Language, Culture and Cognition: An Introduction

Dr. Bidisha Som
Department of Humanities and Social Sciences
Indian Institute of Technology, Guwahati

Module - 07
Part - 2
Lecture - 16
Language and visual attention

Welcome to the 2nd part of module 7, we are looking at the relationship between language and attentional mechanism in humans. Part 1, had given up brief description and overview of how what attention mechanisms are all about and various theories and so on.

Now, we move on to the experimental setup as to how we can show how language and attention are interrelated in various domains of language processing. So, we will start with visual attention in terms of attention, we will start with visual attention today. In the next part, we will look at the other sides of the same story.

So, we will look at language and visual attention with these points in mind. First and foremost we will look at how attention and vision are part of the same broader mechanism, then visual attention in language processing, which is the main topic here with respect to the most commonly utilized paradigm. So, there are couple of paradigms that are utilized for research in this domain, we will look at that as well.

And, how they are different from each other and then we will gradually we will talk more about visual world paradigm, because that is the one we will be focusing on and, then some important studies in this domain using this particular experimental setup.

So, eye movements as an indicator of cognitive processes is something that we take for granted today; as of as of today as we speak eye tracking as a methodology is very well established, with a very rich domain of research output and really very very interesting research findings. However, the this was not taken this was not considered as a thing to be really looked at some time back.

So, the initial experimental evidence for this goes back to 1967 seminal work by Yarbus showed that humans look around space to according to their goal which is something that is almost common sensical today. If we want to see something we want to get more information about something, we look at that particular object.

But he showed this through an experiment in 1967. So, what he basically did was the experimental setup was like this, he asked participants for specific details in a painting. So, the participants were looking at a particular painting and then they were asked specific details. So, when they asked for a detail, the participant would look for that particular information in the display, in the visual display.

So, each time the participant had a different scan path of on the painting. Now, scan path is something that is related to eye tracker as a mechanism, we will look we will look at it in a short while. So, each time there was a different question, the participant was looking for an answer which is the goal directed behavior in our case. So, when we have a goal in mind, basically meaning the top down processing with respect to some amount of a goal, some kind of an intention.

So, that intention is incorporated in the visual search, visual search is what is the paradigm all about and then the participant looks for that information in the scene, in the visual display that is in front of him. This simple experiment is a very very simple and fundamental level of experiment, but it proved a very important thing, which is that eye movements are a direct indicator of the goal directed search, which is now called visual search.

So, eye movements can be taken as a variable in order to find out the online dynamic processing mechanism that the participant is busy in right now. It is almost in films very often in typically in Hindi films, in Bollywood films and other in Indian films, you will notice when somebody is thinking, they are looking skywards, looking somewhere not exactly at anything.

Basically, meaning that you need to stop the input visual input coming from the surroundings in order to for you to be busy in imagery in remembering something. So, memory is if your visual system is busy in getting information from the surroundings, then memory there will be some kind of a that there is probably an interference that they think of. This is not something they have thought about, this is not about their creativity, this is all this is how we always we always behave.

A imagine yourself remembering your childhood, chances are very high we will try to look at something that is not distracting our goal at that given point of time. So, goal directed behavior with respect to visual search goes back to this seminal study by in 1967.

Now, linguistic representation, attentional mechanism, visual input all these integrate at all times during our everyday life. Take a simple example like look at the clouds, it is raining, it is quite cloudy today. So, I imagine I just tell somebody look at the clouds, it is a very simple sentence or pass the salt in a dinner table. When this kind of a sentence is uttered, other people will 'look' at that particular object.

So, when I say I do not even say look at the salt, I just say pass the salt the person, the person whom I am speaking to will look at the salt because there is a goal attached to it, that you have to take that object and pass it on to the speaker. Similarly, 'look at the clouds' is also a simple sentence and it brings together the visual search, the attentional mechanism and then there is of course, a mental representation of what clouds are and so on and so forth. And, there is also an action that is part of this entire system.

So, language and vision interaction is very clear in our day-to-day life, it is there all the time for us to see. However, even things that are very common, things that are almost commonsensical it is sometimes they evade a clear-cut understanding. This is the same in case of language vision interaction as well, though we know and there is an enormous amount of data available out there, that there is an interaction and the interaction is dynamic.

However, the exact nature of it, it still under construction let us say, it is still understudy. There is a lot of work still going on. So, we will start with this slide with a disclaimer that we do not really know all about it till today.

Now, most of the as I said most of these cases, most of these studies today use eye tracking as a mechanism to give us the data that we are utilizing to understand this domain, understand this interaction between language and visual attention. Eye tracking is a machine, there are different types of eye trackers. There is head mounted eye tracker, there is a fixed eye tracker and so on; basically it is a machine that tracks the movement of our eyes as we look at any kind of a display.

So, typically the machine if it is a head mounted eye tracker, it will be placed on the head of the person with some kind of a glass kind of a thing in front of the eyes or there are other kinds of eye tracker as well, which is just put on the table and it tracks your eye movement. While, we look at a particular object, the eye tracker tracks the movement of the eyes and then it gives us the output in various cases.

