

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

Lecture - 13
POS Tagging contd...
Indian Language Consideration; Accuracy Measure

We will continue with our discussion on POS Tagging, which was explain to be a very important task in natural language processing. And what we have seen before is, how to do this task through machine learning methods in particular, through the application of hidden mark of model. We did each mm for part of speech tagging and also saw how the arbitrary algorithm is used for obtaining the correct part of speech tag. In today's lecture, we would like to take up some of the issues in natural language processing, as they appear in Indian language processing.

We will take example from Indian languages in particular Hindi, later we will look at other languages of the country from the south Dravidian languages, from the East Bangla and Tibet, Burman, Oria ((Refer Time: 01:23)) this kind of languages. So, the issue is, how to do part of speech tagging for a non English language, so let us proceed with the material.

(Refer Slide Time: 01:38)

Some notable text corpora of English

- [American National Corpus](#)
- [Bank of English](#)
- [British National Corpus](#)
- [Corpus Juris Secundum](#)
- [Corpus of Contemporary American English](#) (COCA)
400+ million words, 1990-present. Freely searchable online.
- [Brown Corpus](#), forming part of the "Brown Family" of corpora, together with [LOB](#), Frown and F-LOB.
- [International Corpus of English](#)



[Oxford English Corpus](#)
[Scottish Corpus of Texts & Speech](#)

We have seen that, part of speech tagging is done through the annotated corpora, annotated text corpora you remember are those, which have annotations in them. In particular for part of speech tagging, we have a corpora which are tagged with part of speech for example, noun, verb, adjective, and so on. Now, I mentioned here some of the important text corpora for English, American national corpus and British national corpus are extremely large corpora, which are for English and they have lots of texts from different domains, which are tagged with part of speech.

Bank of English is from the financial domain, a number of documents are available which are again part of speech tagged. Corpus of contemporary American English is for modern American pros and they are also part of speech tag. You can see, it is a very large corpora with 400 plus million words, 1990 to present which means, mid to 2000 and this is freely searchable online.


Brown corpus is very famous, because the task of part of speech tagging through machine learning methods was possible, because of brown corpus, this is about 1 million words of English and they have part of speech tags in them. Similarly, international corpus of English, oxford English corpus, Scottish corpus of texts and speech, all these are very valuable data for applying machine learning to natural language processing.

(Refer Slide Time: 03:36)



Indian Language POS tagging

- Happening under a large scale nation wide project called "Indian Language to Indian Language Machine Translation"
- Multiple institutes across the country: consortium mode
- ILILMT POS tags are accepted as standard for tagging of IL corpora



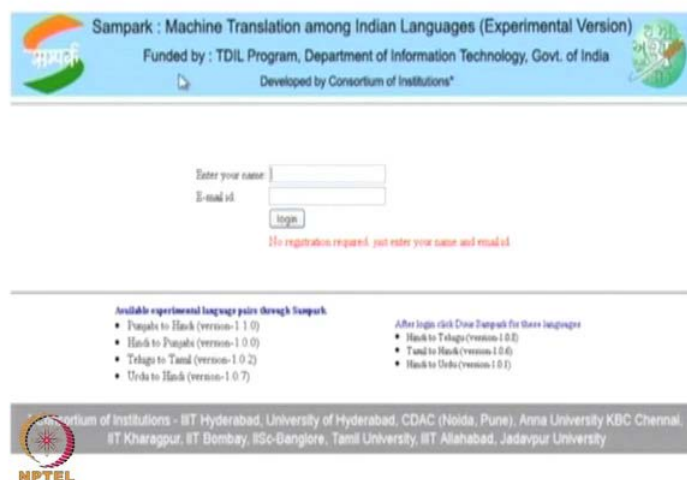
What is the India language scenario, for Indian languages also, there is a lot of initiative from the government to produce natural language processing systems in particular,

machine learning base methods. And we would like to also apply natural language processing techniques for processing of Indian language. Now, in the government of India scenario, most of the documents have to be present in a three languages, the state language, English and Hindi.

So, it is necessary that, translation automated means of translation are developed with all seriousness and natural language processing research with focus on translation happens. So, as mentioned before, part of speech tagging happens to be a starting stage for any natural language processing. So, Indian language part of speech tagging is happening under a large scale nationwide project called Indian language to Indian language machine translation, I mentioned it in my transparencies.

So, Indian language to Indian language machine translation is a very large scale effort, multiple institutes across the country are involved in this task in a consortium mode. The short form is ILILMT and the part of speech tags developed under this consortium have become the standard for part of speech tagging of any Indian language corpus. And this IL corpora will be freely available for research purposes across the country.

(Refer Slide Time: 05:30)



Proceeding further, we find that Indian language to Indian language machine transition system is funded by the Department of Information Technology, government of India. And it is a very large scale effort developed by the consortium of institutes, which involve IIT Hyderabad, University of Hyderabad, CDAC Noida and Pune, Anna

University, KB Chandrasekhar centre at Chennai, IIT Khargpur, IIT Bombay, IISC Bangalore, Tamil university, triple IIT Allahabad and Jadavpur University.

So, one can see that, these institutes cover the length and breadth of the country and have a very large representation in terms of all major languages of the country. One can also see the mentioned of different pairs of language, on which transition systems have been built. So, this system is called Sampark and this has developed lot of large scale corpora with annotation in them in part of speech tag form.

(Refer Slide Time: 06:41)

Noun Tags (Examples in Hindi)

Sl. No	Category		Label	Annotation Convention**	Examples	Remarks
	Top level	Subtype (level 1) Subtype (level 2)				
1	Noun		N	N	ladakaa raagaa kotaaba	
1.1		Common	NN	N__NN	kotaaba kalama cashmaa	
1.2		Proper	NNP	N__NNP	Mohan ravi_rashmi	
1.3		Verbal	NNV	N__NNV	NA	Not Required
1.4		Nloc	NST	N__NST	Uupara rice_aage piche	



We take examples of various tags and as they are used for Indian language processing, the examples are from Hindi and later, we will take examples from other languages also. As is known with a most important category of words in any language are the nouns, nouns carry lot of information. And verbs have prepositions, conjunctions, etcetera serve to link this nouns together and piece this information together to convey a considerable large meaning.

