The coordinated processing of scene and utterance: evidence from eye-tracking in depicted events

Pia Knoeferle (knoeferle@coli.uni-sb.de)
Department of Computational Linguistics, Saarland University, 66041 Saarbrücken, Germany

Matthew W. Crocker (crocker@coli.uni-sb.de)
Department of Computational Linguistics, Saarland University, 66041 Saarbrücken, Germany

Abstract

The monitoring of eye-movements in scenes has revealed that a visual referential context can influence the initial structuring and interpretation of an utterance (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). These and other findings have influenced frameworks of the language system that permit the interaction of visual and linguistic processing (e.g., Bergen & Chang, to appear, Jackendoff, 2002). However, such theories make no explicit predictions about the temporal coordination and relative importance of distinct visual and linguistic processes during sentence comprehension. Recent eye-tracking experiments suggest a tight synchronization between utterance comprehension, attention in the scene, and the influence of scene information on comprehension (e.g., Knoeferle, Crocker, Scheepers, & Pickering, in press), as well as a greater relative importance of scene information than linguistic/world knowledge (Knoeferle & Crocker, 2004). By drawing on theories of language acquisition (e.g., Gleitman, 1990), we propose a first sketch of such coordinated interaction and relative importance of distinct visual/linguistic processes for adult comprehension.

Introduction

People often find themselves in situations where both spoken language and an immediate scene context are available and relevant. When watching movies, for instance, people are able to rapidly integrate both the utterance they hear, and the events they see. The rapid integration of scene and utterance has been demonstrated experimentally in numerous psycholinguistic investigations.

Tanenhaus et al. (1995) demonstrated that a visual referential context influences the initial structuring of an utterance. In instructions such as *Put the apple on the towel in the box*, the phrase *on the towel* can be temporarily analysed as modifier and interpreted as the location of the *apple* (identifying which apple) or it can be attached to the verb phrase and interpreted as a destination (where to put the apple). In a scene containing one apple, the phrase *on the towel* was preferentially interpreted as destination. In a scene with two apples people preferentially analysed the phrase *on the towel* as a modifier of the apple, interpreting it as a location, as evidenced by eye-movement patterns. Crucially, fixation patterns to objects differed from the onset of the utterance depending on the type of referential context (i.e., a one-apple versus two-apple context, supporting a destination versus modifier interpretation respectively). The early difference in gaze patterns suggests that the scene influenced the *initial* structuring of the sentence very early during comprehension of the utterance.

Importantly, the findings by Tanenhaus and colleagues have shown that the informational integration between the language and vision systems is not *informationally encapsulated* in the Fodorian sense (Fodor, 1983). Fodor postulated strong architectural restrictions on the informational interaction between distinct cognitive systems such as language and vision. In his model of the mind, distinct input modules such as language and vision have only access to the *output* of another distinct module, but cannot influence its *internal processes*. The findings by Tanenhaus et al. (1995) provide strong evidence for the view that processes internal to the language system such as the structuring of an utterance can be rapidly influenced by a perceived visual referential context.

Psycholinguistic accounts of language processing

Despite much evidence against a strong version of Fodor's model of the mind, Fodorian views have influenced psycholinguistic theories of on-line language comprehension. Scene information has, for instance, not been explicitly included in most psycholinguistic theories or frameworks of language comprehension (e.g., Forster, 1979; Frazier & Fodor, 1979; MacDonald, Pearlmutter, & Seidenberg, 1994). Furthermore, these approaches do not explicitly specify the visual perceptual system as a cognitive system that might proffer important information for comprehension processes. We argue that as a result such theories do not provide a complete account of language comprehension for situations in which language relates to a scene.

A sub-group of these theories *by definition* exclude the influence of scene information on the initial structuring of a sentence through restricting the informational sources that can influence this process to syntactic information (e.g., Forster, 1979; Frazier & Fodor, 1979).

In contrast to this *restricted* position scene information could, in principle, influence the initial structuring of a sentence in *unrestricted* interactionist frameworks (e.g., MacDonald et al., 1994; Tanenhaus, Trueswell, and Hanna, 2000). These frameworks propose that any available and relevant informational source can influence the initial

structuring of a sentence. However, just as restricted psycholinguistic theories, they still do not explicitly include scene information as an informational source.

