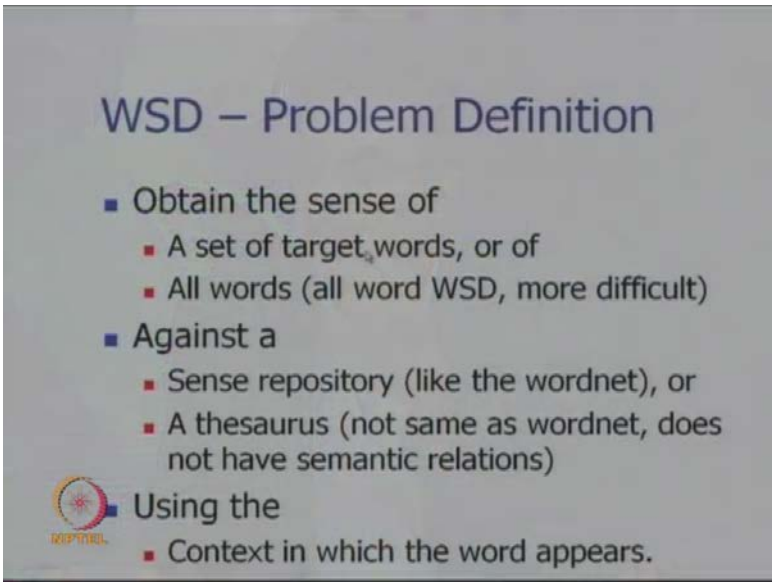
Natural Language Processing
Prof. Pushpak Bhattacharyya
Department of Computer Science and Engineering
Indian Institute of Technology, Bombay

Lecture - 32
Word Sense Disambiguation

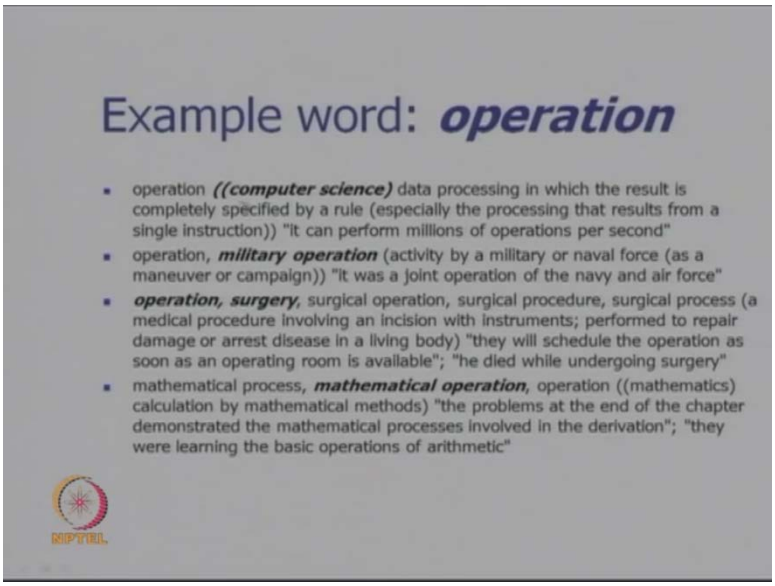
Having discussed word net and with semantic relations synsets and so on, which are absolutely crucial for word sense disambiguation. Let us now move onto approaches to word sense disambiguation its algorithms and different techniques for doing WSD. So, this is arguably the most important topic of natural language processing, which needs word relationships in a context to obtain the definite meanings of the words.

(Refer Slide Time: 00:58)



So, the problem definition is as follows; obtain the sense of a set of target words or of all words all word WSD which is more difficult. Against a sense repository like the word net or a thesaurus which is not same as word net, it does not have semantic relations using the context in which the word appears. So, our goal is to obtain the sense of either a set of target words or all words against a sense repository using the context in which the word appears.

(Refer Slide Time: 01:34)



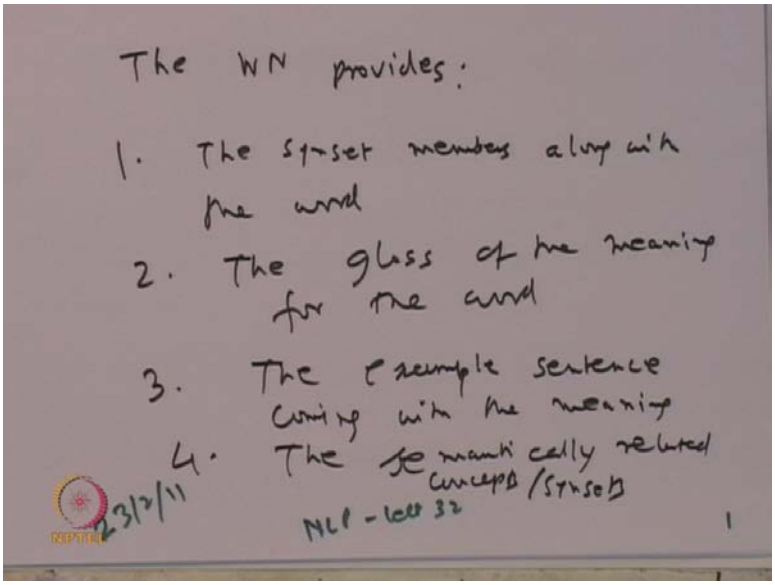
To take an example, we take the word operation which is highly ambiguous 4 senses of this word are shown here. The first sense is in computer science domain and it refers to data processing in which the result is completely specified by a rule. So, the computer can perform millions of operations per second. Then the next sense of operation is in the sense of military operation, this is activity by a military or naval force as a maneuver or campaign. This was a joint operation of the navy and the air force; operation can also be in a medical domain in the sense of surgical operation.

So, the meaning is a medical procedure involving an incision with instruments performed to repair damage or arrest disease in a living body. They will schedule the operation as soon as an operating room is available he died while undergoing surgery. Then comes, the notion of mathematical operation which is the sense of the word operation the meaning is calculation by mathematical methods. The problems at the end of the chapter demonstrated the mathematical processes involved in the derivation they were learning the basic operations of arithmetic.

So, what this examples show is that the word can have multiple meanings. And the senses are exemplified by what is called the gloss which comes as the meaning of the word as shown here. For example operation in the sense of computer science especially the processing that results from a single instruction in case of military operation as a maneuver or campaign by the military.

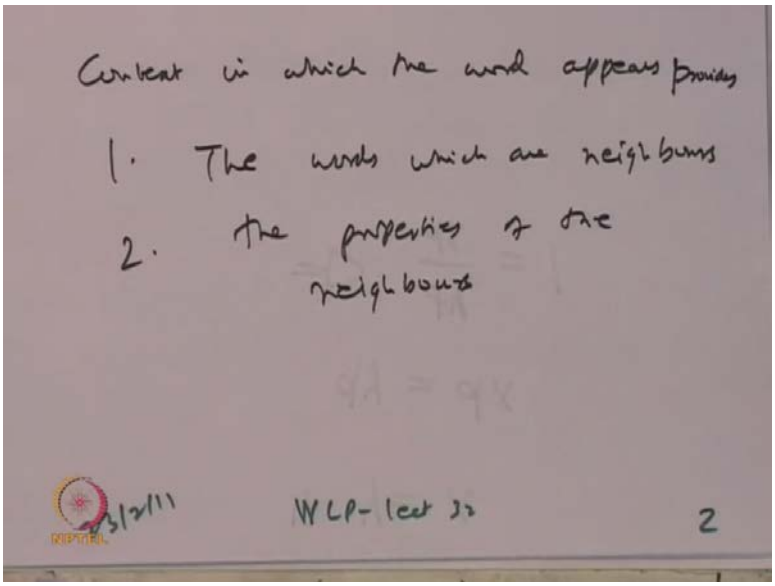
In the sense of medical operation a medical procedure involving an incision with instruments performed to repair damage or arrest disease in a living body in mathematics calculation by mathematical methods. So, these are known as glosses while the example sentence for example, there is an example sentence here that explicates the meaning. Now this particular example is quite illustrative of many important points in the basic idea of how to do word sense disambiguation? The technique of word sense disambiguation, so I will write them down for the purpose of enumeration.

