R.K. Mishra, 'The Interaction between Language and Visual Perception'. *Studies in Humanities and Social Sciences*, Volume XXIII, Number 2, Winter 2016: 41-63.

THE INTERACTION BETWEEN LANGUAGE AND VISUAL PERCEPTION

R.K. Mishra

Introduction

Language is a very complex yet structured symbolic system which humans use to communicate. Much of the twentieth-century linguistics has revealed its internal structural organization as well as its many surface varieties found around the world. Psycholinguists have discovered many psychological and neurobiological aspects of language processing using different methods. However, in spite of the many applications of these techniques that can even map neural functioning as one uses language, our knowledge is limited. Although we can describe the symbolic structures of language using models, we are yet to know how it is supported by our brain's general cognition. Cognitive scientists have made substantial progress in exploring which cognitive systems influence language processing in different modalities. For example, how working memory, vision, and attention modulate the functioning of language. Chomsky has said that the study of language is part of cognitive psychology. By saying so, he situated its investigations within the ambit of psychology and biology. In last few decades, experimental psycholinguists and cognitive psychologists have unearthed which psychological mechanisms influence the processing of language. In this paper, my focus is to examine and narrate how language influences our visual perception. How comprehension of language leads to subtle variations in our cognition. Among many examples of domains where such a thing occurs, I will demonstrate with experimental examples the case of language-driven eye-movements. In particular, I will demonstrate the cross-modal nature of cognition where both vision and language interact dynamically. With the use of sophisticated techniques like the online recording of eye-movements that manifest our visual and attentional shifts, it will be obvious that the processing of language is multimodal. Both speaking and listening to language influence what we see and the concepts we activate. Albeit these conjectures may be valid at least with regard to certain specific experimental procedures this can be extended to the real world situation. The empirical investigation of language processing together with visual processing has led to a clearer understanding of our basic cognition. Although Fodor had declared that language processing is modular, it appears that it is not so. After introducing the fundamental nature of language-mediated eye-movements, I will discuss some specific experimental details that reveal such cross-modal interaction, particularly the activation of phonological and semantic systems during spoken-word processing and speaking. These results have also been extended to the processing of language in bilinguals. In sum, speaking, listening and seeing appear to influence one another dynamically during cognitive processing.

Listening and Looking: The Language Attention Interplay

We create language to describe what is there around us. Our sentences describe the relationships between objects or actions agents perform on objects. Therefore, listening to some language implicitly forces us to search for such objects described by the language. Of course, one can always make a sentence whose meaning is abstract and one cannot readily find any object in the environment that matches to its description. In most ordinary situations, we actively or even implicitly look for objects that we hear about. For example, if you are in a room and you hear the word 'fan', it's likely that you will look up towards the roof. This quick shift of attention towards the roof is an outcome of our sensorimotor experience. A lot has been written on these perceptual and sensorimotor aspects of language in cognitive linguistics (Talmy, 2000). Since sentences are about something, our visual system searches the objects as soon as we comprehend them. While the above description seems intuitive and even simplistic, empirical data demonstrating this has come only now. The precise manner in which spoken language leads us to look for objects or agents in the environment can be studied tracking eye-movements of people in naturalistic contexts. It was Cooper (1974) who first showed that as soon as people hear a word, they shift their gaze towards an object which matches the description. Cooper used eye-tracking as a method to study this dynamic relationship between spoken language comprehension and shifts in attention towards objects in the environment. He presented four or so line drawings on a computer monitor and then a speech fragment about it. For example, when people heard the word ‘Africa’, they looked at the picture of a lion immediately in the display. This shift of attention to an object which was related to the spoken word was very fast. This mechanism can be explained by assuming activation of sensorimotor experiential knowledge triggered by the language. This interaction has been one of the most central tenets of embodied cognition. Many have studied using other methods how people mentally simulate events when they listen to sentences. Thus, Cooper was the first to demonstrate that language comprehension is a dynamic multimodal process. It is not a mere symbolic computation which is amodal. Importantly, that study also opened up the possibility to further examine how and why language comprehenders look for objects in the environment described by sentences. Later researchers termed these studies that mapped language-mediated eye-movements to objects as visual world studies. These studies then claimed the very non-modular and interactive nature of language processing.

The human visual and attention system has evolved to help us find the prey and other objects of interests in the environment. One of the properties of attention is to shift continuously from one point of interest to another (Klein, 2000). Cooper had also demonstrated that people are not merely looking at the object directly referred to by the language but also at related objects. For example, early studies showed that when participants heard a word, they also looked at an object which sounded like it. Many later studies have shown that language comprehenders activate phonological, semantic as well as perceptual information about the object and they actively look for any object which is related in any manner in the environment. In one early study it was demonstrated that when people heard a word, they also looked at an object which was semantically related to that word. A comprehensive review of such studies is beyond the scope of this chapter, and the interested reader is referred to many important reviews that describe them (Huettig et al., 2011; Mishra et al., 2013).

The idea of Cooper was very powerful, at once, both from a methodological and conceptual points of view. Cooper used the commonsense understanding that language refers to things in the world including actions. Before him, the Russian psychologist Alfred Yarbus (1967), in his famous monograph titled “Eye Movements and Vision”, dealt at length with the physiological nature of eye-movements as well as on the top-down influence of context on our visual perception. Yarbus had developed his eye-movement measuring device, although very complicated compared to today’s easy-to-use video-based eye trackers. Yarbus presented participants a painting that depicted a stranger entering a room full of people, and asked participants a few questions as they looked at the painting while he tracked their eyes. His main interest was if people looked at objects of interest in the picture with regard to the questions. That is, if top-down goal influenced visual perception. Of course, the interaction between top-down and bottom-up factors and their role in cognition has a very long and contentious history in cognitive psychology and perception. Simply speaking, top-down goals are endogenous and self-driven while bottom-up forces depend on the saliency of the objects. Philosophically, one can also stretch this line of argument to basic division in human sciences into the rationalist and the empiricist traditions. Yarbus found that people's eye-movements depended on what they were evaluating in each question (Figure 1). Yarbus's technique was excellent, and it showed that where we look often is an outcome of what we want to look for. Although environmental stimuli trigger many a times eye-movements and we look at things as if automatically against our will, soon after top-down factors start playing a role. Yarbus conclusively showed that saccades (very rapid eye-movements that change point of view) and fixations (stable eye-movements) reflect our ongoing cognition. Yarbus measured what is known as a scan path for his studies. A scan path is a chronological record of a viewer's fixations as he inspects a picture. The scan path can give clear ideas about the perceptual and cognitive trajectory of his processing, moment by moment. Ever since then, cognitive scientists have used these observations as the gold standard to examine human cognition in a visual context. Thus, Cooper was already standing on the shoulders of giants when he thought of examining how spoken language may influence ongoing cognition in a cross-modal situation.

