



Brief Article

The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements

Andrea Weber^{a,*}, Martine Grice^b, Matthew W. Crocker^a

^aSaarland University, FR 4.7 Psycholinguistik, 66041 Saarbrücken, Germany

^bUniversity of Cologne, Cologne, Germany

Received 1 July 2005; accepted 4 July 2005

Abstract

An eye-tracking experiment examined whether prosodic cues can affect the interpretation of grammatical functions in the absence of clear morphological information. German listeners were presented with scenes depicting three potential referents while hearing temporarily ambiguous SVO and OVS sentences. While case marking on the first noun phrase (NP) was ambiguous, clear case marking on the second NP disambiguated sentences towards SVO or OVS. Listeners interpreted case-ambiguous NPs more often as Subject, and thus expected an Object as upcoming argument, only when sentence beginnings carried an SVO-type intonation. This was revealed by more anticipatory eye movements to suitable Patients (Objects) than Agents (Subjects) in the visual scenes. No such preference was found when sentence beginnings had an OVS-type intonation. Prosodic cues were integrated rapidly enough to affect listeners' interpretation of grammatical function before disambiguating case information was available. We conclude that in addition to manipulating attachment ambiguities, prosody can influence the interpretation of constituent order ambiguities.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Anticipatory eye movements; Prosody; Constituent order ambiguity; Grammatical function; Intonation; Case marking; Eye tracking

1. Introduction

It has been established that listeners can use prosody to resolve syntactic attachment ambiguities such as early and late closure (e.g. [Kjeelgaard & Speer, 1999](#); [Schafer, Speer,](#)

* Corresponding author. Tel.: +49 681 302 6553; fax: +49 681 302 6561.

E-mail address: aweber@coli.uni-sb.de (A. Weber).

Warren, & White, 2000) and prepositional phrase attachments (e.g. Pynte & Prieur, 1996; Snedeker & Trueswell, 2003). The distribution of prosodic phrase boundaries and their realisation (e.g. boundary tones, final lengthening, pausing), as well as the presence or absence of pitch accents, have been shown to guide listeners' interpretation of attachment ambiguities. In this paper, we further investigate whether prosody can override strong syntactic preferences to resolve grammatical function assignment ambiguities in German.

German uses morphological case to mark grammatical function. Although four cases can be clearly distinguished, the system often features syncretism: In many NPs, nominative and accusative case share surface form. While in general, case marking is available for disambiguation, it often fails to discriminate between the two least oblique, and therefore most frequent, grammatical functions (Keenan & Comrie, 1977), namely Subject (nominative case) and direct Object (accusative case). Furthermore, German is a language with relatively free constituent order. The initial position in matrix declaratives observes very few restrictions regarding the kind of constituent it can host, which includes Subjects, Objects, as well as modifiers. As a result, both Subject-Verb-Object (SVO) and Object-Verb-Subject (OVS) surface orders are possible, and case marking on the Subject and Object may be ambiguous. In addition to the clear Subject-first preference (Hemforth, 1993), incorrect initial interpretation of constituent order ambiguity typically results in a much stronger garden-path effect than modifier attachment ambiguities. One possible reason for this is that reanalysis from an SVO to an OVS structure entails a complete reassignment of the verb's roles to both arguments.

In this study, we examine whether prosody can fill the functional gap arising from a combination of syncretism and free constituent order in German. We investigate whether, in the absence of unambiguous morphological and configurational information, prosody can influence the assignment of grammatical function. To assess this we observed anticipatory eye movements of listeners while they were inspecting a related scene. The timing and pattern of eye fixations to visually displayed referents can be used to draw inferences about spoken-language comprehension (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Not only has it repeatedly been shown that referents in a scene are identified as soon as they are referred to in an utterance, there are several studies revealing that they can be identified *prior* to their mention. Kamide, Scheepers, and Altmann (2003), for example, have shown that unambiguous case marking, combined with verb selectional information, leads to post-verbal anticipatory eye movements in German SVO and OVS sentences. That is, immediately following the verb, and before hearing the second argument, listeners were able to use case to assign the appropriate grammatical function to the first argument and combine this with the semantics of the verb, resulting in increased anticipatory fixations to the appropriate second argument.