(Refer Slide Time: 04:15)



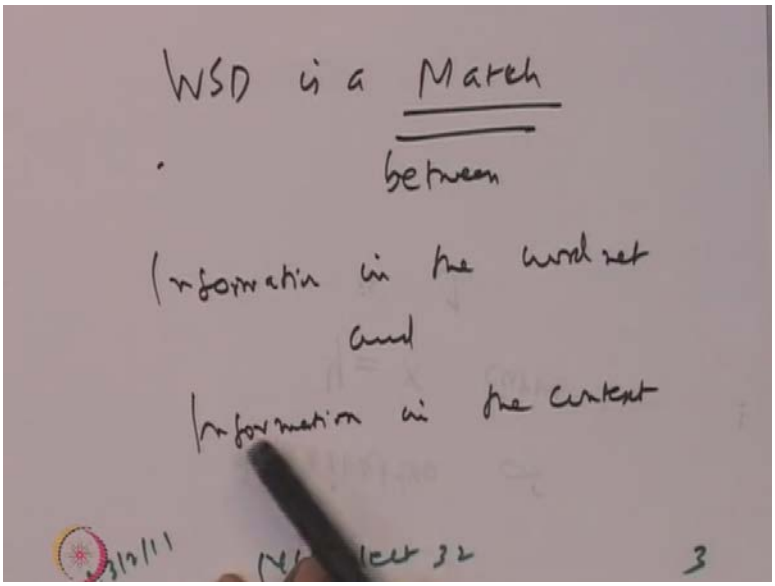
The word net provides one the synset members along with the word, the gloss of the meaning for the word, the example sentence coming with the meaning and 4; the semantically related concepts. So, these are very important and let us go on them once again word net provides the synset members along with the word. The gloss of the meaning for the word the example sentence coming with the meaning, the semantically related concepts or synsets. So, all these are required in the algorithm as information sources for disambiguation.

(Refer Slide Time: 05:50)



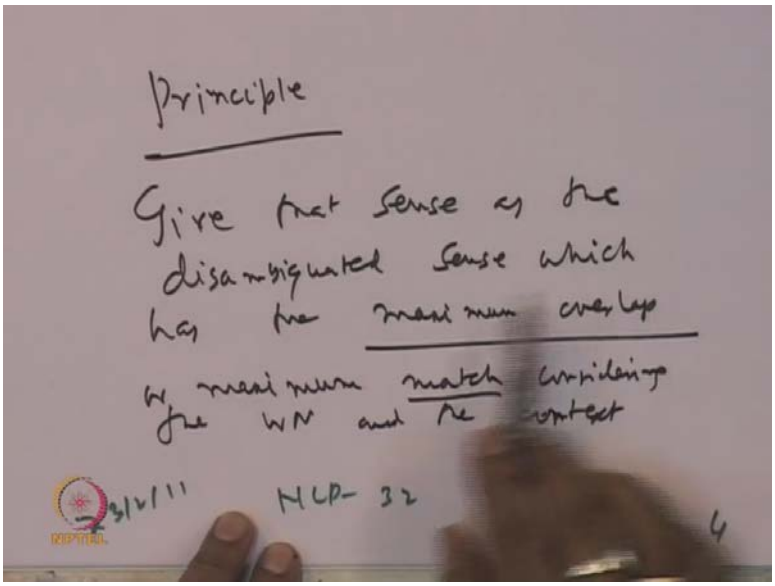
And proceeding forward or we it find is that the context in which the word appears provides 1; the words which are neighbors, 2; the properties of the neighbors.

(Refer Slide Time: 06:33)



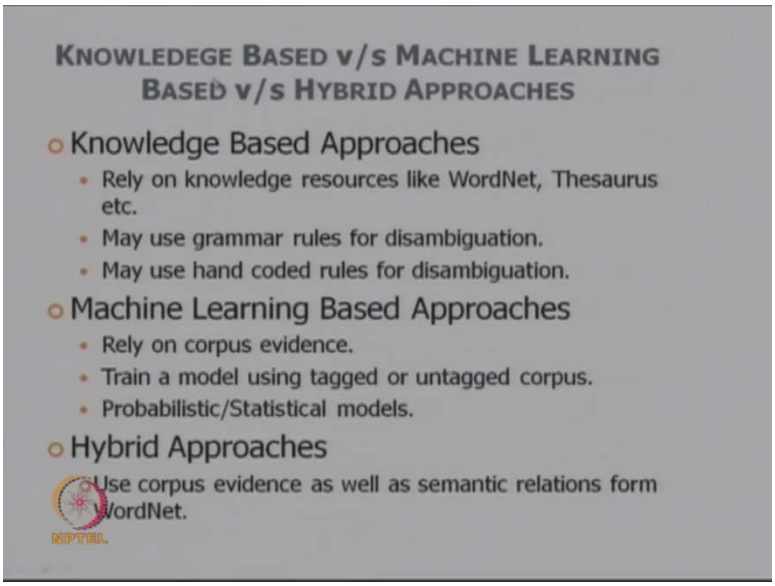
So, WSD is a match between information in the word net and information in the context. So, this is a very important statement we are making WSD is a match between information in the word net and information in the context.

(Refer Slide Time: 07:04)



So, this leads us to the following principle give that sense as the disambiguated sense which has the maximum overlap or maximum match considering the word net and the context. So, this is the main principle give that sense as the disambiguated sense which has the maximum overlap or the maximum match considering the word net and the context. So, going to the slide if we have for example, this as the sentence it can perform millions of operations per second. And here we mean the computer science sense of the word operation. Then this, particular word appearing in this context of perform million per second and so on should lead to this sense of the word operation through a matching process whereby all this information lives out that example sentence. Now which has appeared in the text all this information will be matched and this particular sense will be the winner sense.

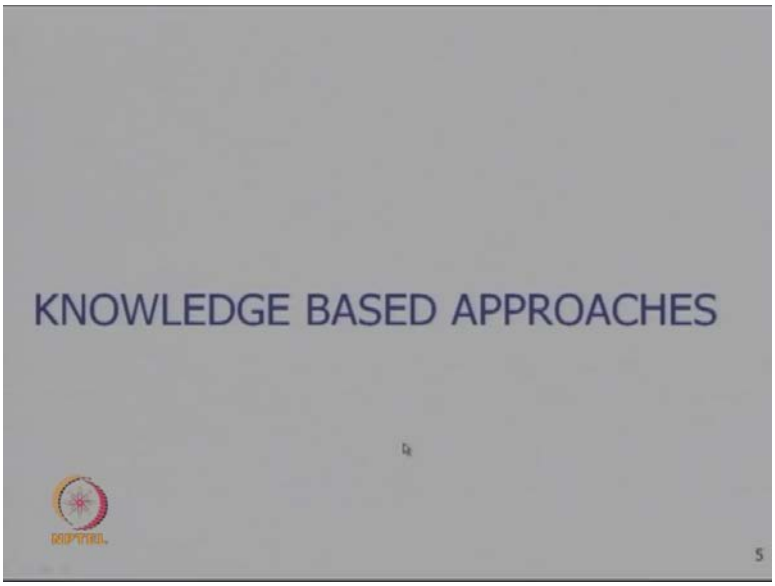
(Refer Slide Time: 08:55)



So based on this particular principle, we have knowledge based versus machine learning based hybrid approaches for word sense disambiguation. Knowledge based approach is rely on knowledge resources like the word net disorders etcetera. May, use grammar rules for disambiguation. May use hand coded rules for disambiguation. In machine learning based approaches one relies on corpus evidence. We train a model using tagged or untagged corpus and probabilistic statistical methods are taken. So, here the main idea has to why machine learning is used is that the system should learn how to perform this match.

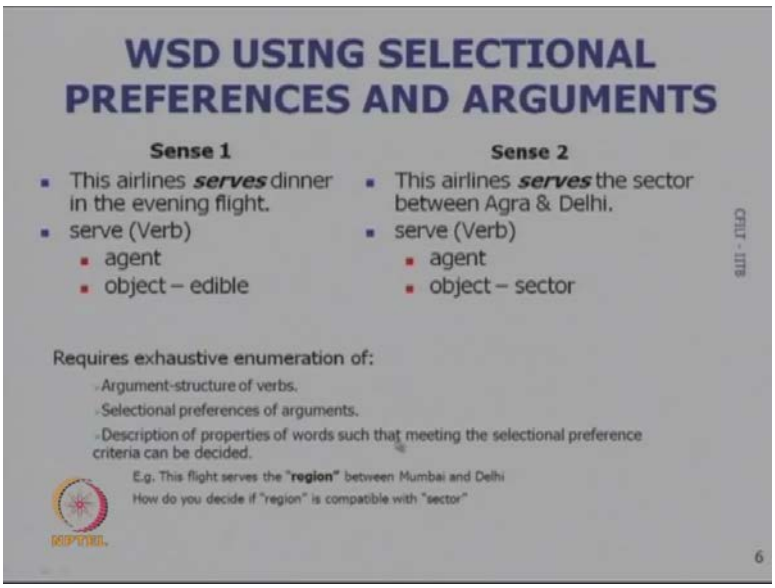
We remember the basic principle the basic principle is that we match the information in the context in which the word appears, with the information provided along with the senses in the word net repository. So, this particular matching has to be done and how to do this matching? How to score this matching? This is learnt by the machine through machine learning methods. So, this gives rise to what is called the machine learning based approach, and then we go to the hybrid approach. Hybrid approach is based on using corpus evidence as well as semantic relations from the word net.