Some technical details must now be given about the main methodological aspects of Cooper's experiments before we start appreciating how influential this development was in the history of most cognitive psychology and psycholinguistics.

Unlike Yarbus, Cooper was not interested in just examining where people look, given some visual stimulus. He was interested in getting hard evidence about the fact that words activate many concepts related to them. Most contemporary psycholinguists believe in the spreading activation of concepts. That is, given one concept, all other related concepts become active during cognitive processing. Cooper also used fragments of spoken words as his primary independent variable in his studies. By giving participants spoken words and then presenting a display that contained line drawings, he made the

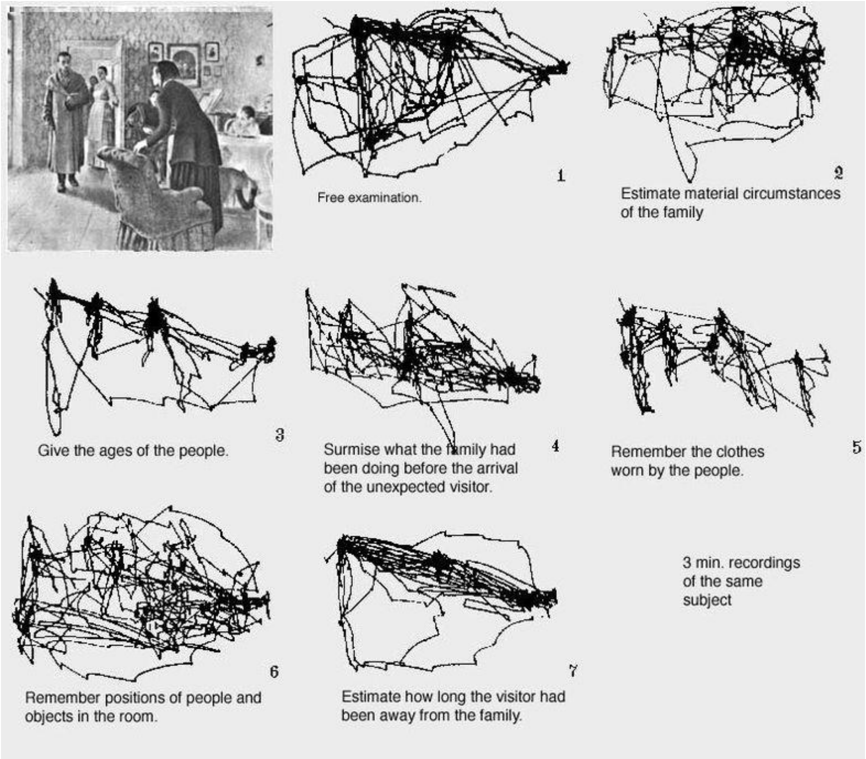
1
Free examination.
2
Estimate material circumstances of the family
3
Give the ages of the people.
4
Surmise what the family had been doing before the arrival of the unexpected visitor.
5
Remember the clothes worn by the people.
6
Remember positions of people and objects in the room.
7
Estimate how long the visitor had been away from the family.
3 min. recordings of the same subject

Figure 1. Russian psychologist Yarbus's experimental results.

experiment cross-modal. His main aim was to examine the magnitude and extent of concept activation as reflected by the eye-movements. It is not enough just to conclude that language perceivers activate concepts related to what they hear but the key question is when do they activate this. Thus, the temporal aspect of the issue became pivotal in all later use of this paradigm. Cooper also very cunningly manipulated the objects and their significance as well as their interrelationship with the spoken word's meaning in his design. While the participants saw the object which the language described, there was also a related object in the display. This allowed Cooper to measure the time course of concept activation in relation to the target. Target in this parlance is what is referred to by the spoken language. In most modern forms of this visual world paradigm, authors have shown that language users also activate objects related to the spoken word in the absence of the referred object in the display (Huettig, Rommers, & Meyer, 2011; Mishra, Olivers, & Huettig, 2013). This makes sense since often in the real world we do not see the object mentioned by the language as such in our vicinity but something that is related to the object. Evolutionarily, this must have been preferred by the system to make us more alert and careful. Cooper also showed that semantic and conceptual information is active instantly with the spoken word presentation. In the many modern uses of this paradigm, dubbed as visual world paradigm, researchers still use the basic assumptions of Cooper albeit with some modifications. They still measure activation of concepts as one listens and looks at a display. The paradigm is used now to answer key questions in psycholinguistics, and the examples are many. The paradigm has not only been used in answering fundamental questions related to time course of concept activations during spoken-word recognition but also more deep questions about mental states (Altmann & kamide, 2009) and many more. Much of this history with regard to current uses and their theoretical significance have been captured in Mishra (2015).

