

## **Fundamentals of language Acquisition**

**Prof. BIDISHA SOM**

**Dept. of HSS**

**IIT Guwahati**

**Week 04**

**Lecture 018**

Lec 18: constructivism

Welcome back. We will start with Lecture 3 today. Within lecture 2, we looked at the variational learning model and discussed what the model basically means and how it works. Now let us look at some of the evidence in favor of the model and then some evidence that does not seem to go along with the model, and then we will look at constructivism as a possible alternative to this theory. So, with VLM, we are still with the nativist perspective. Now, the very VLM primarily posits that we all do have parameters, we do have principles, and we do have parameters, but the usage of those parameters will depend on a probabilistic occurrence.

So, depending on how often a particular format or type of input the person or child gets, that format will be rewarded accordingly. So, let us say the tense marking, plus tense marking, minus tense marking; these two are the on and off conditions of the tense marking parameter. Now, if the child hears, the child switches on the plus tense marking, and then continuously hears plus tense marked sentences, then that format gets rewarded because this fits the analysis of the parameter that has been set. However, if the parameter is set to on and the child listens more often, he or she listens to the sentences that have tense marking, then it will be kind of punished.

So, it will not be used anymore. So, this is how the child is supposedly punished. The child supposedly understands what the parameters are and how to set them. So, it is not a fixed one; it depends on the kind of input the child receives. So, the probability of one type over another is the primary takeaway from this model.

It is a simplified version. Now, does this theory address any of the existing problems? It seems so. Remember, we talked about the co-occurrence of errors and inflection at the same time. So, if going by the nativist theory, if the child has figured out the rule of the parameter of the language that this will be a tense-marked language, then there should not be any option for any error. So it is like either one or the other; you either get the rule or you do not get the rule, but you cannot get the rule and still make mistakes.

So that is where VLM seems to have an answer for that and an explanation for it, because this theory explains the question of the co-occurrence of both inflected versions as well as the errors in this way. The child is initially undecided because both options are readily available to the child: both the plus and the minus tense marking, let us say, or plus or minus argument marking, or whichever way any of the parameters we can take, and the child has both options in the beginning. Now, as the child hears the plus tense structure, that grammar gets reinforced and then the minus tense grammar gets pushed back or, let us say, it is called punished. So, it is punished, the minus tense format is punished. So, over time the child stops using the past tense grammar and sticks to the present tense grammar.

So, that is why initially, if you have set one format, you can see there are problems even though they get it. So, they are in the process of understanding; hence, sometimes they get it correct, and sometimes they do not, so they are undecided. It is kind of a yo-yoing between plus and minus and how the child finally fixes on one. So, that is the time period when we see the co-occurrence of both the errors and the correctly inflected structures. So, that is also why the child gets better over time.

This goes back to our maturation theory within this entire framework. So, while initially in the beginning stages, they do have a problem, but over a period of time, through maturation, they do get better. Similarly, there is another problem of delay in error correction. So, this can also be explained by this theory. So, this depends on the ambiguity and the nature of the input.

This theory is dependent on the child checking the underlying grammar with the input. So, they have the parameters, and they need to check them with the kind of sentences that they come across and then accordingly tick one option or the other. So, now if the problem is that the input grammar and input sentences that the child is exposed to have any ambiguity, then this will create a problem in error correction. So if there is no clear signal that the child receives, then there will be a delay. So the more ambiguity there is in the speech input, the more delays the child will experience.

So that is how this model tries to understand the problem. If the speech input is ambiguous, then there will be a delay. So, it may not always give a clear indication of the correct grammar overtly; not always are we mindful of the grammatical correctness or grammaticality of the sentences, or even the fact that there is a child in the environment who is probably picking up that this is not how. We are not conscious that when adults speak, they are not really aware of whether everything in the environment is input for the child. So, this might create a problem.

So, they give the given example like, "I like you," and you can just change this sweet with anything else. So, I like X, and you also want to have sentences like "I want to like sweets but cannot." In case I have diabetes. Now, the use of the verb in two different forms is present, though without overt marking. So, in this case, I want both: I like this and that.