We see here, noun has been mentioned and there are different categories of noun like common noun, proper noun, verbal noun and something called NLOC, which we are going to explain very soon. So, under noun we have these examples, they are from Hindi, they are in transliterated English. It is written so that, anybody could look at the words of course, the Devnagary strings are also there to represent this words. The first word is

[FL] which means a boys, second word is Raja which means a king, The third word is [FL] which means a book, so all these are nouns.

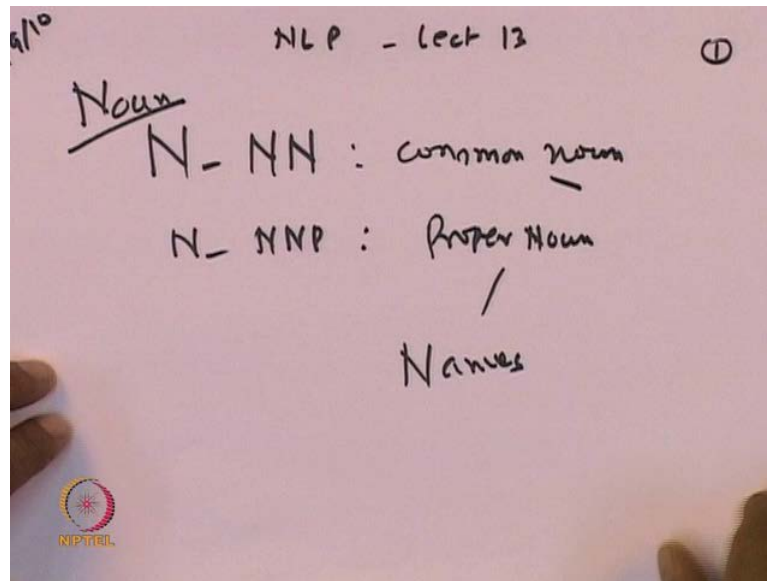
Now, under noun, we have the first category called common noun, so the basic tag here, the basic symbol which expresses this category is N. And then common noun is NN, proper noun is NNP, verbal noun is NNV and location and time specific noun is NST, let us go over them. So, [FL], these are common nouns, [FL] is book, [FL] is a pen, [FL] is the spectacle. So, these are common nouns and the part of speech tag, which should be inserted for this is N underscore NN.

Now, the proper nouns are the names for example, Mohan is a common Indian names, so Ravi and Rashmi and the tag for this is N underscore NNP. The symbol before the underscore is the top level category and the sub category is given after the underscore, NN is common noun, NNP is proper noun, NNV is verbal noun. In Hindi, we do not see this phenomenon, Dravidian languages we see the occurrence of verbal, nouns and we will remark on this category later.

This is a particularly Indian phenomenon, namely NLOC or locative nouns, nouns of time and space. For example, the word [FL] which is above, [FL] which is below, [FL] which is front of, [FL] which is behind, all these are very peculiar words, because they can serve as noun also. They are adverbs, they also can serve as noun, because they can take case marker. For example, one could say [FL], the room above or let say [FL], the staircase below, [FL] the house in front and so on.

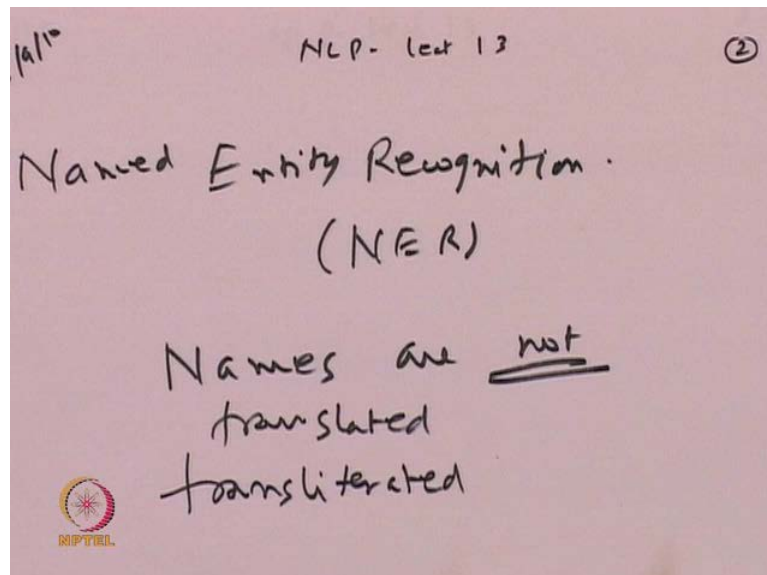
So, all these words behave as nouns and they can take case marker, on the other hand they also act as post positions. For example, [FL] the terrace on top of the house, [FL] the bungalow below the hill, so these kind of usages show the use of these words as post position linking to nouns. So, these peculiar words have double role, they can act as pure nouns and they can also act as post positions, that is why this is special category for them and underscore NST. Now, question that may arise is that, why should we have two categories.

(Refer Slide Time: 11:17)



If I write down two major categories, which are N NN under noun, both are under noun, this is a common noun and N NNP is for a proper noun. Now, these proper nouns are essentially names and these are of course, common nouns. Now, what is the point of distinguishing between them, why should we keep two different tags for this.