Apart from Cooper, it was the eye-mind hypothesis by Just and Carpenter (1976) which brought different variables together in understanding what eye-movements reflect about cognition. Just and Carpenter used eye-tracking in understanding linguistic and cognitive processes in reading. The eye-mind hypothesis claimed that the locus of our gaze reflects what is on our minds at that moment in time. The locus of the gaze that happens for a sustained period also can be taken as the locus of selective attention. Thus, in the context of reading which is an acquired cognitive skill unlike speaking and listening, readers swiftly move across words as they acquire information for comprehension. These rapid movements are nevertheless irregular and uninterpretable. The important work by Keith Rayner (1975) on this has revealed how eye-movements show the very subtle aspects of linguistic comprehension during reading. Note that the eye-mind hypothesis was different from the way Yarbus had explained his results. Of course, the paradigms differed in important ways, but both approaches have used eye-tracking as their main method. Both were interested in knowing how eye-movements reveal about ongoing cognitive processing. Just and Carpenter measured saccades and fixations as readers read the text for comprehension. More importantly, they found that readers are not always looking at each word as they are reading ahead. There is a certain amount of automaticity to be observed during eye-movements in reading. Although what I describe later is more on the lines of Cooper and Yarbus (I discuss top-down goals), I also think that knowledge gained from eye-movement analysis in reading has offered very influential theorization about the dynamics of cognition. Reading being a complex visuo-linguistic process is well suited to measure the temporality of cognition. One important finding in the field has been that fluent readers always acquire information ahead of their eye-movements to those parts through parafoveal processing. That is, they intuitively know what comes next using anticipatory processes, and this often influences how long they are staying at the current location. But reading as a paradigm has its constraints, as is well known. Primarily because a vast amount of individual differences is observed with regard to reading fluency and reader's attention span. Reading is not as easy and natural as listening and speaking. Further, many have developmental dyslexia and also poor reading because of various reasons starting from socio-economic to cognitive. Nevertheless, the eye-mind hypothesis certainly helped pitch the focal point of using eye-tracking to measure active cognition.

If one reads the history of this fascinating multidisciplinary work on perception, cognition, eye-movements and the use of language, it will become clear that many cognitive psychological constructs have been invoked far too often in explaining the results. One such example is the construct called attention. It's clear that when we are inspecting something visually, we are paying attention to it in a layman's terms. However, there is considerable disagreement on this simple theorization till date. Hoffmann and Subramanium (1995) did find evidence for the claim that eye-movements indeed indicate attentional locus. However, what about parafoveal perception and looking around randomly when we acquire information from our surroundings? The point I am trying to bring home is if attention is centrally deployed when I am looking at something with respect to some spoken language. Cooper in his original work had not dealt with this point at length, and it was taken up by psycholinguists only later. For example, it is now known that both working memory and attention (Altmann & Kamide, 2007; Huettig, Olivers, and Hartsuiker, 2011) are involved in such language-mediated eye-movements, as seen in the Cooper type of experiments. The key role of language in centering attention has been well developed by cognitive linguists (Talmy, 2000). For example, Talmy found that a sentence has a figure-ground construct like the way we see natural scenes. What is emphasized in a sentence is what is under attentional focus. Jackendoff (1987) in his many theorizations has also compared linguistic processing to visual processing. According to Jackendoff, language transduces what we grasp from visual processing. Although language used to describe the visual world is not always enough and cannot be so, it helps bring attention to things that matter. For example, as is well-known, the many deictic and referential systems employed by world's languages demonstrate this function of language. This also includes prepositions that tell us where to look for something in the surrounding. Therefore, spoken language channelizes attention in the environment which in turn is reflected by eye-movements. Although the exact mechanism that connects eye-movement programming to the attentional mechanism is beyond the scope of this paper, it can be asserted that human language system is a powerful manifestation of our core attentional system. Language expresses even in dynamic conversational situations what is important then, and speakers and listeners cooperate accordingly. Thus, the role of attention in understanding the key interplay between visual perception and linguistic analysis is very crucial. Finally, eye-movements, the way we measure today in the so-called visual world paradigm, reflect the combined dynamic interplay between language, vision, and attention (interestingly, the theme of the monograph by Mishra, 2015).

Michel Spivey in his 2007 book *The Continuity of Mind* emphasizes the dynamic aspects of cognition which online methods like eye-tracking record. Cognition is never all or none. For example, even during sentence processing our cognitive system entertains many interpretations of the sentence before it settles on one. This has been amply demonstrated in the extensive research on ambiguous sentence comprehension. Since language is often ambiguous and context-bound, what we interpret at which moment in time depends on many factors. Again, Fodor, in his essay on Modularity, had not considered this possibility and had settled for the view that structural interpretations are cognitively impenetrable. However, it has been shown many times that our comprehension system always considers possible alternatives of interpretation before finally rounding off one as the one appropriate for the moment. The visual world eye-tracking paradigm allows us to capture this moment-by-moment nature of competition among the alternatives which is a hallmark of human cognition. Reaction time studies do not allow us this possibility since their data often indicate the very end stages of cognitive decision-making. Tanenhaus and colleagues, in the context of sentence processing using visual world had demonstrated that what we hear and what is in front of us at the moment can dynamically alter our interpretation. Similarly, the use of visual world paradigm in spoken-word comprehension also shows that even when we hear a word, we activate all possible competitors related to this word. This, at once, shows the fluid yet automatic nature of cognitive processes. Although context can modify the interpretations of words, yet competition does occur. This is what I will demonstrate later using one experimental example from my studies where the visual world paradigm was used in ambiguous homophone processing in the Hindi language. Likewise, language comprehension is also all the time massively predictive. Listeners generate many probable representations when they actively process any fragment of language. Typical entities of languages like adjectives and certain case markers can help generating such predictions. For example, in Hindi, adjectives are gender congruent with the nouns they modify. When listeners are presented with such adjectives, they can anticipate the appropriate nouns that are good enough for them. This was also demonstrated by Mishra and Singh (2014) in an eye-tracking visual world study in Hindi. Thus, the visual world method allows us to capture at once the dynamic and evolving nature of cognition and also crucial process like prediction and anticipation.

Both anticipation and simultaneous consideration of alternatives during language interpretation are now considered a regular feature of language processing. Early work by Altmann and colleagues on anticipation using the visual world eye-tracking method showed that human participants could predict an event's outcome by listening to sentence fragment. Listeners can also predict the prototypical attributes of an agent using their contextual knowledge. For example, in one study Altmann and Kamide (2007) presented a display which had a girl and a man with a bike as agents, along with the picture of a candy and another object. When participants heard the sentence that began with the fragment 'the girl will eat …", most participants looked at the candy. Similarly, when the sentence began with the sentence "the man will …" they orientated their gaze towards the motorcycle. This shows that listeners used their contextual knowledge of the real world in shifting their attention. Therefore, this was not just an act of structural interpretation of the sentences or predicting its semantics but situating that comprehension in the environment itself. This embodied, and sensorimotor angle to sentence comprehension in the presence of visual scenario offers rich understanding into the dynamics of language comprehension. Similarly, in another study, Altmann and colleagues examined if listeners use mental simulation to map changes in event state. They presented pictures where either an empty or a half empty glass was seen. Participants listened to a sentence that began "the man will drink the beer…" Immediately listeners started to look at the half empty glass in anticipation. This demonstrates that language users mentally simulate events and change of states as they incrementally listen to the language. Eye-tracking evidence could demonstrate not just when participants started to look at the object that confirmed their predictions but for how long. Such evidence also corroborates other findings that have shown mental simulation during language comprehension. The key point here is anticipation and prediction which seems to motivate the eye-movements towards such objects that are relevant. Thus, eye-movements measured in such a scenario don't just indicate if people are comprehending the language but also their predictive strategies.