So, typically the types of a situations, types of experimental setup that will be used is scene perception, reading, speaking, listening and so on. Comprehension and production of various linguistic types and so on. So, what are the outputs that we get? Primarily saccades, gaze duration, scan path etcetera. Saccades are the ballistic movement of the eyes, how the eyes move from one point to another point on a particular scene is the saccadic movement.

Similarly, gaze duration is where the eyes stay for long time, the what is the time duration, how many milliseconds the eyes look at a particular aspect of a scene is the duration, gaze duration. And, scan path again how the trajectory of the movement eye movements take place. So, these are the indicators in case of indicators of the processing strategy of the participant of the subject.

Because, this is the premise basic premise is that the thing that we are trying to understand clearly or better or simply we are just trying to understand, because we have some goal attached with it, we will be looking at it. And, then the more difficult it is to perceive, the longer duration it needs for us to look at it. So, these are the premi.. basic premises on the basis of which this machine works.

So, there are two main paradigms that are used in visual attention research. The one is called the visual search paradigm, the other is visual world paradigm.

They are similar, but not same there is a slight difference. So, visual search paradigm basically looks at or works on the basic idea that eyes look at things, eyes search for things in the real world. So, this is basically the idea. So, this requires guiding attention to specific location using eye movement. When we search for something, when we look for something, basically we scan the domain, we scan the scene or the environment or whatever the case may be, using our eyes.

So, this needs a guided attention because we are focusing on a particular object with respect to channelizing the attentional mechanism. There are few variables within this paradigm. One is the set size; set size basically refers to how many distractors are there in a particular given scene with respect to the target object.

So, for example, there are two different kinds of this is a typical example of visual search. In the first case you can see it is called feature search. So, in case of feature search, there is only one object that stands out. So, this is different only in terms of the feature which is the red color in this case. So, if you have to search for the red X, you can easily find it out.

In the other case it is conjunction search. So, in this case the sets are higher, number of sets are higher because the distractors are too many. It is not only the feature X, but also in terms of the color, there are many distractors. There are red color O and then green color X and so on. So, depending on the sets and their size, basically meaning the kind of destructors that you have, different strategies will be in place in terms of visual search.

So, there can be serial search and parallel search and how they are defined, how they are you know they how they interact with set size and so on, are the issues that visual search paradigm looks at. So, there are of course, two kinds of attentional deployment; one is top down and one is bottom up and, that is something we have already looked at.

So, in a typical experimental setup, the participants are presented with a display of multiple objects as we have just seen and are asked to find a predefined one, predefined object whether it is a red cross or it is a red square or something, anything. So, there is a visual display and there is an instruction to the participant 'find X', find this thing in the scene.

So, the goal is goal of this kind of a paradigm, in this kind of an experimental setup is to check the interaction between the bottom-up salience of the stimulus and the top-down goal. So, if it is how the how the interaction between your goal as to find the red square or red cross in the scene and how it interacts with the kind of set size that you have, the kind of you know scene complexity that is there, how salient the target object is and so on and so forth.

So, these are the these are the basic things about visual search. This is an example of a simple visual search experiment with two kinds of search: feature search and conjunction search from 2018 paper.

In terms of linguistic research, in terms of language processing findings typically point to a situation that accompanying linguistic input helps in visual search. If the there is a there is a scene that is presented to you and there is an accompanying linguistic input, our search becomes easier typically, in case of feature search if not in case of conjunction search.

So, this is this area of research is called language mediated visual search. Visual search as you can understand is a rather big domain of research and does not necessarily the exclusively depend only on linguist language related notions. So, the language mediated visual search is one part of that visual search. So, vision research is a different domain of research which also depends on visual search method.

So, in language mediated visual search, there are few strands of research; one of them points to the fact that scene perception can constrain real time spoken language comprehension. If there is a sentence that is given to the participant and there is an accompanying scene that is presented to them, there is a display in front of them, then there is a dynamic relationship between these two inputs. So, scene perception can affect language comprehension.

Similarly, language processing can affect scene perception as well, it is both ways. So, the scene can affect the language comprehension, language comprehension can affect scene perception; it works both ways which is why we call it a dynamic relationship. In case of word recognition, now language processing does not only talk about word recognition, it also talks about word production, it also talks about sentence generation, sentence comprehension and so on and so forth.

In the domain of word recognition visual presence of an object similar in name alters comprehension. So, if there is a visual display which is hazy, which is not clear enough, which is under the threshold of understanding and then there is a linguistic input, it helps in comprehending. However, if we tweak that relationship, if it is not representing the object then it also has a negative impact, we will see it in very shortly.

In case of visual world paradigm, there is slight difference with visual search, visual world paradigm the participant looks at a display while simultaneously listening to a spoken utterance. So, there is a the participant has the earphones on and he listens to something and then simultaneously looks at a display. What is looked at here is slightly different, the investigation here typically relates to interaction between the spoken stimulus and the eye movement across the display.