Embedding the language system with vision

Recent research on the language system has, in contrast, begun to take into account the fact that scene information can influence core comprehension processes such as the structuring of an utterance, and explicitly embeds the language system in relation to the other perceptual systems (e.g., Jackendoff, 2002) (see Fig. 1).

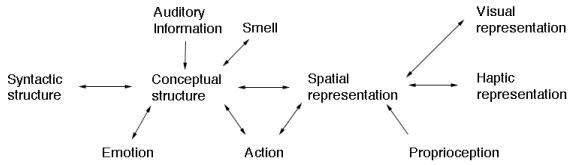

Figure 1: The Jackendovian architecture of the language system (Jackendoff, 1997, p. 44)

Spatial representations provide information about shape and location of objects and are the *`upper end' of the visual system* (Jackendoff, 2002, p. 346). The arrows represent interface modules that provide for the communication between distinct structures. Through communication via the interface modules, information from the immediate visual scene can influence language comprehension and the structuring of a sentence. Thus, such a framework should, in principle, be able to account for the influence of visual scenes on the structuring of an utterance.

A procedural account of language processing

While providing representational interfaces between distinct cognitive systems, the Jackendovian framework makes no explicit predictions about precisely *how* utterance and scene information are integrated. We thus consider whether other frameworks provide a more fully specified procedural account of on-line sentence comprehension.

One such type of framework is the interactionist competitive-integration model (e.g., Spivey-Knowlton, 1994; Tanenhaus et al., 2000). While current implementations do not model the integration of scene information, we propose that such a model might, in principle, be extended to incorporate information that has been derived from a visual scene. A sketch of this type of model is provided in Figure 2.

Processing steps in the model detail how the biases (A, B) are combined, a process that simulates structural ambiguity resolution (between structures A, B). In the first processing step the activations for the two nodes (A, B) of each bias are normalized. In a second cycle the activation of the two structures (A, B) is determined by integrating their respective corresponding bias nodes using a weighted sum. In a third step, structure nodes send feedback to the biases depending on how strongly a bias activated that structure. This type of model thus provides a specification of how diverse informational biases combine, and a linking hypothesis, which predicts processing difficulty when constraints that are very similar in strength compete.

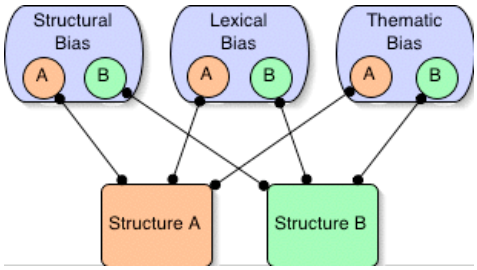

Figure 2: A sketch of the competitive-integration model (e.g., Tanenhaus et al., 2000)

We think, however, that while the model details how *informational biases* combine, it is not a model of *sentence comprehension*. The algorithm in the competitive-integration model does not determine the construction of an interpretation, e.g., how grammatical function and thematic role are assigned to a currently processed phrase. If we want to model not just how to decide between two alternative structural analyses, but if we aim at actively building a structure, then we need a theory that in some form or other permits us to specify conceptual structures, and mechanisms for how conceptual structures combine.

A framework that offers a linguistic formalism for the analysis of meaning, and that still provides for the procedural integration of perceptual and linguistic information is Embodied Construction Grammar (e.g., Bergen & Chang, to appear). Bergen and Chang show how this linguistic formalism can be integrated into a simulation-based model of language understanding. In their model sentence comprehension takes place via the activation of embodied *schemas* (cognitive experience-based structures), and the simulation of motor/perceptual experiences that are associated with these schemas. The Embodied Construction Grammar framework provides a less detailed procedural account than the competitive-integration model. However, it has a stronger linguistic component, which makes it suitable as a framework for a theory of full sentence comprehension.

Neither of these two models, however, has so far specified the *nature* of the interaction between linguistic and visual processes. We therefore take as our starting point the following observation: “We agree that constraint-based models need to be more explicit about the nature of the constraints and how they combine [...]” (Tanenhaus et al., 2000, p. 94). Specifically, we propose that exploring the precise nature of the interplay between the visual system and comprehension could lead to a more fully specified

account (and ultimately theory) of on-line sentence comprehension in visual scenes.