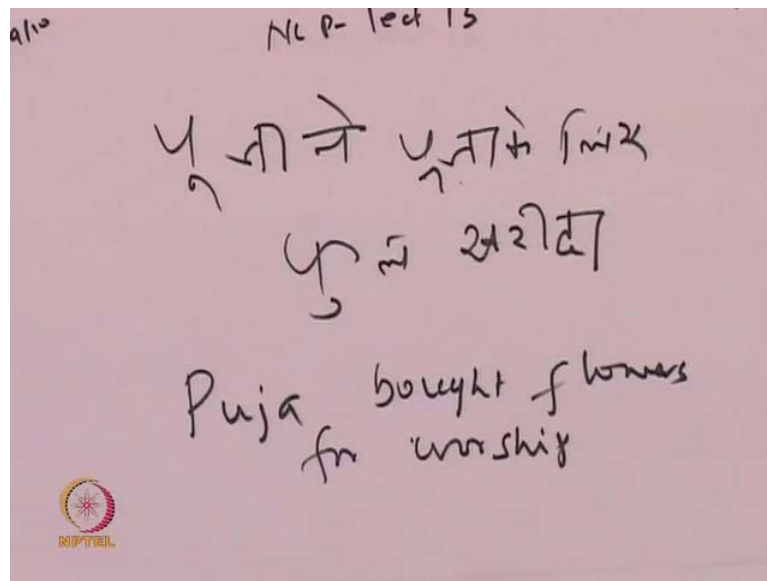
(Refer Slide Time: 12:02)



One important reason is that, in a natural language processing there is a task called named entity recognition or NER, it is important to detect names in Indian languages. That is, because the names are not translated, on the other hand they are transliterated,

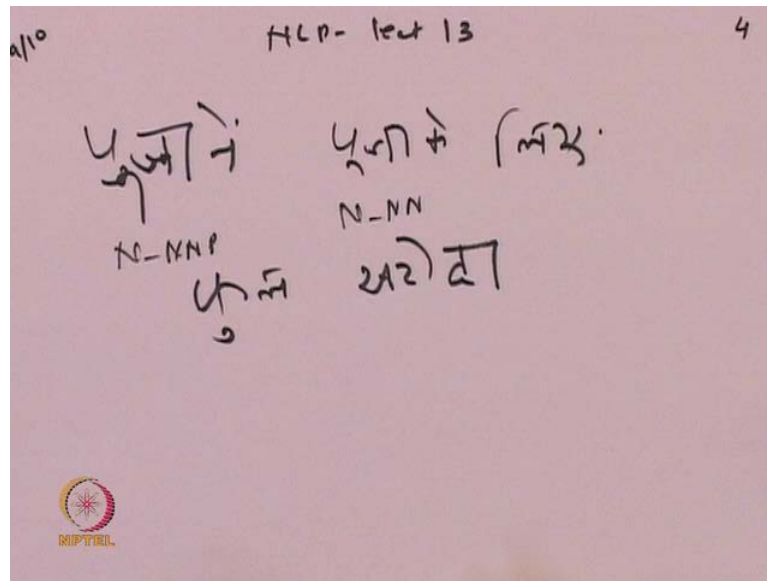
not translate but translated. Therefore, it is essential for us to detect, which noun in the text is a name and which noun is not a name. If a noun is common noun then it should be translated for example the word [FL] should be translated to book. On the other hand, a name like Sita written in English should be transliterated to s i t a, if we are going from Hindi to English and from English to Hindi, it should be transliterated to Sita. One important difficulty that arises is that, Indian languages do not have capitalization for proper nouns.

(Refer Slide Time: 13:23)



So for example, if we take this sentence [FL], so the first Puja is the name of a person that is not worship, the next puja [FL] for worshipping, this puja is a common noun. So, the translation for this sentence will be Puja bought flowers for worship. So, you can see what is happening, the first Puja did not get translated, it got transliterated to puja, p u j a. The second puja which means worshipping, got translate to worship therefore, it is very necessary that, in the natural language text, we detect which is the common noun Puja and which is the proper noun. That is why, this named entity recognition task has become important and the idea is that, when one is doing part of speech tagging, since the annotators are placing marks on the words, there are identifying the nouns. They could do little more work with hardly any additional effort and place the tag of NNP, proper noun with the noun.

(Refer Slide Time: 14:54)



So, in this case, if we take this sentence [FL], if we take this part [FL] then we will have to place the tag of N NNP for this Puja and N NN for this puja, worship. So, when this tagging is going on, we are already doing this task or proper name identification, which requires hardly any additional effort, this is the way it will go. So, that is why, we have a distinction between proper noun and common noun. And if we again look at the slide and see categories under noun, we have common noun, proper noun, verbal noun and N LOC. N LOC is easy to identify, because there is only a limited set of words which appearing in this list.

(Refer Slide Time: 15:57)

Pronoun & Demonstrative Tags (Examples in Hindi)

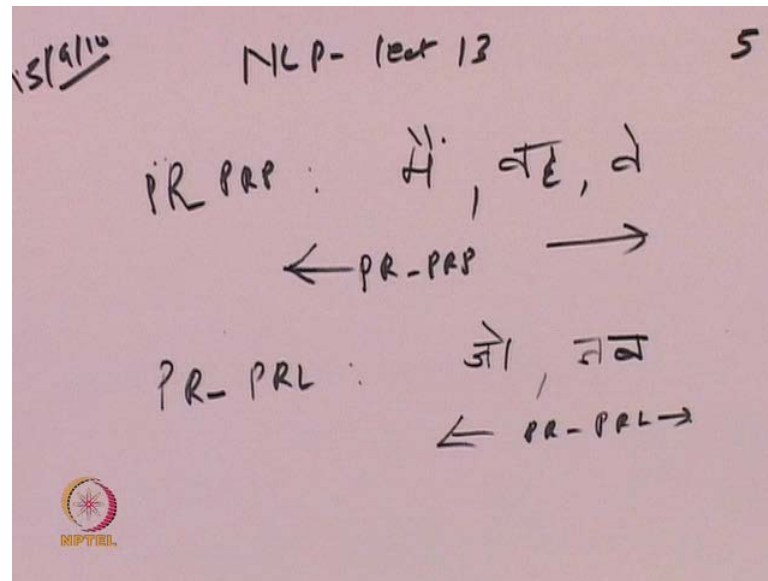
2		PR		PR	
Pronoun	Personal	PRP	PR_PRP	Yaha vaha jo	
	Reflexive	PRF	PR_PRF	Vaha main tum, ve	
	Relative	PRL	PR_PRL	Apanaa swayam khooda	
	Reciprocal	PRC	PR_PRC	Jo jse jab jahaam	
	Wh word	PRQ	PR_PRQ	Paraspara aapasa	
3	Demonstrative		DM	DM	Yaha jo, yaha.
		Deictic	DMO	DM_DMO	Yaha yaha
		Relative	DMR	DM_DMR	jo jse
		Wh word	DMQ	DM_DMQ	kol kis, kaun