(Refer Slide Time: 10:31)



Now we are going to details of various approaches, our first approach is knowledge based approaches.

(Refer Slide Time: 10:38)

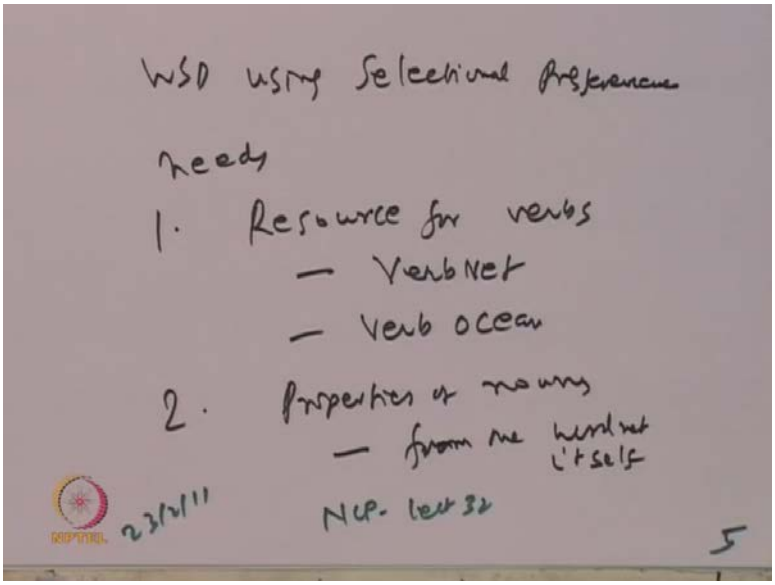


In this set of approaches we now discuss word sense disambiguation using selectional preference and arguments, here is an example. This airlines serves dinner in the evening flight and sense 2 is this airlines serves the sector between Agra and Delhi. So, clearly the 2 senses of this verbs serves are different in this 2 contexts. In fact a translation will make this happen if you translate this sentence into Hindi [FL] airlines [FL] flight [FL].

So serves becomes [FL] whereas, in this particular sentence this airlines serves the sector between Agra and Delhi. The translation will be [FL] airlines Agra [FL] Delhi [FL] [FL] so this will be [FL] and this will be [FL]. So, clearly the 2 verbs have 2 different senses. Now here what we find is that the word serve which is a verb needs an agent it needs an agent in both cases it needs an object what is served. So, here what is served is dinner here what is served is the sector. Now in this case the object is sector, in this case the object is edible. So, here the word sense disambiguation idea says that what you do is that you look at the argument frame of the verb. And wherever, the selection preference, dictates a particular sense pickup that since. So, you can see here again there is the principle of matching coming in from the context we now look at the words which are the neighbors of the word to be disambiguated.

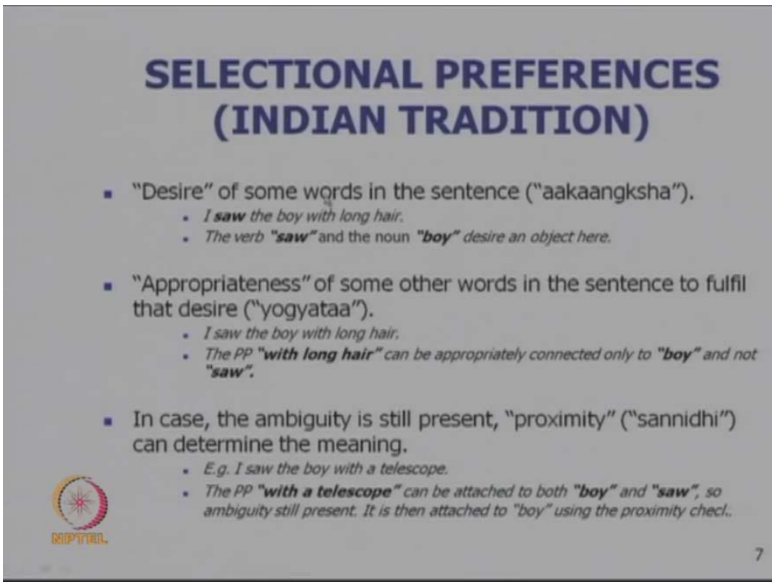
Serve is the word to be disambiguated and in the neighboring words we have dinner. Dinner forms the object and it has the property of the edible. So, here serve has the sense of serving a meal whereas, the object of serve is sector it is not an edible quantity. So, this particular sense of providing a meal is ruled out and sector dictates running a surface which is the sense here. So, this is the basic idea it is a very elegant powerful idea, but it requires many resources. It requires exhaustive enumeration of argument structure of verbs selectional preference of arguments. Description of properties of words such that, meaning of the selectional preference criteria, can be decided. So, all these resources are required, so let me write down what resources are required in for this particular algorithm.

(Refer Slide Time: 14:19)



So, writing on the paper W S D using selectional preference needs 1, resource for verbs. So, this is verb net very important resource, verb ocean another resource properties of nouns from the word net itself. So, these are the resources which we need for doing word sense disambiguation through selectional preference. So, going to the slides once again we find that all these requirements are listed here.

(Refer Slide Time: 15:28)

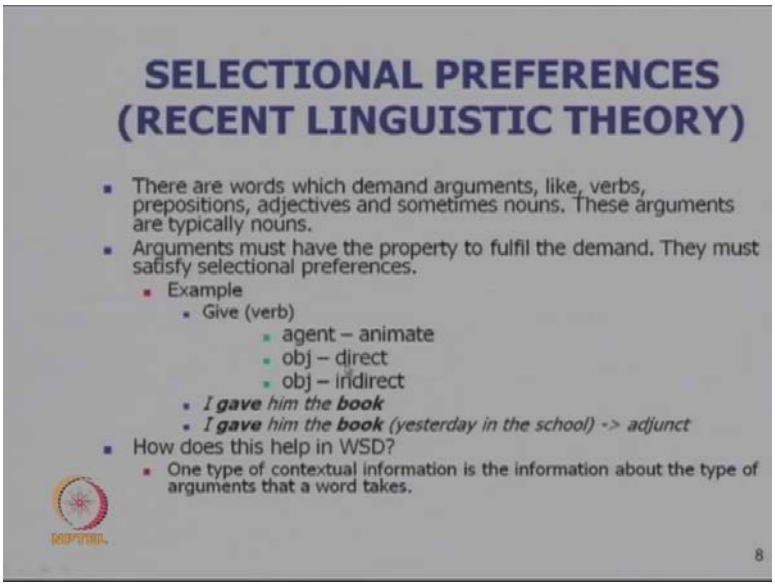


Now, this idea word sense disambiguation using selectional preference leads back to a very ancient Indian tradition for sense disambiguation. So, what these sub body of work is says that some words in the sentence are desire words the word what is called Akanksha which is a Sanskrit word meaning desire. So, if I take this sentence I saw the boy with long then this particular preposition phrase with long hair has to be attached to some quantity in this sentence. It may be attached to saw which means I use with long-hair as the instrument of seeing. And boy it can be attached to boy which means I saw a boy who had long hair.

So, looking from outside we can see that the second meaning is the most appropriate with long-hair should be attached to boy now both see and boy can do with a preposition phrase. Now see definitely has an object here, the boy is the object of seeing with long hair is an adjunct it is non-essential information. It is an additional piece of information in the sentence so it can go either with the boy or with see. So, both see and boy has a certain amount of desire which can be met with this preposition phrase. Now, we come to notion of Yogyathaa or appropriateness of some words in the sentence to fulfill the desire. So, there are desire words and then there are other words that can fulfill that desire. So, those are called Yogyathaa words and those words must have the capability to fulfill that desire. So, I saw the boy with long-hair with long-hair can be appropriately connected only to boy and not to see that.

Because long-hair cannot be an instrument of seeing it does not have that Yogyathaa or capability to be an instrument for seeing. In case the ambiguity is still present proximity or Sannidhi can determine the meaning. So, if I have I saw the boy with a telescope this is an ambiguous sentence both meanings are possible it is possible that I am seeing the boy with a telescope or it is also possible that the boy has a telescope. So, the preposition phrase with a telescope can be attached to both boy and see so ambiguity is still present it is then attached to boy using the proximity check. So, since with a telescope he's proximal to the boy it is nearer to the boy then to see therefore, it will be attached to boy. So, these fundamental principles of disambiguation were discussing in old Indian linguistic tradition namely desire of the word appropriateness of the word to fulfill that desire and proximity on nearness of the word.