So, both are in an unmarked condition in the sense that there is no overt marking on the verbs. So, like in the first case, I like sweets. So, there is no marking after as it will happen for example, if third person likes. So, do you know somebody else who Ananya likes? So, in that case, there will be an "S" marked. That is what we talked about in the previous lecture.

So, that S in the English language stands for many meanings. So, it has many things packed into one. But in the case of the first-person singular, the form remains as it is. Similarly, I want to like sweets, so in that case, it is the infinitive form of the verb; here, it also has no marker. So, these kinds of inputs probably create confusion in the child's mind as to which one is the tensed form, which one is the non-tensed form, which one is the infinitive form, and which one is the tensed form, which is not very clear because the overt marking is not present.

As a result, there may be a delay. So, this problem also has some kind of explanation in the VLM format. Similarly, you have cross-linguistic differences in the error array that were also discussed in the previous lectures, as languages do not seem to follow the same rules. So, if the child has made more errors in language X, it does not mean they will make the same kind of errors in the same category in language Y. Cross-linguistically, there are differences in the kinds of errors that the child makes.

Now, how does that happen? VLM seems to have an answer for that as well. So, because languages differ in terms of ambiguities, we talked about ambiguity in the English language, but not all languages allow that kind of format. There are languages where, for each instance of the verb, there has to be an overt marker; there has to be a marker for the tense or the agreement, whatever it is, that must be present. So, that means there are some

languages that are more ambiguous, some languages that are less ambiguous, and there are more that are in between.

So, it is a spectrum. So, depending on where in that spectrum the language lies, it will determine the kinds of errors and how long the errors last in that language. So, there is the same understanding of ambiguity in the sense of a lack of overt marking of morphological features. That is what we are talking about: the kind of ambiguity. We are not talking about semantic ambiguities here. So, for example, the Spanish children converge on the past tense grammar quickly because, in their case, you do not see the kind of delay that you see in, let us say, English children.

Why does it happen? Because Spanish utterances are unambiguously marked for tense, so there is no confusion. They do mark. As a result, you see Spanish children learning this distinction very quickly, and they converge on it in this language. I have to switch to the perfect tense structure; there is no confusion about it.

So, that is how it is. French has a moderate amount of ambiguity; hence, French children take a moderate amount of time to learn. So, there is a neat format within which the language structure, the amount of ambiguity, and the delay that the children show seem to indicate some sort of equivalence. So, this is fine, too. Even individual differences among children have been explained using this model. So now we see differences in terms of language.

Now we have seen differences in terms of language. Now we see differences even within the same language, with different children having different paces of learning. Not every child attains the same level of proficiency or mastery of language at exactly the same time. So, this has also been explained by the same theory that the parameter here must be the children's home language. If the home language is less ambiguous, if they have the kind of input that is more beneficial for children to learn those structures, and if they are more in the home environment, then they will learn it faster; that is how the correlation has been established.

So, how fast or slow they are able to converge, and how fast or slow they are able to finally switch on the parameter, depends on what kind of input they are getting at home, in spite of, let us say, the language having language x with a very clear-cut plus tense kind of structure and not much ambiguity; but what if the language at home is sometimes more informal? Let us say that some languages have variations, as well. So, the formal variation is something that we use outside the home. The informal variation you use at home is. Now, the informal variation does not mark the tense in all situations on all occasions. What if that were the situation? So, in that case, a child will be at a

disadvantage in converging on the exact structure that this language allows.

So, these are the kinds of explanations that VLM has provided regarding these issues. Now there are some problems as well. There are some issues that have been pointed out. One reason is that different verbs have different error rates. Everything remaining the same, some verbs have more errors than others.

We saw the difference between eventive and stative verbs in the previous lecture. So, if those are the differences, both of them are frequent high-frequency verbs, and still, there is a difference between the nature of the verbs. So, VLM does not seem to be able to answer this kind of problems. Now, let us move on to constructivist theory.