So, what how the spoken utterance impacts the way the person scans the visual display is what we look at. So, what are the typical aspects of that sentence and how the grammaticality at various levels impacts the way we look at the scene is what visual world paradigm looks at. So, what typically happens is that the person will listen to sentence or a word and then there will be display which will contain the target object, but will it will also have various distractors and such and cohorts.

So, a particular word he is listening to and then there will be similar word, that particular object and similar object and there will be a distractor. We will see some examples shortly. So, this is one type, in another type there is no task, in the first kind they he will be asked to pick up the object. So, 'pick up the beaker' that is something one that is a seminal work rather famous one.

So, when you while will listening to pick up the beaker, the beaker is present in the scene. However, there are two distractors as well and there is a fourth filler which has nothing to do with any of the any of the target words. In another kind of visual world studies there is no task, it is just look and listen. So, you only look at the display without doing anything, there is no task to do and you are listening simultaneously.

Here this they have to just simply look at and the eye tracker picks up the eye movement on the scene. So, what happens even when there is no overt task. So, you see the difference in one case, if there is a goal directed behavior; in the other case there is no such goal. And, then you can actually look at how the eye movements differ in terms of attention allocation and so on and so forth.

These are complex domains. There can be many permutations and combinations of each of these designs, but this is the basic level paradigm. The difference between these two paradigms: visual search paradigm and visual world paradigm are like this. Visual world paradigm, the visual display precedes or occurs simultaneously with the utterance. So, typically more often than not it is simultaneous.

So, as you look at the display you hear the utterance at the same time, sometimes the display might precede as well. But, in case of visual search paradigm the instruction precedes the display. So, you listen to the instruction first and then the display comes and that is how you see the instruction has already been given, your intention has been created, top down modality has been switched on. And, then now you look at the scene and depending on how you search, we get our answers.

So, in the visual world paradigm typically the question that we are trying to look at is the linguistic processing, how language processing goes hand in hand with visual attention. And, how what are the modulating factors, what are the interrelationship, how one affects the other and so on. So, the primary focus is on language processing. On the other hand, visual search does not always have language processing as the main goal, there the idea is there to determine the efficiency of search processes.

So, what are the guiding factors in case of a search process? May iy, it may be linguistic search language related such process, it might be scene perception, it might be many other things. So, the primary goal and the way the stimulus is presented is how these two paradigms differ.

Now, let us go on to language the research that in this domain from linguistics perspective. One of the first linguists to talk about visual perception with respect to language, basically bringing together language and visual cognition, visual attention was Ray Jackendoff. Jackendoff compared David Marr's computational theory of vision, this is this precedes Jackendoff.

So, Jackendoff basically used David Marr's understanding of how visual perception really works, what is the theory of vision, it is a very well-known theory of vision by David Marr. So, Marr mentioned that objects are represented at 3D level, that is independent of the viewer's perspective whether or not viewer is you know it is irrespective, independent of the viewers perspective objects are there in the 3D level.

Object identification and vision work through a recursive and hierarchical processing. So, we understand this computer screen in on the table and the table is inside the room, the room is inside and so, we integrate ever higher levels of visual input. So, that is how a vision really works. Jackendoff suggests, Jackendoff take this idea into linguistics and then he says that this modality is not specific only to vision and it is also used by language.

How? One builds overall representation by working recursively on the syntactic structure. This is in fact, a fundamental understanding of Chomsky and linguistics that human syntax is the most important hallmark of language faculty. And, there we can go on embedding ever increasing number of clauses, without really disturbing the fundamental understanding of the sentence.

So, human can produce in finite number of sentences of limitless length and strength which is basically the thesis of Chomsky and Jackendoff does follow that same idea, that it is a recursive pattern of linguistic structure which is similar to how vision really works. So, in vision from the moment light from objects hit our eyes, it is like a process of recursion continuous process of recursion.

Similarly, for hierarchical representation in vision also, Jackendoff finds a comparison between vision and language use, we as we view things as part of a larger scene, ever larger scene. The example he gives is the apple is on the table, the table is in the dining hall, the hall is in the house and so on.

It is always like this, we can always imagine every scene as having ever larger bigger scenes. So, similarly this feature of integration is domain neutral. It is not only respective to vision, but it is basically a primary aspect of human cognition and that is why it is visible in vision as well as in language as well as in many other conceptual practices, many other conceptual processes; that is what is Jackendoff talking about.

So, that is why it is also available to be in language also in language use. So, and because these two are part of general purpose mechanism, because they come together, that is why we are able to say that red apple is on the table and so on, because language and vision are working together.

However, Jackendoff was still follower of Chomsky to a large extent. So, he does not really give a much bigger role to attention in case of language processing beyond this much. And research has of course, moved much ahead of that much beyond that. And, then another linguist much before Jackendoff actually talked about language vision interaction though not really in terms of experimental work, but he did talk about it that is Otto Jesperson.