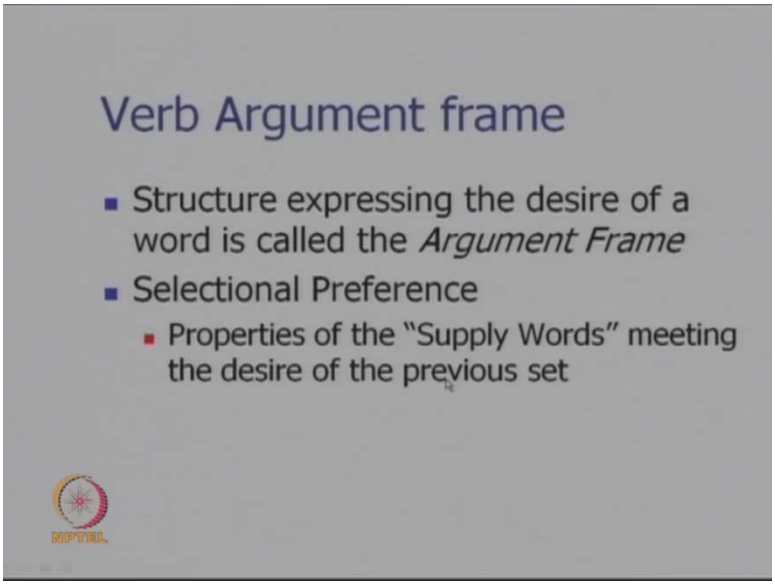
(Refer Slide Time: 18:56)



Now the recent linguistic theory looks up and sectional preferences through the window of arguments and adject adjuncts. So, there are words which demand arguments like verbs prepositions adjectives and sometimes nouns also and the arguments are typically nouns. So, adjectives they qualify something what do they qualify they qualify a noun hence they are their arguments can be nouns prepositions what do they desired again a noun phrase verbs required objects agents and so on. They again can be nouns and sometimes nouns themselves can be desired words nouns will demand nouns. Arguments must have the property to fulfill the demand they must satisfy the selectional preference.

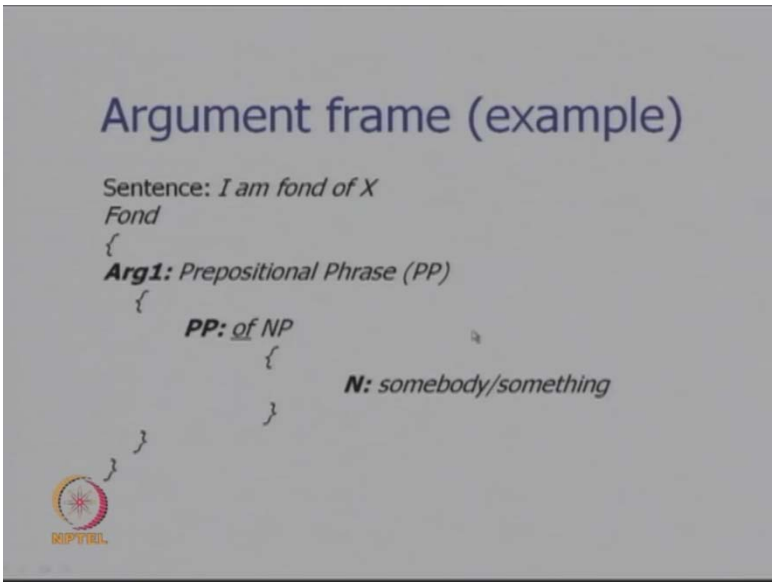
So, for example, give is a verb it has agent and direct object requirement indirect object requirement the agent has to be animate direct indirect objects are inanimate and they are nouns. So, if I take this sentence I gave him the book he is the indirect object, book is the direct object I gave him the book yesterday in the school. So, yesterday in the school is not the essential information it cannot go as argument, but this is the adjunct which is coming in the sentence. So, how does this help in word sense disambiguation one type of contextual information is information about the type of argument that a word takes. So, the type of the argument helps in disambiguating the sense of the word.

(Refer Slide Time: 20:49)



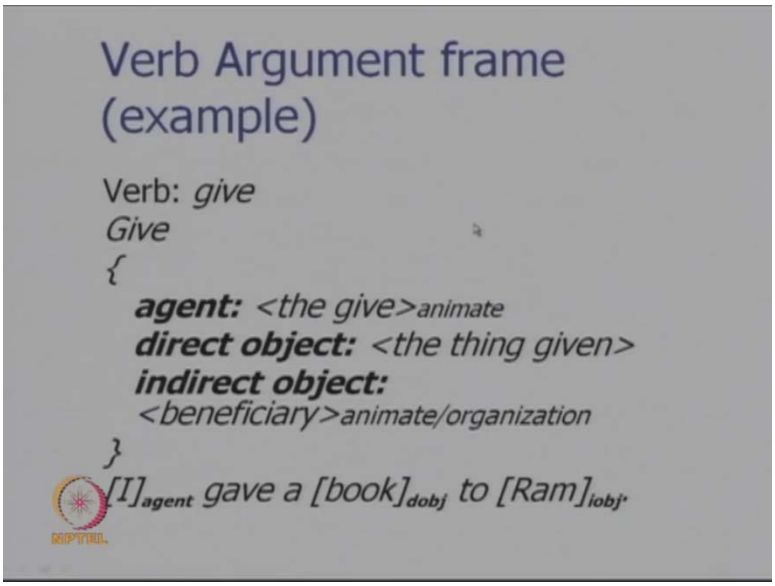
The word argument frame is very fundamental to this technique of word sense disambiguation. The structure expressing the desire of a word is called the argument frame and selectional preference refers to the properties of the supply words meeting the desire of the previous set.

(Refer Slide Time: 21:14)



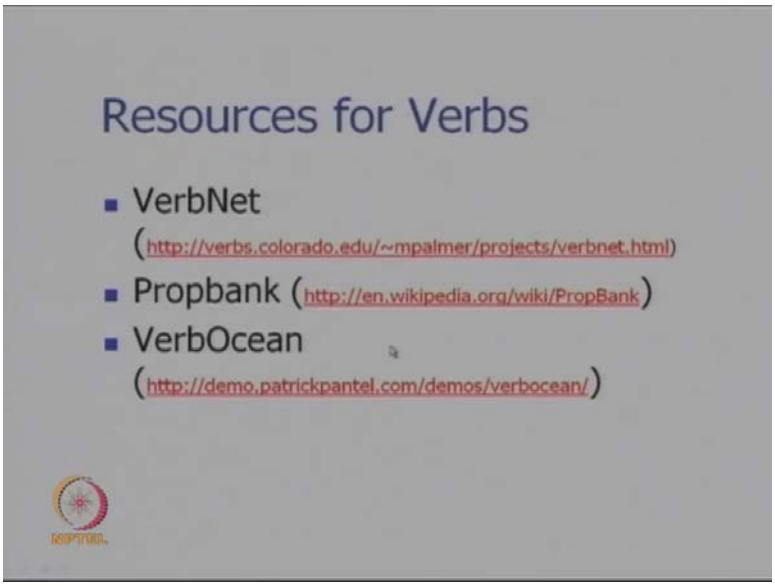
Here is an example of the argument frame, I am fond of X. So, here fond needs an argument the first argument which is a prepositional phrase so I am fond, fond of X of X is required. You cannot finish a sentence by saying I am fond, fond of what is required. And that this shows what is the selectional preference of argument 1? It has must have the following property it must be a prepositional phrase. And the phrase must start with of then there has to be a noun phrase, which is which must have a noun and that is somebody or something. So, I am fond of john this is somebody, I am fond of cooking; this is something; this is the argument frame example.

(Refer Slide Time: 22:09)



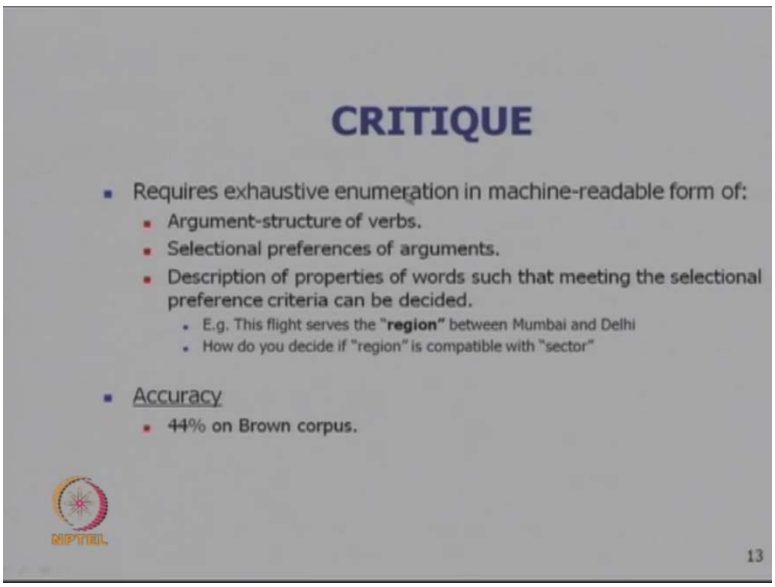
Another example for verb argument frame the last example was for adjective. Give is the word and give as the first argument agent and this is the given which is an animate entity the giver r is missing here. Direct object is the thing given indirect object is the beneficiary of the giving action which can be an animate entity or an organization. So, either agent gave a book the direct object to ram the indirect object. So, this shows that this kind of semantic roles can to be marked on the words in the sentence.