This is another approach. We have been following this two-way distinction in the approach to language acquisition: nativist versus constructivist. So far, we have been looking at the nativist theory and two different paradigms. Now we will move on to constructivist theory. Now, constructivist theory does not believe in any innate grammar. So, they do not believe in this theory, which does not believe in either principles or parameters; forget about switching parameters on and off—there are no parameters at all.

Here, the theory posits that children learn through rote learning from input in the initial stages, and after a point, they are able to generalize across memorized strings on the basis of commonalities. So, a similar kind of word appearing in a similar position in the sentences has a similar kind of meaning, and thereby it is able to generalize across contexts for a category of words, let us say. So, even though they start at the lexical level, over time, they will gradually build up an abstract knowledge of their own language. So, they will start with, let us say, a few words like walk, walked, talked, and then they will notice that this kind of word always appears in a particular position in the sentence and seems to always talk about some kind of action, and so on. Over a period of time, they get the formula and then generalize.

So, when they have achieved that level of understanding and that level of abstraction, they are no longer dependent on lexical entities. Because they have gone beyond that level and reached a higher level of abstraction. Now, in terms of inflections, what do they say? They say that, initially, children start with unanalyzed forms. So, they do not understand at the beginning that "walk" and "walked" are the same word but in different forms; you know there is a modification depending on the tense.

So, that is not how the initial input goes. Initially, they regarded them as two different entities. So, this and that are two different words. However, they start as they receive more and more input because this theory is primarily based on that. So, as children are

exposed to more and more input, they are able to take part in creating knowledge. It is not as if the knowledge were handed down to them.

So, they notice that there is some sort of statistical learning that occurs over various, you know, multiple iterations in which similar kinds of things happen. So the pattern emerges. So, initially, they start with the idea that these are two different words, and then they realize that the position and the context appear to be similar; hence, there is a rule at play. So, there is some sort of understanding that they have reached. They are also able to generalize about the mapping of form and meaning.

So, if the child has already learned words like "walked," "liked," and "wanted," she will be able to extract the word plus the -ed structure, the underlying structure, and apply it to other forms. So, a new word, even a nonsense word, if they hear it, is why they are able to create the inflectional, properly inflected tense forms of that verb. So that is what happens after they arrive at the V-ed structure: the underlying structure. So based on these there has been, multiple studies have been carried out. We are not getting into all the details, but we will discuss only one model here.

So, MOSAIC is a computer model that implements this constructivist theory. This model has no inbuilt knowledge of language, including syntax, morphology, or any other kind; there is no inbuilt knowledge. The entire system works on rote learning. So, depending on the input you give, the machine learns the language and understands how to use the inflections, and so on. So, it rote-learned strings of words from the input; first, it learned the words and memorized them based on distributional commonalities.

That is what we talked about just now: that there is a similarity in terms of position, meaning, and so on. So the same theory applies to this model. So, first they learn, go through a rote learning process, and then understand some distributional commonality; initially, the model makes mistakes as it learns from the beginnings and ends of the sentences. So, the model was created in such a way that, depending on the input sentences, it focused primarily on the first and last parts of the sentences. So the first part is the subject, and the last part is, of course, quoted from the work in a sentence like this.

So, Cicily wants to eat cake. So, what the model memorizes is Cicily eating cake, and hence the model gives an output like "Cicily eats cake." Now, if that is how the model is memorizing, then what will happen initially is that it will create the bare word; there are no inflections present. So, initially, the model made mistakes like this. But after some point, after some time, as the learning continued, what learning continued to mean was that the system had more inputs to work with.

So you can call it testing, trials, and familiarization. So, this model also goes through that kind of system. So, the more training data there is, the better it gets; that is the idea. So, the model, over a period of time, learned to create longer and correct sentences with the proper inflection that were required. So, this study was carried out by this group in 2010. So, they found out that once the system was in place, they compared the performance of the MOSAIC model with the VLM with respect to certain speech data that they collected.

