Developmental changes in allocation of visual attention during sentence generation: an eye tracking study*

Cambios del desarrollo en la localización de la atención visual durante la generación de oraciones: un estudio de rastreo ocular

Recibido: junio 1 de 2012 | Revisado: agosto 1 de 2012 | Aceptado: agosto 20 de 2012

Ramesh Kumar Mishra **
Allahabad University, India

Abstract

To Look and speak requires a dynamic synchronization of both visual attention and linguistic processing. This study explored patterns of visual attention in a group of Hindi speaking children and adults, as they generated sentences to real photographs. Photographs contained either a single human agent performing an intransitive action, an agent performing an action with an object or two actors involved in a mutual action in the presence of an object. The eye movements were recorded as participants generated sentences for each photograph, and several dependent measures were calculated. Eye movements to subject and verb regions in each picture revealed striking differences between children and adults as far as deployment of visual attention was concerned. Adults deployed significantly higher amount of attention to the verb region during the conceptualization process and throughout viewing compared to children. Children had higher number of fixations and saccades to different regions but did not attend to the regions in a stable manner over time. The results suggest that in a verb final language like Hindi, generating sentence requires first allocation of attention to the region denoting action, and children and adults differ from each other in this process.

Key words authors
Visual attention, Scene perception, Sentence production, Hindi, Eye movements.
Key words plus
Perception, Quantitative Research, Cognitive Science.

Resumen

Mirar y hablar requieren una sincronización dinámica de la atención visual y el procesamiento lingüístico. Este estudio exploró los patrones de atención visual en un grupo de niños y adultos hablantes de Hindi cuando generaban oraciones de fotografías reales. Las fotografías contenían un único agente humano realizando una acción intransitiva, un agente realizando una acción con un objeto o dos actores implicados en una acción mutua en la presencia de un objeto. El movimiento ocular fue registrado mientras los participantes generaban oraciones para cada fotografía y se calcularon algunas medidas dependientes. El movimiento ocular para las regiones de sujeto y verbo en cada figura revelaron diferencias notables entre niños y adultos en cuanto al desempeño de la atención visual. Los adultos desplegaron una cantidad significativamente mayor de atención a la región verbo durante el proceso de conceptualización y durante todo el proceso en comparación con los niños. Los niños tuvieron mayor número de fijaciones y movimientos sacádicos de diferentes regiones, pero no atendieron a las regiones de manera estable en el tiempo. Los resultados sugieren que en un lenguaje, en donde el verbo va al final como el Hindi, se generan frases que requieren en primer lugar

Doi:10.11144/Javeriana.UPSY12-5.dcav

Kumar Mishra, R. (2013). Developmental changes in allocation of visual attention during sentence generation: an eye tracking study. Universitas Psychologica, 12(5), 1489-1500. Doi:10.11144/Javeriana.UPSY12-5.dcav

* Acknowledgements I thank Dhruv Raj Sharma who collected data for this study as part of his master thesis in Cognitive Science and Siddharth Singh, who did additional analysis on the data.

* Centre of Behavioral and Cognitive Sciences. Allahabad University. Allahabad 122002. UP. India Corresponding author: Ramesh Kumar Mishra. Centre of Behavioral and Cognitive Sciences (CBCS). University of Allahabad. Allahabad 211002. UP. India. Email:rkmishra@cbcs.ac.in. Ph-91-0532-2460738 (work). Mob-91-9451872007. Fax-91-0532-2460738(work). Home Page :http://cbcs.ac.in/people/fac/30-r-mishra. Personal Page: http://facweb.cbcs.ac.in/rkmishra

distribución de atención a la región que denota acción, y tanto los niños como los adultos difieren en este proceso.

Palabras clave autores
Atención visual, percepción de escenas, producción de oraciones, Hindi, Movimientos oculares.
Palabras clave adicionales
Percepción, investigación cuantitativa, ciencia cognitiva.

Introduction

Speaking comes to us so much naturally that we often fail to appreciate how complicated its underlying cognitive processes could be. We speak sentences in a variety of contexts and situations. A very complicated interaction between visual attention and linguistic processing takes place when we want to describe visual events. Describing events in a scene normally include actors and actions as perceived by the viewer. Therefore, generating sentences while simultaneously processing visual events includes multi-modal interaction of several distinct cognitive processes. To generate a successful sentence one needs to attend to different objects and actions in a specific order to conceptualize and use these perceptions as linguistic referents i.e. nouns, verbs, adjectives etc. Psycholinguistic theories of sentence production have long emphasized on the sequences of cognitive operations that must happen starting from creating intentions until encoding grammatical and phonological forms (Levelt, 1983). However, very little is currently known about how visual attention interacts with the sentence production system, when anyone attempts to speak a sentence while looking at some event. The goal of this paper is to study this matter more throughly with both children and adults, as they describe scenes and their eye movements are tracked in a lesser studied language like Hindi.