(Refer Slide Time: 22:56)



Now the resources for verbs which I mentioned sometime back are the verb net. This is at this url http verbs.colorado.edu under the home page of mpalmer for the verb net project. Prop bank is another resource on which the information can be found in English Wikipedia. Work Ocean is a project enumerating and storing the verbs this is available in this url.

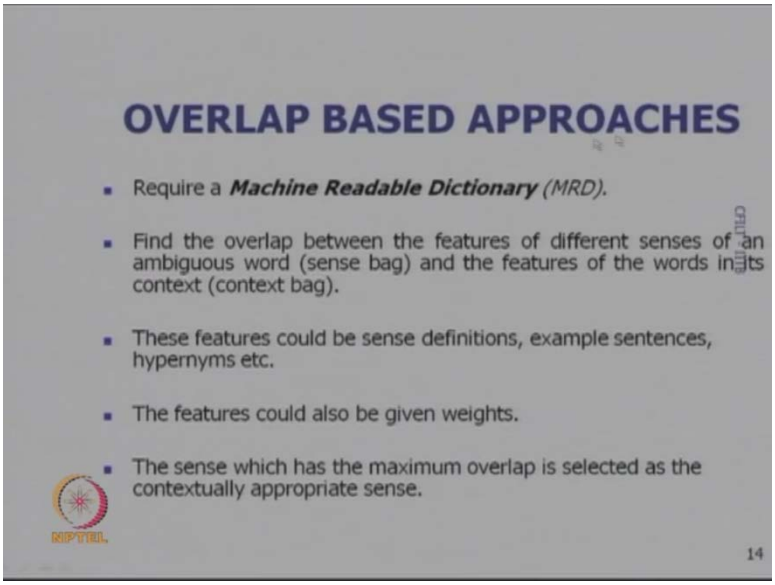
(Refer Slide Time: 23:35)



Proceeding further the critique of selectional preference based word sense disambiguation is that. It requires exhaustive enumeration in machine readable form of

argument structure of verbs, the agents, objects, etcetera. Selectional preferences of arguments, which is nothing but the properties of the nouns coming as agents, objects, etcetera, then description of properties of verbs such that meaning of the selectional preference criteria can be decided. For example, the flight serves the region between Mumbai and Delhi how do you decide if region is compatible with sector? So, this has to be done by means of a lexical resource. And some reported to work using brown corpus which is a well-known purpose in an lap says that about 44 percent accuracy can be achieved using selectional preference based WSD.

(Refer Slide Time: 24:33)

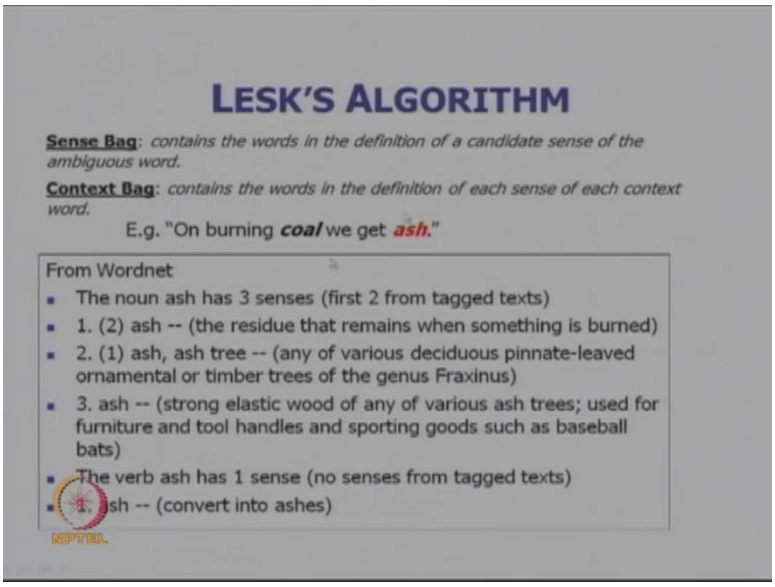


Now we come to another approach which is extremely popular and well-known this is known as overlap based approaches. So, here there is a here is an approach, which makes the idea of matching between the context words are the information in the context along with the information in the sense in the word net extremely explicit. This is clearly based on the idea of matching. So, this is what is written here overlap based approach is need machine readable dictionary. It finds the overlap between the features of different senses of an ambiguous word and this difference senses is known as the sense bag.

And the features of the words in the context of the word being disambiguated this is called context bag. So, sense bag and context back and the overlap between these 2 bags decides the winner sense of the word. So, these features could be sense definitions

example sentences hypernyms etcetera. And the features could also be given weights the sense which has a maximum overlap is selected as the contextual appropriate sense.

(Refer Slide Time: 25:55)



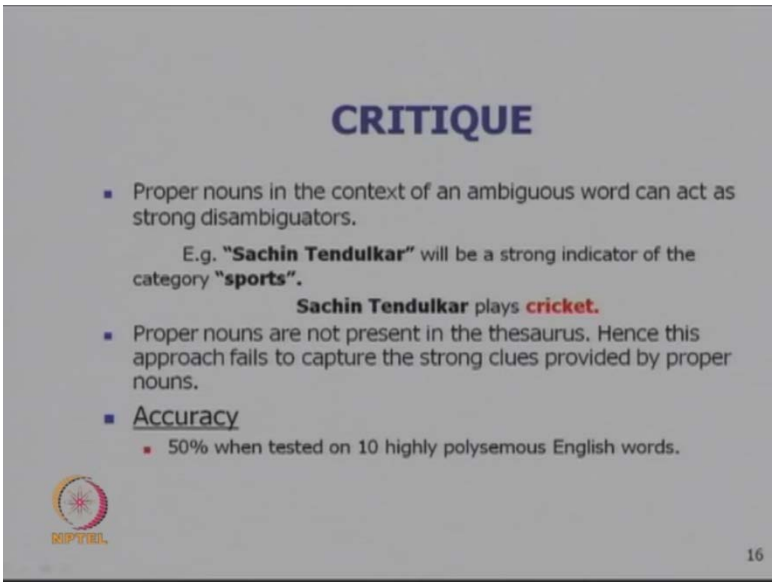
Here is an example and the example is known as a Lesk algorithm this is based on the overlap idea. Lesk algorithm is one of the most famous algorithms in word sense disambiguation. So, sense bag contains the words in the definition of a candidate sense of the ambiguous word. And the context bag contains the words in the definition of each sense of each context word. So, the words in the context provide the context bag both directly and indirectly. So, they provide the elements for matching and they are also providing their own senses and those words in the sense in their senses as the elements of matching. So, if I take this sentence on burning coal may get ash, and suppose ash is the word which needs to be disambiguated. So, from the word we find that the noun ash has 3 senses ignore for the moment these phrases in the bracket. Ash, as the first sense the residue, that remains when something is burnt the most frequent sense of ash.

Second sense is ash tree, any of the various deciduous pinnate leaved ornamental or timber trees of the genus Fraxinus. So, this is the tree sense of ash this was the residue sense of ash. The third sense is strong elastic wood of any of various ash trees used for furniture and tool handles and transporting goods such as baseball bats. So, here this ash has a sense of wood, so what is the 3 senses ash as residue, ash as tree, ash as wood, these are the 3 senses the word. Ash has one sense ash is to convert into ashes we do not

have to look at the word sense. Because now the part of speech direct disambiguation has become quite match here. And given the words in a context a number of words in the context it is possible to disambiguate the words in terms of their parts of speech direct with high accuracy. It is common to see English part of speech taggers which have the accuracy of 97 to 98 percent. Indian language part of speech taggers report accuracy of above 93, 94, and 95 percent for various Indian languages.

So, if I take this sentence here on burning coal we get ash let us see how to disambiguate this. What happens is that we find here the glosses of the words, so here this is the gloss the residue that remains when something is burned. This word burnt has an overlap with the word burning in the context of ash burning and burnt no other sense has anything overlapping with the context. So, we will flag the first sense of ash as the target sense as the disambiguated sense and we will be writing here, because this is the sense of ash.

