet is a lexical knowledge base based on conceptual look up, it organizes lexical information in terms of word meaning rather than word form. And the main use in natural language processing for wordnet is the word sense disambiguation. So, several important points are made in this slide, wordnet first of all is a lexical knowledgebase; that means, it is a knowledgebase where information properties and linkages of word are stored. And it is based on conceptual look up which

means that the fundamental unit which is searched in the wordnet is a concept rather than a word we will elaborate this particular point.

As we go ahead, but add oppose to a dictionary what happens is that the fundamental unit of information that is obtained from the wordnet is a meaning unit, and we have to of course explicate what do you mean by the meaning unit and how it is represented. But, that will come little later and this is the distinguish point of the wordnet, the look of is the terms of meaning unit rather than word. And the lexical information is organized in terms of meaning rather than the word form, again to be elaborated as we go ahead. And of course, the main use in the natural language processing of these very important resourse has been word sense disambiguation.

(Refer Slide Time: 03:10)



Now, this whole structure is based on a interesting psycholinguistic theory, so as this particular slideshows a human lexical memory for nouns at least, it is stored as a hierarchy all right. So, the factors the word in the human memory is of course and appears fact, but how at the stored, at the stored as unconnected unit of information or is there is structure in the way they are stored.

So, here is a structure shown on the screen canary is the name of a bird c a n a r y that is the word form, and it is a type of bird and that is why this is the know which is about canary, and bird is a kind of animal which therefore, is a chain now canary bird animal. The properties are stored alongside the notes animal for example, can move animal has

skin birds, they can fly not all animal can fly. Therefore, only these note is associated with this information can fly, canary in the bird, which can sing not all bird can sing therefore, we store this information can sing with canary.

Now, I experiment to us perform well a human subject was asked these questions can canary sing, can canary fly, does canary have skin. So, you can see in this information based if we look for the pieces of information, then the first question can canary sing, can be immediately answered from the canary note itself, because it stored directly the fact that canary can sing. Can canary fly, this is not stored in these particular nodes, it is stored in the node above, which is the hyponymy of canary or the general is the more general concept in which the concept of canary is subsumed.

So, can fly is stored in the bird node and it can be answered as yes, because canary being a bird inherits all the properties of bird. Finally, does canary have skin? These can not be answers from canary nodes nor from the bird node, but only from the animal node, where has skin is stored. And bird and canary being the descendent the animal inheritor of all the properties, in that notes and therefore, we can answer to the affirmative that canary does have skin.

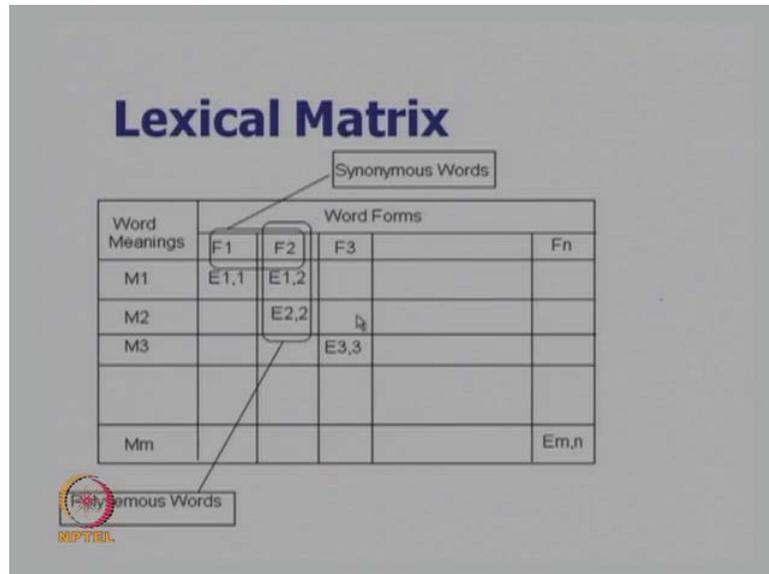
Now, these three questions you can see are answered from the canary node, than one level of ascendance, and than two levels of ascendance for the third question. So, when we ourself answered this question by looking at this diagram, we take some time for answering to either canary can sing, but this is a immediate look of from the canary nodes, than we need sometime ascending to bird, and more time for ascending to animal.

So, this particular fact also holds true, when a human beings answer these question by accessing their own memory in their brain. So, it is seen that can canary sing elicits a fast response a pretty fast response, because this is being access from the notes canary itself. Can canary fly slightly slower response, because ascending the hierarchy is involve in this, does canary have skin, has the slowest response, because one needs to go up the hierarchy up to animal.

So, there is lot of evidence to say that, the nouns in human memory are organized has a hierarchy, which is why we have fast to slower to slowest response in answering this three questions, so time for response cannot be explained unless we have these hypotheses that the nouns are organized as a hierarchy. So, from these one can now take

a look into the wordnet lexical knowledge base and it is a lexical reference system, which is based on psycholinguistic theories of human lexical memory.

(Refer Slide Time: 08:24)



Now, proceeding further, we now discuss a very fundamental concept, in a wordnet or in computational semantics or lexical semantics the concept of lexical matrix, so this is a matrix where the columns word form, which means word like dog, cat, bank, river, procrastination, difference all these words. Stay as word form, know references to their meaning, but badly the word form are placed here one after the other in the column, one could follow some order, it could be for example, the lexical order we have the word starting with A in the first column then in the second column A in a lexical matrix order and so on, and so forth.

What is more complex? Is the row, row store the meanings M 1 M 2 M 3 up to M m, so there F n word forms and M m meanings, now the rows indicate meaning. So, for example, suppose the first row gives the meaning of apple, so there each some identify for the apple meaning of the apple, fruit meaning. Suppose, that identify it 0 0 0 0 1 the first meaning from the meaning representatory, and this meaning is manifested as apple here let see, this meaning is manifested as another word on the same row, which is synonymous as to apple and has the meaning of fruit apple. So, all those words which turn for this meaning appears here, all replaced some indication 1 0 or tick, to say that for this meaning this is the word, this meaning is the word and so on.

(Refer Slide Time: 10:48)

Hand-drawn lexical matrix diagram. The horizontal axis is labeled 'Form' and contains the words: Apple, Bank, Cat, depend, Eye, Figure. The vertical axis is labeled 'Meaning' and contains binary strings: 00001, 00002, ..., 00099. A handwritten note on the left says 'fruit meaning of apple' with an arrow pointing to the first row. Another note says 'depend meaning of bank' with an arrow pointing to the 'depend' column. Ticks are present in the cells (00001, Apple), (00002, Bank), and (00099, depend). A logo with 'KIPTTEL' and '18/1/11' is in the bottom left. The text 'NLP- lect 29' is at the bottom center.