He talked about when he will talking about language vision interaction, he mentioned that linguistic labels refer to visual objects. Of course, everything that has a name basically has a linguistic level. What we mean by name is a label, we need to give a label to something a visible object in order to identify it and in order to have a representation in our mind.

So, for example, he gives this example of apple when a child starts learning language, he sees a red apple and says ‘apple’ to identify the object that is how children are taught. Again, he sees a green apple and says apple to identify the object again. So, what is happening here is that the visual perception of shape, color and other features or of the objects are different.

A green apple is not the same as the red apple, but how we, at the beginning how we integrate this information with the visual input with the linguistic input is that we create a category information. We have this we have already seen in the section on categorization how category information takes the help of linguistic labeling. So, this is where it comes from, Otto Jesperson was one of the first to talk about it. So, he says the child identifies the category to which the object belongs.

All the other complexities are for the time being kind of maintained, but not used in the label overt level. Thus, language encodes and expresses visual perception directly. In fact, this is where a lot of cultural differences also occur, which we have already seen which means that language takes the visual inputs and transforms them to be used keeping its constraints into account.

Constraints as in the other features, the green apple being green apple has you know can be eaten in particular, where the apple can eaten in a different way and so on. All those things are also there; however, the labeling depending on the categorization is how it works. So, visual perception that do not have readily available linguistic tags are not expressed. This is a very interesting area of research even today we will see.

So, now we move on to the experimental work as to how, this is not very old; experimental work in language vision interaction has started only a few decades back. And however, we have made a lot of progress even in the short span of time, typically with the because of the invention of eye tracker as a mechanism.

So, one of the very famous very landmark studies in this domain in one of the beginner and on of the initial studies, that showed that hearing a verbal cue can make visible an otherwise invisible image. This study used what is called ‘continuous flash suppression’, the technique to make an image sort of invisible.

It creates a disturbance on the picture which is why you cannot really readily identify the object. So, it is degraded in the vision in terms of clarity. So, CFS is a technique of doing that. So, valid what they did was they used, there were many conditions in the study. This is how it worked.

So, what they did was the study the this is how temporarily temporal sequence it had, this is how it moved. So, at the beginning what happens on the screen, there will be a fixation, fixation cross on the screen. So, that the eyes do not distract here and there and then there is an auditory cue. Again there is a delay of 450 millisecond and this is the stimulus that we are talking about. So, the visual display is sort of covered by some kind of a disturbance because of which you do not see.

Now, depending on what they hear; so, there they had valid label and invalid label. Valid label is the actual name of the object that is hidden here, in other cases it is not the object that is hidden there. So, if there is a match let us say there is a pumpkin here and they hear pumpkin before, they are able to see the pumpkin more clearly. However, if you say there was a house, you just say is house and then show see the picture then you will be less able to identify the pumpkin right here.

And, then there is a question ‘did you see an object’? And, depending just to mention just to know that you know this is a, this is a question that is asked to ensure that you actually paid attention and then this is the validity prompt that is given. So, was the object you saw a pumpkin and so on. If they say no that is one answer, if they if it is yes then they are the validity is ‘what did you see, did you see a pumpkin’?

So, if they have heard a pumpkin, the word pumpkin and then the visual display was actually a pumpkin which they saw and then this is how they validate. So, the finding it is a simple study, but the findings were enormous the in terms of a significance that linguistic labels help us help in visual perception even when the object is visually degraded. The perception is difficult due to the way the continuous flash suppression works and, but then we having a label that is connected to it helps in perception.

So, this is what it goes in the way of saying that language and vision interaction is dynamic and it works both way. In the in other words, what they basically found was, at a theoretical level, that language compresses and brings back awareness in the absence of an explicit visual cue. This is what the this is a seminal finding a very important work in this domain and this is what they found.

That language helps aids in language can aid, it can also decrease the possibility of finding the visual cue. Because, when there is an invalid cue, invalid cue in this case meant which did not match the visual display, they actually did not find the object as well.

So, there are many other studies by the same group, they have found that linguistic words lead to heightened perception of visual objects in many other studies as well. Thus, visual attention in the external world often gets guided by language input. So, language input actually has a guiding role in visual perception. This is one strand of research.

Similarly, Michael Spivey's work has been extremely important and Michael Spivey and his group and his work has it actually spreads across various domains including fictive motion understanding. So, he has his studies have brought into focus a very fluid interactive relationship between language and vision it's not one way, it is actually both ways. So, findings from many of his studies reveal the current understanding of what we know for language vision interaction and the role of it.

So, all these studies many of these studies primarily ask some simple questions. The questions are whether the nature of the visual environment has an impact on the way we process concurrent spoken language, how what is the nature of the interaction is what we are trying to figure out.

Of course, there is an interaction, numerous studies have proven. I have just given you a couple of examples, but there are many others that has proved beyond any doubt that there is an interaction that is dynamic, that language guides visual perception and then visual perception also helps in comprehension of sentences and so on.