A powerful way to study overt visual attention and its shifts during cognitive processing is to record eye movements. Eye movements offer real time data about the nature of the locus of cognitive processing in a variety of task situations (Rayner, 1998; Liversedge & Findlay, 2000; Mishra, 2009; Huettig, Mishra, & Olivers, 2012). Fixations provide us details about the locus of cognitive processing (Irwin, 2004). Several researchers in the past have explored how eye movements can inform us about spoken language processing during simultaneous processing of visual and linguistic information (Allopenna, Magnuson, & Tanenhaus, 1998; Huettig & Altmann, 2005; Altmann & Kamide, 2007; Mishra, Pandey, Singh & Huettig, 2012). However, not much has happened in the area of sentence production and visual processing (See Mishra, In Press). Although there are some studies that have attempted to link eye movements with noun phrase production, or even single object naming (Griffin & Bock, 2000; Meyer, Sleiderink & Levelt, 1998; Griffin, 2001).

Many of these studies began using eye movements to understand time scale of conceptual and phonological processes during name generation (Griffin & Davision, 2011). A consistent finding in many of these studies has been that speakers always look briefly at objects and events before talking about them (Griffin & Bock, 2000). This brief pause has been thought to indicate a stage that includes conceptualizing for speaking. It has been observed that speakers gaze systematically at objects while preparing their names (Griffin, 2004). This time spent while looking at particular objects contributes to extract phonological and syntactic forms of names. Moreover, this visual attention that speakers direct towards the referents is often without any overt association between the forms and their names (Griffin & Oppenheimer, 2006). It has also been observed that if there are two objects, speakers move their eyes towards the second object when they have already retrieved the phonological form of the first object. This suggests an intricate dynamic interaction between visual attention and linguistic processing during sentence generation.

However, one problem with these studies is that they have used line drawing of simple objects and these objects have appeared out of context. Moreover, speakers had to generate just a noun phrase in a repetitive manner on most trials. This lack of flexibility could have compromised their generalisability. This is not what we see in life around us where objects appear embedded in rich scenes, and are surrounded by several other objects. Speakers are very creative in terms of their sentence generation. It is also an everyday observation that, by giving the same picture different speakers generated different types of sentences. Therefore, one has to explore how visual attention is programmed when we describe complex natural scenes and in a more unrestricted way. There is a much rich body of work on eye movements during scene perception (Henderson, 2003; for a sensori-motor account see also Mishra & Marmalejo-Ramos, 2010).

There are two important issues when we begin to explore shifts in visual attention as they correspond to linguistic processes during sentence generation. One is how speakers divide their attention among several potential visual referents, and use the extracted information for constructing sentences. The second one is how speakers of different languages, that have different grammatical and typological features, do so. This issue is crucial to our study since we examine this with Hindi speakers that has a flexible word order and some typical linguistic features. Recently there has been some effort to directly examine these issues with innovative designs and using eye movements as dependent measures. Coco and Keller (2012) examined if speakers produce similar scan paths during viewing scenes when they produce similar sentences. In this study, participants were given scenes to produce sentences and eye movements were recorded. Interestingly, the authors found a good correlation in scan paths during planning and speaking stages. Participants fixed similarly, for similar durations of time, on objects when they produced sentences of similar length and structures. This suggests that speakers extract information during fixations and also order this information while speaking.

This study is not clear explaining what will happen in cases where speakers from a language, as it is the case with Hindi, use the verb at the end of the sentence. Nevertheless, it might not be the case that there is always a one to one match between where we look and what we produce in the sentence. Kuchinsky, Bock and Irwin (2011) asked speakers to generate phrases mentioning time while they watched a clock. Some speakers were asked to generate phrases when the clock's hands denoted the usual meanings (i.e. the short hand for hour and the long hand for minute), but for some participants it was reversed. The results showed that this altercation did not influence fixation durations as such. The authors argued that participants exercise a top-down control during language generation in such a situation and, more importantly, the visual context is modified according to the linguistic demands. Recently there has been use of eye movements in the study of sentence production in aggrammatic aphasic patients (Cho & Thompson, 2010).

Another crucial drawback of many of the aforementioned studies is that not many of them have used real world scenes. Real world scenes are far more complex, and pattern of eye movements during scene perception has provided some excellent data regarding interplay of attention and perception (Henderson, 2003; Johansson, Holsanova, & Holmqvist, 2006). Further, not many have manipulated scene complexity. By scene complexity we are referring to the number of actors and the type of action that is depicted (i.e. intransitive, transitive, etc). These appear to be crucial scene elements that can affect linguistic processes during speaking. Finally, it is not very clear how children and adults differ in their planning and conceptualization while exercising viewing and speaking. We examined these issues in both children and adults while they saw photographs of real world scenes and generated sentences.

One critical aspect of our study was the use of Hindi language. In English, verbs come immediately after the first noun phrase, while in Hindi they come at the end of the sentence, for canonical structures. If the serial ordering of visual referents influences where speakers see and in what sequence, then it is somewhat problematic for Hindi. It is of course currently not very clear from studies where do participants look to retrieve the verb information. This information is crucial, as there are verbs that often determine the number of arguments that the sentence will have, as well as the agreements used in sentences (Kempen & Hoen Kamp, 1987; Bock, 1995; Bock, Irwin, Davidson, & Levelt, 2003). Therefore, we think it is interesting to explore where do Hindi speakers look when they formulate and speak sentences when one manipulates the types of action depicted in the scenes.