So, they analyzed the spontaneous speech of 14 mothers and their children in five different languages. So, English, Dutch, German, French, and Spanish are all the languages from which we have given examples. So, some of the major European languages from which they collected data were the mothers' speeches to the children. Speech by the mother and children, and then how the process went. So, at first, the variation model was checked for the predicted errors, what kinds of errors the model produced, and then the same speech data was used for the MOSAIC model to see if the model would make the same kinds of mistakes or if it would have a better result with respect to the same speech data.

And so the same speech data was used for the types of errors it would produce. So what was the result? The result was very interesting; it accounted for most of the child's errors. Both VLM and Mosaic performed similarly. So, whatever VLM predicted, the children actually made that kind of mistake. Similarly, MOSAIC the model whatever kind of mistakes the model made a similar kind of mistakes the child also made.

So, broadly speaking, there is a similarity between the two models with respect to the data. But only MOSAIC could predict which verbs were more likely to have errors. So, this is something that the VLM could not predict, which we have already seen; and as predicted, the VLM could not foresee this kind of difference between the various verbs having different levels of errors. So, that was predicted by MOSAIC.

So, this is an interesting case. Another case where this model provides a satisfactory answer is the case of a differential rate of error in some auxiliaries and inflections compared to others. This is explained by the concept of frequency. So, the higher the frequency of occurrence of a particular kind of structure, the better the model performs. So, it is kind of understood that high-frequency morphological structures will be learned more easily than those that are of low frequency. So, for example, if the child continuously hears the past tense in the household, they are more likely to learn the past tense more easily compared to, let us say, the future tense hypothetically.

So, the more of that kind of training data you provide, the better that machine will become in that aspect. So, it is entirely dependent on the input, and that is why it can also

address the differences. Finally, the gradual acquisition of inflection morphology is also why it happens over a period of time; this is the same reason that the earliest forms are restricted to rote learning. Because they did not say that the machine learns the system in the very beginning, as soon as it is exposed, it already knows what is understood according to nativist theory. Nativist theory says that the system is already in place, so there is no room for development or improvement, but constructivist theory says that there is always ample room for improvement because initially there is rote learning; over a period of time, students will look at similarities, the distribution of words, and the structures and meanings, and over a period of time, they will master it.

So, naturally, this model provides a better explanation for the gradual acquisition of inflectional morphology. Why is it that some morphological structures are acquired earlier than others? So, there were many studies, but one of the best-known ones is about Spanish-speaking children. Initially, children understudied produced only one form of inflection for each word. So, there was whatever was of high frequency.

So, let us say that "walked" was the most frequent, but not "walking." So, they kept producing walks in many cases even when they were not relevant. Over time, they realized that each word could have different kinds of inflectional structures, and they then started producing them, but this depended on frequency. So, if the past tense marker is high-frequency, it will be the first one to be acquired; the future tense will be acquired later if it is less frequent, and so forth. So, that is exactly what they found out as well.

So, gradually, the children learn to use other forms of the same verb. It makes sense, but there is also some evidence against MOSAIC. Evidence exists in some cases where MOSAIC could not answer the questions that arose. One error was that creating a bare form is more difficult than expected, given the inputs. Because this model is based on comparing the occurrences of words in different formats across contexts, it is useful for understanding language use. So, given that as a constant, the use of bare forms continues for quite a long time among children.

This has already been attested to. We have discussed it as well. That bare form continues to figure in the child's speech for a very long time. Why does it happen? Because the input is not always in its bare form. So, that is one problem. Also, the same group found that MOSAIC had more errors in Dutch than in English.

Now, this model predicts more errors in Dutch than in English. However, in reality, children make errors in the opposite way. Children make more errors in English than in Dutch. So, this is another area that they found has a problem with the mosaic. So, this is, in a nutshell, what the model is about. Secondly, pronouns are reported to have fewer

errors than nouns, as they are said to be learned using high-frequency slots because they are of high frequency.

I, you, and he are very high-frequency words. So, naturally, pronouns are supposed to have fewer errors than nouns. But in the controlled study, this was also contested in a control study by Guo et al., in which they looked at a very narrow age range of children and used the elicitation technique for children to produce sentences. They found that a similar kind of error in nouns and pronouns was produced by the children. So, this is also another domain to which the MOSAIC model does not seem to provide answers.