Context and Ambiguity in Spoken Sentence Processing

Much of what I have said so far dealt with both methodological and conceptual aspects of the visual world paradigm. Along with it, I mentioned that this allowed the study of contextual language processing in a systematic manner where eye-tracking data provide valuable online measures. More importantly, using information processing cross-modally also allows us to study anticipatory and predictive processes during language processing. Furthermore, this can be useful to study how language users consider alternative meanings that may not be appropriate during the comprehension process.

Ambiguity during sentence processing may arise because of the way we process words. Take, for example, the English word "pen", which has a dominant meaning of a writing instrument and a non-dominant meaning of 'enclosure'. Dominance here is linked to frequency of use in everyday speech. A long-standing debate in the psycholinguistics of lexical ambiguity resolution has been the effect that the sentence context has on lexical ambiguity resolution. In the context of lexical homophones, one may wonder if the primary (dominant) meaning and secondary (non-dominant) meanings of homophones interact differently with context. People access the dominant meaning, e.g. pen as writing instrument instantly when they find it in sentences. However, they also access the non-dominant meanings of such homophones. Many studies have found that both dominant and non-dominant meanings are active at the same time, however, to different degrees. Already it has been elaborated that such joint activations of concepts during language comprehensions are a norm of such cognition than deviance. The question is, do the subordinate meanings of ambitious homophones get activated even when the sentence context is further biased towards the dominant meaning? Does enriching the context towards one meaning stop the activation of the other irrelevant meaning?

Mishra and Singh (2014) examined this issue using ambiguous homophones in Hindi and manipulating the sentence context. For example, a word like '*choti*' has two meanings. One meaning of '*choti*' is 'hair lock', and another is 'hilltop'. Other researchers had observed that the dominant meaning is active regardless of context (Kambe, Rayner, and Duffy, 2011). Also, even after any contextual bias, the non-dominant meaning is still activated to some extent (Duffy, et al. 1998). Others have argued that if prior context is sufficiently biased towards one meaning of any ambiguous homophone, then the other non-dominant meaning may not be active at all (Simpson, 1981). Many researchers had studied these using reading as a model. The visual world eye-tracking paradigm has been used to measure online activation of dominant and non-dominant meaning activations in the case of ambiguous homophones. For example, Huettig and Altmann (2007) presented participants a display containing line drawings that had a shape competitor of an ambiguous homophone "pen" with its dominant meaning along with distracters. Critically, the presented spoken sentences had a boosted activation of the subordinate meaning. For example, the sentence: "the welder locked up carefully, but then he checked the pen." The question was if such a strong contextual bias towards the subordinate meaning will eliminate the activation of the dominant meaning. The visual world paradigm allowed the experimenters to map activations online in the form of the proportion of fixations to different objects over time. The data showed that even when the sentence was biased towards the subordinate meaning, the shape competitors of the dominant meaning were still not ignored. This showed the pervasive nature of lexical activation even when the context clearly mandates the activation of one.

Mishra and Singh (2014) wondered what if one gives a further boost to the dominant meaning of an ambiguous homophone in the sentence; will it completely subside the activation of the subordinate meaning? If subordinate meaning activation persists, then it will indicate a complete context-independent mechanism of lexical activation. They too used the eye-tracking visual world paradigm like Huettig and Altmann (2007), and explored the effects using shape and semantic competitors. The distinction between them is perceptual, not conceptual. Shape similarity is based on low-level perceptual analysis, for example, the similarity between a coin and the moon. They both are roundish objects, but they don't share any conceptual or lexical similarity. However, there are objects that share semantic similarity but are perceptually different, e.g. a goat and a cow. Thus, Mishra and Singh (2014) wanted not just to explore if the contextual boost to the dominant meaning overrides any subordinate activation, but how such a competition will be seen in eye-movements when we have objects that are either perceptually or semantically matching. Huettig and Hartsuiker (2007) earlier had established the time course of activations of such concepts during spoken-word recognition. Mishra and Singh (2014) used Hindi homophones. One of the meanings could be considered as dominant and the other subordinate. This fact was tested through ratings done by Hindi native speakers. Below I give sample sentences taken from Mishra & Singh (2014)

Example of Sentences Used:

1. Neutral sentence

   'Bato bato me choti ki charcha hone lagi.'
   Talk in mountain peak/hair lock about discussion began

   'While talking the discussion on ***choti*** (mountain peak/hair lock) began'

2. Biased sentence

   Himalaya parbat ki choti aath hajar meeter uuchi hai
   'Himalaya mountain's peak eight thousand meter high is'

   'The ***choti*** (peak) of Himalaya is at the height of 8000 mts (biased sentence with dominant meaning 'mountain peak')'

As is evident, one sentence was biased towards the dominant meaning while the other was neutral. The use of the word "neutral" here as far as activation goes can be tricky. Even for neutral it's assumed that the dominant meaning will get most activation compared to the subordinate meaning. The following figure (Figure 2a) represents a sample trail used in the experiment.