Brown-Schmidt and Tanenhaus (2006) studied the role of visual attention on construction of particular types of sentences (i.e. active vs passive). It was observed that participants produced more active sentences when their visual attention was initially drawn towards a particular agent. This suggests that the initial locus of attention on a scene significantly affects the choice of linguistic structures during sentence planning, i.e. subjects (agents) Vs. objects (themes). This question is of crucial importance when we consider the issue of language structure and how it might affect this process. For example, many contemporary linguistic theories assume that verbs affect the generation of sentential frames strongly. In a verb final language, like Hindi –with SOV word order—, subjects should look at the regions of action in a photograph to derive the syntactic structure of the sentence during the period of conceptualization. In the current study we particularly explore this process with children and adults.

Events or actions are often encoded in a verb. As noted earlier, verb lemmas specify the arguments that are required for sentence planning (Pickering & Branigan, 1998). Then, in a visual context, these arguments would be represented by different objects and actors involved in different actions. Therefore, sentence planning would require surveying them in a particular order. However, the relationship between visual attention and sentence production may not be direct, in the sense that speakers look at objects in an order that corresponds to the way they appear in the sentence. For example, different typological and grammatical patterns that languages employ may play a role in this aspect. For English, when one has to generate an active SVO sentence, one may look at the first noun phrase and the verb, as well as the second noun phrase in a sequence. However, for Hindi, which is an SOV language, it is not known if subjects look at the verb region first or at the objects.

It is clear from the current research that speakers deploy visual attention to objects in a visual field in a certain order during the preparation of linguistic levels. Past studies have mostly studied the naming of objects presented in isolation and not in complex scenes. From this research, it is not very clear how visual attention is controlled during sentence generation when one is presented with natural scenes. In literature of eye tracking there is substantial evidence that suggests that viewers selectively attend to certain aspects of scenes when viewing. However, literature on the control of eye movements during speaking has not, until now, considered the subtleties that scenes bring in when one plans to talk about its contents. Further, such studies have not been addressed with a language other than english. In this research we are interested in the developmental aspects of shifts in visual attention during sentence generation.

We examined eye movement patterns of children and adults as they produced sentences for pictures differing in their complexity, i.e. number of agents and objects that they contained. We have specifically examined the pattern of deployment of visual attention to different aspects of the picture during the processing of speaking in children and adults in the Hindi language. Most importantly, we examined how the complexity of pictures (i.e. the presence of more number of agents and objects) affect attention systems while speaking. Unlike previous studies in this domain, we used subjects speaking Hindi, a free word order language with SOV as the canonical structure.

Method

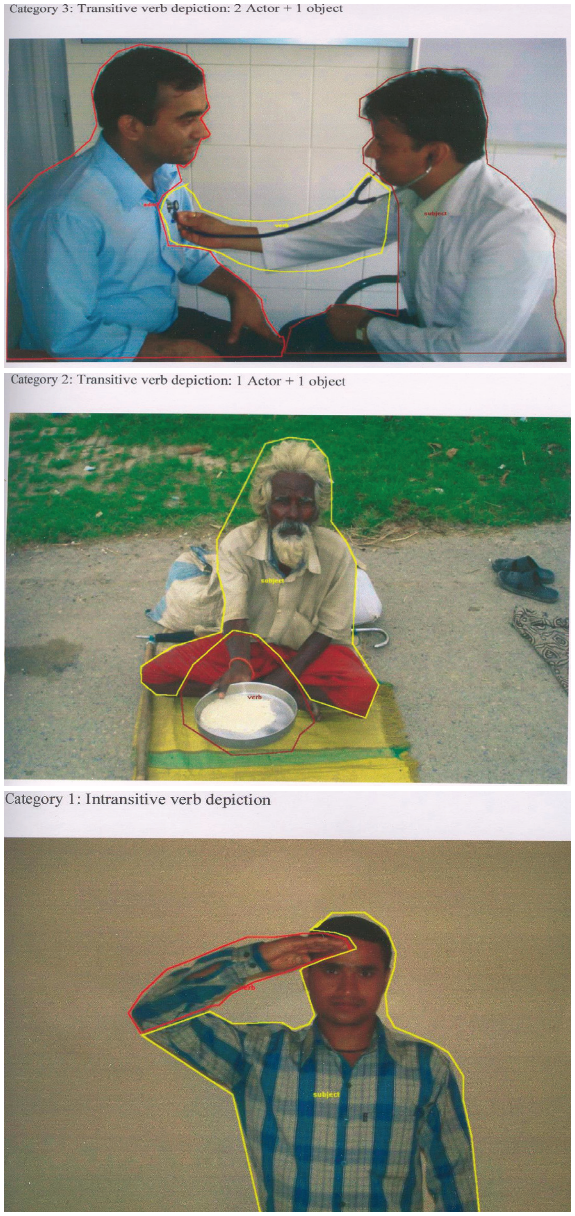
Participants

25 Children (7-11 years of age) and 25 adults (18-30 years of age) participated in this eye tracking study. All participants were native speakers of Hindi and were from Allahabad. Children were also sampled, based on teacher recommendation (in communication) in schools. None of the participants had any known neurological or behavior condition, and none wore glasses. The study was approved by the ethics committee of Allahabad University. All participants were naïve towards the purpose of the study.