So, let me write it down and take an example, so I am draw a lexical matrix here, I just put a few word, I put apple here is the form axis certain say, and this is the meaning dimension, so we have here meanings, so let we have apple, bank, cat, depend, eye, figure and so on and so forth. All the English word appeared in the column, and here of the meaning say 0 0 0 0 1, 0 0 0 2, 0 0 99 and so on, suppose this is the fruit meanings of apple then here will put an put a tick here, suppose the this is the depend meanings of bank. So, we have put a tick here and tick here, this means that this is the depend meaning of bank and the meaning is expressed by bank here and depends here.

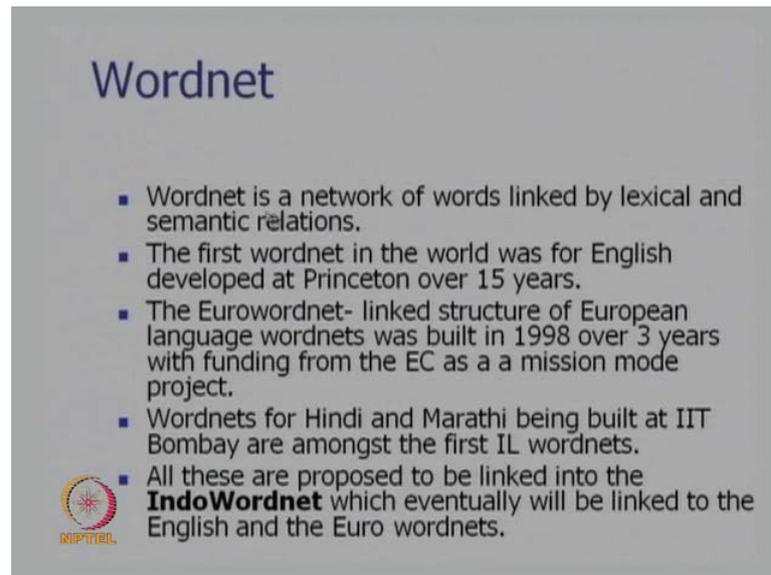
(Refer Slide Time: 12:27)

Lexical Matrix

Printed lexical matrix diagram. The horizontal axis is labeled 'Word Forms' and contains F1, F2, F3, ..., Fn. The vertical axis is labeled 'Word Meanings' and contains M1, M2, M3, ..., Mm. The matrix cells contain: (M1, F1) = E1,1; (M1, F2) = E1,2; (M2, F2) = E2,2; (M3, F3) = E3,3; (Mm, Fn) = Em,n. A box labeled 'Synonymous Words' points to the E1,1 and E1,2 cells. A box labeled 'Synonymous Words' points to the E2,2 cell. A logo with 'KIPTTEL' and '18/1/11' is in the bottom left.

So, if you look at the slide once again if I concentrate on a column then for a particular word these are the meaning, so there is a cells in that column below the word express what is called the polysemi of the word that is multiple meanings of the word and words there on the row for a particular meaning express synonymy. So, this is expressed here synonymous word along the row, polysemi along the column.

(Refer Slide Time: 13:00)



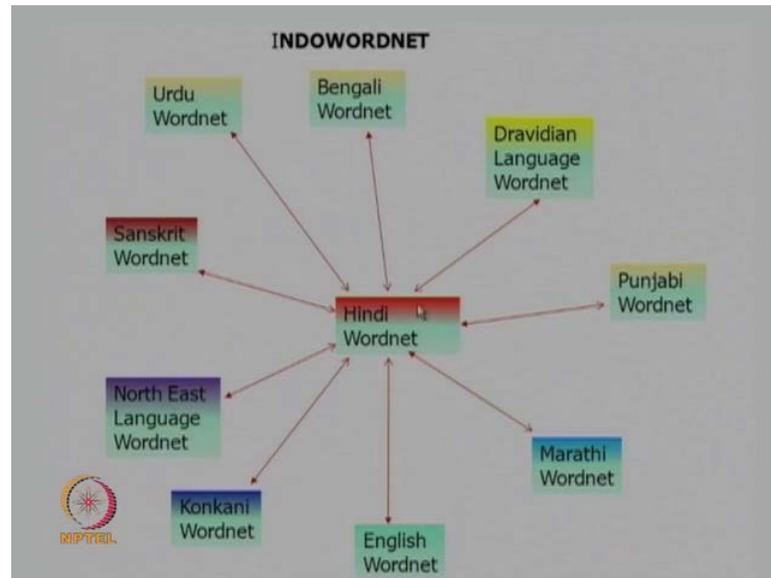
Wordnet

- Wordnet is a network of words linked by lexical and semantic relations.
- The first wordnet in the world was for English developed at Princeton over 15 years.
- The Eurowordnet- linked structure of European language wordnets was built in 1998 over 3 years with funding from the EC as a mission mode project.
- Wordnets for Hindi and Marathi being built at IIT Bombay are amongst the first IL wordnets.
- All these are proposed to be linked into the **IndoWordnet** which eventually will be linked to the English and the Euro wordnets.



So, what net is a network of words linked by lexical and semantic relations, here some fact about various wordnet in the world, the first wordnet in the world was for English which is developed at Princeton over 15 years. The euro wordnet which is a linked structure of European language wordnets was build in 1998 over 3-years with funding from the European commission. As a mission mode project wordnets for Hindi and Marathi are being built at IIT, Bombay, Hindi wordnets is very match here, and he is among the first Indian language wordnets which is being widely use along the world. And all these wordnets of Indian languages are proposed to be linked into indo wordnet, which is the eventually will linked to English and euro wordnet.

(Refer Slide Time: 13:55)



So, here is a picture, which hold for the wordnet build in activity across the country in India, so the first wordnet to build to us the Hindi wordnet, which is shown at the centre. Hindi wordnet has been linked for many of the inset with the English wordnet, Marathi wordnet is being created at IIT Bombay, Sanskrit wordnet is being created that IIT Bombay.

Urdu wordnet is being created at a many places named Jawaharlal Nehru University, University of Hyderabad in Lahore, Bengali wordnet is being created in West Bengal in Indian Statistical Institute, also Bengali wordnet is created at IIT kharagpur, than Bangladesh and also by a group in Japan. Dravidian language wordnet have been created in southern part of the country, Punjabi wordnet is being created in Patiala, North East Language wordnet are being created in the North East the part of the country, Konkani wordnet is being will at Goa. So, these are Indian language wordnet all of them are linked and aligned with be Hindi wordnet, and these utilize to a massive wordnet structure for Indian language is called that indo wordnet, and this will finally, be linked with European wordnet and fully with the princeton wordnet for English.