The results in the form of eye-movements over time to different objects for different sentence conditions showed a very interesting pattern (Figure 2b). The initial bias towards the dominant meaning did lead to low activation of the subordinate meaning. Listeners looked at the shape competitors of the critical homophone words in both sentence contexts more than the unrelated distractors. However, such eye-movements were higher for the neutral condition than the dominant bias condition. This pattern of results suggests that while context may modulate activation of irrelevant lexical items, it cannot eliminate them. Language users, thus, seem to activate all possible meanings of ambiguous words simultaneously and after some competition settle for one. As has been described before,

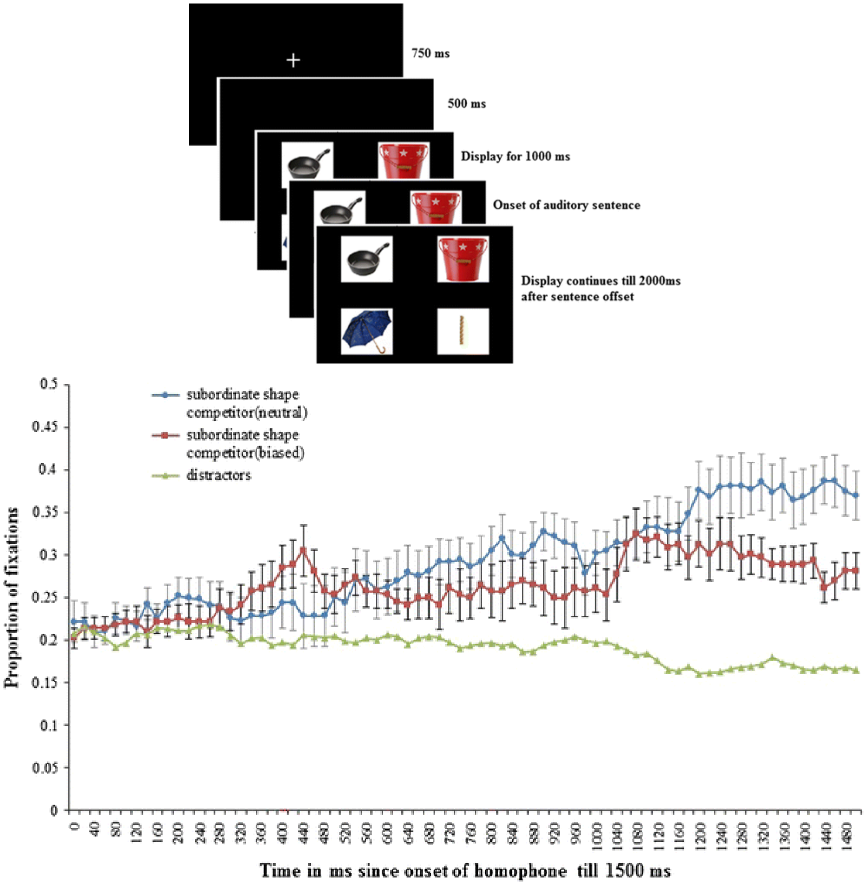

Figure 2. (a) Sample display showing four objects and trial sequence. After a central fixation, participants see four pictures on the computer screen. After a delay of 1,000 ms, a spoken sentence containing an ambiguous homophone is presented. (b) Proportion of fixation plots for different objects in the display after critical spoken word onset.

the main contribution of the visual world eye-tracking data lies in capturing the online dynamics of lexical competition in the form of eye-movements towards visual objects. Mishra and Singh (2014) also did a second experiment where they used semantic competitors of the subordinate meanings. This experiment was similar in every aspect compared to the first one. The results showed that just like the first experiment, listeners still looked at the semantic competitor of the subordinate meaning when the context was biased towards the dominant meaning. These results thus demonstrate that both perceptual information and semantic information are activated for subordinate meanings even when the context supports the dominant meaning.

Of course, there are some caveats with the visual world method, and they often may be considered as its limitations. One argument has been that here we present the pictures in an artificial manner and in the real world spoken words occur amid a wider array of visual objects. One of the objects which are related to the spoken word is strategically placed among distracters. This may lead to strategic processing and looks among the listeners. They may know that one picture is related and therefore may look at it preferentially. For instace, in the above example, the picture that was related to the shape of the subordinate interpretation. However, this language-mediated eye-movements occur at such rapid time scale that it's difficult to conceive of any strategy (Salverda and Altmann, 2011). Secondly, when asked later, most participants seem not to note any relationship. Since there are many participants and many trials, any such strategy in some is cancelled out when grand averages are prepared. The activation patterns then give an unambiguous record of lexical activation. Such activations suggest that lexical activations during language processing are unconstrained and if at all the constraints show their effect later, in the course of time.

Individual Difference and Language-Mediated Eye-Movements

At this point, most researchers are concerned about accounting for their data with regard to individual differences. One major objection to much of the data in psycholinguistics and cognitive psychology in the last several decades has been that they come from only very particular types of populations. In other words, invariably researchers take their participants from university students, who of course, have a very high degree of literacy and language skills. Furthermore, basic cognitive systems like attention, working memory and visual perception including familiarity with computers (most experiments are computer-based) is very high with this population. Therefore, based on this data, we do not know how such results will be with other populations, for example, say whose who lack formal literacy or even those who have no computer training. Individual differences include an understanding of such basic cognitive factors of an individual and how together they influence that individual's cognition. Of course, healthy children and healthy adults differ on a wide range of tasks because children's cognitive system is still at an evolving stage. We see also a wide range of scores among adults, since not all have similar working memory capacity and attention abilities. Much research that has explored sentence processing with regard to working memory has shown this already. Similarly, adult students who have any developmental reading impairment tend to perform poorly on many psycholinguistic tasks. Therefore, any deep and wholesome understanding of cognition, and when it comes to language, processing has to include an in-depth appreciation of individual differences. It's only now that such comparative studies are taking place and the differences obtained are stark.