Stimuli

Natural images were selected representing humans in two types of actions: (a) those depicted through transitive verbs. (b) Those depicted through intransitive verbs; involving 3 kinds of actor depiction: (a) 1 actor + no object, (b) 1 actor + 1 object, (c) 2 actors + 1 object. There were thirty images in total and ten of each type. These were photographs taken specifically for the occasion with actors representing several actions (Appendix1). These images were presented to the participants for a period of 10 seconds each. Eye movements were recorded as participants visually saw the images.

Images were captured using a Sony cyber-shot 7.1 mega-pixel camera, in two kinds of modes: (a). indoor; and (b). outdoor. For the indoor images, actors were invited to either, (a) a particular room in the Centre of Behavioral & Cognitive Sciences, Allahabad, or (b) a photography studio, where they were photographed while they performed the actions (primarily for the intransitive verb category). The outdoor images were clicked primarily for the transitive verb category. For these, the researcher went to the original settings where the actions normally took place: example, a pan shop for a photograph depicting (*ek admi paan laga raha hai "One man is preparing the betel leaf"*). All photographs were clicked from a front-on perspective; the mode of the camera was kept constant on "intensive sensitivity". Figure 1 shows examples of each type of images and the AOIs, considered for analysis. (see Figure 1)


Category 3: Transitive verb depiction: 2 Actor + 1 object

Category 2: Transitive verb depiction: 1 Actor + 1 object

Category 1: Intransitive verb depiction

Figure 1. Three types of pictures used in the sentence generation task. Fixations for the subject region.
Source: Own work.

Apparatus & Procedure

Participants were part of the experiment in an individual way. The experimental room was dimly lit and was sound proofing. At the beginning of the experiment, after placing the head-mounted earphone-and-microphone on their head, they were instructed on the following issues. First, they were instructed about basic eye-tracking, issues that they needed to know in order to participate (for example, they were told to avoid moving their head out of the eye-tracker once the calibration has taken place; also, they were requested to not make too many eye blinks, and so forth). A chin rest was used to stabilize the head movements. Second, they were asked to view the photograph on display and start describing it in a single sentence as soon as possible. Third, the sentence must necessarily be produced in Hindi, and should be the one that best describes the image that is on display. Fourth, participants were asked to keep viewing the photograph while they described it, and not shift their gaze out of the visual display. Finally, participants were instructed on practice tests, and how to take breaks between the experiment, when necessary. Eye movements were recorded as participants viewed the images.

Stimuli presentation was through the PRESENTATION software (Neurobehavioral systems). The photographs were presented on a 17-inch LCD screen. All sentences produced were recorded using the Goldwave software (version 5.23), with a Philips SHm7405, 30,000 Hz head-mounted microphone. Eye movements were recorded using an SmI EyeLink hi-speed, 1250 Hz. eye-tracking system (Sensomotoric Inc. Berlin). Participants generated sentences as soon as the picture was on the screen and the display went off with the sentence being recorded. The next test began after a delay of 1000 msec. There were thirty trials consisting of thirty pictures in total for each participants. The order of presentation was randomized for each participant and was counterbalanced for adults and children.

Data analysis

Each photograph was divided into two areas of interests, for the purpose of measuring eye movements. A verb region contained zones of action or instruments used in performing the actions themselves. The subject region contained the face and body regions of the human actors. For the photographs containing the intransitive actions, the subject regions and verb regions contained the body and the hands, which denoted actions respectively. For the transitive photographs, the human body was the subject region, while hands and objects together denoted the verb region. For two agents with an object acting out an action with each other, bodies served as subject region and the instrument was the verb region. In this case eye movements made to both the regions were averaged for the purpose of this analysis.

We measured both fixational and saccadic eye movements on these regions during the act of speaking. Dependent measures were the total number of fixations, the total number of saccades, the average of duration of fixations, as well as the total gaze durations. These dependent measures were selected as they indicate different aspects of visual attention deployed on photographs during sentence generation. Measurement of eye movements was made using BGaze software.

Results

Proportion of fixations

Proportion of fixations on two areas of interest, verb and subject region, for all three sentence types, was measured. Time window from the onset of image until its offset was considered, spanning the production duration. To be precise, the total time of recording i.e. 8000ms was divided into slots of 50 ms bins, and fixing proportions averaged for all the subjects in children and adults group were accounted. Figure 2 shows the fixation proportion of each individual plot, showing the fixations towards an AOI (verb or subject) for children and adults. The upper panels show proportion of fixations to the verb regions for three different types of images for children and adults. The time course plot begins from the onset of the picture on the computer screen until the sentence has been spoken. We calculated the proportion of fixations for each bin measuring 50 ms for the entire duration. The x-axis shows the time in milliseconds from this onset for 8000 ms.

For statistical analysis, voice onset latencies i.e. the exact time of utterance of each subject from the onset of the image was calculated and later on averaged for all the subjects (Voice Onset Latency). VOL for children was 2000ms (approx.) and for adults 1500ms (approx). We compared the average fixation proportion on each AOI during the Voice onset latency window for children, as well as for adults, in order to do a group comparison. Independent sample t-test was conducted in order to see whether at VOL window, any difference between adults and children existed, or to establish whether modulation of visual attention differs significantly for particular AOI's in children and adults, during the conceptualization phase. We assumed the VOL time period as the conceptualization period.