The nature of the interplay between scene and utterance processing

One way of specifying what we term the 'nature' of the interplay between visual perception and comprehension is by detailing how mental representations of utterance and scene derive and combine. An important aspect is whether the mechanisms of the systems that process utterance and scene are coordinated or not.

Linguistic and visual processing might, for instance, be *asynchronous*. (i.e., not require a common timing reference in order to communicate). Alternatively, the interaction of distinct cognitive processes might be of a more *synchronous* nature (i.e., requiring a common time signal to coordinate them). Support for a tight coordination of linguistic and visual processing comes from Tanenhaus et al. (1995), and Zelinsky and Murphy (2000). Their results show linguistic processing to serve as a "signal" for other cognitive processes such as attention in the scene (e.g., Tanenhaus et al., 1995), and duration/frequency of object inspection (e.g., Zelinsky & Murphy, 2000).

To investigate the time-course of scene-sentence integration during the structuring of an utterance, we re-consider findings by Tanenhaus et al. (1995). Recall that eye-movements to objects differed from the onset of the utterance depending on the type of referential context in their studies. Eye-movements showed furthermore that shortly after people had heard a word in the utterance they inspected the real-world referent of that word. The fixation patterns in their studies thus provide evidence for the view that the utterance directs attention in the scene. Second, they demonstrate that there is an early influence of the scene on the initial structuring of an utterance.

What the fixation patterns in the studies by Tanenhaus et al. (1995) do not permit to determine is *when exactly* the scene influenced structuring and interpretation of the utterance. This is because eye-movements for the two context conditions (one apple, two apples) differed from the very start of the utterance. As a result, there are at least two ways of interpreting their data.

One interpretation is that people acquired the referential context *prior* to hearing the utterance, and that there was from the start only one way in which the referential context permitted structuring of the utterance. The second possible interpretation is that comprehension of words during presentation of the utterance directed attention in the scene, and that this guiding function of the utterance was necessary for making scene information available in the first place. Under this view, only *after* relevant scene information had been identified through the utterance, was it able to influence the structuring of the utterance. While the findings by Tanenhaus and colleagues are compatible with the second interpretation they do not permit teasing apart the two interpretations.

Further to the coordination of the interaction between scene and utterance processing, a fundamental and little-studied question is how important scene and utterance processing are relative to one another in the integration process. Teasing apart the relative effects of perceived scene information and stored linguistic/world knowledge in online sentence comprehension is of relevance for frameworks of the language system (e.g., Bergen & Chang, to appear; Jackendoff, 2002). Such endeavor may ultimately allow us to propose a more fully specified model of processing mechanisms for real-time sentence comprehension in visual scenes.

Experimental evidence from eye-tracking in depicted events

Experimental findings have indeed demonstrated a tight coordination of when in relation to the mention of scene information that scene information influences structural disambiguation. Further, there is evidence that suggests a greater relative importance of immediately depicted events than linguistic/world knowledge during thematic interpretation of an utterance.

Findings by Knoeferle et al. (2004) revealed the coordinated influence of depicted events on structural disambiguation and incremental thematic role assignment by monitoring people's eye-movements in visual scenes during comprehension of an utterance that related to the scene. Specifically, they examined the time-course with which depicted events (e.g., princess-washing-pirate, fencer-painting princess, see Fig. 3) enabled structural disambiguation and incremental thematic role assignment of initially structurally ambiguous German subject-verb-object (SVO)/ object-verb-subject (OVS) sentences. Listeners heard *Die Prinzessin wäscht/malt den Pirat/der Fechter* ('The princess (amb.) washes/paints the pirate (obj./patient)/the fencer (subj./agent)'). Once the verb had identified the relevant depicted action anticipatory eye-movements in the event scenes provided evidence for expectations of a patient (the pirate) and agent role filler (the fencer) for initially ambiguous SVO and OVS sentences respectively.

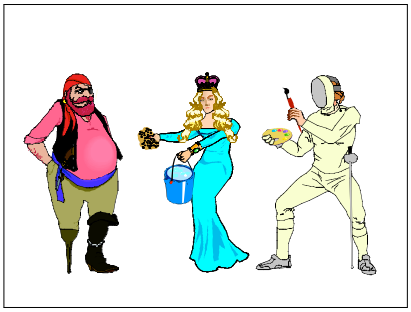

Figure 3: Example image from Experiment 1 (Knoeferle et al., in press)

Findings from the third experiment reported by Knoeferle et al., (2004) demonstrated that even when the main verb was sentence-final and did not establish early reference to the depicted events, linguistic cues (temporal adverbs) that appeared prior to the verb still enabled disambiguation. Crucially, this shows that the rapid coordinated influence of depicted agent-action-patient events on on-line utterance comprehension does not depend upon reference by the main verb. Even when the verb did not make the depicted actions available for early disambiguation and incremental role assignment, soft adverbial cues were sufficient to make the relevant scene information accessible.