A recent eye-movement study by Knoeferle and colleagues (Knoeferle, Crocker, Scheepers, & Pickering, 2005), investigated the interpretation of German SVO and OVS sentences with case-ambiguous initial NPs. Structural disambiguation took place only at a second NP that was clearly case marked as either nominative (Subject) or accusative (Object). In their scenes, however, depicted actions were potentially able to resolve the ambiguity as soon as the verb was encountered. Their findings again revealed anticipatory post-verbal eye movements to the appropriate second argument, indicating that listeners

were able to use depicted events to resolve the ambiguity, and assign grammatical functions appropriately.

In the present study, we build upon the studies of Kamide, Scheepers, and Altmann, (2003) and Knoeferle et al. (2005), by considering the interpretation of German SVO and OVS structures with sentence-initial NPs that were ambiguously case marked as nominative or accusative. Morphosyntactic disambiguation of grammatical functions took place at the second NP that was clearly case marked as either accusative or nominative. Scenes accompanying the sentences showed plausible Patients and Agents for the referent of the first NP in relation to a given action.¹ Actions were not depicted. Thus, even though the scene presented potential referents it could not help with disambiguating thematic roles. In contrast with previous studies, however, prosodic cues could potentially help listeners resolve the temporary SVO/OVS ambiguity.

Post-verbal anticipatory eye movements to the appropriate Patient or Agent were expected to inform us about listeners' interpretation of the sentence-initial NP: looks to the Patient imply the initial NP was interpreted as Subject (and therefore Agent), whereas looks to the Agent imply the initial NP was interpreted as Object (and therefore Patient). If intonation patterns influence the interpretation of grammatical functions, then listeners in our study should anticipate more often the Patient when the sentence begins with an SVO-type intonation, and the Agent if it begins with an OVS-type intonation.

2. Experiment

2.1. Method

Participants. Thirty-two native speakers of German, all students at Saarland University, took part in the experiment for monetary compensation. They had not participated in the pre-testing of the materials, and they had normal or corrected-to-normal vision and normal hearing.

Materials. Sixteen German nouns referring to animates were chosen as ambiguous characters (e.g. *Katze*, 'cat'), which could serve as either the Agent or the Patient of a given action (e.g. *jagen*, 'chasing'). For each ambiguous character and its specified action, two further animates were selected: one was a suitable Patient (e.g. *Vogel*, 'bird'), provided the ambiguous character was the Agent (i.e. cats chase birds); the other was a suitable Agent (e.g. *Hund*, 'dog'), provided the ambiguous character was the Patient (i.e. dogs chase cats). In a rating study with ten participants it was established that both thematic roles for the ambiguous character were equally plausible for the given action in combination with the two further animates (e.g. it was as plausible that a dog would chase a cat as it was that a cat would chase a bird).

¹ Throughout the paper we will refer to the grammatical functions Subject and Object when describing the linguistic input and to the thematic roles Agent and Patient when describing the entities in the scene. In the present study, Subjects are always Agents and Objects Patients.

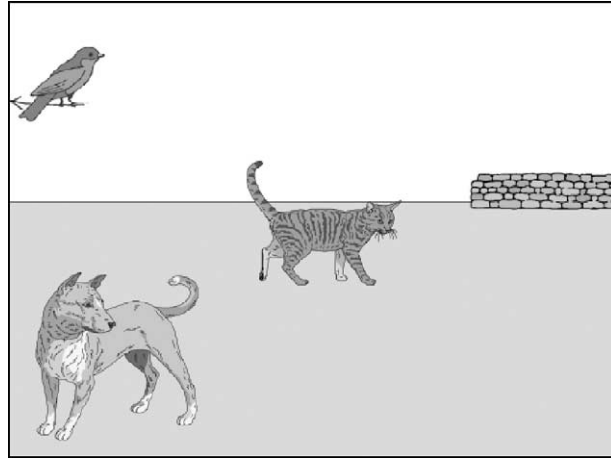


Fig. 1. Example of visual scene presented to participants.