For the intransitive depictions during the conceptualization phase fixations on verb region for children (M = 0.25, *SD* =0.14) was not significantly different compared to adults (M = 0.3, *SD* = 0.14), *t* (24)= -1.15, *p* > 0.05. Similarly proportion of fixations on subject region for children (M = 0.24, *SD* = 0.1) as compared to adults (M= 0.28, *SD* = 0.09) during the conceptualization phase, was also insignificant *t*(24)= -1.39, *p* > 0.05. However, for pictures with one object and one actor (transitive pictures) adults (M =0.36, *SD* =0.09) deployed higher visual attention towards the verb region compared to children *(M =0.29, SD =0.12), t* (24) *=1.67, p <0.05.* For subject region also visual attention deployment by adults (M =0.39, *SD* =0.11), was higher than the one of children (M =0.29, *SD* =0.11) and *t* (24) = 1.79, *p* <0.05. For image depicting 2 actors + 1 object, the mean comparisons on the verb region (M =0.12, *SD* =0.05) and subject region (M =0.37, *SD* =0.11), *t* (24) = -0.01(ns) for children was again insignificant to the fixation proportions of adults at verb (M =0.12, *SD* =0.04) and subject regions (M =0.39, *SD* =0.1) at VOL window. (see Figure 2)

We also analyzed the total number of fixations, total number of saccades, and the total gaze duration, as overall measures of visual attention. For total number of fixations, the overall affect of age, *F* (2,48) = 43.101, *p* < 0.05, was significant while no other interaction was found to be significant. For intransitive pictures adults had higher number of fixations (M =7.52, *SD* =3.23) than children (M =2.817, *SD* =2.89), *t* (24) = 21.22, *p* <0.05. For the verb region of transitive images children had higher number of fixations (M =9.751, *SD* =4.33) then adults (M = 6.543, *SD* =3.11), t (24) = 14.46, *p* <0.05. For subject region this difference between children (M = 8.86, *SD* =3.11) and adults (M =8.66, *SD* =2.1), were not significant t (24) = 11.618, *p* <0.05.

For images with two actors and one object, Post hoc comparisons of verb children (M =5.592, *SD* =1.23) and verb adults (M=4.032, *SD* =1.42) was significant *t* (24) =6.59, *p* <0.05. Fixation towards the subject region for children (M =5.98, *SD*=2.11) and adults (M =4.16, *SD*=1.09), *t* (24) =8.215, *p* <0.05, also show significant (Figure 3).

Mean comparisons for the number of saccades came out to be insignificant, as there was no in-

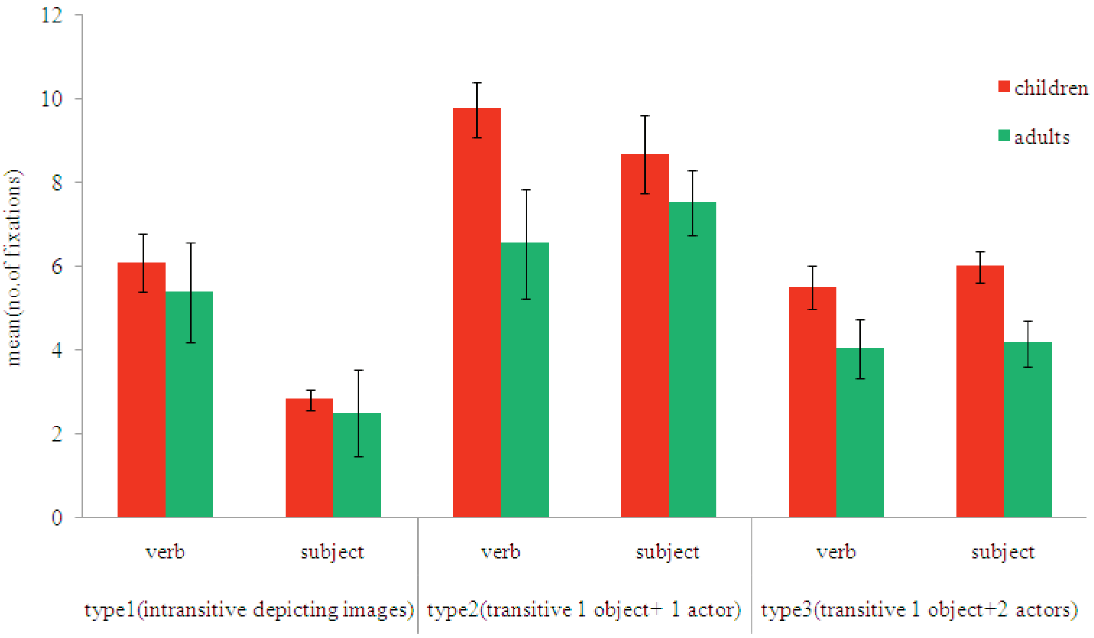


Figure 2. Mean number of fixations for children and adults for different pictures and AOIs
Source: Own work.

teraction effect on the age with AOI'S or sentence types. We conducted ANOVAS with Subjects – Adults and Children—, Type of Pictures (Three types) and region of interest (subject and verb) as factors on dependent variables. However, there was a significant two way interaction of image types with AOI's, *F* (2,48) = 47, *p* <0.05.