The coordinated influence of depicted events on incremental thematic role assignment importantly generalizes to another language and sentence construction as revealed by Knoeferle, Crocker, Scheepers, and Pickering (2003). Results reported in their paper demonstrated the early verb-mediated influence of depicted events for on-line disambiguation of the English main verb/reduced relative ambiguity, thus extending and generalizing findings from the comprehension of German SVO/OVS sentences.

Further to examining the *temporal coordination* of scene and utterance processing, Knoeferle and Crocker (2004) investigated the *relative importance* of depicted events by directly comparing their influence on incremental thematic role assignment to that of stored thematic role knowledge. It has been found that these two informational sources each affect incremental thematic role assignment on-line. Prior research by Kamide, Scheepers, and Altmann (2003) demonstrated the rapid influence of verb-based thematic role knowledge on incremental thematic role assignment. As detailed above, studies by Knoeferle and colleagues showed the rapid verb-mediated influence of depicted events on incremental thematic role assignment and structural disambiguation.

Based on these two sets of findings, Knoeferle and Crocker (2004) carried out an experiment that examined two research questions. The first aim was to see whether it was possible to replicate that stored knowledge of thematic role relations (that were not depicted) and immediately depicted events (that were non-stereotypical) each allow rapid thematic role interpretation.

To investigate this issue, participants were presented an utterance that *uniquely* identified only either a stereotypical agent or an agent of a depicted event. An example scene showed a wizard, a pilot and a detective serving food (see Fig. 4). Sentences were all in object-verb-subject order. When people had heard 'The pilot (object/patient) *serves food to* …', perception of the detective who was depicted as serving food to the pilot enabled early incremental thematic interpretation of the detective as the agent of the food-serving action after the verb (no other agent in the scene was a plausible agent for a food-serving action, or depicted as serving food). In contrast, after people had heard the beginning of a comparison sentence ('The pilot (object/patient) *jinxes* …') eye-movements revealed that the wizard was identified as the most plausible agent on the basis of stored thematic role knowledge (no other agent in the scene was a plausible agent for a jinxing action or depicted as performing such an action). This was revealed by a higher proportion of anticipatory eye-movements to the stereotypical agent (wizard) than to the other agent when the verb was 'jinx'. In contrast, when the verb was 'serve-food-to', there were more inspections to the agent of the depicted event (detective) than to the other agent.

Figure 4: Example item from Knoeferle & Crocker (2004)

The second goal of this study was to determine the relative importance of depicted events and verb-based thematic role knowledge. To test this issue, participants heard an utterance in which the verb did not determine uniquely whether the comprehension system should rely on verb-based thematic role knowledge (identifying a stereotypical agent) or on depicted events (identifying an alternative, agent of a depicted event) for thematic interpretation of the utterance ('The pilot (object/patient) *spies-on* …', see Fig. 4). Both stereotypical thematic role knowledge and scene events provide relevant information about thematic relations. The detective is a stereotypical agent of the spying action identified by the verb (but is not depicted as performing such an action). The wizard, in contrast, is depicted as involved in a spying action but is not a stereotypical agent for a spying action. In this case, there was a strong preference of the comprehension system to rapidly rely on depicted events over stored thematic knowledge for processes of incremental thematic role assignment. Evidence for this came from a higher proportion of anticipatory eye-movements to the depicted agent (the wizard) in comparison with the stereotypical agent (the detective) shortly after people had heard the verb.

The case for *coordinated* processing

In sum, the two main findings reported above are first that shortly *after* a cue in the utterance mediated a relevant depicted event, the eye-gaze pattern provided evidence for rapid structural disambiguation and incremental thematic role assignment. Second, depicted scene events when identified by a verb have a greater relative importance in assigning thematic role relations than stereotypical thematic role knowledge associated with a verb. In what follows, we discuss the implications of these two findings for accounts of on-line incremental utterance comprehension in situations

when both utterance and scene are relevant for comprehension.