Because we are talking about frequency, we are discussing what is easier to learn and so forth. So, this is also a recent finding that does not seem to provide any answers. Thus, if we need to conclude the comparison between these two theories, both nativist and constructivist theories have proven useful in explaining various issues regarding the acquisition of inflection markers. However, both have a few issues, only a few of which seem to have been unexplained thus far.

So, of course, the debate does not end here. There are many more works that have been published, and there are still debates going on. For those who are interested, I will add all the references. You can gain an advanced understanding of the topic. But this is the baseline. These are the fundamental ideas regarding the nativist and constructivist debates on this topic.

Now, let us look at another interesting debate within this domain, called the past-tense debate. Now, the English past tense has been the topic of a very long debate regarding its processing. How do children learn? How do children learn the past tense, specifically that of irregular verbs? Now, the English language has two different kinds of verbs. The regular verbs have a very simple lifespan.

They just take the -ed morpheme and become past tense. But irregular verbs have a completely different life journey. So, this is where the problem lies. How do children understand, how do children figure out what happens, and what is the underlying mechanism? So, the regular inflection of -ed is learned easily by children and is even overgeneralized quite often. On the other hand, there are about 180 irregular verbs in English.

They cannot be generalized in the same way. In fact, they are often victims of overgeneralization by regular verbs. So, there are different theories; we will discuss only two of them here. One of the important theories is the words-and-rules theory. Words and rules theory states that the lexicon and grammar are two different parts of the language

faculty of the human mind.

So, it is not only grammar; there is also lexicon. The lexicon consists of all the words that the language has, and it also includes irregular verbs. So, irregular verbs are considered lexical entities and not the result of any rule-governed phenomenon. So, that is how the difference is; even though they have a grammatical feature of past tense incorporated into them, according to this theory, they are processed as lexical entities rather than as grammatical entities. On the other hand, regular verbs are generated productively by the rules.

So, these two categories of verbs should not be considered the same. So, on the one hand, regular verbs are created by adding the inflectional morpheme, and on the other hand, irregular ones are stored and memorized as lexical entities. This is a very well-known publication by Pinker and Ullman in which they describe it. So irregular forms do not require an exceptions module.

So it is not an exception. They are a different category in their own right. That is the idea here. Two subsystems can overlap in their expressive powers. So, what they mean by expressive power is the grammatical function. So, the two subsystems are the lexical and grammatical subsystems; both are part of the superordinate node of the language faculty. So, some entities in the lexicon and some entities in the grammatical category can have a similar kind of grammatical purpose, or what is called "expressive power.

" So, even if they belong to two different subdomains, they have a similar kind; thus, overlap occurs in terms of the functions that they perform in the sentence. So, a given combination of features or syntactic roles can be expressed either by words or by rules. The role is similar; it is to express the time of the event with respect to the time of speaking, which is, let us say, the past tense. So, this happened before that before now.

So, this idea can be expressed either by a word or by a grammatical rule. They do not have to be from the same stock, and if they are not from the same stock, then there is no question of an exception. That is what they mean when they say that the exception module is unnecessary. Because only when you put everything under one rule do you say that these are the exceptions to the rule; hence, this is a list of words that do not follow the rule. Here, that question does not arise at all.

They can be in different categories but still perform the same function. Now, if we compare this theory with classical generative theories, there are some differences. In the traditional generative theory of Chomsky and Halle, the irregular forms are generated by affixing an abstract morpheme to the stem and applying rules that alter the stem's

phonological composition. So, from this perspective, most irregular forms are not entirely arbitrary. So, there is some sort of rule applying there, as well.

So, they call it falling into families that display certain kinds of patterns. So, they give examples like ring, rang; sing, sang; sit, sat; like this. So, this is another group. This is another family of, let us say, families of categories of words and categories of formats. So, there is a rule, there is an affixation of a morpheme, and thereby you change the phonological nature and composition of the words.