Comparison of means for the total dwell time also shows the interaction of age with sentence type, as well as AOI's. Three way ANOVA (2 participant types, three picture type, two AOIs), *F* (2,48) = 47, *p* <0.05, shows significance of this interaction, and generally total dwell time of children on all AOI's is more than that of adults on all the AOI's (Figure 4).

Post hoc analysis was done to compare individual mean differences within and between groups. For image type 1, dwell time for children and adults on the verb region was found out to be insignificant. For the image type 2, dwell time on verb region for children (M =2469.45, *SD* =1132.11) and adults (M =1873.24, *SD* =995.23) show significant difference, *t* (24) = 9.04, *p* <0.05. For subject region interestingly, dwell time for adults (M =2284, *SD* =1175) was significantly more than children (M =1873.79, *SD* =896.44), *t* (24) = 7.85, *p* <0.05.

Discussion

The eye movement patterns of children and adults during sentence generation, while viewing a static photograph, revealed a strong effect on age and also on the type of picture. Children in general deployed higher visual attention to the subject and object regions of the pictures for all three picture types. Thus, it seems that children needed more time to look at the corresponding portion of the pictures, to retrieve the conceptual knowledge and also to transform that information for the ongoing activity of sentence generation. Thus children made consistently higher number of saccades to both the subject and object regions while producing the sentences, and their overall gaze duration was higher. This could mean that sentence generation with simultaneous viewing of a complex scene requires consistent allocation of visual attention to the relevant regions for extracting syntactic structures.

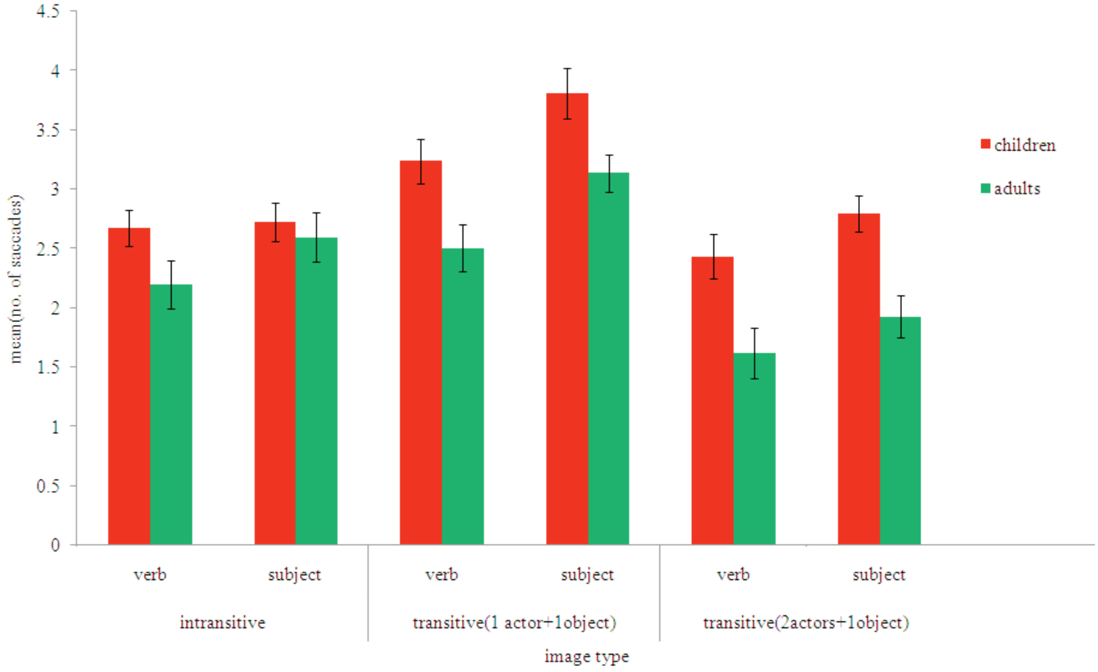


Figure 3. Mean number of saccades for children and adults for different pictures and AOIs
Source: Own work.

Many previous studies that have investigated time course of phonological information retrieval during object naming, have found the fixation durations on the pictures to be synchronous with naming latency (Belke & Meyer, 2007; Meyer & Van Der Meulen, 2000). Further, most previous studies have revealed a tight coupling between visual attention and name retrieval and a sequence of visual and linguistic information as naming is in progress. However, none of these studies have investigated what is the pattern of visual attention and its alignment with conceptualization during scene viewing and speaking. Therefore, allocation of higher visual attention in our study by children could not be attributed to the fact that this time was used for lemma and phonological form retrieval, but for sentence construction: since sentence generation is not just a sequential compilation of phonological information from several entities. Since syntactic structure generation would include both an update of form and content, in a holistic way. Adults, on the other hand, did not require longer deployment of visual attention, probably because of their more proficient strategies and experience. However, interesting difference appeared depending on the number of actors and actions involved.