Let proceed to the next category of words, these are pronouns and demonstrative tags, let us look at the two top level categories, which are pronoun and demonstrative. There are symbols for them, PR is the basic symbol, DM is the basic symbol here, for pronoun we see that the sub categories are personal pronoun, reflexive pronoun, relative pronoun, reciprocal pronoun and WH word. Let see the examples, personal pronoun, [FL] etcetera, [FL] is first person singular number, [FL] is third person singular number, [FL] is second person singular number, [FL] is the third person plural number, all these words refer to persons and these are pronouns, that is why these are personal pronouns. Reflexive, reflexive refers to the pronouns referring to self for example, [FL] my own house, [FL] my own house, [FL] understand yourself, [FL] go yourself and see these are pronouns which are reflexive, they refer to self.

And therefore, one has to place the symbol here, underscore PRF that is reflexive, this PR before underscore is the basic category, which is pronoun. Relative pronoun refers to those pronouns, which link classes for example, [FL] the boy who came yesterday seems well, so this Jo is a relative pronoun qualifying standing for a noun. So, this is the relative pronoun PR underscore PRL expresses this fact, reciprocal is a kind of pronoun denoting two persons with reciprocal action between them.

[FL] both of you know each other [FL] means, each other [FL] what are you discussing between yourselves. So, [FL] etcetera are reciprocal pronouns, they involve two persons and here, the symbol is PR underscore PRC, WH word is [FL] etcetera and these are also pronoun, when they uses a question. [FL] who is coming today, [FL] when will you go, [FL] where do you stay, so these are pronouns they are used in question, they refer to nouns. Thus, we have this five categories under pronoun, sub categories under pronoun and their symbols are these, PR underscore PRP, PR underscore PRF, PR underscore PRL, PR underscore PRC, PR underscore PRQ. Let us now spend a little time in understanding, is it possible by a machine to distinguish between them, what clue is there.

(Refer Slide Time: 19:47)



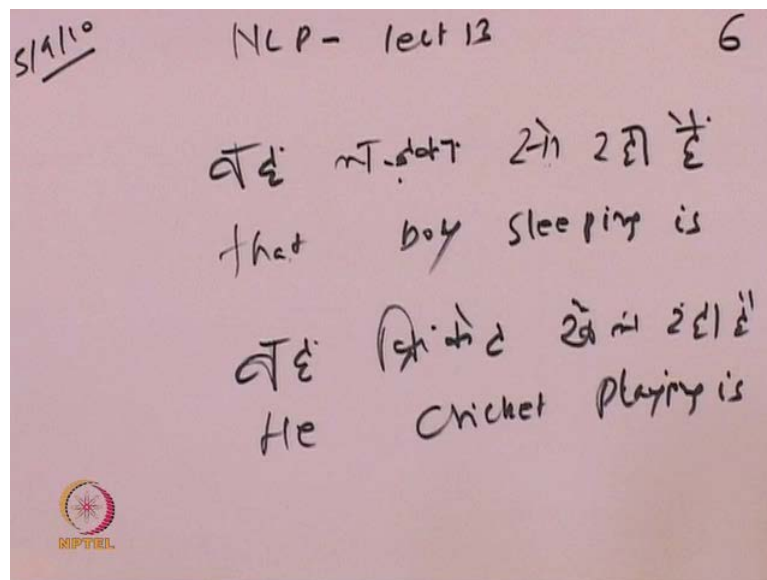
We write, let us say, two or three of this categories, so let us take the PR PRP case of let say, [FL] and so on. Can a machine detect that, these are personal pronouns and can it place this kind of tags, which is PR PRP for all of them, can the machine do it. The another tag let me take, which is let say relative pronoun, which is PR PRL and there we have examples like [FL] and so on. And here, the tags should be PR RPL, can a machine detect this, yes the immediate impression is that, the machine can detect them, because [FL] is very unique.

Look at the word and you know that, the tag is PR PRP, look at the word [FL], the tag is PR PRP, [FL] all of them is the case. If we take the word let us say, [FL] which is reciprocal, [FL] PR PRC, it is easy to detect them. The reason is that, they forms at this, now we come to a difficulty, we come to this category call demonstrative and we will see that things are not so simple. A demonstratives are those entities which denote a particular noun, the basic symbol is DM.

So, there are three kinds of demonstratives say deictic relative and WH word, so the symbols are DMD, DMR and DMQ. So, DM underscore DMD, we have examples [FL], what is the example here, [FL] this [FL] is a demonstrative, because it is denoting a particular boy, these has to be distinguish between [FL] here this is a pronoun. But, [FL] here it is a demonstrative, now the question is, in which is case is [FL] a pronoun and in which case, is it a demonstrative.

We have discuss this before, this is not a simple problem, you cannot simply say that, if [FL] is followed by a noun then it will be a demonstrative, otherwise it is a personal pronoun. Similar is the case with relative demonstrative, [FL] or WH word demonstrative [FL] and so on. So, you see the difficult here is that, the categories pronoun and demonstrative share a lot of words. In other words, a word like [FL] can be both pronoun and demonstrative, therefore the question that arises is that, how can a machine distinguish between these two words. And it has to distinguish these two words for the categories from the context, from the symbols or the words which appear around them, so we discuss this with an example.