However, what is the nature of that interaction is what we are trying to look at. So, if there is an interaction how increased visual complexity affects the likelihood of word object mapping at various levels of language mediated visual search. This is something we will look at shortly.

So, to understand how exactly language and vision interacts, we must know which knowledge types are retrieved while processing both language and visual input. It is fine we know that there is an interaction, but what kind of language, what kind of input system is retrieved, what is activated while we are looking at a particular scene is what takes us to the finer nuances of this relationship.

So, visual world paradigm has yielded a huge amount of data, a large amount of empirical evidence to look at this domain more carefully.

One area is word object mapping, word object mapping as in you hear a word and there are objects displayed on the scene and how our various grammatical processes work. For example, the phonological understanding, the syntactic, the semantic, the perceptual understanding of that object and so on and so forth. How it guides our attention, visual attention on those display, on those objects that are displayed on the scene and how does the interaction really work.

For example, when one begins hearing a word a simple word, that begins with let us say the word ca, the cat the word ‘cat’. As we start hearing the word in the with the onset of the first phoneme /ca/ that is immediately what happens is that we not only our brain not only activates the word cat, but all the other possible words that starts with that phoneme.

This is this finding actually goes back a bit few years couple of until ‘98 in fact, is the famous study where when it took place, we will see shortly. So, when we start listening to any word, it not only activates; this is what we call phonological cohorts. So, every word that starts with the same phonological input will be activated, this has been already proved.

So, when we hear before we have heard the whole word cat, the moment we hear /ca/ then cab and carrot and many other things are also active in our brain. By the time we go to the last part of the word ‘cat’ and then we have gradually come back to only one and then focused on the word ‘cat’.

So, this kind of activation happens as a in terms of a rapid spreading activation style and this not only takes into account the sound aspect of it, but also many other aspects as we will see. So, as auditory processing advances, meaning when we are hearing, when we are hearing an auditory input and as it progresses over you know temporal progression takes place, the irrelevant activations die out.

So, /ca/ activates ‘carrot’ and ‘cab’ both so, there are three activations at the very least. But, as we progress the irrelevant ones die out and finally, the target is perceived which is remarkably fast process. It does not take too much of time to listen to the word cat few milliseconds, but even within that time we have already activated many other competitors of this of the word and then settled on the cat.

So, recorded eye movement data show quick orientation of attention to word related objects within the first 100 millisecond. 100 millisecond is a remarkably short span of time. So, the moment we hear /ca/, our eyes already scan if the if the display has all these three words ‘cat’, ‘cab’ and ‘carrot’ and a distractor like ‘bottle’ let us say. So, the eyes have already scanned all the three possible objects, by the time auditory input goes till the word cat, till the final phoneme eyes come back and settle down on the actual object.

But, this is the progression. So, this is how quick our orientation of attention really happens. So, this clearly suggest that linguistic input has an immediate, immediate as in terms of 100 milliseconds. So, that is how immediate it is, immediate visual search.

So, we will look at one famous study, perhaps the most famous study on word object mapping at phonological level, I had just talked about this. So, participants looked at a computer screen showing pictures of beaker, beetle, speaker and carriage. So, the carriage is the unrelated distractor, it has nothing to do with the target and beaker is because this is the sentence ‘pick up the beaker’.

So, they had to choose the beaker and so, beaker is the target object here and these two are somehow related, similarly these two are related.

In a remarkable finding this is how the display, this is how the study actually works. This is the head mounted eye tracker, the person listens thats the scene camera, this is the display. So, you see this is beaker, this is beetle, this is a carriage and this is a speaker.

Now, what happened was as they listened to the word, as the what sentence progressed with the first sound /bI/, the eyes moved to both ‘beetle’ and ‘beaker’ because they were competitors, phonological competitors. They started with the same sound. So, fixation on pictures of ‘beaker’ and ‘beetle’ was very high in the initial phonemes of the spoken word beaker. Now, after the be part is over and the second syllable is pronounced /ker/, by now you already know that it is beaker right.

But, even then fixation tended to increase on weaker and then speaker as the end of the word beaker unfolded. This is remarkable finding, this is still fine because you still do not know what is happening, what is coming in your way whether it is beaker or beetle. But, here we have already listened to you know moved on to the later part of the word, the participants still look at the ‘speaker’. Because, this part of the final phoneme of these two sentences are words are similar.

So, this is how remarkable and this is all this happens within remarkably short span of time, this is what we mean by word object mapping at phonological level. So, because there is a similarity of sound between these objects that are displayed, the eyes go and scan and take information from all the possible cohorts, all the possible similar objects to look at and find before finally, settling down on ‘beaker’.

The funny thing is that participants who take part in this kind of experiments, if you ask them ok you heard you pick up the beaker, why did you look at the speaker? Speaker has nothing to do with beaker, they will actually say I did not even think about the speaker. But, the eye movements have been recorded and that is how we know that this is how actually spread activation works.