What is theoretically interesting in these results is to note how visual attention is directed towards the verb regions, representing the action zones and subject regions. As noted earlier, most contemporary linguistic theories believe that verbs are important structures in a sentence whose transitivity determines how many possible arguments can accompany it in a sentence. Therefore, for sentence generation, the transitivity of a verb could have important influences on the structural arrangements of the other arguments. Assuming that one produces uttarances in a sentence in a linear fashion, information from verbs must be derived first for further realisation of it's arguments i.e. noun phrases and other constituents. However, world's languages differ from one another in terms of the word order that they manifest. Thus word order (i.e. where the verb appears in a sentence), must play a crucial role in the timely composition of the

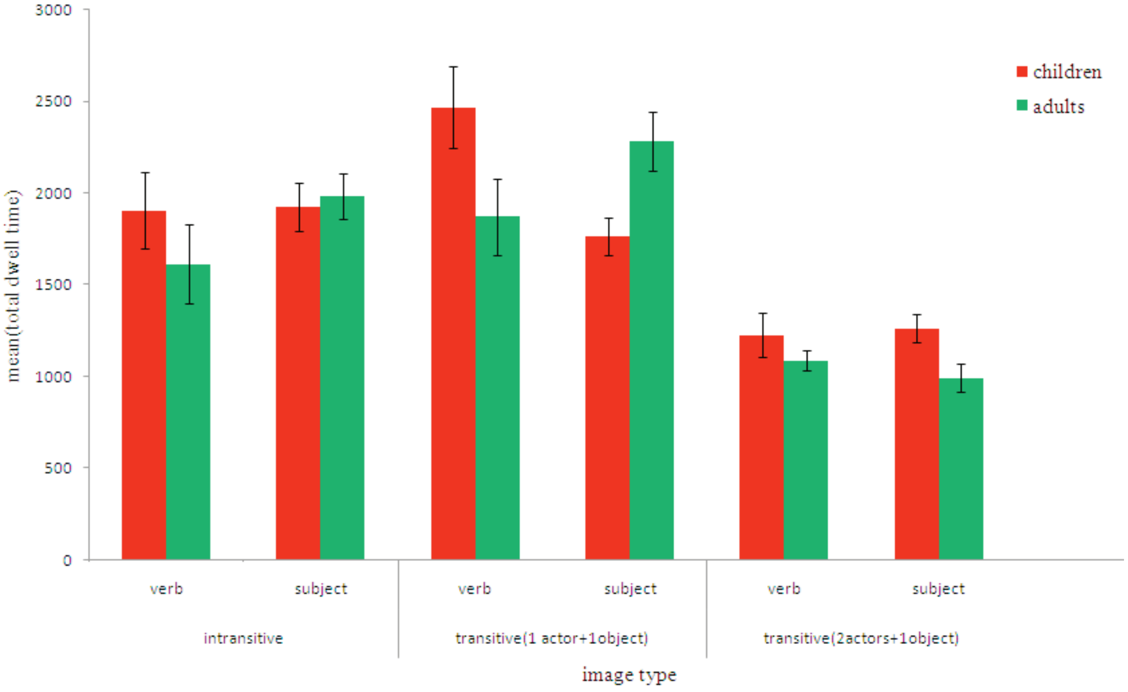


Figure 4. Total gaze durations for children and adults for different pictures and AOIs
Source: Own work.

sentence. Thus, we had hypothesized that in Hindi, being a verb final language, in its canonical sense, speakers must devote maximum visual attention to this region early on during the conceptualization process, so that they can figure out the other arguments and their position in the sentence.

Interestingly, the proportion of fixations to subject and verb regions in our pictures show that adults deployed consistently higher visual attention to the verb region, than to the subject region in all cases, compared to children. This is interesting from a developmental and linguistic point of view. However, this does not mean that children produced wrong sentences or were unable to produce structures, since they were gazing at the subject region. The explanation could lie with a more efficient adult system of language processing where information from verb could be used quickly to compute online the sentence. Children, on the other hand, might have followed a less canonical pattern of sentence construction (i.e. S-V-O). This also has a resemblance to other data from Hindi, where the non canonical SVO pattern has been found to be less depending in processing Mishra, Pandey & Srinivasan, 2011). However, at this point of time, without further controlled studies, this approach remains as a hypothesis. However, the noticeable difference between children and adults in their looking behavior towards the verb and subject regions during conceptualization, does suggest a basic difference in planning strategy.

From a vision-language interaction perspective the differences between children and adults in their deployment of visual attention is important. Observer's goals and top-down knowledge affect the way linguistic knowledge maps onto visual information (Salverda, Brown & Tanenhaus, 2010). Much of eye movements one sees in such cross-modal scenarios are anticipatory (Altmann & Kamide, 2007) and reflect sensori-motor systems (Mishra & Marmalejo-Ramos, 2010). This means, the overall experience with the visual context and rapid generation of syntactic structures will determine attentional mechanisms. Thus, when encoding the visual material in the scene, with the goal to articulate a sentence, subjects have to look at those locations more. For example in our case, subjects are mostly going to look at the faces and bodies of the agents and the actions they are engaged in, rather than to other objects that are in their environment. This observation fits nicely with recent other data from vision research that suggest top down control of visual attention and eye movements (Nuthmann & Henderson, 2010).

Thus, visual attention and its deployment as linguistic encoding in progress is controlled in a top down manner by the goal of the subject. However, during speaking, an object based attention is constrained by the type of linguistic material being processed. Thus, children and adults differ in terms of eye movement behavior. For example even between the subject and verb regions we see a very different type of fixational eye movements for different type of pictures. Interestingly, fixations to the subject and verb regions for both children and adults were more or less consistently deployed, though variably throughout the act of speaking. Thus, our results show a very subtle and systematic difference between children and adults in their allocation of visual attention to subject and verb regions of pictures in a sentence generation task. This difference may have a developmental cause, but more so it tells about the systematic development of multimodal interaction (i.e. between vision and language).