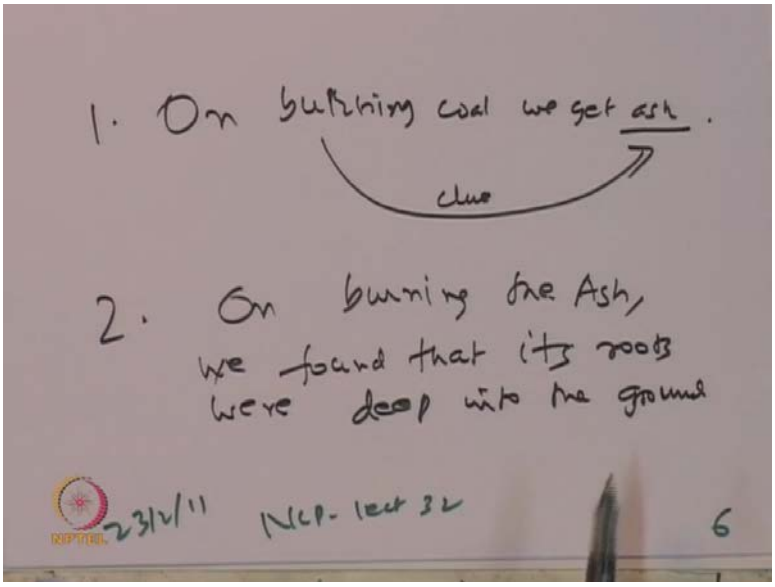
(Refer Slide Time: 29:51)



So, these are very simple idea which shows how the algorithm works and however the method is not free from its critiques. Proper nouns in the context of an ambiguous word can act as strong disambiguates. For example Sachin Tendulkar will be a strong indicator of the category sports. So, in the sentence Sachin Tendulkar plays cricket the sense of cricket becomes very apparent because of the presence of Sachin Tendulkar. Cricket has 2 senses, 1 is sports sense the other is the inset sense.

So, here the word play is disambiguated Sachin Tendulkar is a strong disambiguator. Now proper nouns are not always present in the word net, hence this approach fails to capture the strong clues provided by the proper nouns. The accuracy reported is about 50 percent when tested on ten highly polysemous English words, but there are other problems with this approach.

(Refer Slide Time: 30:56)

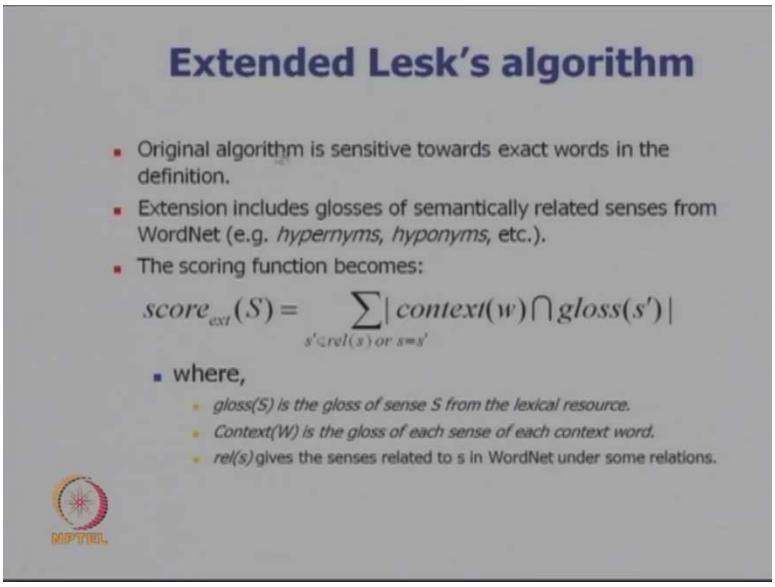


Let me write it down so our sentence was on burning coal we get ash. This is the one to be disambiguated and the clue was coming from burning, so this is the clue word clue. Now we can see that this is sentence 1 we can have another sentence where on burning the ash we found that it its roots were deep into the ground. These are sentence on burning the ash we found that its roots were deep into the ground. So, here we clearly mean ash as a tree because the clue here is roots, but if you see the definition of ash here the tree sense does not have the word root. So, the word net based approach will in no way disambiguate this if it goes by direct surface level matching.

Now, the interesting thing here is that if you latch on to the sense backwards and if we look at the parts of a tree we take the mirandame information for tree. Then we will get the word root and then this will have a matching with the word root in the given sentence. So I think you get the idea, the idea is that here the word burning is misleading clue it will take it take the algorithm to the first sense which is the residue sense of ash. Whereas, the intended sense is the tree sense and that clue is coming from the word root.

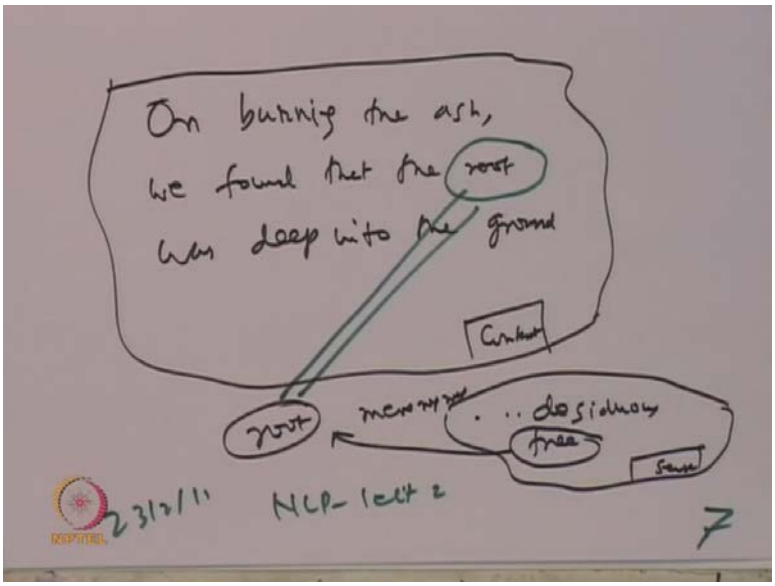
So, the word net based approach can be misled into a wrong sense if one is not careful with the context. So, this was the critique.

(Refer Slide Time: 33:43)



Now we take extended Lesk algorithm and for this we have already laid the platform. Because we have said that the previous example had the word root as the clue. The word root was not present in the direct definition of the tree sense of ash, but if we take the word tree from the sense of ash. And go to the word net to look at the senses of tree and gets semantically related words. Tree has the word root along with it and that will help in the disambiguation so pictorially this can be represented in the following way.

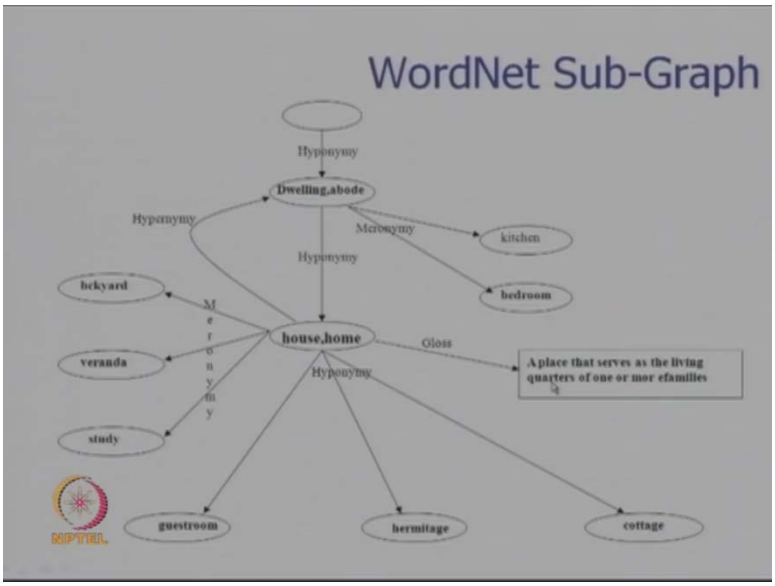
(Refer Slide Time: 34:41)



So, what we are saying is that on burning the ash we found that the root was deep into the ground. So, this is the clue word is a clue word and this is the context now when we pick up the sense bag. So, that had the word tree deciduous tree now the word tree has the word root through meronyme, which is a part of relation. And now we can establish a matching between these 2. So, this root from the context matches with root from the word net. The sense repository and that will give rise to the disambiguated sense, so that is the main idea of the extended Lesk algorithm.

