Fundamentals of language Acquisition

Prof. BIDISHA SOM

Dept. of HSS

IIT Guwahati

Week 05

Lecture 022

Lec 22: Reference problem and extension problem

Welcome back. Today, we will have the second lecture of this module, Module 5, where we are looking at the word learning process and the theories with respect to word learning. In the first lecture, we looked at how the phonological word form is mapped onto the semantic part of the word, the meaning mapping, and the various types of theories and ideas that have been put forward. Today we will talk about two specific kinds of problems that the children are understood to be facing: one is called the reference problem, and the other is called the extension problem. So, the reference problem is a very simplistic kind of problem, simplistic in nature. Quine talked about the radical translation problem, and this is where we take the idea of the reference problem in language acquisition.

So, what did Quine say? He talked about a person; his idea pertained to someone translating a previously unknown, radically different language. Now, how would he go about it? He would probably start by asking what word the natives used for a certain event. So, it gives the example of a white rabbit scurrying by, and the native says, "gava gai." The natives gave only one word, and this is what happened.

There is an incident; there is a native output, native as in a person who is a native speaker of the language. So, they said, "Gavagai." Now, what does it really refer to? This is the scene, but to which aspect of the scene is the word "gavagai" referring? Is it the rabbit? White? Scurrying? Animal? Is he seeing the event? There are so many aspects of the same scene. So, which aspect of that entire scene is the word referring to? This is the reference problem. Even when children are learning words, they have understood a word; they have identified a word from the speech stream.

How do they pinpoint which aspect of the referent it is referring to? What is it talking about? So, our problem statement is that a novel word may have an unlimited number of potential references. Let us try to understand this problem; it might not look like much of a problem if you know the language, but if you do not, then you will see what happens. So, to make things easier, I have quoted from a very well-known book by Daniel Everett, "How Language Began: The Story of Humanity's Greatest Invention." The book's name is "Do Not Sleep, There Are Snakes," in which Daniel Everett talks about his work among the Pirahãs in the Amazon, where he spent decades learning the language and studying the people. So, in one of his initial encounters with the Pirahã people, during which he is trying to get the words and sentences of the language, he is armed with his pen and paper.

In doing that, he does this. He, Daniel, drops a stick on the ground. The Pirahã man says this. I am not sure how to pronounce this, so this is what the Pirahã sentence is. Now what Daniel thinks is that because he dropped the stick on the ground, he thought what the native said probably meant 'you dropped one stick' or 'the stick falls to the ground'.

It was much later that he found out, when he was more familiar with the language, that it meant 'stick it ground falls.' This is how the word-for-word translation will go: "stick it to the ground falls." Similarly, when he was at this stage, it was the beginning. So, he dropped one stick; that was the output. Then again, he dropped another stick.

So, he drops two sticks to the ground, and then there is another sentence that comes from the Pirahã person, and now Daniel thinks that this means "two sticks fall to the ground." However, what the native actually said was, "A slightly larger quantity of stick falls to the ground." So, this is the reference problem that Quine was talking about. Now, the reason I have taken the example of this particular language, Pirahã, is that it does not have much similarity to languages like English, French, or any other languages that we know. For example, this language does not have any words for hello, goodbye, thank you, I am sorry, you are welcome, and so on.

Some courtesy-related words that we are very familiar with and that we almost consider fundamental things that are very common must be present in the language, but this language simply does not have these expressions. For example, again quoting Daniel Everett, when a Pirahã arrives in a village from one village to another, he might say, "I have arrived," but no one says anything in reply; forget about greeting him or doing any of those things that we typically associate with receiving a guest; nobody would even reply. Similarly, if one gives something to someone, some kind of gift, no "thank you" is used. Sometimes they say something that literally means, "Transaction acknowledged." That is what Daniel's translation is.

So, the act might be reciprocated later in kind, but never in verbal terms. So, they will never say thank you, and they will never say you're welcome, and so on. Something similar happens to the child too, but in the child's case, it is a more difficult problem. A child learning his first language, first words, and their meanings has a similar but complicated problem because there are no pre-existing languages. So, for Daniel, after he has collected enough sentences, he can go over them one by one, word by word, verify them, and then come back to their meaning.

For the child, there is no first language to come back to; there is nothing; it is a blank slate. So, when he or she hears their mother say that it is a dog, the meaning is not exactly clear; we think it is clear that there is a dog and that they are calling it a dog. From a child's perspective, there may be many possible meanings. The dog of that referent and the child have a very limited understanding of the language to find their way out of this. Also, linguists know that there might be words for objects like rabbits in a language, and they look for them.