Nouns referring to ambiguous characters were always of feminine or neuter gender, since surface forms of feminine and neuter NPs do not distinguish between nominative and accusative case in German. Nouns referring to the two further animates were of masculine gender, where there is a clear distinction between the two cases. The pictures of an ambiguous character, its Patient and Agent, and an additional inanimate distractor object were displayed together in a scene (see Fig. 1). The position of different character types and their orientation was varied across scenes.

For each of the 16 experimental scenes, two sentences were created. The first type of sentence had a temporarily ambiguous Subject-Object ordering of arguments, starting with a case ambiguous NP1, followed by a present tense monotransitive verb in active voice, an adverb, and an unambiguously accusative NP2 (see Table 1). The second type of sentence had a temporarily ambiguous Object-Subject ordering of arguments, starting with the same case ambiguous NP1, followed by the same verb and adverb, but ending with an unambiguously nominative NP2. Crucially, syntactic disambiguation was not available until NP2. When case-marking on NP2 assigned an Object function, the function of NP1 was disambiguated as Subject; conversely, when case-marking on NP2 assigned a Subject function, the function of NP1 was disambiguated as Object.

Table 1
Sentence pair accompanying the scene in Fig. 1

SVO	<i>Die Katze jagt womöglich den Vogel.</i> 'The cat (NOM, amb.) chases possibly the bird (ACC).' The cat is possibly chasing the bird.
OVS	<i>Die Katze jagt womöglich der Hund.</i> The cat (ACC, amb.) chases possibly the dog (NOM).' The cat is possibly chased by the dog.

Sixteen filler trials were added. Similar to the experimental scenes, the filler scenes were also composed of four pictures. Sentences accompanying the filler scenes, however, could vary in different ways: unambiguously case-marked NP1s, different verb forms and positions, more than two pictures referred to, Objects with prepositional phrase, and dative Objects. Filler sentences had both Subject-first and Object-first orderings of arguments.

Coloured pictures were selected from a commercially available collection of coloured line drawings (IMSI MasterClips, 1990) and further processed using Canvas[®]. The sentences were read aloud and recorded onto DAT in a sound-attenuated room by a phonetically trained female native speaker of German. The speaker was made aware of the local structural ambiguity in the context of the visual scene, and was asked to distinguish the two structures prosodically. No explicit instructions were given as to the prosodic structure. The two intonation patterns employed are transcribed using GToBI (Grice & Baumann, 2002; Grice, Baumann, & Benzmlüller, 2004) and presented in stylised form in Fig. 2a and b. In SVO sentences, the nuclear accent (sentence stress) was on the verb, whereas in OVS sentences it was on the first NP. In both cases the accent placement was marked. Due to the various syntactic structures employed, nuclear accents in filler trials appeared in varying positions.

Two lists were constructed. Both lists contained experimental and filler trials in pseudo-random order, such that before each experimental trial there was a filler trial. Experimental trials appeared once in a given list, either in SVO or OVS condition. Each list contained an equal number of SVO and OVS trials. Two representative practice trials were added at the beginning of each list.