So, that is exactly what is happening here. So, this is also not an outlier in that sense. They are not completely arbitrary. The problem with this position is that there are far too many irregular verbs in this language; in English, there are too many different types of irregularities, shall we say. So, if you need to have a theory for each of them, you need too many theories, and you also need to have too many exceptions. So, that is a problem with looking at irregular verbs from the traditional generative framework. Now we have another model, which is the RMM model that actually comes from a connectionist model proposed by Rumelhart and McClelland.

This model basically looks at the entire problem in relation to a pattern-associator model. So, there is a pattern that gets associated, and this is how it counters the word-and-rule theory. So, this is the debate: the past tense debate is between the RMM and the word-and-rule theories. So, according to this RMM model, this learning takes place through the association of the phonological features of the stem with those of the past tense form. So, it is not that the association is between words, but that the association is between the phonological form of the word and its past tense forms.

So, that is what the association is at that level, not at the word level. So, the pattern associator links sounds to sounds rather than linking words to words that are stored in memory. Now, in order to understand what this theory actually means, we need to go back one step because this theory is part of the overall connectionist model proposed by Rumelhart and McClelland. This theory is also known as parallel distributed processing. It is not possible to describe that entire theory in a short time like this, but let me just say that this theory opposes the ideas of the linear and modular theories of mental functions.

Mental functions: there are different theories about how the human mind and brain function. One of the theories states that the mind is modular. So, there are different domains and different modules in the brain, each of which takes care of a separate function. Connectionist models, on the other hand, state that mental functions arise from a complex network of connected units. That is the most important aspect of this thing; that is why it is called the connectionist model. So, there are different, you know,

different domains in the brain; different regions work together; this is a connected and well-connected network of systems that gives rise to any kind of mental faculty.

So, including language, it is another mental faculty. So, this also has the same kind of understanding. So, learning in this model is primarily based on associations and patterns, and that is what we just talked about. So, the patterns and associations here are based on the sound system rather than on the words, as far as this model is concerned. So, this is a simplified version: similar words have sounds that are alike, and that is what is meant by similar. So, similarity is based on sounds, and their representations are also partially overlapping.

So, when you figure out the similarities, the representations are superimposed. Now, this allows such models to acquire families of similar forms more easily than arbitrary sets and to generalize the patterns to new similar words. For example, if the system is exposed to words like "fling," becoming "flung," and similarly "cling," becoming "clung," then this is because they sound similar. So, similar-sounding words have similar systems and patterns. So, that kind of gets superimposed that if it is fling and becomes flung, then cling will become clung, and now you give another similar-sounding word like spling; it does not exist, but the system will be able to create splung as well.

So, that is how the system apparently learns, as far as this theory goes. Of course, this is a very watered-down analysis of the model, but you get the point. Similarly, flip becomes flipped, and clip becomes clipped; if that is how the system works, then it will result in another nonsense word like plip: it will become plipped. So, that is how this model is supposed to work. However, like all other models, this one also has its share of problems.

For instance, later studies have pointed out that there are some odd structures that the model generates. For example, mail, and it becomes mumbled tribal treelit. So, for novel words, they sometimes create a really odd kind of output or no output at all. So, there are some limitations to the model in that it does not seem to function well. Of course, the model has gone through it.

So, the initial model was given in the 1980s, and eventually, there were updates to the model. So, later models tried to solve this by adding some rules to the model regarding how it works. So, even more recent models, to quote Pinker, have given up on generating past-tense forms altogether. Their output layer now contains one unit for every past tense change, turning inflection into a multiple choice test. So, basically, this model has also run into issues and encountered some problems regarding the past tense marker and how the past tense forms work with irregular verbs. So, this is, in a nutshell, the comparison between the RMM, which is part of the connectionist model by Rumelhart and

McClelland, and the words-and-rules model proposed by Pinker and others.

So, these are the two different sides of the past tense debate: past tense as in how regular versus irregular verbs correctly get their past tense marking. What happens? What is the rule, and how is it represented in the mind? So, this is where we will conclude the lecture. In the next lecture, we will take up some other issues, and eventually, we will also look at the blocking hypothesis and other cases. Thank you.