That is why Daniel was trying to see what you would say when I dropped one stick on the ground and then two sticks on the ground. So, he was expecting to get the numerals; probably you know the numerals are 1, 2, 3, and 4. How do you count in this language? So, we know as linguists or as normal people that, normally, a layperson also knows that certain things might have a word in that language. So, we look for it: children who do not have that kind of, you know, background information. So, that is another problem.

For example, for the children, the word Gavagai might even mean that we do not have any word for that animal in our language. This entire thing might also be its meaning. So, you show, for example, that when you go to some natural forests or encounter an animal for the first time, you have no idea what it is called because you have never seen it before. You might just say something. And if it is a language that you do not know, this could be one of the meanings of the sentence, right? That is the kind of problem that a child faces.

Thus, the child may have in mind a limitless array of objects, relations, and states as possible referents for the word and, of course, their possible linkage to the conceptualization. So, for example, take the words "toy mouse." Now, which aspect of it should the word refer to: mouse, mammal, animal, toy, wooden object, cat prey, furry things, and so on? So, there are many things that the conceptualization aspect might map onto, you know? The word, the referent, and the conceptualization; there is a three-way transaction happening here. So, what is it that the child will finally pick up? So, ambiguity is present even under the simplifying assumption that the intended referent is available. When it is not available, things only become more difficult.

But even if it is available, this is the kind of problem that you might encounter. Worst case scenario is learning that word meaning mapping is difficult for concrete words, but imagine the abstract words. How difficult it is to talk about an abstract word! In linguistics, we have this issue in cognitive linguistics; for example, we have something called categorization, and we study categorization. How you categorize an abstract concept in your language might not be the same as how you categorize the same concept in another language, even if they are similar languages. And that is when we talk about first and second languages.

Just to give a small example, the word 'abhiman' in Hindi means pride—some sort of, you know, some kind of pride; when you consider yourself important enough, that is 'abhiman'. But another language that is very closely related to Hindi, and that you know, traces its history to the same ancestor language as Bangla. Bangla also has the same word "abhiman," but it is not the same word because the meanings are not exactly alike. So, abhiman does not mean pride in Bangla; it means something very different. So, this is something that we are talking about regarding cross-language mapping.

So, it is very difficult for abstract notions to be learned and understood; from a child's perspective, it is even more difficult. So, these are problems that we think exist and that we know exist, but do children really face these problems? Does it hamper their word-learning capacity? It does not seem so. So, despite all the problems, children actually learn words impressively fast. So fast that Pinker once called children "lexical vacuum cleaners"; that is how fast they learn and acquire words. Because after 1 year and between 1 year and 2 years, it depends, of course, on the children; some children learn it faster, but they learn a lot of words in that phase.

So, one important study in 2012 tested whether 6- to 9-month-old infants understand words denoting body parts and food items, because these are the things that are most commonly talked about in a child's environment. They studied this using eye-tracking as a method. The basic premise was that people tend to look at objects whose names they hear. You hear our name, and then immediately try to look for that object in the environment. If it is there, you fix it here.

That is the idea. So, the infants looked at a computer screen displaying objects while their mother would say, for example, "Look at the apple." The child looked at the apple. So, this kind of study that they carried out had many predecessors; we have also discussed this particular one in the subjects that looked at the intended object more times than could be attributed to chance. So, children keep moving their heads. Some of the orientations might be attributed to chance; they just happened to look that way, but in this case, they looked at the intended objects more often than they could have.

So, that means they could map the name onto the referent, and we are talking about 6- to 9-month-old children. Another study reported a slightly older study from 1994 that found the average American child, a 12-month-old, understands 50 words. A one-year-old American English-speaking child understands 50 words. Thus, it is not surprising that children understand a large number of words by the time they are 2 years old, and, of course, they keep adding to their vocabulary. So, there is a slight difference between production and comprehension vocabulary because comprehension happens before production.

So, the number of words a child is able to produce will always be less than the number of words they understand at a particular stage. So, if they are able to produce 50 words, they probably know more than that. If they comprehend 50, they are probably able to speak less than that. That is the mismatch that is typically found in all children.

But eventually, they found out. And then, of course, there is this idea of fast mapping that we already talked about. Here, I have quoted the chromium experiment. This is one of the oldest experiments, dating back to 1978. Now, one of the theories put forward to address this problem is what is called cross-situational statistics. The proposal of the idea is that even though there are an unlimited number of possible references for a label in any given situation and across situations, So, one situation may also have different kinds of output, but across situations, there are, of course, lots of variations.