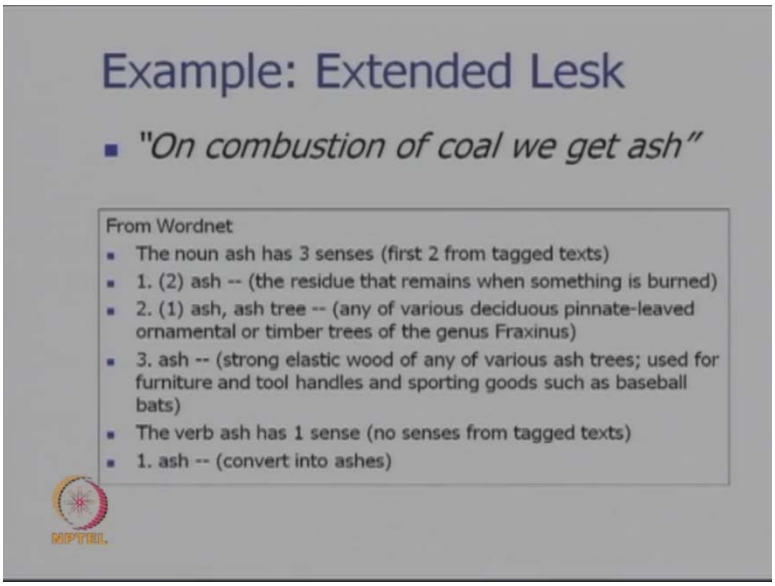
Original algorithm is sensitive towards exact words in the definition extension includes glosses of semantically related senses from the words. For example, hypernyms, hyponyms, meronyms etcetera. The scoring function now becomes score of a sense is equal to context of the word which needs disambiguation W. And the gloss of S dashed where S dashed is semantically related to S or S is equal to S dashed. So, get all those words in those senses which are related to S all are equal to S were gloss S is the gloss of senses from the lexical resource context. W is the loss of each sense of the context word and rel S gives the senses related to S in the word net under some relations.

(Refer Slide Time: 37:02)



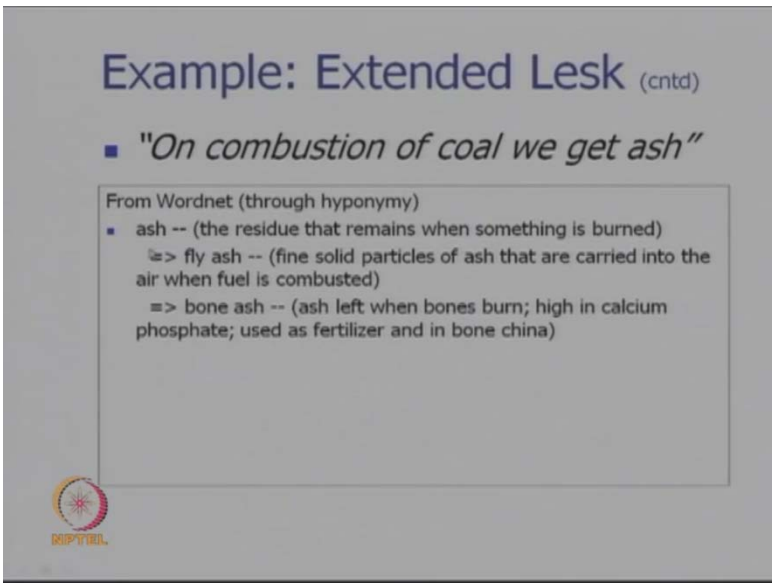
So, this actually reminds us that we have to make use of the structure of the word net. And the structure of the word net is essentially graph if we were to disambiguate a word like house appearing in the context. Then we will when you go make use of the sense bag of house we have to consider these whole sub graphs for a particular sense of house. So, if I take a sentence for example, the house was demolished during road broadening. The role of house was demolished during road broadening and this is the correct sense of house. So, we look at all possible sense vocalized as all possible sources of information around house which comes from gloss example sentence, hypernyms, hyponym meronyms and so on. So, extended Lesk makes use of the word nut structure and its graph.

(Refer Slide Time: 38:17)



Here is an example on combustion of coal we get ash. From the word net noun ash has 3 senses and we find that the word combustion does not appear in any of them. So, none of the senses will return its score all of them will return its score 0.

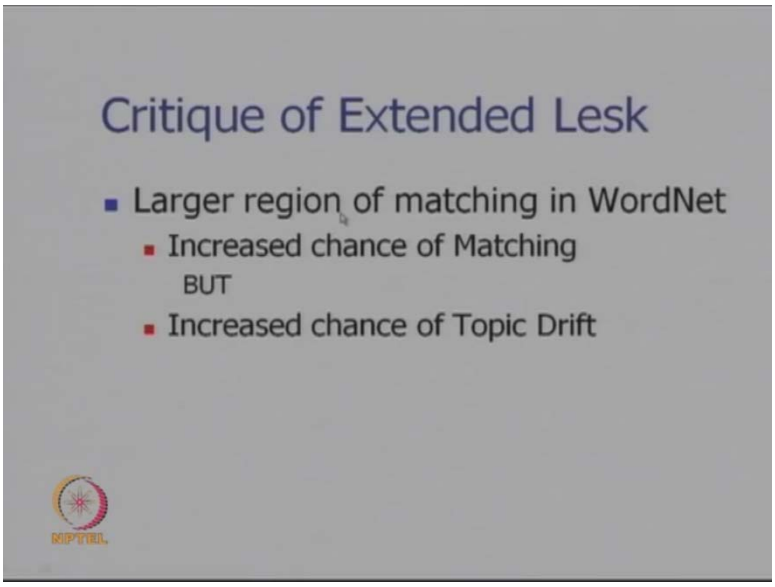
(Refer Slide Time: 38:38)



However, if we use extended Lesk then we go to semantically related senses. So, ash has hyponym fly ash is a kind of ash, bone ash which is another kind of ash. And we find that we get an overlap with combustion when we go to the hyponym of the sense which is the fly ash here. So, combusted and combustion have an overlap so the residue sense

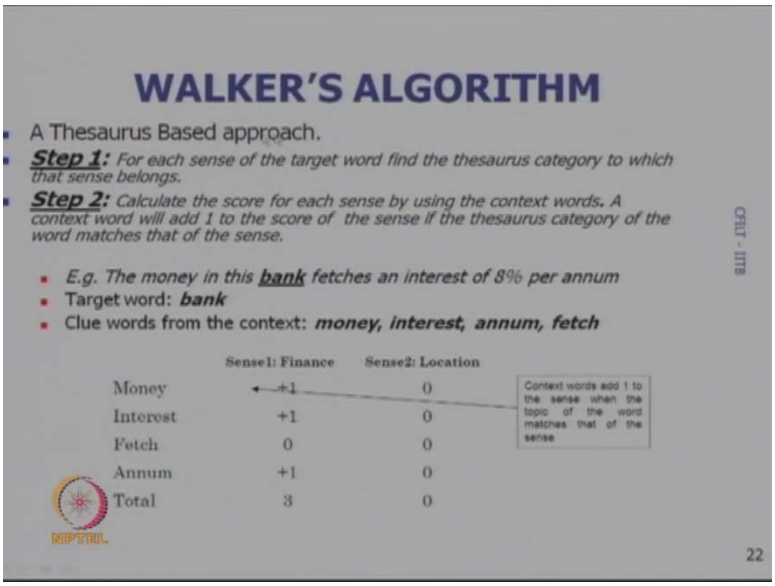
of ash becomes the winner, because of this deeper level overlap. So, this is extended Lesk.

(Refer Slide Time: 39:23)



Now, the critique of extended Lesk is that there are larger regions of matching in the word net. It has increased chance of matching but it also has the increased chance of topic drift, because we can match with words which have nothing to do with the particular sense. So, for example, the word combustion was used as the clue for matching. Now, in this case the word combustion was, but supposes the word combustion had another sense. Another additional sense and then this overlap or this matching will produce a wrong sense.

(Refer Slide Time: 40:13)



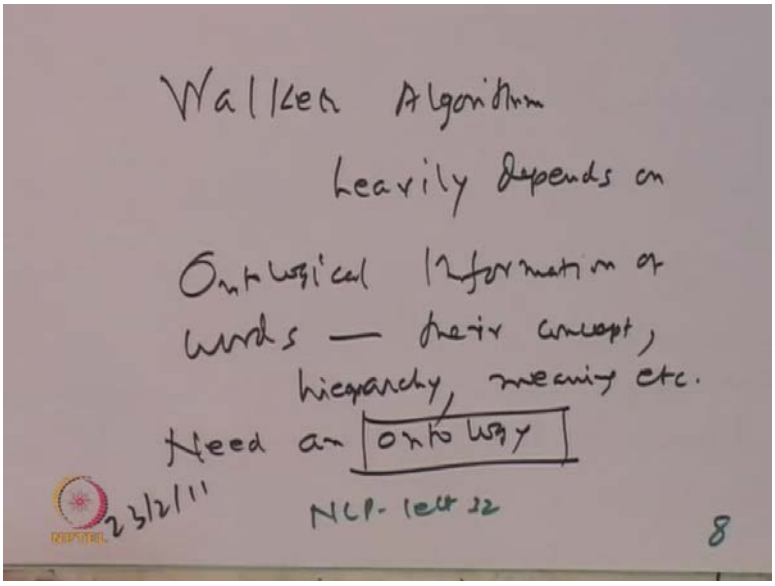
Next algorithm is called Walker's algorithm this is a thesaurus based approach. So, the depth makes of a thesaurus the first step is that for each sense of the target word find the thesaurus category to which that sense belongs. And step 2 is calculating the score for each sense by using the context words. A context word will add 1 to the score of the sense if the thesaurus category of the word matches that of the sense. Here is an example the money in this bank fetches an interest of 8 percent per annum. So, the target word is bank and clue words which explicate the sense of the word are money, interest, annum and fetch. So, what we find here is that there are two sense counters for bank sense 1 and sense 2.