(Refer Slide Time: 15:39)

Fundamental Design Question

- **Syntagmatic vs. Paradigmatic relations?**
- Psycholinguistics is the basis of the design.
- When we hear a word, many words come to our mind *by association*.
- For English, about half of the associated words are *syntagmatically related* and half are *paradigmatically related*.
- For *cat*
 - *animal, mammal*- paradigmatic
 - *mew, purr, furry*- syntagmatic



Now, when building wordnet we ask a fundamental design question, this question pertains to Syntagmatic versus Paradigmatic relations, so we will explain the meaning of Syntagmatic and Paradigmatic relations, actually will write down the definition.

(Refer Slide Time: 16:00)

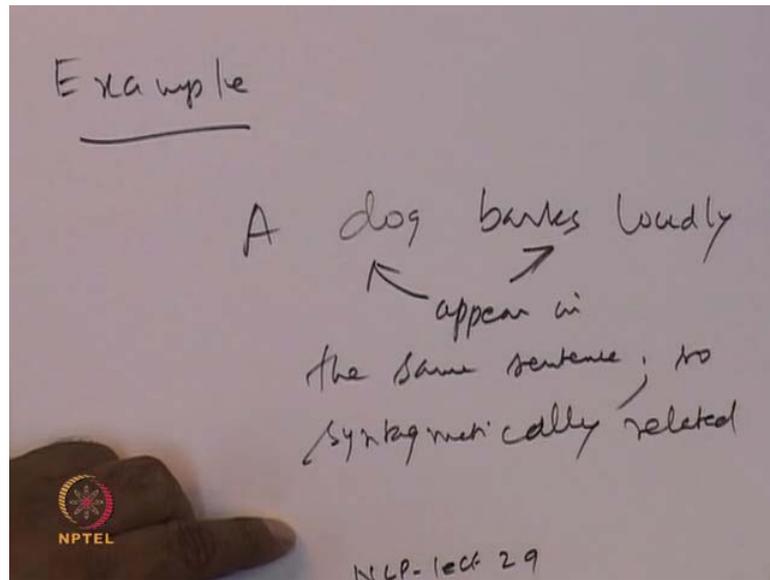
Words are syntagmatically related iff
they appear together in same
syntagma, (i.e.) syntactic unit,
typically a sentence



NLP-lect 29

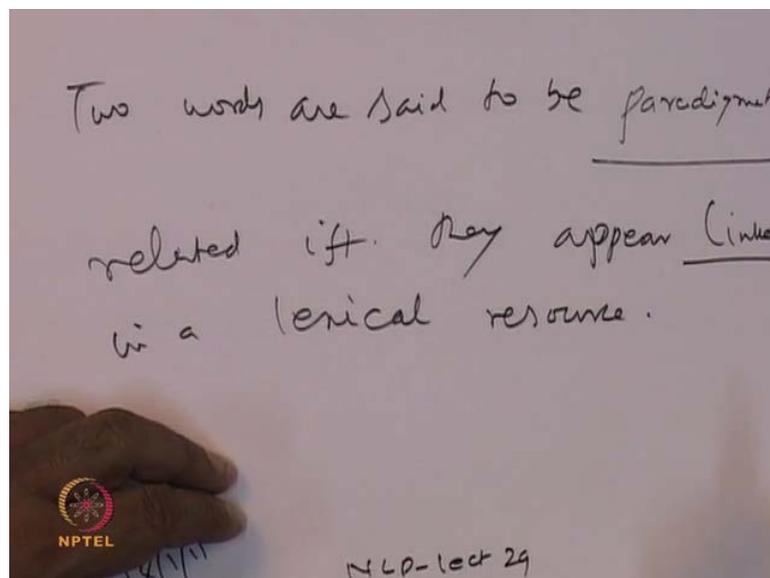
So, we say words are syntagmatically related, if they appear together in same syntagma, that is syntactic unit typically a sentence.

(Refer Slide Time: 16:57)



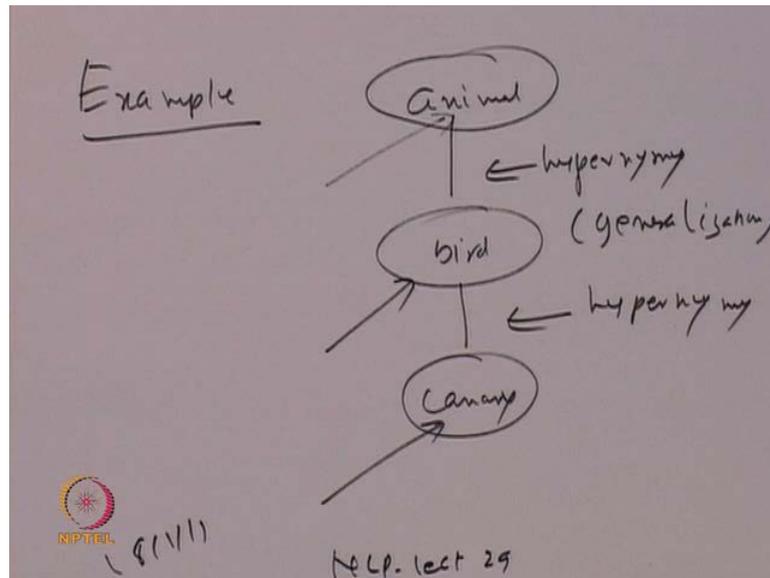
For example, a dog barks loudly, so dog and barks appear in the same sentence, so syntagmatically related.

(Refer Slide Time: 17:28)



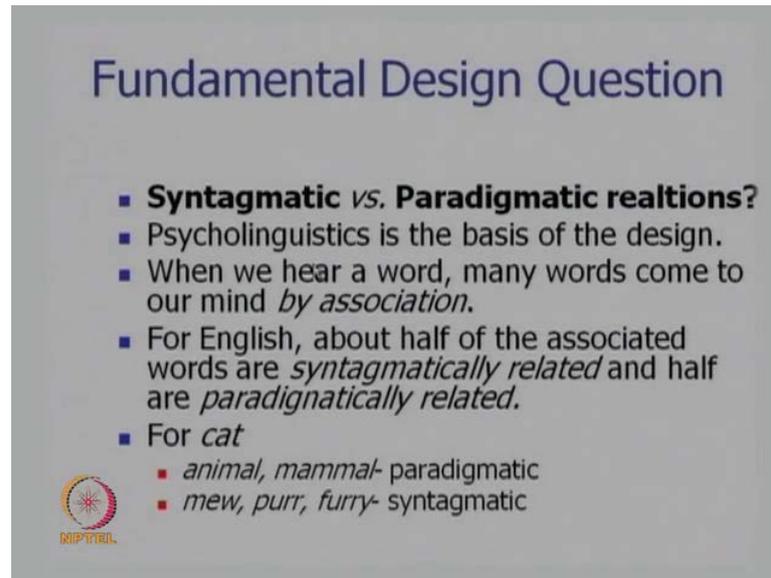
Similarly, two words are said to be paradigmatically related, if and only if they appear linked in a lexical resource.