However, the referential uncertainty will vary. This is an important concept that this particular theory addresses regarding referential uncertainty. So, referential uncertainty basically means that the learners—that is, the children we are talking about, infants—need to track the invariance; that is, which aspect remains constant across scenarios. If they can track that there are variations and certain things that are comparatively invariant, that is the particular thing they need to track. That is what cross-situational statistics is.

So, how many times does this appear in an invariant format, and how many times does this vary? So, if they can track the invariance, then they can link the label to the invariant entity as well. So, the potential importance of this possibility has also been studied by a number of researchers. I have added them. This is a more recent theory. So, there have been many studies that took place in the 2000s and slightly before.

That was the reference problem. So, how do they map the word form to the meaning, what problems do they probably face, and so on? So now we have the problem of extension. So, it's fine once they have found out that the dog refers to the pet dog; now this is the reference problem. So, the child understands that this is the referent for the word "dog." So, let us say now that the reference problem we have seen is that the child has to map the word form

onto the referent. Now, once we have solved that problem, how do we solve it? The child figures out that the word "dog" refers to her pet dog.

Now that solves the basic problems. Now, the word "dog" does not refer to a single instance of dogs. That means the word "dog" does not refer only to the child's pet. It refers to all the other dogs in the world. Now, this is the extension problem. So, it also refers to the Labrador and other breeds, you know, Chihuahua and all of that.

How does the child learn to extend the meaning of the category? Because the word "dog" is not about a particular dog, but rather a category name than an entity name. Of course, the entity is also a dog, but the child has to understand that this particular word and its meaning extend to all the other members of the same category: dogs. How does the child do that? For example, they also have to remember where to stop while they are at it. So, for example, a dog cannot be used for catching cats and rabbits. So, you know about other animals that might look similar.

So, a small white dog and a rabbit might look similar, but they do not refer to the same thing; they do not mean the same thing. So it cannot be extended to rabbits. nor with things that are thematically associated with dogs. So, the dog's leash and a bone are things that are thematically related, but again, they are not part of the category "dog.

" So, they have to stop there. Thus, in some sense, the child needs to create an understanding of a category. When figuring out that "dog" refers to all animals in the dog category and not to anything beyond. So, when do the children figure out the category information in their childhood, and when are they able to label the words in the category rather than beyond the category? Similarly, this is not only about specific objects or animals or something; it is also applicable to grammatical categories. For example, in the English language, you have the preposition 'in'; Indian languages do not have prepositions; we have postpositions, but either way, they refer to something: they refer to a particular arrangement of objects in space. So, in English, for example, 'in' is applicable when the toy is in the basket, the butter is in the fridge, and so on.

However, when we use a sentence in English like "The ball is under the cup," it is a very different kind of concept. So, the child who is growing up with an English-speaking child has to figure out which configuration of the objects can be labeled "in" and which can be labeled "under." So, this is another kind of difficulty. Now, what makes this more difficult is that these kinds of categories are not exactly universal; words do not map onto the categories in the same way across languages. So, I just gave an example: I have given an example of the word 'abhiman'; that could be one example of this kind of different mapping between words and concepts.

Even when the words sound similar, they are both emotion words; still, they do not mean the same thing, but something similar. There are languages in the world that have completely different kinds of words, and even the realm of mapping does not match. So, one cannot depend on any universal innate knowledge to help one. So, this means that a child needs to learn to extend words in different ways based on the language-specific preferences. So, here we will not be able to fall back on some sort of basic, innate knowledge.

An important study in this regard showed how English and Korean differ in describing the location of an object in relation to other objects that are in or under a kind of configuration. So, for example, in English, the distinction between "in" and "on" is important; it shows the difference between containment and support. So, you can keep the flower vase on the table and the book on the table, but you can put the food in the fridge or, you know, all those other kinds of examples that we have seen. However, in Korean, there are no words for "in" or "on." This language does not distinguish between containment and support the way   English   does.

So, the idea here is completely different. The conceptualization is completely different. In this language, they distinguish between tight fit and loose fit; it does not matter whether it is in or on; that kind of understanding is missing in this language. So, there are all these different types of concepts that they have. So, "kkita" in Korean, for example, is used to describe objects in a tight-fitting condition, like putting a ring on a finger.