Individuals may differ very significantly on their cognitive abilities based on their literacy. Many studies have shown that illiterates perform poorly on tasks that require visual discrimination and also language-based tasks. The relationship between acquisition of literacy and overall cognition is well established (Huettig and Mishra, 2014). Unfortunately, in many countries in the world as in India, a very large percentage of the general population is illiterate. Studies have shown that acquisition of reading enhances visual attention and working memory. Therefore, a fundamental difference in cognitive processing may emerge between a literate and an illiterate person. Mishra and colleagues have been studying psycholinguistic processing among illiterates as compared to literates, using methods such as eye-tracking (Huettig, Singh, and Mishra, 2011; Mishra, Singh, Pandey, and Huettig, 2012; Olivers, Huettig, Singh, and Mishra, 2014) and also brain imaging (Skeide, et al. 2017). The limited amount of psycholinguistic studies that have happened in India are with again university students. Therefore, these studies can't reveal how the findings apply to illiterates. Although previously researchers had studied cognitive deficits in illiterates, very few had studied online language processing in such a population. In the first study of its kind, Mishra and colleagues (Huettig et al., 2011) compared illiterates and literates using a visual world task semantic and phonological activation. In that study, literates and illiterates were presented with spoken sentences and pictures on a computer monitor. One of the pictures was a phonological neighbor in one study, and another was a semantic neighbor in a different experiment. The results showed that illiterates were slow in activating these related words as compared to the literate as revealed in eye-movements. This slowness can be interpreted as the result of either poorer working memory or ability to integrate visual and linguistic information online. Most recently Huettig and Janse (2016) have shown that working memory capacity influences the magnitude of language-mediated eye-movements. It's not as such a slowness of spoken language processing but slowness in activating the many related concepts dynamically. Such data reveal that psycholinguistic processing in the illiterate may suffer as a result of the absence of literacy. More recently, brain-imaging data has also shown that functional connectivity in the illiterate brain among areas that process language and visual information is weak (Dehaene et al., 2010). A more recent work that examined the effect of long-term literacy training on illiterates' brain networks shows increasing functional connectivity. Taken together, literacy can be considered as a major factor indicating individual difference when it comes to explaining psycholinguistic and other cognitive processing. Below I describe a study where illiterates and literates were compared in a task to measure the difference in anticipatory eye-movements using eye-tracking.

Mishra, Singh, Pandey and Huettig (2012) examined language prediction in illiterates and literates using the visual world eye-tracking paradigm. Prediction has now been understood as a major mechanism which explains language processing (Pickering and Garrod, 2007). Prediction arises from experience with language use and helps in anticipating further during language processing. For example, most listeners and readers can anticipate a word that is yet to come in a sentence using their knowledge attained so far. The classic 'cloze' task used in sentence processing examines such predictive processing. It's important to note that prediction is not only used in language processing but most other types of cognitive processing. For example, Singh and Mishra (2016) demonstrated that bilinguals could anticipate a future motor action based on their current understanding of the task in an oculomotor attention task. Prediction during sentence processing is context-bound. Often users of a certain language can predict with a great amount of certainty upcoming words, taking cues of elements specific to that language. For example, the case markers present in Indo-Aryan languages like Hindi can alert listeners what to expect further. For example, at a syntactic level, case markers in Hindi such as 'ko', 'se' and 'ne', when attached to a head noun (agent/subject) at the beginning of the sentence can predict further verbal additions appropriate for such a construction. Similarly, gender markers in Hindi can help listeners to anticipate other nouns that agree with such genders during spoken language processing. Mishra et al. (2012) examined if illiterates and literates differ in their prediction during spoken language processing using eye-tracking. They exploited the fact that in Hindi, adjectives that come before nouns and modify them also agree in gender with those nouns. Further, additional elements like 'wāla' and 'wāli' used in Hindi constructions have to agree in gender with adjectives and nouns. Below, a sketch is given of the stimuli used, and the logic pursued (Figure 3).

In Hindi, nouns are gender-marked. Adjectives modify nouns, and they also copy the gender endings of the nouns. For example, the adjective 'uncha' (high) modifies the noun, e.g. 'darwaja' (door). So when someone utters a sentence like "woh uncha wala…", it is likely that listeners will search for a noun which is masculine and for which such an adjective is appropriate. Thus, it's a simultaneous evaluation of physical attributes and gender agreement. The study asked if by listening to such fragments, Hindi listeners could predict the correct nouns when they see objects on a computer screen. If listeners can anticipate the appropriate nouns, then they will look at such objects preferentially which can be tracked using eye-movements. The design was simple enough to be used with illiterates, as it did not involve any knowledge of written language. Figure 3 shows sample trial, showing the spoken language fragment used with such a display. To generate any useful inference, these kinds of studies require very rigid control of the stimuli. For example, we normed all the adjectives and the nouns for their gender by Hindi native speakers

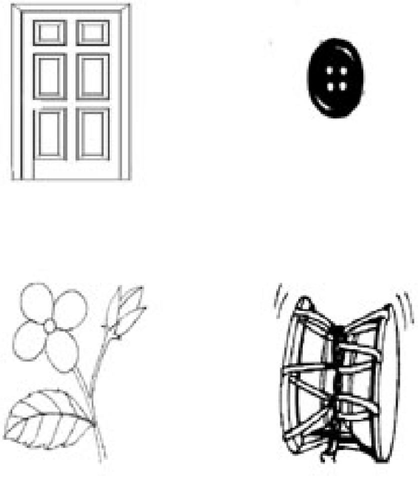

Figure 3. The figure shows four line drawings. One of the drawings is that of a 'door'. Participants listened the sentence fragment 'abhi aap ek uncha waala darwaaja dekhenge', literally: Right now, you are going to a high door see —You will now see a tall door. We measured eye-movements starting with the adjective towards the objects. The other three objects are unrelated distracters.

who did not participate in the main experiment. We asked these native speakers which noun they will choose as the most frequently used noun with such adjectives. It was observed that participants were choosing nouns that were gender congruent. Further, the line drawings were also rated for their acceptability with the names used for them. The task was simple enough where they had to just listen to the sentences and look at the computer monitor. Eye-movements were measured continuously.

The figure below (Figure 4) shows the proportion of fixations for targets and distracters for illiterates and literates. It is evident from the data that soon after the particle onset, literates started orienting their eyes towards the correct noun that was appropriate. This deployment of attention kept on increasing for the literates as time passed. However, for the illiterates, we do not see such bias in attention emerging after the particle onset, and also the overall fixations are low. We can conclude that the literates could predict fast the appropriate noun on the display with the adjective and particle, whereas the illiterates only look at the nouns when they heard this. Therefore, it appears that literacy levels influence predictive processing during spoken language comprehension among other factors. Such online data provide strong evidence first and foremost of the fact that listeners indulge in predictive processing using the information processed from current input. They do not wait till the complete information arrives in the acoustic stream. Why do the