(Refer Slide Time: 23:25)



So, [FL] that boy sleeping, that is a clause that boy sleeping is, that boy is sleeping, so this [FL] is that, it is denoting a particular boy [FL] therefore, this is a demonstrative. And what will a machine do, the machine will see that, this [FL] is followed by noun and therefore, this is a demonstrative. But, this simple rule will fail in many cases for example, look at this [FL] cricket, [FL] he cricket playing, he cricket playing is. So now, look at this sentence, look at this sentence and these two sentences similar in terms of syntactic function, [FL] cricket, both nouns.

So, [FL] both verbs [FL] playing, sleeping, present tense continuous aspect, so how will a machine know that, this [FL] is demonstrative where, is this [FL] is a pronoun, how will it know. You can see that, these requires deep semantics, you can see that this [FL]

cannot be a qualifier for cricket, it cannot denote this noun, it cannot be demonstrated for this noun, because of deep semantic reasons. Therefore, a simple rule like noun following a demonstrative or pronoun, will tilt the balance in favor of noun, that you can say probabilistically.


But, this rule is not completely infallible then a simple minded rule could be, this [FL] is a demonstrative and not a pronoun. But, we see that, this rule fails in this example [FL] cricket [FL] this [FL] is not a qualifier for cricket and there are deep semantic reasons, why this was not a demonstrative for cricket. There are deep semantic reasons and those clues are not available to the machine at this level, at the level of POS tagging. So, this is the main challenge of part of speech tagging where, we know that different categories for a particular word cannot be obtained with complete certainty from a limited context.

So, if you look at the slide once again, there are these pronoun categories and these demonstratives and we find that, both these categories share many words between them. In fact, everything in demonstrative is also contained most of the words here and also contained in the pronoun category words and it requires large context and complicated processing to distinguish between the categories, identifying the correct tag. So, I think, I have impressed upon you this point that, the part of speech tagging within a category or across categories is often not a simple problem to solve, because of this kind of ambiguity.

(Refer Slide Time: 27:02)

Verb Tags (Examples in Hindi)

4	Verb			V	V	gaa gaa soaa haMaa ha rahaa	
4.1		Main		VM	V_VM	gaa gaa soaa haMaa	
04/01/01			Finite	VE	V_VM_VE		This subtype WILL NOT be used for Hindi as Hindi does not have enough information at the word level
04/01/02			Nonfinite	VBI	V_VM_VBI	--do--	
04/01/03			Infinitive	VBB	V_VM_VBB	--do--	
04/01/04			Gerund	VNG	V_VM_VNG	--do--	
		Auxiliary		VAUX	V_VAUX	ha raaha hoaa	



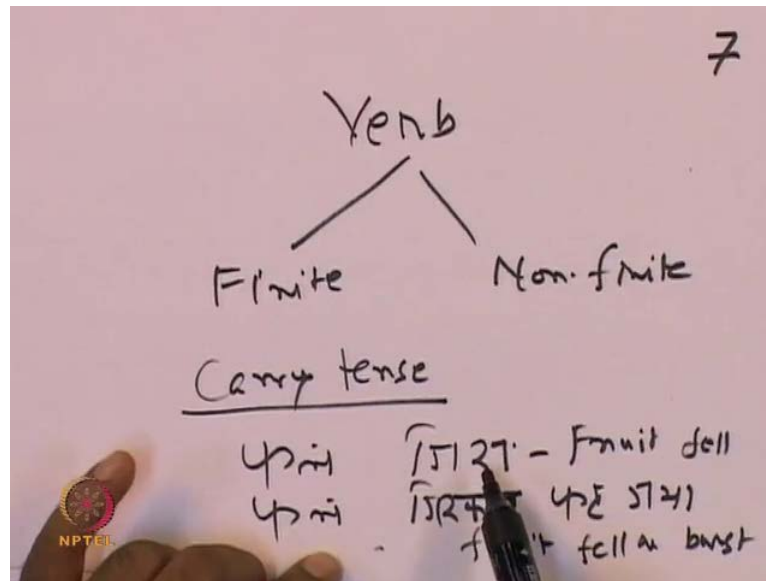
We proceed further, look at other tags, we come to the second most important category of words called verbs, verbs denote actions and the basic symbol for verbs is V. There are many examples here from Hindi like [FL], these are different verbs, [FL] is to fall, [FL] is to go, [FL] is to sleep, [FL] is from the root laugh, [FL] is an auxiliary, copular verb [FL] is to stay and so on. So, all these are verbal categories and under this, you see there are many many different categories, sub categories, but Hindi retains only a few.

There are again deep reasons for them, we will discuss them slowly, first is the category the most important category the main verb. So for example, [FL] a fruit fell from the tree, [FL] Ram school [FL] Ram went to the school, [FL] one should sleep at night, [FL] he is laughing. Now, there are other subtypes, which are shown to be cancelled I have purposefully shown them to be cancelled, but before we discuss them, we take the next most important category in Hindi, namely auxiliary verbs [FL], etcetera.

Auxiliary verbs are also known as helping verbs, they go with the main verb and produce some attributes on them, technically called they assign aspectual information and moved information sometimes on the main verb. So for example, I can say [FL] fruit fell from the tree, this is past tense, now I can make use of [FL] to change the tense. [FL], so this is the present continuous tense and showing an activity, which is taking place now, [FL] the fruit is falling from the tree.

So, I have made use of two auxiliaries here [FL] also similarly, can be an auxiliary, so these are helping verbs, which change some of the aspects of the main verb. Now, we come to these categories, which are shown to be cancelled, so finite verb, non finite verb, infinite verb and a gerundial verb. So, finite verbs are ordinary verbs, verbs the technical meaning of, the technical definition of finite verb is the following.