So, the brain is remarkably fast in giving, it is like you know giving you on a platter ok these are the these are the objects that you have, these are the you know ready plate, now you pick and choose whichever you want. So, this is how the process really works. Similarly, word object mapping also works at semantic level.

Semantic level another yet another study showed that participants directed their overt visual attention towards a depicted object, when a semantically related, but not associatively related target word was unfolded. So, if you hear ‘piano’, you also look at ‘trumpet’ that is the thing.

So, there are same if the objects are not phonologically related, but are semantically related in the sense that they are in the same semantic field, meaning wise there is some similarity. So, all of these both of these are musical instruments. So, if you hear piano, you will also look at the trumpet.

The likelihood of fixation was proportional to the degree of conceptual overlap. So, the there is there are also finer nuances, how much of overlap conceptually speaking between the words are depends tends to show the higher degree of fixation on those non-target objects.

It is quite overt, piano and trumpet are similar in many ways, but even at a higher level of mapping, at the level of perceptual understanding even there you will find this kind of word object mapping. This was yet another very important study by Huettig and Altmann, 2004. They showed that even when they heard, even when the participants were presented with the picture of a cable, while listening to the spoken word snake they are looking at it.

So, this is there is no phonological similarity, there is no semantic similarity, but perceptually there is a similarity between a cable and a on an a snake, a rope and a snake and so on. So, here in this case also they found a lot of mapping, a lot of fixations on the picture of the cable.

So, what this basically means, all these findings basically show us that we are at every given point, every given moment of linguistic comprehension, each and every word that we uttered or that we comprehend also activates a huge number of competitors in our brain ready to be used.

Yet another interesting study by Huettig and McQueen 2007 showed that listener’s fixation behavior during mediated visual search can be characterized by a tug-of-war basically. So, we have seen there is a phonological mapping, there is a semantic mapping, there is even a perceptual mapping, but what is the degree of fixation among these.

So, what gets preferential treatment and what does not, what is the you know at temporarily speaking, as the word unfolds auditorily, what where do you find the gauge duration higher, does it is it the phonological cohort or is it the semantic cohort or is it the perceptual cohort that gets higher attention? So, among them what is the relationship? So, they find that it is not very clear-cut. There is actually a tug of war between phonological, semantic and visual levels of representation.

So, there are it is again as I said a dynamic relationship. So, typically at the word onset the phonological cohort gets higher attention in all these cases. However, as they progress so, temporality matters a lot in depend in activating the mappings. So, this is these are the findings in case of word object mapping.

Then anticipatory eye movement is yet another very fascinating area of research. Anticipatory eye movement basically refers to moving the eyes to a place where you anticipate something to happen, it is a context dependent processing, visual processing. So, this is this has been utilized in many other areas of vision research, a lot of anticipatory movements have been reported in a all kinds of different tasks like the tea making, sandwich making and driving and so on and so forth.

Driving of course, is a very interesting area where anticipatory eye movements are very very crucial. So, what you anticipate? What you expect to happen and depending on that your eyes will move to a particular part of the scene. So, in case of language processing also it has been found out in many domains of language processing, whether it is know reading or many other areas.

So, predictable words are read faster than unpredictable words. For example, in a sentence if we are if we are looking at a sentence, 'I will go to I will go to office this morning' versus 'I will go to the moon' this morning. And, then there is a there is a predictability difference between where you go to and then that will affect the how fast you read the sentence. Similarly, many other such studies are there.

The idea in such studies is to find out if the perceivers can consider various aspects of visual scene and linguistic information to make predictive eye movement even before the language has fully described them. So, half of the sentence you have heard and then depending on the scene whatever display is given, you already look at the object that best fits the sentence, that is even before you have heard the whole sentence.

So, more recently eye tracking studies using spoken language have shown that participants can use semantic as well as syntactic information to anticipate an upcoming visual reference. So, this is not dependent only on the semantic aspect of the language, but also the syntactic aspect.

In one such study this is again a very well known study, participants were presented with a semi-realistic visual scene, which had a boy, a cake and some toys. This was a visual display, while they heard a sentence like the boy will move the cake or the boy will eat the cake. As the display was there, eye movements were and they were listening to the sentence, eye movements were recorded.

What do we find? How do we find the pattern of eye movement? The eye movements to the ‘cake’ started significantly earlier in the ‘eat’ condition compared to the ‘move’ condition. So, these are the two different kinds of sentence with respect to cake of course, there are many sentences, each of these experiments use a lot of the stimulus set. So, when there is a there is a cake, they will have move the cake versus eat the cake and there are many other versions.

So, in case when they heard the word eat, the eyes quickly move to the cake much before the word cake actually where was heard. But, this did not happen when they heard the word move because movable objects are toys are also movable and so on, but when it is ‘eat’, only edible object on the screen was the cake. So, the eyes will move; so, this is a predictive processing the anticipatory eye movement.