So, that is where Kkita will be used. Nohta, for example, is like putting a cup on a table on a horizontal surface, and similarly, they have three more categories. So, you see how the difference works, but they conceptualize the relationship between two objects by breaking it down into five different categories. So, this is an example of the overlap: this is the study that I am talking about, a very well-known study. So, I have quoted the page number;   you   can   also   just   go   check.

This is how they talked. So, in the case of English, this is the domain of "in," and this is the domain of "on." So, input rings in the basket, put Legos in the box, put books in the box, and so on and so forth. But one will put rings on the pole and Legos on one another. Only in this particular domain, which is the domain of "kkita" in Korean, is it relevant. So, which has nothing to do with either in or on, but is some sort of fit; which in this case is the   tight   fit   condition.

So, when you put pegs in a hole, they have to be tightly fitted; boxes in books and the covers will have to be a tight fit, and so on. So, these are the language-specific differences

in concepts and with respect to word meaning that we are talking about. Now, within the extension problem, we have something called overextension. Now, there are many instances of overextension that children typically exhibit. Overextension means using the word far beyond the members of a different category or even within the same category; it involves extending the meaning beyond what it is supposed to mean.

So, dog is for all four-legged animals. So, they are going from category dog to category animal; sometimes they use it. So, that is a kind of overextension we are talking about. Similarly, they use something called diffuse syncretic overextension; for example, something called a chain. Complex words like 'quah' might be applied to a duck in a pond, starting with, you know, the word 'quack.' So, it could probably mean a duck in a pond, but it could also mean water; it could mean milk; it could mean a picture; and so on.

This is the study that discovered this. Then, there are also other kinds of overextensions. So, there is a developmental sequence of over-extension. For example, early idiosyncratically, there were loosely associated complexes like quah, then correct generalizations and later categorical overextensions. So, these are the kinds of differences that we find in a sequence. Similarly, the associative complex, rather than the chain complex,   is   characterized.

And also overextension based on associations through contiguity. So, daddy for Father's shoes, and so on. In this domain, one of the most well-known studies is Rescorla's 1980 study, in which he distinguishes between types of overextensions and categories of overextension. One of them is called "Analogical." This is where there is a perceived similarity between the actual reference and the current reference of the word, such as "comb" for "centipede." Now, you can easily imagine how a centipede might look like a comb to a child because there is no similarity in the real world, but there is a perceived similarity since the teeth of the comb might look like the legs of the centipede.

In his study, he found that the variations of the word "ball" applied to a range of round objects. So, various kinds of round objects are called balls. So, he categorizes them; one is the close relation of two balls, which are a balloon and a marble, that are very closely related. Then comes the clear member of another conceptual category.

For example, apples and eggs are also called balls. And then, more ambiguous cases like wool pom-poms, spherical water tanks, and everything else were called "balls." So, this is what he calls analogical. So, because of the underlying perceived similarity across categories and objects, that is where children tend to overextend. Another kind of extension that he talked about is the categorical extension. The use of a word to label a referent is close to the standard referent in some clear higher order of the taxonomy.

For example, the word "dada" is used for men generally among family members. So, they can probably start saying "dada" to their father and then extend it to all the other members; you know there is some sort of similarity within the same category. So, all the male members of the family will be called, you know, across family boundaries as well. So, all the men's items of that category will be called 'dada'; this is called categorical extension. Apple will include oranges as well.

So, the distinction between apples and oranges comes much later. Hat for many similar things, like crowns and so on. So, they are all in the same category. And then there is something called a predicate statement. A predicate statement is when the use of words conveys information about the relationship between the immediate referent and some kind of absent person, object, location, property, or state. So, starting with one word, they actually try to convey the entire sentence, the entire scenario.

So, it is a one-word-sentence kind of thing. So, "doll" means the usual location of the doll in the crib, "key" means a door, and so on. So, these utterances are basically like one-word sentences. Now, why do children overextend? There are various kinds of theories, as we have all just seen, that find some kind of perceptual similarity, or, you know, some other kind of similarity; these have been put together here. So, there are many that have been put forward; perceptual, of course, is the most common similarity that they find. Then there are functional similarities, affective response action schemes, and situational context, which is another reason that has been given for overextension.

Also, it could be a combinatorial utterance; a combinatorial utterance is when they say "Daddy" to mean "Daddy's shoes." So, where a relational or syntactic meaning is basically conveyed by a single word, So, these are the things that children typically do. Now, all of these extension problems and types of overextensions that we talked about are in the domain of word production. So, when children speak, these are the kinds of so-called errors that they make and through which we are trying to understand how they map words onto their meanings and conceptualizations. But overextension can also happen in the case of comprehension, which is a little difficult to identify and is less common.

