

Introduction to '*Attention and Vision in Language Processing*'

Falk Huettig (Max Planck Institute for Psycholinguistics)

Narayanan Srinivasan (University of Allahabad)

Ramesh K Mishra (University of Hyderabad)

in press, in Mishra, R. K., Srinivasan, N., & Huettig, F. (Eds.) *Attention and Vision in Language Processing*. Springer.

It is perhaps obvious that language interacts with vision and attention. In many everyday situations people give or receive directions or instructions for action or talk about things in their surroundings, say, a photo or a painting, the state of the kitchen sink, or an unfolding sports event. Moreover, our shifts in eye gaze are often (consciously or subconsciously) directed by spoken language input. A parent may tell a child to look at the beautiful leaf or a visitor may ask about a new gadget he has spotted in our living room. In short, the ability to integrate visual and auditory input with stored linguistic and non-linguistic mental representations is a hallmark of human cognition. It follows from these observations that it is very likely to be a mistake to study language, vision, and attention independently of one another. It is surprising therefore that many of the cognitive processes occurring under such circumstances traditionally have been investigated in isolation although they are all involved when language is used.

The traditional approach to investigate language separately from other cognitive systems is a consequence of the theoretical development in the language sciences. Hockett, for instance, developed his list of features of human language (Hockett & Altmann, 1968) and *displacement* (the fact that concepts need not refer to an object that is physically present) was considered a (or the) key feature of human language. Recent eye tracking research however suggests that when the object of spoken language *is* actually physically present individuals express a strong tendency to refer to it. The cognitive system does this by orienting its visual sensory apparatus towards the object thereby linking a linguistically activated type representation to a specific perceptual token in the outside world. Cooper (1974), for instance simultaneously presented participants with spoken fictional stories such as a safari in Africa and a visual display containing nine line drawings of concrete objects (e.g., a lion, a zebra, a tree, a camera, etc.). Participants in the study were asked to just listen to the stories. Cooper found that listeners' fixations of the objects were very closely time-locked to the unfolding speech input. Whenever a spoken word referred to an object participants rapidly shifted their overt attention to the object or similar objects even though this was not required for any task. Studies within the visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) later confirmed that participants tend to very rapidly fixate objects which are visually similar (Dahan & Tanenhaus, 2007; Huettig & Altmann, 2004, 2007), similar in meaning (Huettig & Altmann, 2005; Yee & Sedivy, 2006) or whose names are consistent with the unfolding speech signal (Allopenna, Magnuson, & Tanenhaus, 1998). In other words, although language need not refer to an object that is physically present it is often used in such a way. Moreover, spoken language often guides visual orienting without volitional control. Language-mediated eye movements appear to be fast, unconscious, and largely overlearned (i.e. automatized through extensive practice, cf. Logan, 1988). It seems that some prior conditions need to be met for language to be able to drive eye movements (for instance to actively listen to the speech and a willingness to make eye movements, i.e. to look around rather than to force oneself to fixate one location). But once these conditions are met the available experimental evidence suggests that the

integration of language with oculomotor behavior may be unstoppable (Mishra, Olivers, & Huettig, 2013, for further discussion).

Many researchers in psycholinguistics in the 1980s and 1990s have ignored vision and attention because of the widely held view that vision and language are separate informationally encapsulated modules (Fodor, 1983). If the language system is encapsulated then there may be few reasons for language scientists to read work on vision and attention and study how people produce and understand utterances about objects and events they see. Fodor's notion of the modularity of mind however has been strongly challenged in the last twenty years. Psycholinguists have shown how linguistic and non-linguistic processes jointly determine how the language user understands language (see for instance the special issues in the *Journal of Memory and Language*, Ferreira & Tanenhaus, 2007; and *Acta Psychologica*, Hartsuiker, Huettig, & Olivers, 2011).

This volume is timely in the best sense in that it explicitly puts the focus on the interactions between language processing and other cognitive processes such as attention and vision. Key issues concern the functional architecture of the mind, how linguistic and non-linguistic processes jointly determine language comprehension and production, and how the linguistic system interfaces with perceptual systems and attention.

Section 1 focuses on attention and vision in spoken language comprehension and production. Rigoli and Spivey (Chapter 1) set the stage by arguing that language is part of a continuous perception-action loop that a person develops with his/her environment. They make the case for a fundamental extension of cognition onto the surrounding environment. Rigoli and Spivey argue that by defining language and cognition as separate from other individuals and from our actions in the environment much of the field has been overlooking the very essence of language itself.

Altvater-Mackensen and Mani (Chapter 2) present the results from a novel experiment using a visual priming paradigm to assess the extent to which toddlers retrieve sub-phonemic detail during lexical access. They observe that both the retrieval of an object's label and toddlers' recognition of a word involve activation of not only phonemic but also sub-segmental information associated with the lexical representation of this word.

Hintz and Huettig (Chapter 3) present three eye-tracking experiments investigating the impact of the complexity of the visual environment during language-mediated visual search. Their results demonstrate that phonological word-object mapping is dependent on the nature of the visual environment. These findings add to a growing body of evidence that the nature of our visual surroundings induces particular modes of processing during language-mediated visual search.