(Refer Slide Time: 30:10)



I will write it down, verb, finite and non finite, finite verbs the most important property of finite verbs is, they carry tense. So, if I say, [FL] this is carrying the tense information namely the past tense information in the form, so it is a finite verb. So, this is the finite form, but if I say, [FL] means fruit feel, [FL] is fruit feel and burst, so this [FL] does not carry the tense information, the time information is not containing this. So, time information is contained in [FL] which is the past tense activity.

And in fact, it is contained only in [FL], shows the activity took place in the past, [FL] does not show it. So, that is why, it is called a non finite form and [FL] is the finite form, it is carrying the tense information. (Refer Slide Time: 31:31) If we come to the categories once again, now VF is finite form, nonfinite is the form which does not carry tense, infinity form is also a kind of non finite form, gerundial verb is again nonfinite form, we will discuss them later.


For the moment, it suffices to say that, verbs have two different forms, finite and nonfinite. Now, why is it that, Hindi does not want to distinguish between finite and nonfinite form, it does not show different categories, by which marking can be placed for finiteness and non finiteness. Let us discuss this, we have to say that, the finite form and non finite form are important for carrying the text information. Now, in Hindi in general, from a small context of the words around it, it is not possible in general to distinguish between the non finite category and the finite category.

We will have to take these examples little later, the reason is that, we have to understand the properties of the verbs and their forms to be able to correctly identify finite and nonfinite categories. So, we will postpone this discussion for a moment and we will back to this when we discuss the properties of the verbs. We proceed further with the remark that, in Hindi we have only two major categories, main verb and auxiliary verb.

(Refer Slide Time: 33:18)

**Adjective, Adverb
and Conjunction Tags
(Examples in Hindi)**

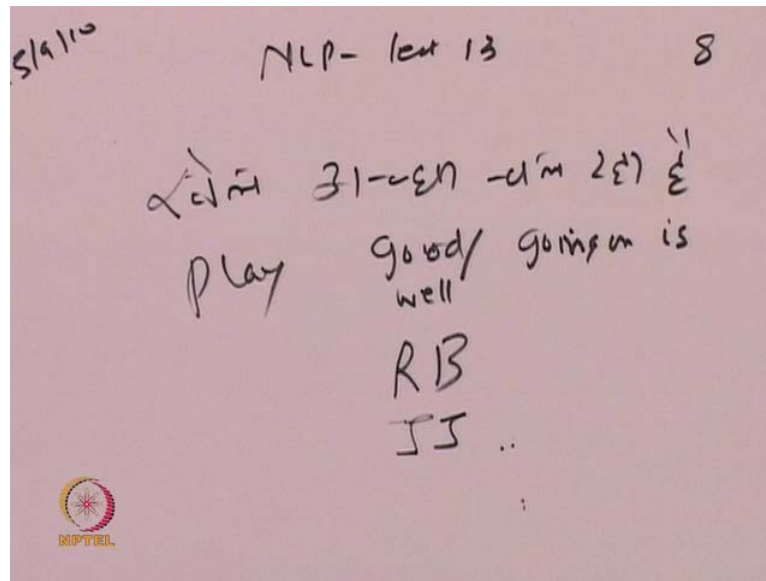
6	Adjective			JJ		sundara, acchaa, baRaa	
6	Adverb			RB		jaldi, teza	
7	Postposition			PSP		ne, ko, se, mein	
8	Conjunction			CC	CC	aur, agar, tathaa, kyonki	
8.1		Co-ordinator		CCD	CC__CCD	aur, balki, parantu	
8.2		Subordinator		CCS	CC__CCS	Agar, kyonki, to, ki	
			Quotative	UT	CC__CCS__UT	---	Not required

 NPTEL

Now, we come to relatively easier categories, namely adjectives, adverbs and conjunction tags. Again the examples are from Hindi, adjective is given the symbol JJ, this is inconsistency with the Penn tree bank tag. Penn tree bank tag is for English and this has become the most popular tag for producing annotation of English language corpora. The examples of adjective tags are [FL], [FL] is beautiful, [FL] is good, [FL] is big, so these are the adjectives which qualify nouns.

Next category is the adverb category, now here we have words like [FL] and [FL] which means quickly, both these words mean quickly, they indicate the manner in which an action is taken taking place. So, they qualify verb therefore, they are adverbs, but there is a problem here, which is a classical problem for adjective and adverb, adjectives can function as adverb.

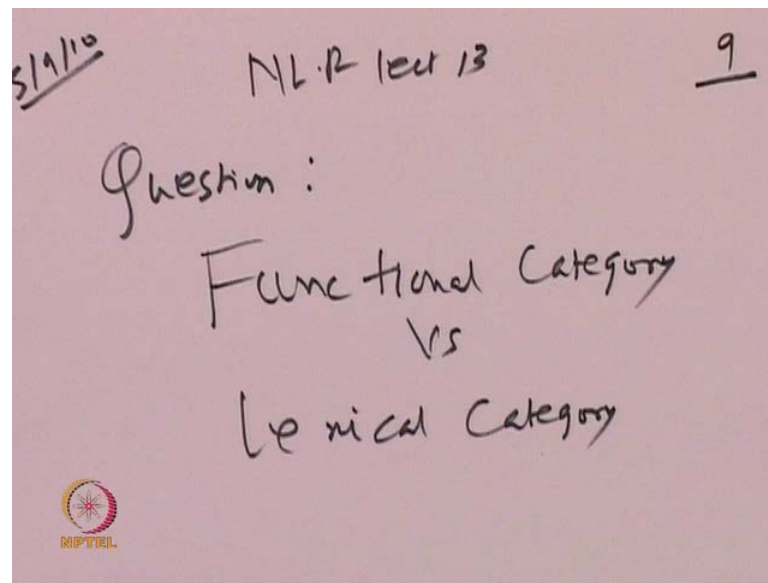
(Refer Slide Time: 34:42)



So, for example, [FL] if I take the sentence [FL], I will write it down [FL] play good going on is. The play is going on well or good, in this case this good or well is specifying the manner, in which our action is taking place therefore, it is an adverb. So, the correct tag for this is RB and if it is qualifying a noun then the correct tag is JJ. So here, we encountered another case of a difficult ambiguity, if a word is qualifying a noun then it is an adjective.