(Refer Slide Time: 17:49)



Again I take an example here is the example, so we saw that bird canary and animal, so the word animal bird canary, they appear linked, these linked that is called hypernymy, which will explain very soon means generalization. So, animal is the more generalize concept than bird, bird is generalize concept then more generalize concept canary, so the words canary bird animal are linked by means of generalization specialization linked and therefore, there are paradigmatically related. So, syntagmatically related word appears together in sentences, phrases, large textual units maybe larger textual unit, and they are set to occur in a corpus together, on the other hand paradigmatically related words appears together in lexical resource. Let us say resource of synonyms for example, thesaurus or the wordnet which is a complex structure of meaning unit.

(Refer Slide Time: 19:01)



Fundamental Design Question

- **Syntagmatic vs. Paradigmatic relations?**
- Psycholinguistics is the basis of the design.
- When we hear a word, many words come to our mind *by association*.
- For English, about half of the associated words are *syntagmatically related* and half are *paradigmatically related*.
- For *cat*
 - *animal, mammal*- paradigmatic
 - *mew, purr, furry*- syntagmatic

 NIPTEIL

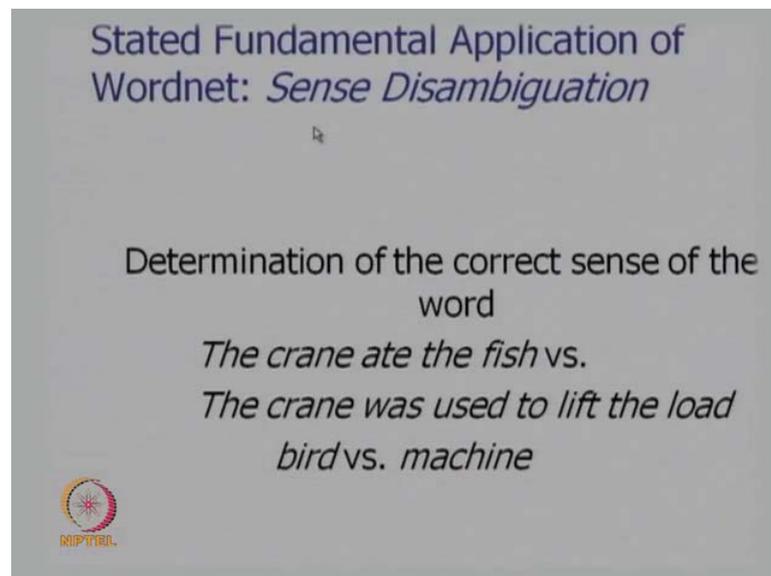
So, Syntagmatic versus Paradigmatic relations is a fundamental design question and psycholinguistics is the basis of the design of the wordnet, so the whole part is that when we hear a word many words come to mind by association. For example, when I hear the word cat the word animal, mammal, they come to the mind words says mew, purr, furry, these words also come to the mind.

The first two words animal and mammal, they can because the paradigmatic association cat animal and mammal they appear together in a lexical knowledge base whereas, the word mew, purr and furry they come, because the syntagmatic association. So, we have seen phrases like or sentences, like a cat mew, a cat purr or phrase like a furry cat, so these are examples of phrases and sentences form by word which appears syntagmatically together.

So, now, the question that arises is a should, one store syntagmatically related or paradigmatically related words in the wordnet, because word association is what is captured in the wordnet, meaning association, word association. Therefore, the question that naturally arises is syntagmatic association or paradigmatic association should be stored, now you might say there is circularity in the Hindi statement, because we are defining paradigmatically related words as the words, which appear together in a lexical resource.

And now, when asking do should lexical resources hold paradigmatic or syntagmatic words is not it true that lexical resources by definition is a repository of paradigmatically related words. So, this is correct; however, the point being made is that one as to be conscious a wordnet builder has to be conscious that paradigmatically related word should go into the lexical resources; however, we do find syntagmatically related words also, because the wordnet also links words which are across parts of speech. So, the word mew, would be linked with the word cat, barks would be linked with the word dog by a number of syntagmatic relations, which we have to see go ahead. So, the wordnet primarily hold paradigmatic relations, but it is also store some syntagmatically related words.

(Refer Slide Time: 21:50)



Now, the stated fundamental application of wordnet is sense disambiguation, which in shortest the determination of the correct sense of the word. So, here there are two sentences the crane ate the fish versus the crane was used to live the load, and in the first case the crane ate the fish, crane is used in the sense of the bird, and the crane was used to lift the load here the crane was used in the sense of a machine.

(Refer Slide Time: 22:23)

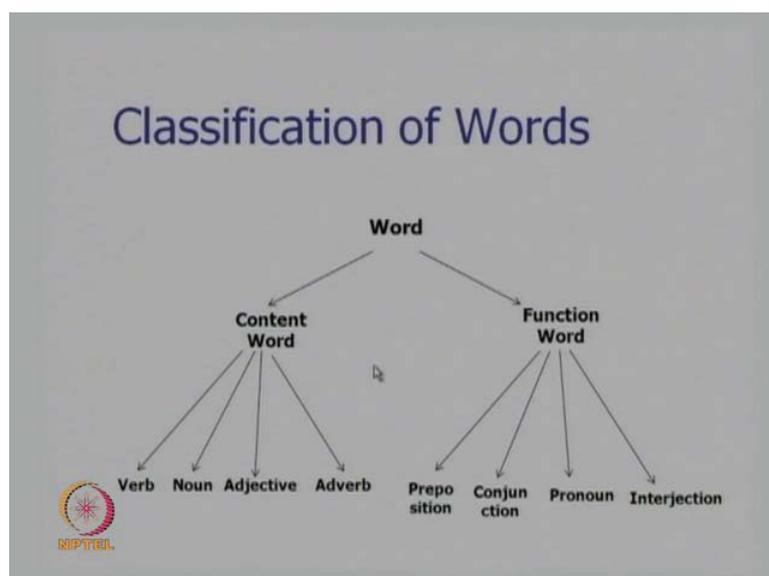
The problem of Sense tagging

- Given a corpora **To Assign correct sense to the words.**
- This is sense tagging. Needs **Word Sense Disambiguation (WSD)**
- **Highly important for Question Answering, Machine Translation, Text Mining tasks.**



Now, we come to an important problem in this area, the area of lexical semantics problem of sense tagging, given a corpora we have to assign correct sense to the words and these particular problem is called sense tagging. Is thus needs word sends disambiguation and it is highly important for question answering machine translation text mining tasks etcetera, etcetera. Now, from the tagged corpora up senses by applying machine landing algorithms, so we can build a word sense disambiguation engine.