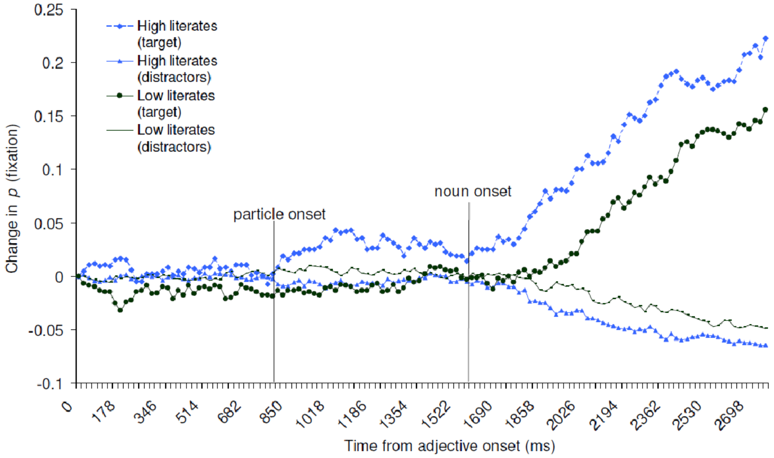

Figure 2. Changes in fixation proportions on the target objects and (averaged) unrelated distractor objects for low literates and high literates. Zero on the timeline is the acoustic onset of the adjective.

Figure 4. Change in proportion of fixations to targets and distractors for high and low literates.

illiterates show deficits in such processing when the task was easy and non-demanding?

The slowness that we saw in the illiterates in their predictive eye-movements may have many reasons. It's unlikely that they did not know the nouns or their genders. Researchers have found that illiterates are slower in naming line drawings (Reis, Petersson, Castro-Caldas and Ingvar, 2001). Is it so that illiterates did figure out which object to look at but were uncertain with regard to its phonological form? Notice that the task used involved only comprehension, not production. In the context of visual world paradigm, a long-running issue has been if listeners name objects covertly even when no production is called for, and this retrieval of phonological form mediates the eye-movements. If illiterates are slow in this mechanism, then this can explain why their eye-movements were slower. Similarly, it is possible that illiterates are slower in basic visual perception. For example, the Russian Psychologist Luria (1976) had found evidence that illiterates are bad with optical illusion. Huettig and Mishra (2014) offer an extensive historical review of the literature on the linguistic and cognitive deficits that have been found with illiterates. No one before Mishra, et al. (2012) had examined cross-modal processing in illiterates. Therefore, we can't be entirely certain without additional evidence that the slowness we found with illiterates is because of their difficulty in naming line drawings or visual perception alone. It has been suggested that the fixation proportions measured as the dependent variable in the visual world task are an outcome of both linguistic and visual processing (Huettig, Mishra, and Olivers, 2011; Huettig et al., 2011). Working memory capacity plays a role in strengthening this connection between visual and linguistic processing in such a cross-modal situation.

Discussion and Conclusion

I began this paper with an introduction to eye-movements and particularly the visual world paradigm which has been used quite extensively to study cross-modal cognition. The paradigm's strength lies in the fact that it captures moment-by-moment the online nature of cognition. It captures the alternative considerations on the minds of subjects during processing which traditional methods like reaction time could not capture. Further, using this we not only can know how visual information influences linguistic processing but how linguistic information influences visual cognition. The data provide very rich information through mapping of eye-movements if language users are activating lexical units that are not task-related.

This is what Spivey has described in his book, *The Continuity of Mind*. Language processing is then essentially cross-modal and situated. Much of what I have said also coheres with theories from situated and embodied cognition. Classical linguistic and psycholinguistic analysis studied language in isolation from other sensory effects. However, today it is well-recognized that what we speak and what we understand through language uses rich sensory data from other modalities. Many have also studied the conceptual basis of language production using the visual world paradigm. This was not discussed at length in this paper as it was beyond its scope. The paradigm has now been applied successfully in child language research (Holzen and Mani, 2012) and also to understand disordered speech. Below, I make some general observations regarding the data from the two experiments which were presented, and their underlying theory, if any.

Experiment one explored the dynamic influence of prior information in the sentence and its effect on activation of lexical items. In this case, when the sentence was further biased towards the dominant meaning, listeners still activated the subordinate meaning of an ambiguous homophone. This more or less happened when the competitor was presented either as a shape similar or a semantically related object. The eye-movements demonstrate that listeners could activate both a perceptual feature and a semantic feature of the task-unrelated word as they listened to sentences. This kind of data could not be captured with more traditional methods like reaction times or sentence recall, without sacrificing the precision. Thus, in this experiment, the language-mediated eye-movements revealed the online tussle between different representations that we entertain as we listen to words in a sentence. Sentential syntax had no role to play either facilitating or constraining such an effect. Such spurious activations are ever present during everyday language processing. This evidence is in sharp contrast to the assumptions of the more traditional psycholinguistics. We still do not know why such spurious activations happen during language processing when extraction of one meaning appropriate for the context is the key to successful comprehension. Thus, in sum, language-mediated eye-movements as seen in a visual world task reveals how human language processing device entertains all kinds of considerations as it finally settles for one meaning. Similarly, when we do a traditional sentence-comprehension task and ask if the reader agrees or disagrees with certain interpretations, we do not know if he can consider the alternative interpretation at some point in time. This was shown first with the classic ambiguous sentence “the horse raced past the barn fell”. There, the explanation was more syntactic. However, it’s possible that even when syntax constraints structural interpretations, listeners may still activate all other meanings momentarily.

How these results change our views about language processing in cross-modal context? First and foremost, they suggest that language processing is not modular in the usual sense of the word. Language processing uses all those cognitive processes that are used for other non-linguistic processes. For example, prediction and anticipation are processes that are used by other action systems. Language users use their everyday knowledge to predict and anticipate events or objects to be described by language. We also saw that language processing is constrained by several individual difference factors. Those who are highly literate and whose are illiterates process language differently. Although at this point it’s not possible to pinpoint the exact factors, it is clear that there is large variation among the population. This should alert us to how we do our psycholinguistic experiments and to what extent we can expect homogeneity.

References

Altmann, G. T. and Y. Kamide. 2007. “The Real-Time Mediation of Visual Attention by Language and World Knowledge: Linking Anticipatory (and Other) Eye Movements to Linguistic Processing. *Journal of Memory and Language, 57*(4), 502-518.