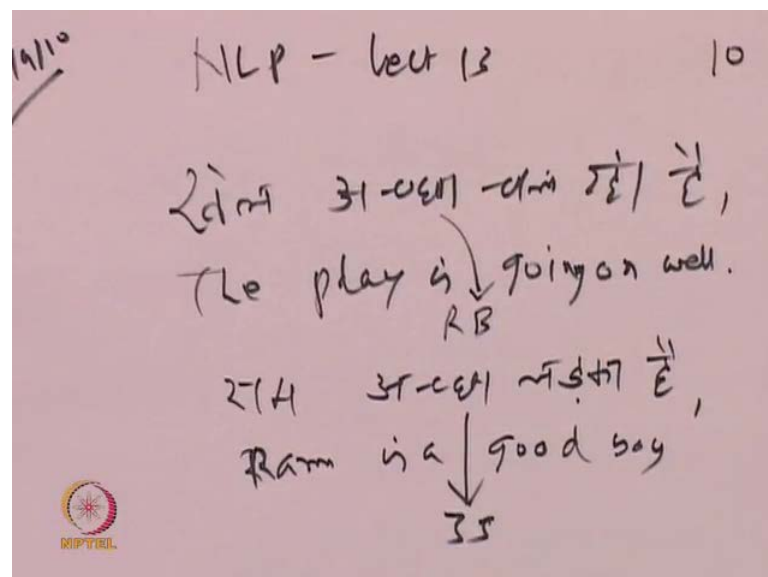
If the word is denoting the manner of an action then it is an adverb, but the problem is that, in Hindi and in many languages, the same word can function as both adjective and adverb. Now, it is nontrivial to identify, if there adjective or adverb is general, because the noun can be at a distance from the adjective [FL] and therefore, the adjective and the noun may or may not be adjacent to each other. Therefore, it may be difficult to immediately identify, whether this adjective is actually an adjective or an adverb. This brings us to an important discussion on part of speech tagging, let us write this down and it is a very fundamental point in part of speech tagging. This is so important that, we would like to spend some time understanding this issue.

(Refer Slide Time: 36:36)



The problem is, good important question, functional category versus lexical category, what is the functional category of a word and what is it is lexical category. Let us understand this point, which often lends to complications in part of speech tagging.

(Refer Slide Time: 37:07)



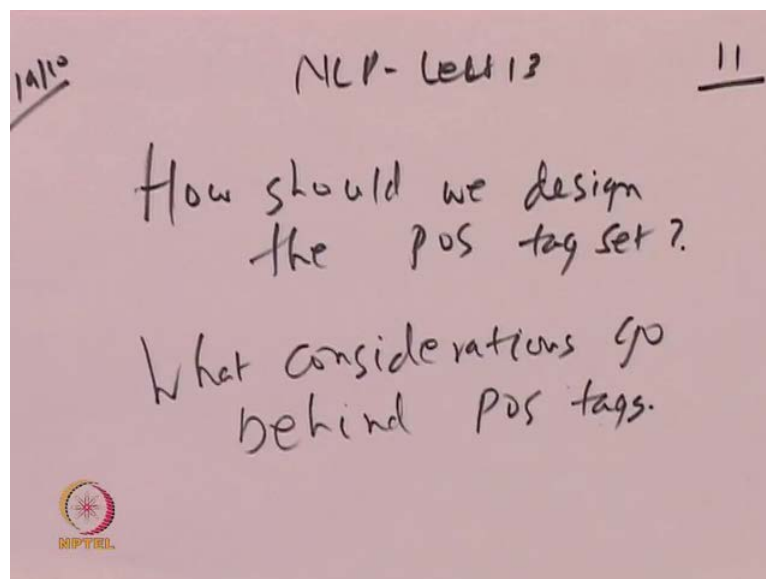
We took this example of [FL] the play is going on well and this sentence, Ram [FL], Ram which is a common Indian name, Ram is a good boy. In this case, [FL] qualifies [FL], which is boy therefore, it is an adjective, we will call give it that tag JJ. This [FL] is a manner for the verb going on, so this is the tag RB adverb, so when we will look at

[FL] in isolation, we do not see it in context the neighboring words. we just look at the word in isolation. What is the category that comes to our mind when you look at [FL], when you look at [FL], the first impression is that, here is an adjective.

And when [FL] is recorded in the dictionary, when it appears in the lexicon, the first category that will be mentioned for it in the dictionary will be adjective. So, this is known as the lexical category, some dictionaries will also mention this as adverb. Many dictionaries actually will mention this as adverb also, but the first and foremost category for the word is adjective, this is known as the lexical category. When it plays a particular role in the sentence, it may or may not retain its lexical category.

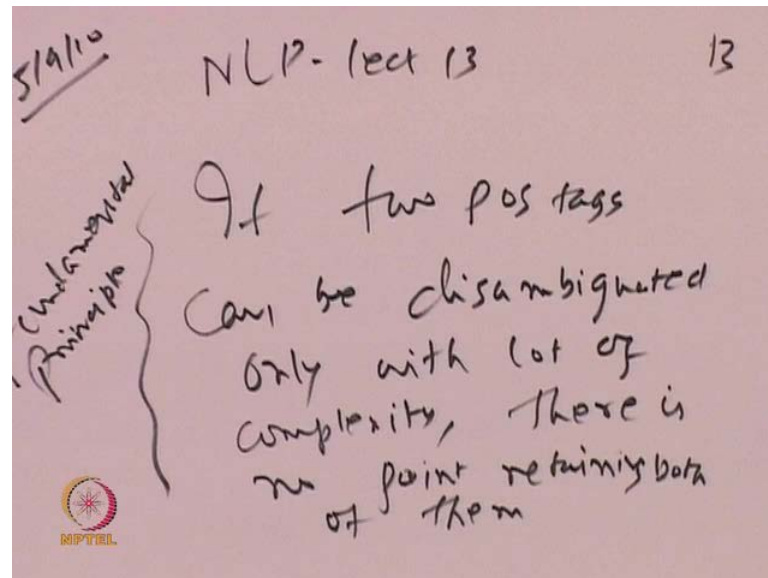
We took this example here, [FL] is adjective, but in this particular sentence, it is functioning as an adverb. So here, [FL] is functioning as adverb, though its predominant lexical category is adjective. So, this kind of adjective, adverb ambiguity is very common in a natural language and the machine has to resolve from the contextual words. Many times, these kind of function versus lexical category, this ambiguity is not possible in a limited context and then there is no point in having two separate tags. So, this is an important point, so we write down one of the considerations for part of speech tags.