(Refer Slide Time: 23:00)



So, now, when you look at a word you have to assign sense it to them in the corpus, so one question that arises is that what kind of word should be given this kind of sense identification. Now, it is known that words in the language are divided into two categories, namely content word and function word, so content words are like verb noun adjective and adverb. These are content words function words are preposition, conjunction pronounced, interjection, so a language definitely has a ((Refer Time: 23:42)) of content word and function word the main information load is carried by the content word. And out of the again nouns carried the maximum amount of information load followed by verbs adjective and adverb come a qualify for nouns and verbs respectively.

Function words on the other hand are close category word, such word define a language languages are very, very sensitive together, function word that the oldest units in a language and one does not see very easily the addition or deletion of function word. Whereas, content words get created very frequently especially, through the exigency of a new technological development, so for example, a new verb came after the Xerox machine was introduced, namely the verb Xerox.

He Xeroxed a paper, so here Xerox means copied made a copy of, but made a copy of through Xerox machine and therefore, Xerox as a word has come to stay. Similarly, new nouns get created for example, I introduced some new term when in the introduction lecture from this course, a word like communifaking where one is taking communication by holding their cell phone on to on ears, while not actually speaking into the phone is called communifaking.

So, this new word, which is a verb it came into existence, because of the advent of the mobile technology, similarly because of the existence of the Google, there is a new word discom googolation. Discom googolation is phenomenon of a person feeling very uncomfortable, when he or she is not able to acts of the internet quickly, this is called discom googolation, discomfort due to not being able to Google. Prepositions are what like with, in by, to, and so conjunction, and but while pronoun are I, me, he, they interjection is oh, ha, etc, with explain expressive motion.

This are lower level note show the content word and functions word, and we are saying that function words are very much the defining unit of the language, they do not change

frequently. Now, it is true that sense parking is the main or content word or particular sense for example, if I take the note noun here the world bank when it is a noun, can a true senses riverbank or the place of finance, where money is withdrawn from or deposited at. So, noun like that would require a marking for the meaning identity, when it appears in the corpus, similarly adjective the word red for example, red as an adjective, it is a qualifier for a noun a red car for example, adjective is not a point, but adjective also can be ambiguous ((Refer Time: 27:12)).

(Refer Slide Time: 27:15)

Example of sense marking: its need

एक_4187 नए शोध_1138 के अनुसार_3123 जिन लोगों_1189 का सामाजिक_4354 जीवन_125623 व्यवस्था_48029 होता है उनके दिमाग_16168 के एक_4187 हिस्से_120425 में अधिक_42403 जगह_113368 होती है।

(According to a new research, those people who have a busy social life, have larger space in a part of their brain).

नेचर प्यूरोसाइंस में छपे एक_4187 शोध_1138 के अनुसार_3123 कई_4118 लोगों_1189 के दिमाग_16168 के स्कैन से पता_11431 चला कि दिमाग_16168 का एक_4187 हिस्सा_120425 एमिगडाला सामाजिक_43540 व्यवस्थाओं_1438 के साथ_328602 सामंजस्य_166 के लिए थोड़ा_38861 बड़ा_25368 जाता है। यह शोध_1138 58 लोगों_1189 पर किया गया जिसमें उनकी उम्र_13159 और दिमाग_16168 की साइज के अंकड़े_128065 लिए गए। अमरीकी_413405 टीम_14077 ने पाया_227806 कि जिन लोगों_1189 की सोशल नेटवर्किंग अधिक_42403 है उनके दिमाग_16168 का एमिगडाला यावा हिस्सा_120425 बाकी_130137 लोगों_1189 की तुलना में_38220 अधिक_42403 बड़ा_426602 है। दिमाग_16168 का एमिगडाला यावा हिस्सा_120425 आयन_1992 और मानसिक_42151 स्थिति_1652 से जुड़ा हुआ माना_212436 जाता है।

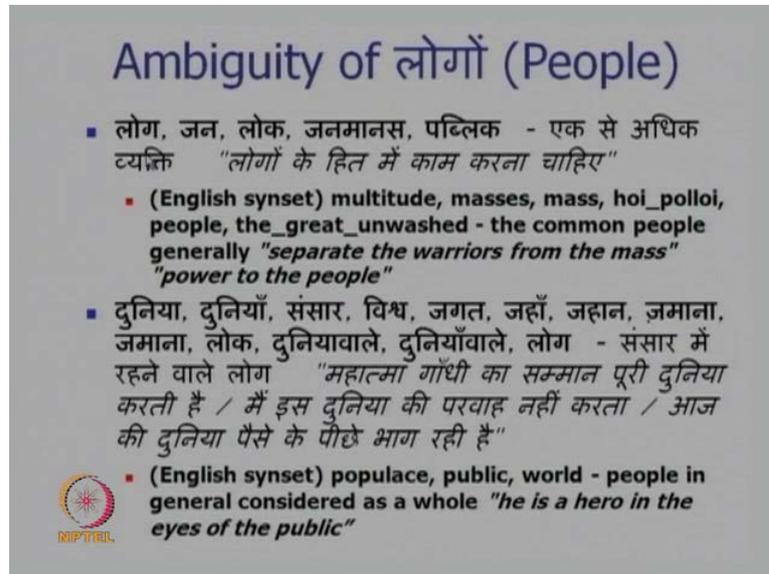


Examples of all those, I will shown here in this slide, this is in Hindi and Hindi transaction is given below, the sentence is according to a new research, those people who have a busy social life have larger space in a part of their brain. So, there is in this new search will say that those people who have a busy social life have lodger space in a part of their brain, and the corresponding Hindi is [FL], so these are the sentence [FL].

So, these sentences given and you see that there are some marks here in numerical entities place decide the word and what these are simply the id's the meaning id's, as there has a stored in the wordnet and they are attached to the word. So, the word [FL] is for example, is ambiguous [FL] can multiple meaning the number [FL] or one [FL] can be [FL] qualify as it is the case here [FL], this new research is being qualified by one. Similarly, [FL] can have multiple senses the sense which is are interest is 1138 number

sense [FL] can be ambiguous 3123 as a founder of regret here , [FL] as come in the plural form here and [FL] can be very ambiguous.

