

Preface

Twenty years ago, the research area which we know as 'visually situated' language comprehension was taking its first steps, enabled by the re-emergence of a measure that continuously tracks eye movements to objects during the comprehension of related spoken sentences. Since its introduction to the study of language by Richard Cooper in 1974 and its re-discovery and successful utilization in 1995 by Michael Tanenhaus and collaborators, the measure and paradigm has become known as the 'visual world paradigm'. Tanenhaus and his students employed the technique to assess whether real-time syntactic structure building during spoken comprehension is modular, or whether the unfolding syntactic structure and semantic interpretation is rather modulated by a suitable referential visual context. Strikingly, eye movements during the processing of local structural ambiguity depended upon the referential visual context. Contexts containing two apples, led comprehenders' to prefer attaching a locally structurally ambiguous prepositional phrase (*Put the apple on the towel...*) as the modifier of the noun phrase *the apple*, overriding the default preference for attaching into the verb phrase. The effect of the visual context was such that it seemed on a par with linguistic disambiguation through a relative pronoun (*Put the apple that's on the towel...*).

Since these early days, the visual world paradigm has been popular among both psycholinguists and cognitive scientists for investigating real-time language processing. The multi-faceted research questions that have been studied using the visual world paradigm speak to its usefulness, as does the substantial, and ever increasing, number of published visual-world studies. Indeed, it has become the dominant paradigm for investigating visually situated language use.

Reflecting the widespread and multi-faceted use of the paradigm, the present volume offers a collection of reviews on visually situated language processing research from experts in the field. It hopes to excite interest in linguists, psycholinguists, and cognitive scientists, who have not yet used the visual-world method and showcases not only the diverse topical research questions and the historical development of the paradigm but crucially provides a solid introduction to the methods in the first three chapters. Chapter 1 presents an in-depth introduction to the by-now well-established field of visually situated language (Spivey and Huettenlocher). Michael Spivey and Stephanie Huettenlocher outline the historically contributing factors that gave rise to the field of visually situated language processing and discuss some of the strengths and weaknesses of the visual world methodology, as well as the implications of results on visually situated language processing for our view of the cognitive processing system. Their chapter is matched by an expert introduction to the perception of the visual environment (Chapter 2), in which Ben Tatler highlights constraints imposed on cognitive processes through our visual interrogation of different sorts of environment (e.g., arrays of objects, photographs of real-world environments, and natural task settings). Tatler further discusses the role of task-based expectations in the encoding and retention of visual information. From Tatler's review on visual perception we turn our attention back to the visual world paradigm and its methodological properties, laid out by Pyykkönen-

Klauck and Crocker in Chapter 3. They review the opportunities and challenges in using overt visual attention (and in particular different gaze measures) as an index of the cognitive processes and mechanisms implicated in a variety of language processing tasks. In doing so, they review key assumptions in linking visual attention to cognitive processes (the so-called 'linking hypotheses'). For experienced researchers these chapters present a state-of-the-art overview of methodological aspects; teachers will find them useful as introductory materials for their course, and students for familiarizing themselves with the methodology.