Similarly, yet another study by another group found that participants look at the motorbike after hearing a sentence fragment ‘the man will ride’ and that is there are of course, many displays. So, among them the ‘motorbike’ gets the highest fixation. But when the sentence is different, look at a when the sentence is having a girl as a subject, ‘the girl will ride’ they do not look at the motorbike, they look at the girl. The picture of the girl on the display.

So, this is this means that it is not only the sentential aspect, the linguistic aspect of the sentence that is the girl will ride of course, she can ride a bike motorbike. But, then even the contextual information, real world information, pragmatic information how the society really works that also plays a role in terms of this interaction between linguistic cue and the visual display.

So, that is what they found out with respect to the two different sentences and how the visual attention really worked. Another yet another study by Mishra et al in 2010, found that Hindi speakers anticipated and looked ahead towards an object that shared gender with the adjective. In languages that have the grammatical gender, we know that inanimate objects take gender.

So, that if there is [wahan ek chhoti-si] and then [chidiya baithi hai] let us say the sentence is like this, the auditory input is [wahan ek chhoti-si chidiya baithi hai], but by the time you are listening to [chhoti-si] and the display shows both animate and inanimate objects both which can share the gender, the eyes will scan all the possible objects.

So, the grammaticality of the gender information in this case grammatical aspect of gender also helps us anticipate the in anticipatory eye movement.

So, we have looked at the word object mapping, we have looked at anticipatory eye movement. And, then now let us look at ‘looking at nothing’, when there is nothing we still look at something. In real life everyday cognition objects referred by language may not always be present in the visual field; we talk about. In fact, that is one of the most interesting aspect of human language.

We can talk about things that are not present right in front of us, this is one of the differences as well between human language and the communication system of many other animals. So, we can talk about the Harappan civilization, we can talk about the mars mission and so on and so forth, none of is none of which is present in front of us.

So, we can talk about objects which are not there in the visual field. A very important finding in this regard shows that viewers look at a blank location, where an object was previously, if the sentence mentions it. Say if you are listening to a sentence or an object or a word depicting an object and you still look at a screen where the object previously was.

So, what is happening here is in this case there is no visual perception, there is no visual you know direct visual input. So, why this why then we still look at that area? Because, we are probably engaging in mental imagery by guided by language.

So, this is one very important very interesting study by Huettig et al 2016, what they did was there were two conditions in this experiment. In one case again there is a fixation cross, there is a blank screen followed by that then there are three objects. A monkey, there is a tambourine and there is a banana, probably no it is a boat, it is a canoe.

So, there is a canoe, there is monkey and there is a tambourine there. So, in one case they when the display is there they also hear the word banana. So, they look at the what they look at the display while they look at listen to ‘banana’ and then they the closest cohort is monkey because monkeys are semantically associated with bananas; so, they look at the monkey.

But, even when there is the there is the in the second condition was this where the display was not there and they look at and the listen to ‘banana’, the fixation should have been more on these places where the monkey was. However, they do not find that much of, but if there was a banana actually there and in yet another condition of the experiment, if they have actually banana there then there will be they will find more fixation on the area where banana was.

So, this is how the finding is. So, they for their finding is twofold. One is if the object depict object was the target object was really depicted in the scene then they will look at where it was, when it was when it has been already taken off. So, if the banana was there in the scene let us say it was here and it remains here and they will look at the banana.

However, and then in the case of absent target, they will look at this particular area in the screen even when no image is there because banana was there. But, they do not find similar amount of fixation for the cohorts for either for monkey or for canoe because canoe is a shape cohort, monkey is a semantic cohort, but they do not find equal number of fixation on this.

So, there is a difference as to how much of mental imagery is guided by language input in this particular case.

So, this is what is with what is called looking at nothing, this is one this is a. In fact, there are this study actually builds up on the older study where they showed that the object with the placement of the object even when the object is not there is looked at. Now, they try to show Huettig and et al try to show whether it is only the object, but also the cohorts and they find there is a difference between.

So, when there is the display we see that the cohorts also get a lot of visual attention, but in case of missing target, the cohorts do not get equal amount of fixation. So, that is about looking at nothing. Similarly, there are also a lot of studies looking at how culturally sensitive cues from the real world dynamically interacts with language processing.

One such study was carried out by Hartsuikar 2015. It is a it is a landmark study where they look at bilingual population and this the study was very interesting, this had a display like this. So, the they had a display and there was a sentence which they had to describe the display.

So, in one case there were three American personalities, film stars. So, personalities let us call them. So, Elvis Presley and then this is Eddie Murphy, this is Jennifer Aniston and then they also had some Dutch personalities. So, they had Tom Boonen, Kim Clijsters and King Albert and so on.

So, they had many displays like this. What happened here was the manipulation here was like this, all these people are Americans, all these Tom Boonen, Kim Clijsters and King Albert are Dutch. The participants were Dutch English bilinguals. The study was, the task was 'just describe the pictures' and the template was like this. So, two of the pictures will move down and one picture will stay put.