(Refer Slide Time: 29:11)



Ambiguity of लोगों (People)

- लोग, जन, लोक, जनमानस, पब्लिक - एक से अधिक व्यक्ति "लोगों के हित में काम करना चाहिए"
 - (English synset) multitude, masses, mass, hoi_polloi, people, the_great_unwashed - the common people generally "separate the warriors from the mass" "power to the people"
- दुनिया, दुनियाँ, संसार, विश्व, जगत, जहाँ, जहान, ज़माना, ज़माना, लोक, दुनियावाले, दुनियाँवाले, लोग - संसार में रहने वाले लोग "महात्मा गाँधी का सम्मान पूरी दुनिया करती है / मैं इस दुनिया की परवाह नहीं करता / आज की दुनिया पैसे के पीछे भाग रही है"
 - (English synset) populace, public, world - people in general considered as a whole "he is a hero in the eyes of the public"

 NIPTRIL

So, we see it in this slide one meaning of [FL] is [FL] the English synset is a multitude masses mass, [FL] people the great unwashed, here common people generally separate the warrior from the mass power to the people. These are example sentences, so the first meaning of flow is the common people masses mass issue, another sense of flow is the world populace public world people in general considered as a whole, he is a hero in the eyes of the public. So, [FL] he is a hero in the eyes of the public here in the sense is [FL].

(Refer Slide Time: 30:13)

Example of sense marking: its need

एक_4187 नए शोध_1138 के अनुसार_3123 जिन लोगों_1189 का सामाजिक_43540 जीवन_125623 व्यस्त_48029 होता है उनके दिमाग_16168 के एक_4187 हिस्से_120425 में अधिक_42403 जगह_113368 होती है।

(According to a new research, those people who have a busy social life, have larger space in a part of their brain).

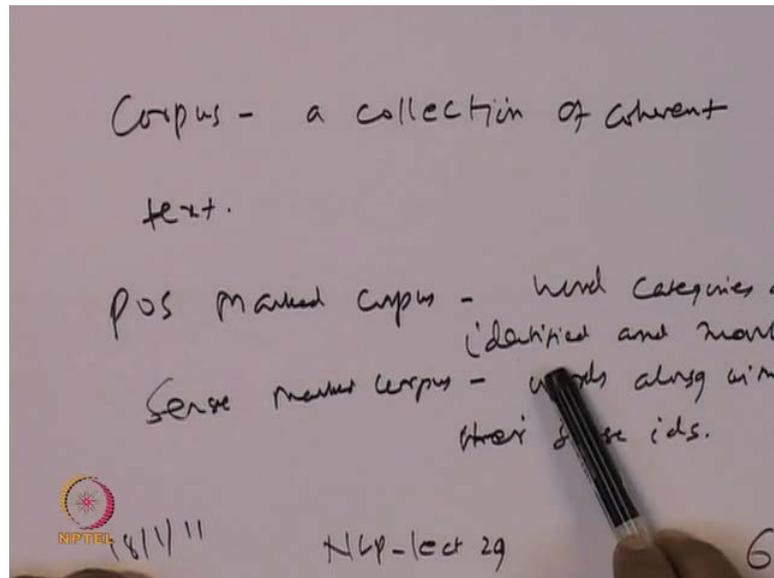
नेचर ब्यूरोसाइंस में छपे एक_4187 शोध_1138 के अनुसार_3123 कई_4118 लोगों_1189 के दिमाग_16168 के स्कैन से पता_11431 चला कि दिमाग_16168 का एक_4187 हिस्सा_120425 एमिगडाला सामाजिक_43540 व्यस्तताओं_1438 के साथ_328602 सम्बन्धित_166 के लिए थोड़ा_38861 बड़ा_25368 जाता है। यह शोध_1138 58 लोगों_1189 पर किया गया जिसमें उनकी उम्र_13159 और दिमाग_16168 की साइज के आंकड़े_128065 लिए गए। अनुरोधी_413405 टैग_14077 ने पाया_227806 कि जिन लोगों_1189 की सोशल नेटवर्किंग अधिक_42403 है उनके दिमाग_16168 का एमिगडाला वाला हिस्सा_120425 बाकी_130137 लोगों_1189 की तुलना में_38220 अधिक_42403 बड़ा_426602 है। दिमाग_16168 का एमिगडाला वाला हिस्सा_120425 भावनाओं_192 और मानसिक_42151 स्थिति_1652 से जुड़ा हुआ माना_212436 जाता है।

NIPTTEL

So, you see the word [FL] is a and ambiguous, which two meaning and the lexicographer the person who has mark the ((Refer Time: 30:21)) put an identification here 1189, which is the meaning. He things he is after pray for [FL] in this context, so from this it should be apparent to you that sense marking is a tough task, it is an involve complete task, because the person marking a senses will have to understand the meanings of the word. Absolutely, accurately will have to understand the sentence as a whole, will have to understand the domain of the goal, the exact domain in which the text is written.

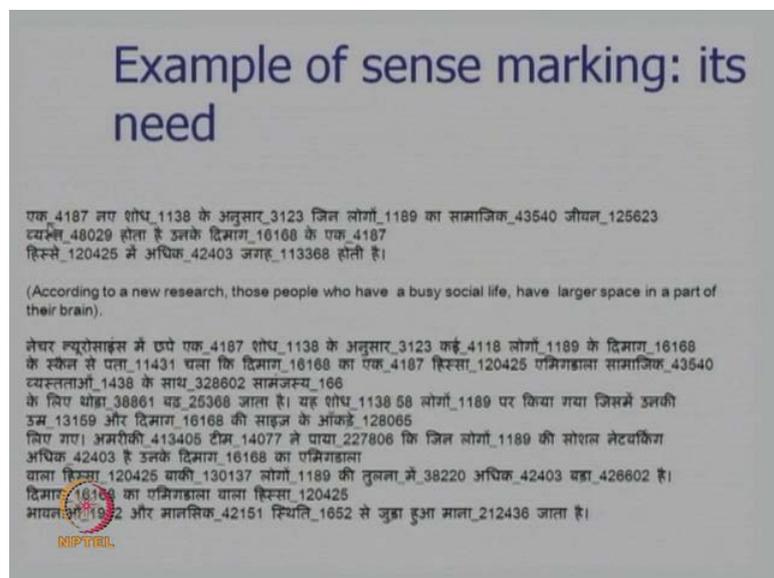
So, having the understand all this points the person is in a position to correctly identified the sense of the word and correctly put a tag, expressing the meaning of the world. So, these entity does to see in front of you namely sentences along with this kind of numbers that as two words and sentences formed in paragraph, paragraph formed many chapter and so on. These example, is known as the sense mark corpus to be used in natural languages, so first of all corpus is an extremely valuable entity in a actual language is processing, corpus means a coherent and piece of text. So, let me define purpose which is very important notion for us, it was mentioned in part of speech tagging.

(Refer Slide Time: 31:56)



So, corpus is a collection of coherent text, then you have pass marked corpus, where word categories as identified and what sense marked corpus, ((Refer Time: 32:27)) words along with their sense id's.