Following these three foundational chapters, the remaining nine chapters provide in-depth reviews and discussions of key research themes in visually situated language research. They cover different foci, from using visual attention as a window into linguistic processes (Chapters 4-6), to looking at the active contribution of the visual context (Chapters 7-9), to interactive dialogue (Chapters 10-11) and the interaction of language with action (Chapter 12). In more detail, Roger van Gompel and Juhani Järviö review the role of syntax in sentence and referential processing (Chapter 4). They argue that non-structural factors (e.g., the referential context, action-based affordances, verb biases and prosody) can all rapidly affect how adults process structurally ambiguous sentences. In some cases, young adults can even use syntactic cues such as case marking for visually anticipating objects, meaning that they begin to inspect an object before it has been named. Children, by contrast, appear to ignore the referential visual context in their processing of local structural ambiguity. Just as syntactic cues are integrated rapidly and can inform expectations in adult comprehension, so do semantic cues. Paul Engelhardt and Fernanda Ferreira (Chapter 5) review among other topics how conceptual knowledge is accessed during lexical processing, how compositional semantic interpretation can elicit expectations (manifest in anticipatory object-directed gazes), how comprehenders compute event interpretations for metonymic sentences (serially), and the principles involved in speakers's use of referential expressions. Overall, they conclude that the linguistic input is not just mapped onto individual objects but onto event representations gleaned from both the visual scene and from long-term event experience. From the semantic interpretation of sentences we move on to discourse processes, a topic reviewed by Elsi Kaiser in Chapter 6. Kaiser presents theoretical approaches to information structure, and highlights the advantages of the visual world paradigm for research on discourse-level processes. She discusses the relation between prosodic cues and information structure, how information structure is encoded in linguistic structure (e.g., prosodically prominent nouns tend to be associated with newness or contrast, while syntactically prominent nouns are more associated with given than newness). Complementing Chapter 4, she reviews processes of reference and pronoun resolution at the discourse level and discusses the effects of factors such as discourse coherence.

The visual world paradigm has further been employed to examine figurative language processing as discussed by Matlock and Huettenlocher in Chapter 7. They review evidence on the processing of 'fictive motion' sentences including a motion verb without describing motion (e.g., *A road goes through the desert*). Their chapter embeds fictive motion processing in interactive dynamical systems

and concludes that their processing is highly similar to actual motion, thus extending the notion of representations to what has been termed 'embodied' representations (i.e., conceptual representations are related to the perceptual states from which they originated). Craig Chambers (Chapter 8) reviews evidence on the role of 'affordances' (potential ways of physically interacting with an object) in language processing. Visual-world findings clarify that affordances – much like syntactic or semantic cues - influence language comprehension and the resolution of structural ambiguity rapidly. However, while Matlock and Huette in embrace an embodied view of language processing in Chapter 7, Chambers argues in Chapter 8 that it remains to be seen whether affordances as a kind of perceptually-based information play a privileged role in language processing. He concedes that the rapid integration of linguistic and visual information may be partly subserved by shared systems or representation formats. Adding to the theoretical discussion, Chapter 9 by Pia Knoeferle reviews visual context effects on language processing and argues that these are important for language processing in light of their pervasiveness across reading and spoken comprehension, different types of scenes (cliparts, photographs, and real-world), different aspects of the visual context (a speaker's eye-gaze, mimics, and gestures) and both concrete and abstract language.

Chapters 10 and 11 by Dale Barr and Sarah Brown-Schmidt both review the state of the art in visually situated studies on dialogue. Dale Barr presents an account that aims to resolve ongoing controversies about the extent to which common ground (the shared knowledge between two interlocutors) matters for guiding interlocutors' visual attention during communication. While some studies have observed immediate effects of common ground and credited this to the task (interactive rather than passive), others have failed to do so. Barr proposes that these apparent differences do not result from the (interactive) task but instead from inconsistent analysis methods and interpretation, and that once we take this into consideration, the discrepancies can be reconciled. Sarah Brown-Schmidt also reviews the investigation of interactive dialogue and highlights the role of the visual environment for establishing joint domains of reference. She argues that in conversation language is created by interlocutors with partially overlapping contextual representations and emphasizes that the extent to which an interlocutor's perspective matters for communication has emerged as a central research question. The volume concludes by bridging language to motor actions in a chapter by Thomas Farmer, Sarah Anderson, Jonathan Freeman, and Rick Dale. These authors provide an overview of evidence suggesting a co-extensive relationship between language and action. They support their claims about language processing and the implicated (embodied) representations with insights from studies testing the embodiment of sentence processing and the involvement of manual motor movements. They highlight that tracking computer-mouse movements around a visual display can complement the eye-movement record, whereby mouse movements are argued to provide insight into linguistic processes that other (discontinuous) measures may not provide.

Pia Knoeferle

Pirita Pyykkönen-Klauck

Matthew W. Crocker