So, there are important behavioral differences in terms of overextensional production and comprehension. So, children often overextend in production even when they correctly infer the appropriate adult words in comprehension. So, even when they are not making any errors in comprehension, they might still overextend while producing the same word. So, for example, identifying the correct referent upon hearing the word strawberry, overextending the word to apple, and extending the word apple to refer to the strawberries in production is important. So, when you show them strawberries and they understand

strawberries, they can link it in terms of comprehension, but when they have to produce a word, when they have to say it, then they overextend the word "apple" to include strawberries.

So, this kind of mismatch is also quite common between comprehension and production regarding overextensions. The asymmetry is basically due to the difference between production and comprehension, which is very common not just in terms of overextension, but also in terms of all other kinds of linguistic capacities and categories. Because typically, the general trend is that comprehension precedes production. Comprehension is always better than production in the initial stages of life, not only for typically developing children but also for atypical children.

Children tend to do better in comprehension tasks than in production tasks. So, that is probably the reason we see the asymmetry. So, it is nothing very unusual. So, this has already been taken care of. This could be one reason. And there are other reasons that have been put forward: one of them is that comprehension and production rely on two separate systems.

So, some have said that there might be two separate systems in the mind: one taking care of comprehension and the other of production; others say it is a single system, and so on. So, these are some of the reasons for the asymmetry. Other views are also present; for example, the incomplete conceptual system, the pragmatic choice, and, of course, retrieval error. An incomplete conceptual system is easily understandable, as children, when they are learning to speak, are still going through very important changes; they are in a crucial phase in terms of their growth: conceptual, physiological, and cognitive growth.

So, that is when we are; that is the time window we are looking at. So, because they are not yet there, they are still growing. So, probably that is the reason why they overextend words: because developing conceptual systems cannot yet distinguish concepts to the extent that adults do. Another thing is that a pragmatic choice is made because they have a very limited vocabulary. So, they are doing the best they can. They are optimizing the vocabulary they have to extend it to other possible candidates as well.

This is also understandable. And then, retrieval error is another thing that has been put forward. This is a performance error caused by the cognitive effort of retrieving unfamiliar words. Because unfamiliar words are more difficult to retrieve, familiar words are easier to retrieve.

So, they use and overextend familiar words. So, these are some of the views. So, an incomplete conceptual system theory, however, can explain only the problem of

overextension; but it cannot really explain the asymmetry between production and comprehension. If the system is still not fully formed, then the conceptual system is not yet fully formed, and it should reflect similarly in terms of comprehension and in terms of production. That does not happen. What happens is that comprehension always receives higher marks compared to production, so to speak.

So, that is why this theory does not really address it. Also, these theories do not propose a formal model to explain the conceptual leaps that children make when they overextend. So, this is not a very simple thing. They are actually making conceptual leaps. So, of late, there have been some proposals, and some computational models have also been proposed. One of them was proposed recently by Xu and Pinto Junior (2021); they discuss a general computational model that could account for all of these overextensions and other kinds of inconsistencies.

So, along with overextension, there is also the concept of underextension, which is, of course, less common, but it is also present. It is basically defined as the use of lexical items in an overtly restrictive manner. So, for example, you know that referring to the dog as only one's own pet dog and not to other dogs.

So, the dog on the street is not a dog; only the dog at home is a dog. So, that is under extension. So, we will find similarly similar examples in many other locations. So, "kitty" refers only to the pet cat. So, O'Grady says that underextension errors reflect children's natural tendency to focus on the prototypical member of a category. For children, this will be reflected in the members that are around them in their environment, you know, in the environment in which they live that might be considered a prototypical member of the category.

So, that is why they stick to that underextension. Similarly, other theories have also been put forward. For example, this refers to children's dependence on their cognitive abilities and their adjustment to the conversational situation. So, there are many such theories. So, overextension and underextension are the two kinds within the extension problem. So, we have seen two kinds of problems, and this is how we can summarize them so far. To learn the meaning of new words, children need to isolate the correct reference for a word, which is the reference problem, and then they need to learn to extend it appropriately for their own language, which might be different from other languages, while avoiding both over-extension and under-extension.

So, this is how we take care of the reference problem and the extension problem in our

course. In the next segment, we will talk about other theories, other problems, and other theories. Thank you.