(Refer Slide Time: 32:49)



All right, this is words we shown in front of you, the in the sentences there words and that sense identity.

(Refer Slide Time: 32:58)

Basic Principle

- Words in natural languages are polysemous.
- However, when synonymous words are put together, a unique meaning often emerges.
- Use is made of *Relational Semantics*.
- *Componential Semantics* where each word is a **bundle of semantic features** (as in the Schankian *Conceptual Dependency* system or *Lexical Componential Semantics*) is to be examined as a viable alternative.



Now, we discussed the basic principle which is used in creating the wordnet, so first point to observe from which all these discussion is started, if there words in natural languages are polysemous; however, when synonymous words or put together a unique meaning often emerges. So, this approach for disambiguation is the approach of relational semantics oppose to relational semantics is called what is known as the componential semantics, where each word is a looked up on a bundle of semantic features. So, this is similar to the Schakian conceptual dependency framework or the lexical componential semantics framework, so the competitor to wordnet basic principle which is relational semantics is componential semantics.

(Refer Slide Time: 33:46)

Componential Semantics

- Consider *cat* and *tiger*.
Decide on *componential attributes*.

Furry	Carnivorous	Heavy	Domesticable
-------	-------------	-------	--------------

- For *cat* (Y, Y, N, Y)
- For *tiger* (Y, Y, Y, N)

**Complete and correct
Attributes are difficult
to design.**



Just understand componential semantics first we suppose consider the two words cat and tiger, we first decide on what is called the componential attributes of the word. So, the attributes, we have thought really when are furry whether the animal as per or not, carnivorous whether the animal eat meat or not, heavy whether the animal is heavy or not, domesticable whether the animal can be made effect in the house. So, for cat the answer to this four question or as follows cat is a furry animal it is a carnivorous animal it is not heavy, and it is domesticable, for tiger this is Y Y for Y Y Y N; that means, a tiger is a furry animal, the carnivorous animal, it is a heavy animal, but it is also carnivorous. So, the complete and correct attributes are difficult to design when we take the approach of componential semantics, we have to come up with the correct set of attributes.

Here, on the attribute is shown as furry, carnivorous, heavy and domesticable, the question that may arise is how do you know that this is the attributes setup attributes, and this is for the purposes the question of a correct and complete set of attributes matches. These attributes also are expressed by words and these words themselves can be in a ambiguous. So, componential semantics though attractive in concept is not, so easy to implement, because of the falling two reasons, one is that the set of attributes need not to be complete and correct it may be difficult to come up with the set-top attribute, which distinguish the words properly.

And the next difficulty is that the attribute them self are expressed by a language word, which is a own ambiguity. So, understanding the ((Refer Time: 35:57)) and attribute, which will be than exactly the same way by different person is not in esay matter, so this was componential semantics.

(Refer Slide Time: 36:11)

Semantic relations in wordnet

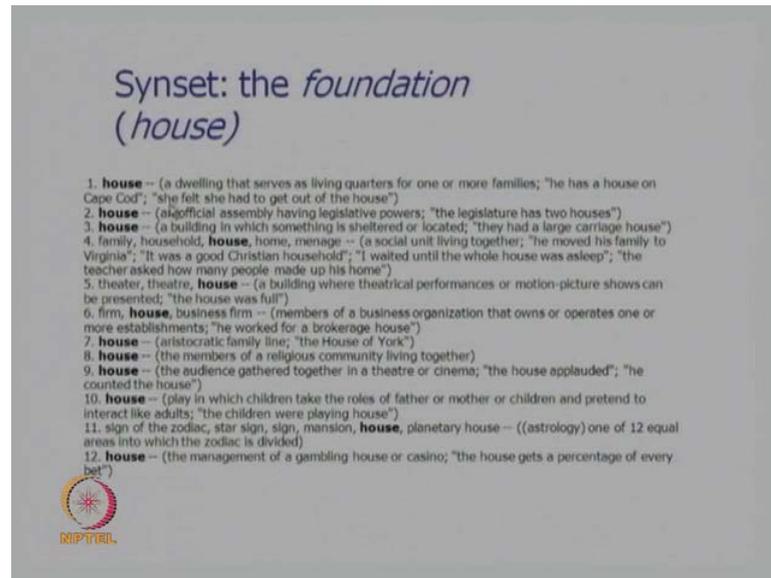
1. Synonymy
2. Hypernymy / Hyponymy
3. Antonymy
4. Meronymy / Holonymy
5. Gradation
6. Entailment
7. Troponymy

1, 3 and 5 are lexical (*word to word*), rest are semantic (*synset to synset*).



Now, we mention and not elaborate relationships in a wordnet, so synonymy is the synonymy relationship between word hypernymy, and hyponymy are generalization and specialization linkages, they hold between meaning in the wordnet. Antonymy is the opposition, meronymy is part of relationship, graduation is the graded and antonymy relationship, entailment is the implication relationship, troponymy is the manner of the relationship, so 1 3 and 5 that is Synonymy, Antonymy and Gradation word to word.

(Refer Slide Time: 37:05)



These are lexical relationship are word to word relationship and the rest are semantic there synset to synset relationship, so now, we discuss this very fundamental building blocks of wordnet, namely the synset, this is the foundation for the creation of the wordnet. Here, we have as example the word house, which is a very ambiguous word, the first meaning of house is a dwelling that serves as leaving quarters for one or more families, he has a house on cape cod she felt, she had to get out of the house, these are the example sentence.

So, here is the meaning of house which is a structure, the second meaning of house this is an official assembly having legislative powers, the legislature has two houses, so in England for example, we one has House of Flat and House of Common. In India there are again two houses Lok Sabha and Rajya Sabha, so this is the second meaning, the third meaning of house is a building in which something is sheltered or located, they had a large carriage house. So, this again the structural meaning of the house and it is not easy to see what the difference between 1 and 3 are, this meaning of house is very interesting in the sense of family. So, these words family, household, house, home, menage these are synonymy word, expressing the concept of family, the glass here that is the explanation of the concept is this.

A social unit living together, he moved his family to Virginia or he moved his house to one Virginia it was a good Christian household, so you could also say it was a good

Christian house, I waited until the whole house was asleep the teacher asked how many people made up his home. So, one-minute agree with all the example sentences, but what is clear from the these assembly of word is that one means house to be a family, theatre house, this is another meaning of house a building where theoretical performances motion picture shows can be represented.

So, theatre house this becomes concept, house can also mean a business form where the synonymy form, business form, members of a business organization that owns or operates one or more establishments, he worked for a brokerage house. House can also mean there aristocratic family line the house of York, so I gained 4 and 6, there is a slight difference and 1 ((Refer Time: 39:43)) 4 7 and one ((Refer Time: 39:44)) argue, that these sense of house in the sense of aristocratic family line could come below the fourth sense, which is a sense of family.