Altmann, G. T. and Kamide, Y. 2009. “Discourse-Mediation of the Mapping between Language and the Visual World: Eye Movements and Mental Representation”. *Cognition, 111*(1), 55-71.

Cooper, R. M. 1974. “The Control of Eye Fixation by the Meaning of Spoken Language: A New Methodology for the Real-Time Investigation of Speech Perception, Memory, and Language Processing. *Cognitive Psychology, 6*(1), 84-107.

Dehaene, S., F. Pegado, L. W. Braga, P. Ventura, G. Nunes Filho, A Jobert,... and L. Cohen. 2010. “How Learning to Read Changes the Cortical Networks for Vision and Language. *Science, 330*(6009), 1359-1364.

Hoffman, J. E. and B. Subramaniam. 1995. “The Role of Visual Attention in Saccadic Eye Movements”. *Attention, Perception, and Psychophysics,* 57(6), 787-795.

Von Holzen, K., and N. Mani. 2012. “Language Nonselective Lexical Access in Bilingual Toddlers. *Journal of Experimental Child Psychology,* 113(4), 569-586.

Huettig, F., J. Rommers, and A. S. Meyer. 2011. “Using the Visual World Paradigm to Study Language Processing: A Review and Critical Evaluation”. *Acta Psychologica,* 137(2), 151-171.

Huettig, F. and G. T. Altmann. 2007. "Visual-Shape Competition during Language-Mediated Attention is Based on Lexical Input and not Modulated by Contextual Appropriateness". *Visual Cognition,* 15(8), 985-1018.

Huettig, F., and E. Janse. 2016. "Individual Differences in Working Memory and Processing Speed Predict Anticipatory Spoken Language Processing in the Visual World". *Language, Cognition and Neuroscience, 31*(1), 80-93.

Huettig, F., N. Singh, and R. K. Mishra. 2011. "Language-Mediated Visual Orienting Behavior in Low And High Literates". *Frontiers in Psychology,* 2, 285.

Huettig, F., and, R. K. Mishra. 2014. "How Literacy Acquisition Affects the Illiterate Mind: A Critical Examination of Theories and Evidence". *Language and Linguistics Compass, 8*(10), 401-427.

Huettig, F., R. K. Mishra and C. N. Olivers. 2011. "Mechanisms and Representations of Language-Mediated Visual Attention". *Frontiers in Psychology,* 2.

Huettig, F., C. N. Olivers and R. J. Hartsuiker. 2011. "Looking, Language, and Memory: Bridging Research from the Visual World and Visual Search Paradigms. *Acta psychologica, 137*(2), 138-150.

Jackendoff, R. (1987). "On beyond Zebra: The Relation of Linguistic and Visual Information". *Cognition.* 26: 89-114.

Just, M. A. and P. A. Carpenter. 1976. "Eye Fixations and Cognitive Processes". *Cognitive Psychology,* 8(4), 441-480.

Kambe, G., K. Rayner and S. A. Duffy. 2001. "Global Context Effects on Processing Lexically Ambiguous Words: Evidence from Eye Fixations". *Memory & Cognition,* 29(2), 363-372.

Klein, R. M. 2000. Inhibition of Return". *Trends in Cognitive Cciences,* 4(4), 138-147.

Luria, A. R. 1976. *Cognitive Development: Its Cultural and Social Foundations.* Harvard University Press.

Mishra, R. K., N. Singh, A. Pandey and F. Huettig. 2012. "Spoken Language-Mediated Anticipatory Eye-Movements are Modulated by Reading Ability-Evidence from Indian Low and High Literates. *Journal of Eye Movement Research, 5*(1).

Mishra, R. K., C. N. Olivers and F. Huettig. 2013. "Spoken Language and the Decision to Move the Eyes: To What Extent are Language-Mediated Eye Movements Automatic?" *Progress in Brain Research: Decision Making: Neural and Behavioural Approaches.* Elsevier. 135-149.

Mishra, R. K. 2015. *Interaction Between Attention and Language Systems in Humans.* Springer.

Mishra, R. K. and N. Singh. 2014. "Language Non-Selective Activation of Orthography during Spoken Word Processing in Hindi–English Sequential Bilinguals: An Eye Tracking Visual World Study. *Reading and Writing, 27*(1), 129-151.

Mishra, R. K. and S. Singh. 2014. "Activation of Shape and Semantic Information during Ambiguous Homophone Processing: Eye Tracking Evidence from Hindi". *Cognitive processing*, 15(4), 451-465.

Olivers, C. N. L., F. Huettig, J. P. Singh and R. K. Mishra. 2014. "The Influence of Literacy on Visual Search". *Visual Cognition*, 22(1), 74-101.

Pickering, M. J. and S. Garrod. 2007. "Do People use Language Production to Make Predictions during Comprehension?" *Trends in Cognitive Sciences*, 11(3), 105-110.

Rayner, K. 1975. "The perceptual Span and Peripheral Cues in Reading". *Cognitive Psychology*, *7*(1), 65-81.

Reis, A., , K. M. Petersson, A. Castro-Caldas and M. Ingvar. 2001. "Formal Schooling Influences Two-But not Three-Dimensional Naming Skills". *Brain and Cognition*, 47(3), 397-411.

Salverda, A. P., and G. Altmann. 2011. "Attentional Capture of Objects Referred to by Spoken Language". *Journal of Experimental Psychology: Human Perception and Performance*, 37(4), 1122.

Singh, J. P. and R. K. Mishra. 2016. "Effect of Bilingualism on Anticipatory Oculomotor Control". *International Journal of Bilingualism*, *20*(5), 550-562.

Simpson, G. 1981. "Meaning Dominance and Semantic Context in the Processing of Lexical Ambiguity". *Journal of Verbal Learning & Verbal Behavior*, 20, 120-136.

Skeide, M. A., U. Kumar, R. K. Mishra, V. N. Tripathi, A. Guleria, J. P. Singh, ... and F. Huettig. 2017. "Learning to Read Alters Cortico-Subcortical Crosstalk in the Visual System of Illiterates. *Science Advances.* Vol(No)

Spivey, M. 2008. *The Continuity of Mind.* London: Oxford University Press.

Yarbus, A. L. 1967. *Eye Movements during Perception of Complex Objects.* US: Springer171-211