Knoeferle (Chapter 4) reviews the literature on real-time visual context effects and visually situated language comprehension in children and in healthy young adults. She argues that visual context should play an active role in child language comprehension and that children benefit from a similarly rapid interplay of visual attention and language comprehension as young adults.

Norcliffe and Konopka (Chapter 5) consider the extent to which the planning processes involved in producing sentences are fine-tuned to grammatical properties of specific languages. They argue that incrementality is a general principle of production that applies cross-linguistically and that incremental encoding can be controlled by those aspects of the language that are responsible for linearization, namely grammar. Norcliffe and Konopka conclude that differences in language-specific grammatical constraints on word order result in differences in the order of encoding operations performed to produce grammatically correct utterances.

The second section of the book focuses on attention and vision processes in reading. Saint-Aubin and Klein (Chapter 6) discuss whether printed words are identified the same way

when presented in isolation and in connected texts. They review the literature on the cognitive processes involved in reading connected texts with a special focus on the combined use of eye movements and the letter search task.

While there have been a larger number of studies on visual and attentional processes involved in reading of European languages using predominantly the Latin script, there is very little work on writing systems from other regions of the world. Winskel (Chapter 7) addresses this gap by focusing on visual and attentional processes involved in reading Thai script. Using Thai she illustrates how the distinctive features of an orthography (reading without interword spaces) can influence the visual and attentional processes involved in visual-word recognition and reading.

Section 3 contains chapters which focus on attention and vision in bilingual language processing. Hartsuiker (Chapter 8) asks how bilinguals select the right language for the particular context they are in and how do they do this. He first discusses the evidence on the degree to which bilinguals activate multiple languages and how this is constrained by linguistic cues. He then presents new lines of research that investigate visual language cueing (such as the language associated with familiar people's faces, with the race of unknown faces, and with cultural icons).

Chabal and Marian (Chapter 9) provide an informative review of eye tracking work on bilingual spoken word processing and parallel language activation. They make the case that multi-modal investigations of language processing (e.g. using eye-tracking) are not only ecologically valid as they closely resemble real-world multi-modal situations but also demonstrate how language interacts with other cognitive and perceptual systems in a non-modular mind.

Mishra and Singh (Chapter 10) present evidence for language non-selective activation in Hindi-English bilinguals using an oculomotor task. They show that bilinguals suffer interference during a simple visual task suggesting that they activate translations of spoken words unintentionally. Mishra and Singh discuss the detrimental influence such non-selective activation may have on cognitive processing.

Finally, Section 4 focuses on language processing in a social context. Lev-Ari (Chapter 11) discusses how the identity of interlocutors influences which cues we attend to during language processing. She points out that adjustments in the allocation of attention can have cascading linguistic and social consequences. Lev-Ari argues that it is indispensable for our understanding of language processing to study cognitive processes in different social contexts.

Vinson, Dale, Tabatabaiean, and Duran (Chapter 12) further take up the issue of language processing and social context. They review literature suggesting that social cues (including low-level perceptual variables, perception of another's gaze, knowledge of another's belief states) influence language processes. They make the case for a systematic research agenda to uncover how various processes work together to bring about multimodal coordination between two or more interacting persons.

This volume emanates from the first "Attentive Listener in the Visual World" workshop held in 2012 at the Centre of Behavioural and Cognitive Sciences at the University of Allahabad, India. Studies of language, vision, and attention are intrinsically related. We hope that this volume will encourage further workshops and research on this crucial topic for the language sciences.

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, 419-439.
- 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, 84-107.
- Dahan, D. & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12, 453-459.
- Ferreira, F., & Tanenhaus, M. K. (2007). Introduction to the special issue on language-vision interactions. *Journal of Memory and Language*, 57(4), 455-459.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.
- Hartsuiker, R. J., Huettig, F., & Olivers, C. N. (2011). Visual search and visual world: Interactions among visual attention, language, and working memory (introduction to the special issue). *Acta Psychologica*, 137(2), 135-137.
- Hockett, CF, & Altman, S. (1968). A note on design features. In TASEbeok (Ed.). *Animal communication: Techniques of study and results of research*. Bloomington: Indiana University Press. (pp. 61-72).
- Huettig, F., & Altmann, G. T. (2004). The on-line processing of ambiguous and unambiguous words in context: Evidence from head-mounted eyetracking. *The online study of sentence comprehension: Eyetracking, ERPs and beyond*, 187-207.
- 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, B23-B32.
- Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, 15, 985-1018.
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137, 151-171.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-527.
- Mishra, R. K., Olivers, C. N. L., & Huettig, F. (2013). Spoken language and the decision to move the eyes: To what extent are language-mediated eye movements automatic? In V. S. C. Pammi, & N. Srinivasan (Eds.), *Progress in Brain Research: Decision making: Neural and behavioural approaches* (pp. 135-149). New York: Elsevier.
- 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.
- Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1-14.