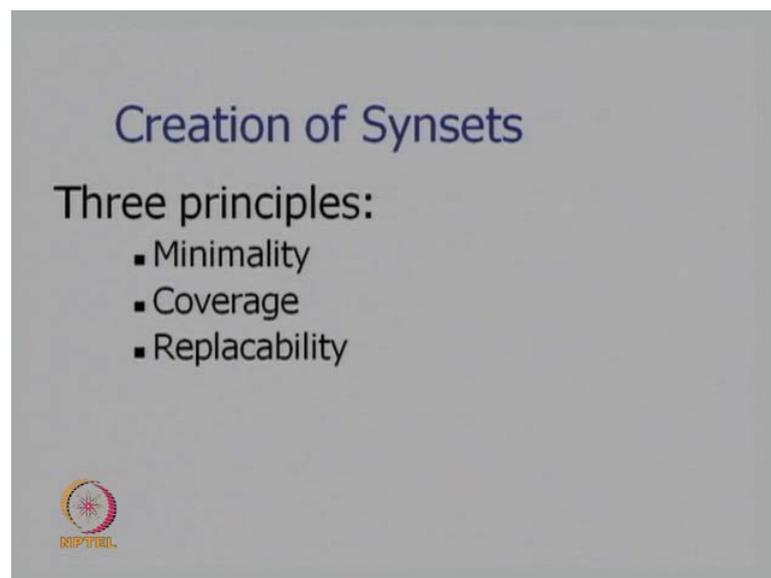
So, aristocratic family line is a kind of family therefore, it 7 could be a hypernym of fourth, anyway this cratic always appear for the synthesis of any wordnet they proceed with other house. House can be the members of a religions community living together, house can be the this meaning audience gathered together in a house or cinema, this theater sense of house, and look like it is a reputation comparator to the sense number 2.

House also can mean a play in which children take the roles of father or mother, for children and pretend to interact like adults, the children were playing house, when this case house, becomes like a game play by a small children. House can mean the sign of the zodiac, star sign or sign for example, this is in astrology one of the 12 equal areas into which the zodiac is divided, so for example, that the house of mercury, the place which means the place for mercury house of ((Refer Time: 40:00)) house of Jupiter and so on, and so forth.

This is a completely different meaning of house, which is the zodiac meaning, astrological meaning, house can also means the management of a gambling house or casino the house gets a percentage of every bet. So, when once see a 12 sense of houses and one-minute agree with all these senses and their location and their arrangement; however, it shows how much of care in is to be taken to express the find meanings advances of a word.

It is also possible that some sense actually does not belong there, this is a more restricted concept just like we saw the aristocratic family is a restricted concept compared to family, so it possibly should not appear at the same level as a sense of family. So, anyway, so these are the difference synset of house, and is the sensors are extremely finding then it required synset of meaning a lot of other machinery to get it the correct new hands of the word.

(Refer Slide Time: 42:21)



The finite the synset are separated, the more difficult it is for algorithm to ((Refer Time: 42:19)), now we discuss the creation of synset in the wordnet, and the creation of synset follows three principles, principle of minimality, coverage and replacability.

(Refer Slide Time: 42:35)

Sunset creation (continued)

Home

John's home was decorated with lights on the occasion of Christmas.
Having worked for many years abroad, John Returned home.

House

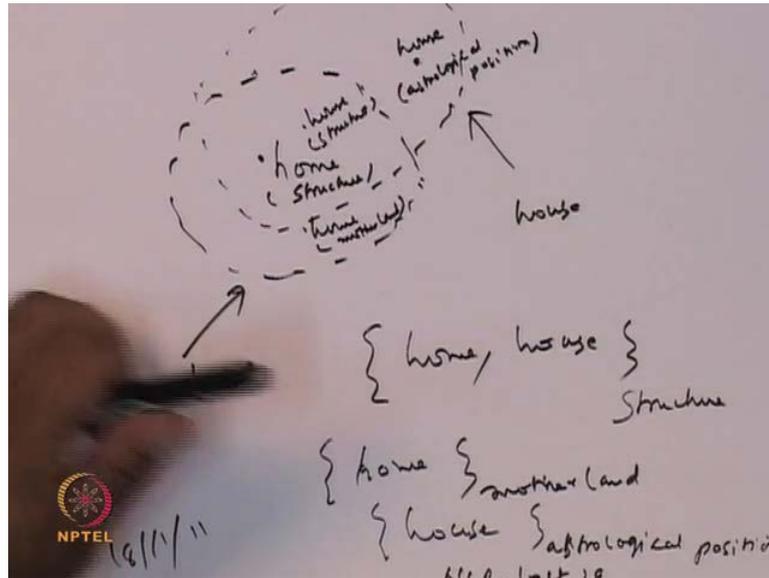
John's house was decorated with lights on the occasion of Christmas.
Mercury is situated in the eighth house of John's horoscope.



Whereas, understand what the principal mean, suppose we have the word home, home it is shown with two differentiate by means of two example. John's home was decorated with lights on the occasion of Christmas, so this is one sense of home as you can see we should compared, this to the next sentence having worked for a many years abroad, John returned home. So, this home has a meaning of motherland once own country, and these home has a meaning of building.

The house, John's house, John's building here means a structure, so this house was decorated with lights on the occasion of Christmas and these home as a meaning of motherland. Now, we take the word house John's house was decorated with lights on the occasion of Christmas, you can see here the word home can replaced with house for conveying the same meaning and here house as the meaning of a structure. Mercury was situated in the eighth house of John's horoscope and this is a completely divergences a sense of house, this does not agree with the motherland sense of home, so we can express this by means of a diagram.

(Refer Slide Time: 44:13)



So, we have here home, and home as a region of meaning around it, so one meaning of home is the structure, another meaning of home is motherland, and then house has another region of ((Refer Time: 44:34)) this is for home and this is for house. So, this part is common, house as a structure and this part has no commonality with home, where house is as a astrological position, so from these diagram is not it obvious, that we can put home and house together to means the structure sense.

And we can put home as a separate set to mean motherland, and we can put house as a separate set to mean astrological position, so this is a fairly important concept let me take to the diagram once again as I drawn it. This is the meaning region around the world house, house as a meaning of astrological position which is shown here, house has a meaning of a structure, which is shown here, these interjects with the structure meaning of home shown here. But, then there is another meaning of home is motherland, which does not intersect in any meaning of house. So, home and house as put together as synonymous word, a set of synonymy meaning structure home is as a single set with a motherland meaning, and house a is a single set with astrological position meaning. So, these form of ((Refer Time: 46:21)) wordnet will liberate in the next class.