At most, the findings of this study should be considered as preliminary, since the study has its own methodological limitations and could not answer many important questions. The results obtained show that children require to pay higher attention during sentence production compared to adults. However, the analysis could not reveal how visual attention was used for formulation of different sentential constituents. This is because it is nearly impossible to control sentence production during free scene viewing, unless once uses few line drawings and rigidly controlled the order of production of some noun phrases, as has been done in previous studies. Further, for an SOV language, since the verb comes at the end, it is very difficult to pin point in time when during viewing verb was conceptualized. Since, many psycholinguistic studies have indicated that speakers first generate verb information and then attach relevant arguments to construct sentences. Future studies in this direction should use more refined design and control above factors to draw any meaningful conclusion. From this study's results, we can safely conclude that children and adults differ from one another in terms of their viewing strategy, when they are asked to process pictures and produce sentences.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, *38*(4), 419-439.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language a world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. doi:10.1016/j.jml.2006.12.004

Belke, E., & Meyer, A. S. (2007). Single and multiple object naming in healthy ageing. *Language and Cognitive Processes*, *22,* 1178-1211.

Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition*. Vol 11: Speech, language, and communication (pp. 181–216). Orlando, FL: Academic Press.

Bock, K., Irwin, D. E., Davidson, D. E., & Levelt, W. J. (2003). Minding the clock. *Journal of Memory and Language* , *48,* 653-685.

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes while talking about size: An investigation of message formulation and utterance planning, *Journal of Memory & Language*, *54,* 592-609.

Cho, S., & Thompson, C. K. (2010). What goes wrong during passive sentence production in agrammatic aphasia: An eyetracking study. *Aphasiology*, *24*(12), 1576-1592.

Coco, M. I., & Keller, F. (2012). Scan Patterns Predict Sentence Production in the Cross-Modal Processing of Visual Scenes. *Cognitive Science*. 36 (7), 1204–1223.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*(1), B1-B14.

Griffin, Z. M. (2004). Why Look? Reasons for eye movements related to language production. In J. M. Henderson, & F. Ferreira, *The Interface of Language, Vision, and Action: Eye movements and the visual world*. 213-248. New York: Psychology Press.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, *11*(4), 274-279.

Griffin, Z. M., & Davison, J. C. (2011). A technical introduction to using speakers eye movements to study language. *The Mental Lexicon*, 6(1), 53-82.

Griffin, Z. M., & Oppenheimer, D. (2006). Speakers gaze at objects while preparing intentionally inaccurate labels for them. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 943-948.s

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences*, *7*(11), 498-504.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition,* 96 (1), 23-32. doi:10.1016/j.cognition.2004.10.003

Huettig, F., Mishra, R. K., & Olivers, C. N. (2012). Mechanisms and representations of language-mediated visual attention. *Frontiers in Psychology, 2*, 394.

Irwin, D. E. (2004). Fixation Location and Fixation Duration as Indices of Cognitive Processing. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*. (pp.105-133). New York: Psychology Press, xiv.

Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. In: *Cognitive Science*, 30 (6), 2006, S. 1053-1079.

Kemper, S., Herman, R., & Lian, C. (2003). Age differences in sentence production. *Journal of Gerontology: Psychological Sciences*, *58B*, P260–P268.

Kuchinsky, S. E., Bock, K., & Irwin, D. E. (2011). Reversing the hands of time: Changing the mapping from seeing to saying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 748 -756.

Levelt, W. J. (1993). *Speaking: From intention to articulation*. MIT press.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in cognitive sciences*, *4*(1), 6-14.

Meyer, A. S. (2004). The use of eye tracking in studies of sentence generation. In J. M. Henderson, & F. Ferreira, *The Interface of Language, Vision, and Action: Eye movements and the visual world* (pp. 191-212). New York: Psychology Press.

Meyer, A. S., Belke, E., Häcker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language, 57,* 210-231.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: eye movements during noun phrase production. Cognition, 89, 25-41.

Meyer, A. S., & Van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review, 7,* 314-319.

Mishra, R. K., & Marmolejo-Ramos, F. (2010). On the mental representations originating during the interaction between language and vision. *Cognitive Processing. 11*(4):295-305.

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (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), 1-10.

Mishra, R. K., Pandey, A & Srinivasan, N. (2011). Revisiting the Scrambling Complexity Hypothesis in Sentence Processing: A self-paced reading study on anomaly detection and scrambling in Hindi. *Reading & Writing, 24*, 709-727.

Mishra, R. K. (2009). Interface of language and visual attention: Evidence from production and comprehension. *Progress in Brain Research, 176,* 277-292.

Nuthmann, A., & Henderson, J. M. (2010). Object based attentional selection in scene viewing. *Journal of Vision, 10(8): 20*, 1-19.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory & Language, 39*, 633-651.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2010). A goal-based perspective on eye movements in visual world studies. *Acta Psychologia*, *137*(2), 172–180.

Tanenhaus, M. k., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.