So, this is how they had to describe. So, Elvis and Eddie Murphy move down, but Jennifer Aniston stays put. Similarly, two pictures of the Dutch personalities will move down and one will stay put simple. Just look at the picture and describe what is happening. The animation consisted of two pictures moving and one picture staying put. Manipulation here was the combination of the language in which they had to describe and the kind of pictures they had.

So, in some cases they looked at these pictures of Elvis Presley and Jennifer Aniston and so on and they had to describe it in English. In another situation they had to describe it in Dutch, in an other case similarly they look at all these displays and they just they are they describe it in both English and Dutch.

What happens is that they play a role in the processing. So, when they see when there is a mismatch, mismatch of the visual display and the language that they in which they have to talk about; there is an interference in their linguistic output, interference typically in this case. So, that the conjunctions will get affected.

So, in case of if the see Dutch people and then they have to speak in English, the Dutch conjunction for 'and' gets into the English output. So, this is how visual attention is not only refers to the picture that is present, but it also activates a cultural background for us and which actually interferes, which has a dynamic relationship with the language that we are utilizing at that given point of time.

So, when you look at the pictures that there is nothing there is absolutely no information given about the pictures we, but we all know that these are American people. So, American people, English American English is the language that it incorporates and this is the entire frame that you activate the moment you see this.

But, you are have to describe these things in your own language in Dutch, the reverse is also true. So, you are looking at Dutch personalities, but you have to describe them in English. So, there is a tussle there and this is what they found this was a very interesting finding and it has been replicated by many other groups of researchers as well and with varying degrees of interference.

In some cases even in case of Chinese English bilinguals, we see a lot of interference of the image on the language processing, even though the image themselves have nothing to do with the task as such. But, the image is not just an image, it is also a, it is it also brings into focus the frame within which the of which the language is only a part.

Here another interesting domain of language processing with respect to vision research is that of figurative language processing. So, it is not only a simple thing like 'pick up the beaker' where you are looking at you know word object mapping or even a sentence reading, even in case of figurative language processing; for example, fictive motion sentences. What is a fictive motion sentence?

Fictive motion refers to a particular case where we describe a static scene in a dynamic way. A static scene in dynamic way for example, the there are two ways of describing a road. There is a road between these two cities, that is one way that is the literal way, the other way is the road goes from x to y. So, for example, in Guwahati there is a road called GS Road, GS Road stands for Guwahati Shillong Road.

So, we can just say that Guwahati-Shillong road connects Guwahati and Shillong that is the literal way of talking about it, but we can also say GS Road goes from Guwahati to Shillong. This second part, second example is a fictive motion sentence. We are assigning some sort of a movement on the static scene, the roads do not go anywhere right. So, but we give this kind of a dynamicity to a scene by using this fictive motion sentence.

This is abundant, this is abundantly available in all languages of the world and more often than not, for roads, for river and so on and so forth. For fence so, the fence runs along the boundary of the garden, the river goes from here to there, the road goes from here to then and so on.

What happens is that an interesting finding is that even when we know that the road is not going from anywhere from any place to any other place, if we just tweak the nature of the sentence we can see a very different pattern of visual attention ascribed to the same scene. This has been found out by many different manifestations of the same primary premise.

So, what in the research in this domain typically do is that they will give a scene of a particular of whatever the sentence talks about. So, depicting that there will be a picture. So, let us say there is a fence on the ground and there the scene will appear with two different sentences in two different conditions. So, in one case they will listen to the sentence, there is a fence here, there is a fence or there is a road or there is something. There is a river between the valleys and so on.

The same scene will be also shown with a fictive motion sentence like the fence runs along the boundary of the garden or the river goes from you know one valley to another or whatever. So, the scene remains the same, but there are two accompanying sentences in two different conditions. The remarkable finding is that, when the scene is accompanied by a literal sentence, simply saying that 'it is there' then the eyes the eye movement fixes on the middle of the scene typically more often than not.

So, this is a fixation duration is higher, more focus on the center of the picture. However, when the same scene is depicted, same scene is accompanied by fictive motion sentence, the eyes actually scan the along the trajectory of the thing, that is mentioned, be it a fence, be it a river, be it a road or whatever. So, you see more fixations along the trajectory than in the center. This is also a very very important finding, that how language can guide our visual attention.

How simply tweaking the sentence, we are not changing the meaning, we are not changing the objects in the scene, nothing. It is just a different depiction of the same thing and visual attention gets guided by that.

So, these are some of the areas that we looked at. This is of course, not exhaustive, there are many many other domains and many nuances within each of these sub areas. But, this gives you an idea as to how language and vision interact in a dynamic way in various types of language processing. We have primarily looked at the word processing level and also at the level of other areas like fictive motion figurative language processing.

We will also take up the sentence processing in the next part, where we will look at other kinds of mechanisms that are used to look at this particular interaction. We will look at brain mapping techniques that are used with respect to EEG and fMRI and so on. So, this was about visual perception, visual attention with respect to language where we looked at the role of eye tracking as a mechanism; part 3 we will look at sentence processing with respect to various brain mapping techniques.

Thank you.