(Refer Slide Time: 39:53)



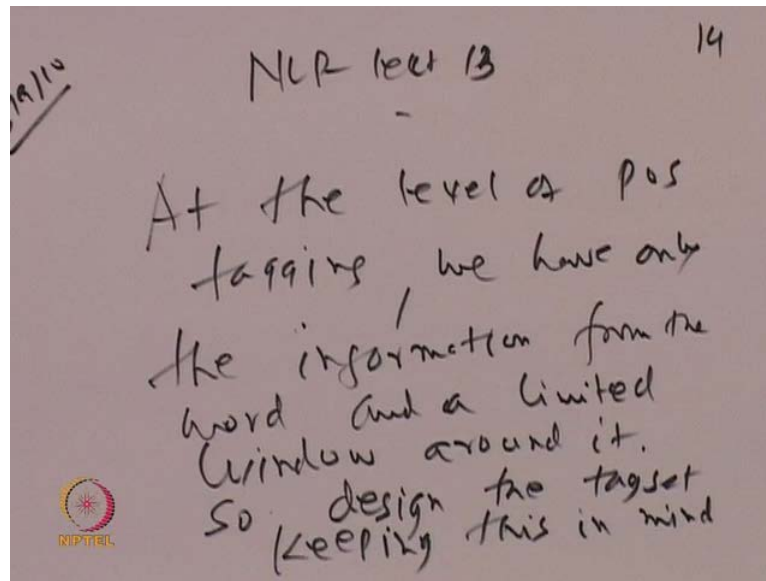
The question we were asking is, how should we design the POS tag set, what considerations go behind POS tags. How should one design the POS tags, what considerations go behind them.

(Refer Slide Time: 40:27)



Clearly you can see that, if two POS tags can be disambiguated only with lot of complexity, there is no point retaining both of them. This is a very very fundamental principle of part of speech tag design, we call it fundamental principle. I read it once again, if two POS tags can be disambiguated only with lot of complexity, there is no point with any both of them. So, what we are saying is that, at the part of speech tag level, this is the second principle.

(Refer Slide Time: 41:29)

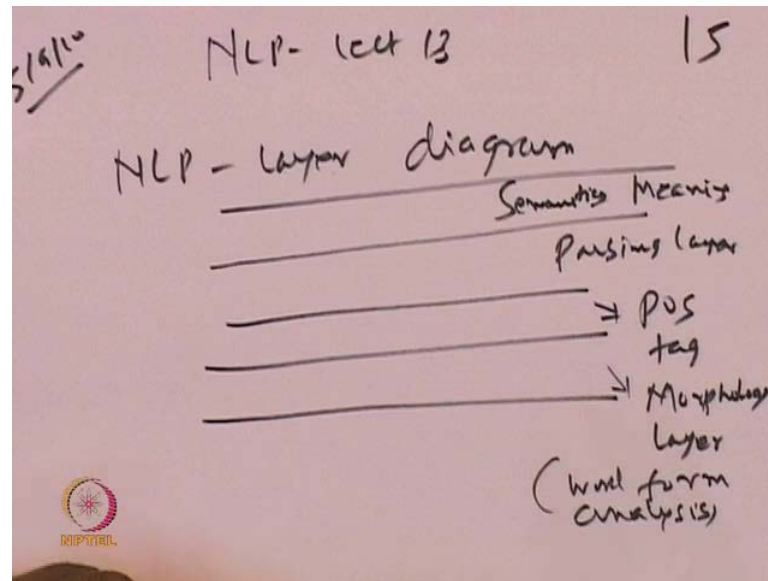


At the level of POS tagging, we have only the information from the word and a limited window around it, so design the tag set keeping this in mind. So, if we have two POS tags, t_1 and t_2 and in general, t_1 and t_2 cannot be distinguished only by looking at the word or from a small window around the word then there is just no point keeping those POS tags, because at the level of POS tagging, you will not be able to distinguish one from the other.

So, the example that we took of demonstrative at verb pronoun, for some words it is very difficult to decide, whether they are demonstrative or pronoun, just from the word or from a limited window around the word, it is very difficult. So, at the level of part of speech tagging, there is no point giving those words or keeping for those words two tags. And these two tags will only be there for idealistic reasons, they are not practical, practically they are impossible to distinguish from each other.

So, these consideration goes into parts of speech tag design, so the two principles that we have talked about are, at the level of POS tagging we can only distinguish from the word and a limited context. And the other principle is that, the two POS tags can be distinguished only with lot of complexity then there is no point keeping both of them. So, a diagram is helpful at this point of time, which I will draw sketchily.

(Refer Slide Time: 43:49)



This is a NLP layer diagram, part of speech tag is a layer which is placed on the text and it is supported by the morphology layer which means, word form. Morphology layer is word form analysis, above that is the part of speech tag layer, above that is the parsing layer. So, part of speech tag layer can only make use of information below it or at the same level, it cannot make use of information which is above it like parsing and meaning semantics. So, the part of speech tag has to be designed keep this point in mind that, it has to make use of words and a limited window around the word. We will take up these important issues in the next lecture.