Recall that while findings by Tanenhaus et al. (1995) reveal a close time-lock between the mention of a word and the time when listeners establish reference to relevant scene objects, they do not allow to clearly determine a closely time-locked reciprocal influence of the perceived scene in determining the structuring of the utterance. Results from Knoeferle and colleagues thus importantly extend the findings by Tanenhaus et al. (1995). The rapid verb-mediated influence of the depicted actions on on-line sentence comprehension importantly revealed a tight coupling of visual and linguistic processing. This was apparent from the eye-movement records: Unlike in the Tanenhaus et al. (1995) studies, fixation patterns between the SVO and OVS condition did not differ from the start of the utterance. Rather, fixation patterns in the experiments by Knoeferle et al. (in press) only diverged *after* people had heard the verb that identified the relevant depicted action. The finding of a close coordination of structural disambiguation with the time when the verb identified the action and its associated thematic relations allows us to exclude a procedural account in which the timing of the influence of scene events is underspecified (e.g., Tanenhaus et al., 1995).

Let us now in more detail consider the extent to which the findings from experiments by Knoeferle et al. (in press), and Knoeferle and Crocker (2004) are compatible with the other lines of research that we briefly introduced further above (e.g., Bergen & Chang, to appear; Jackendoff, 2002; Tanenhaus et al., 2000). The finding that depicted events rapidly influence on-line structural disambiguation and incremental thematic role assignment seems to be broadly compatible with all of these frameworks. Note, however, that none of them predicts a close synchronization in the integration of visual and linguistic processing.

Results from Knoeferle and Crocker (2004) (i.e., the preferred reliance on depicted events as opposed to stored thematic role knowledge for incremental thematic interpretation) do not appear to be straightforwardly accounted for by these frameworks. The relative priority of depicted events during comprehension cannot, for instance, be fully explained by a Jackendovian or Embodied Construction Grammar framework. Neither of these predicts a greater relative importance of depicted events than of verb-based thematic role knowledge during comprehension.

Interactionist models can also not account for the greater relative importance of depicted events (e.g., Tanenhaus et al., 2000). Recall that this type of model would predict processing difficulty if constraints that are comparable in strength compete. Findings from Knoeferle and Crocker (2004) importantly showed that the competing informational sources - stored thematic role knowledge and depicted events - were comparable in strength: Each of these two types of information enabled thematic interpretation rapidly when the verb uniquely identified them. The model further would appear to predict that when equally strong constraints compete, they cannot rapidly be applied during processing. This was clearly not the case in the experiment by Knoeferle and Crocker (2004). When the verb identified two informational sources (verb-based thematic role knowledge and depicted actions) that were comparable in strength, comprehension processes rapidly relied on depicted events for incremental thematic interpretation.

Since existing frameworks do not predict findings of the relative importance of depicted events, the first step is to identify the factor(s) that caused the preferred reliance of the comprehension system on depicted events in comparison with stored thematic role knowledge. The origin of the preferential reliance on scene events might, for instance, derive from developmental and/or evolutionary comprehension strategies. The rapid impact of explicitly depicted thematic relations between entities identifies the comprehension system to be highly adapted towards acquiring new information from its environment rather than always relying on linguistic and world knowledge.

Such a proposal is compatible with theories of language acquisition such as Gleitman (1990). In her account of language acquisition Gleitman argues that a child can extract event structure from the world around it. When the child perceives an event, the structural information it extracts from it can determine the interpretation of a sentence that describes that event. The interpretation of a sentence can in turn direct the child's attention within the visual environment. The fact that the child can draw on two informational sources (sentence and scene) enables it to infer information that it has not yet acquired from what it already knows. Assume, for instance, the child perceives a girl throwing a ball, and hears the sentence *The girl is throwing the ball*. Assume further, the child knows the word *girl*. Even if it does not know what throwing means, perception of the event can enable it to deduce the meaning of the unknown verb it just heard. It can further identify *girl* as the thing performing the throwing-action.

Such a developmental account offers one explanation for why the comprehension system relies in preference on depicted events over stored thematic role knowledge. Indeed, the observed priority of depicted events may have developed in the course of language acquisition. There is furthermore clearly an important interplay between the function of language in guiding attention in the immediate scene during language acquisition and the influence of scene information on language understanding. The developmental account that Gleitman proposes motivates thus not only the relative priority of depicted events, but also the close temporal coordination between utterance comprehension and the use of scene information during comprehension that Knoeferle et al. (in press) observed in their experiments.