The first sense is in the domain of finance, the second sense is corresponding to the sense of location. The word money increases the count in the sense 1 counter it does not add contribute anything to the second counter. Interest also adds 1 interest is in the finance domain fetch does not add anything to any of the counters. Annum adds 1 to the finance counter nothing to the location counter and thereby we find that the finance counter has a value of three location counter as a value of 0 and therefore, the finance sense of bank here is the winner sense. And that is appropriate in this particular sentence I find that the finance sense of bank is used.

Now the question of course, that arises is that for Walker's algorithm the crucial resource is the thesaurus it is from the thesaurus that this kind of counts come in and get added to

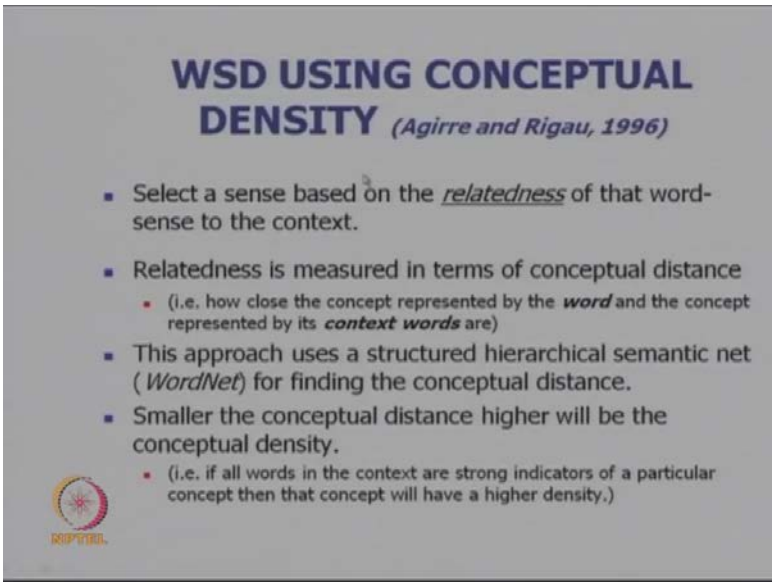
the appropriate counter. But the question is how do I know that interest, annum, etcetera are in the finance domain and not in the sense of location? So, interest is an abstract noun, location is concrete, annum is an abstract noun again location is a concrete entity tangible entity. So, this means that we need a resource where this kind of information is kept the fact that interest cannot be a location interest in the finance domain and so on. So, this algorithm is heavily dependent on this kind of information and they are known as ontological information.

(Refer Slide Time: 43:22)



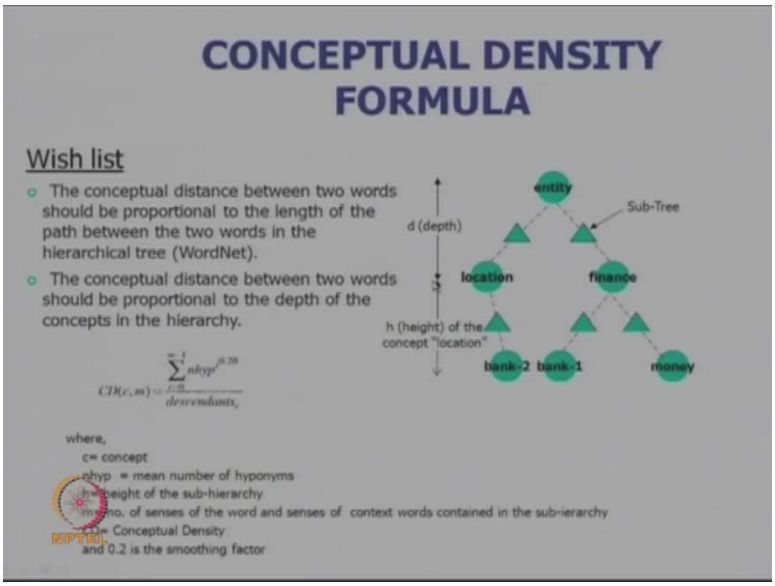
So, I would like to write it down and say the following that walker's algorithm heavily depends on ontological information of words their concept hierarchy, meaning etcetera, so need an ontology. So, let us remember this walker's algorithm crucially depends on ontology. Now we see the slide again and this fact is expressed in walker's algorithm.

(Refer Slide Time: 44:34)



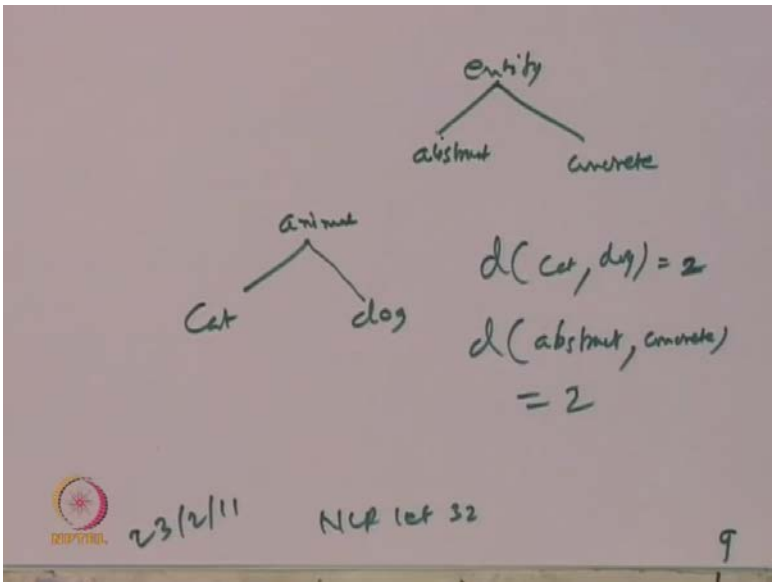
We move onto W S D using conceptual densities please remember that all these approaches are knowledge-based approaches. And in particular they are overlap kind of approaches. So, this conceptual density idea was proposed by Agirre and Rigau. Here the idea is to select a sense based on the relatedness of that word sense to the context. And relatedness is measured in terms of a technical term called conceptual distance. So, this represents how close the concept represented by the word and the concept represented by its context words are. This approach uses a structure hierarchical semantic net word net for finding the conceptual distance, smaller the conceptual distance higher will be the conceptual density. If all the words in the context are strong indicators of a particular concept then the concept will have a higher density.

(Refer Slide Time: 45:43)



So, here is a wish list for conceptual density formula, the conceptual distance between 2 words should be proportional to the length of the path between the 2 words in the hierarchical tree. Now, since it is the distance which measures the overlap and which is the clue for disambiguating the word. Therefore, we say that the conceptual distance is proportional to the length of the path between the 2 words in the hierarchical tree. And what are the 2 words? The 2 words are the word which needs disambiguation and the clue word from the context. The conceptual distance between 2 words should be proportional to the depth of the concepts in the hierarchy. Now, this is very well illustrated by the following example which I will write.

(Refer Slide Time: 46:41)



And so one can see here suppose I take the word cat and dog whose parents are animal suppose. And we also take 2entities which are very general and that they are also immediate children of a very high level concept called entity, which transfers everything. So, suppose we have abstract and concrete, the distance between abstract and concrete is 2 you go to the parent and come down to the simile. Here also the distance between cat and dog is 2, so d cat, dog is 2 and d abstract, concrete is also 2. But would we say that the similarity between cat and dog is equal to the similarity between abstract and concrete would we say that?

We will not be able to say that because abstract and concrete are extremely large concepts and their hardly similar there is hardly any similarity between them. But cat and dog both are 4 legged carnivorous animals there are familiar differences between them. But they are much more similar than the than are abstract and concrete. The similarity between cat and dog is much more than the similarity between abstract and concrete which are hardly similar. So, we should not be misled by the distance between 2 nodes as such in the word net we have to consider the level at which these concepts exists, how general, how specific these are. So, we will continue with these ideas in the next lecture.