The experiments that we have discussed in exploring the interplay of linguistic and visual processing share one fundamental aspect: the utterances are about the immediate visual scene. In language acquisition, when parents talk to their children, language is likely often about the immediate scene which children typically explore. In these situations, it

serves a specific function, namely making the immediate scene accessible for the child, and identifying objects that it perceives (see, e.g., Roy & Pentland, 2002, for related research in modeling).

During adult language comprehension, however, this is not true to the same extent. Language often sub-serves other functions and is only used to refer to entities in the immediate scene for part of the communication we engage in. Much of our day is spent talking, reading, or writing about things not immediately present. Examples that come to mind are the expression of abstract ideas, or the narration of past events. As a result of this observation, we deem it necessary to qualify the findings by Knoeferle and colleagues. We acknowledge that in situations where the utterance does not directly refer to the immediate visual environment, depicted events will almost certainly not have the importance that is suggested by the above-discussed findings on the influence of depicted events since the scene is irrelevant. It is the immediate presence and relevance of both utterance and scene during comprehension, which enables the rapid interplay between these two informational sources.

We do expect, however, that in situations where the utterance is about the immediate environment, the findings of the priority of depicted events over stored thematic role knowledge in thematic interpretation will hold true. We propose that while language is often not about the immediate scene in adult life, we have spent a substantial part of our lives acquiring language. We suggest that this period may indeed have shaped both our cognitive architecture (i.e., providing for rapid closely temporally coordinated interaction between cognitive systems such as language and vision), and comprehension mechanisms (e.g., we rapidly and in preference avail ourselves of information from the immediate scene when the utterance identifies it). Specifying in more detail the nature of multi-modal comprehension (the temporal coordination and relative importance of distinct cognitive processes) may eventually lead to a more fully specified procedural account within a framework of comprehension such as Jackendoff (2002) or Bergen and Chang (to appear).

Acknowledgments

This research was funded by a PhD scholarship to the first author, and by SFB 378 project "ALPHA" to the second author, both awarded by the German Research Foundation (DFG).

References

Bergen, B. & Chang, N. (to appear). Embodied Construction Grammar in simulation-based language understanding. In J. Ostman & M. Fried (eds), *Construction Grammar(s): cognitive and cross-language dimensions*. John Benjamins.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Waler (Eds.), *Sentence processing: psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Lawrence Erlbaum.

Frazier, L., & Fodor, J. D. (1979). The sausage machine: a new two-stage parsing model. *Cognition*, 6, 291–325.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1, 3–55.

Jackendoff, R. (1997). *The architecture of the language faculty. Cambridge*, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, 32, 37–55.

Knoeferle, P., Crocker, M.W., Scheepers, C., & Pickering, M.J. (2003). Actions and roles: using depicted events for disambiguation and reinterpretation in German and English. In: *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 681-686), Boston, MA.

Knoeferle, P., Crocker, M.W., Scheepers, C., & Pickering M.J. (in press). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*.

Knoeferle, P., & Crocker, M.W. (2004). Stored knowledge versus depicted events: what guides auditory sentence comprehension? In K. Forster, D. Gentner, & T. Regier, *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 714-719). Mahwah, NJ: Lawrence Erlbaum.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.

Roy, D., & Pentland, A. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science*, 26, 113–146.

Spivey-Knowlton, M. J. (1994). Quantitative predictions from a constraint-based theory of syntactic ambiguity resolution. In M. Mozer, J. Elman, P. Smolensky, D. Touretzky, & A. Weigand (Eds.), *The 1993 Connectionist Models Summer School* (pp. 130–137). Hillsdale, NJ: Erlbaum.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 632–634.

Tanenhaus, M. K., Trueswell, J. C., & Hanna, J. E. (2000). Modeling thematic and discourse context effects with a multiple constraints approach: implications for the architecture of the language comprehension system. In M. W. Crocker, M. J. Pickering, & C. Clifton (Eds.), *Architectures and mechanism for language processing* (pp. 90–118). Cambridge: Cambridge University Press.

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: an effect of object name length on eye movements. *Psychological Science*, 11, 125–131.