Procedure. At the beginning of a session, participants received written instructions telling them to listen to the sentences and to inspect the images. Participants were also informed that the session would end with a brief sentence-recognition test. After calibration of the head-mounted SMI EyeLink eye tracker, each participant was presented with the trials from one of the trial lists. Images were presented on a 21" colour monitor at a resolution of 1024 × 768 pixels. Sentences were presented auditorily over headphones and started 1000 ms after the appearance of an image on the screen. A camera on

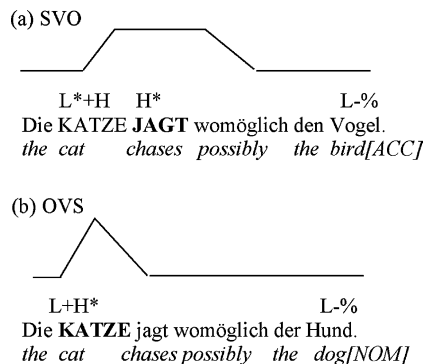


Fig. 2. Stylised contours of intonation patterns for SVO (a) and OVS (b) sentences. Capital letters are used on accented words, bold caps on those bearing the nuclear accent. SVO sentences had an initial accent on NP1 and the nuclear accent on the verb. OVS sentences had the nuclear accent on NP1.

the participants' dominant eye provided the input to the tracker. Onset and offset times and the spatial coordinates of fixations were recorded (250 Hz sampling rate). Between trials, participants were told to fixate a dot in the center of the screen for an automatic drift correction. For the analysis, custom-made graphical software was used to display locations of participants' fixations as dots superimposed on trial displays. Fixations were coded as pertaining to the ambiguous character, the Patient, the Agent, or the distractor. Fixations outside any displayed object were coded as looks to the background. Blinks were added to previous fixations. Consecutive fixations on one object (i.e., no intervening saccade to another object) were counted as one inspection.

3. Results and discussion

One scene was displayed distortedly during the experiment and had to be removed from analysis. For the remaining 15 items, we computed proportions of inspections launched during the presentation of the verb and the adverb. The two regions following the ambiguous first NP are critical because during that time eye movements to anticipated Patients or Agents occur before the disambiguating second NP refers to them. We report hierarchical log-linear models with the two factors character type (two levels *Patient* and *Agent*) and sentence type (two levels *SVO* and *OVS*; following the analysis in Knoeferle et al., 2005).

Fig. 3a shows average proportions of inspections that started during the presentation of the verb (e.g. *jagt*, 'chases'), the region immediately following the case-ambiguous NP1. As can be seen in Fig. 3a, during the verb listeners inspected the suitable Patient (*Vogel*, 'bird') more often than the Agent (*Hund*, 'dog') in both *SVO* and *OVS* sentences. Log-linear analyses revealed a main effect of character type, with the Patient being preferred over the Agent ($\text{LR}\chi^2_{(\text{part.})} = 6.73, P < 0.01$; $\text{LR}\chi^2_{(\text{item})} = 6.73, P < 0.01$), but no interaction between character type and sentence type ($\text{LR}\chi^2_{(\text{part.})} = 2.03, P > 0.1$; $\text{LR}\chi^2_{(\text{item})} = 1.35, P > 0.2$). More inspections of the Patient imply that listeners preferred to interpret the case-ambiguous first NP (*die Katze*, 'the cat') as Subject and were therefore anticipating the Patient as upcoming Object. This reflects the well known preference for the canonical Subject-first structure in German.² The absence of an interaction between character type and sentence type shows that prosodic marking of the structural ambiguity had not influenced its interpretation at this point of the sentence.³ In fact, the distinctive prosody of the sentence can be discerned only towards the end of the verb, by which time either a fall

² Planned comparisons showed that there were not significantly more looks to the Patient than the Agent for the *OVS* alone ($\text{LR}\chi^2_{(\text{part.})} & \text{LR}\chi^2_{(\text{part.})} < 1$). This could be seen to imply that the prosody begins to affect eye-movements already during the verb.

³ It typically takes about 150–200 ms before a programmed eye movement is launched (e.g. Fischer, 1992; Matin, Shao, & Buff, 1993), although considerable variation has been observed (Altmann & Kamide, 2004). Observed fixations are thus triggered by acoustic information that has been presented somewhat earlier. The lack of a prosodic effect during the earlier region of the verb, could therefore be due to fixations during that region being partly triggered by acoustic information from NP1.

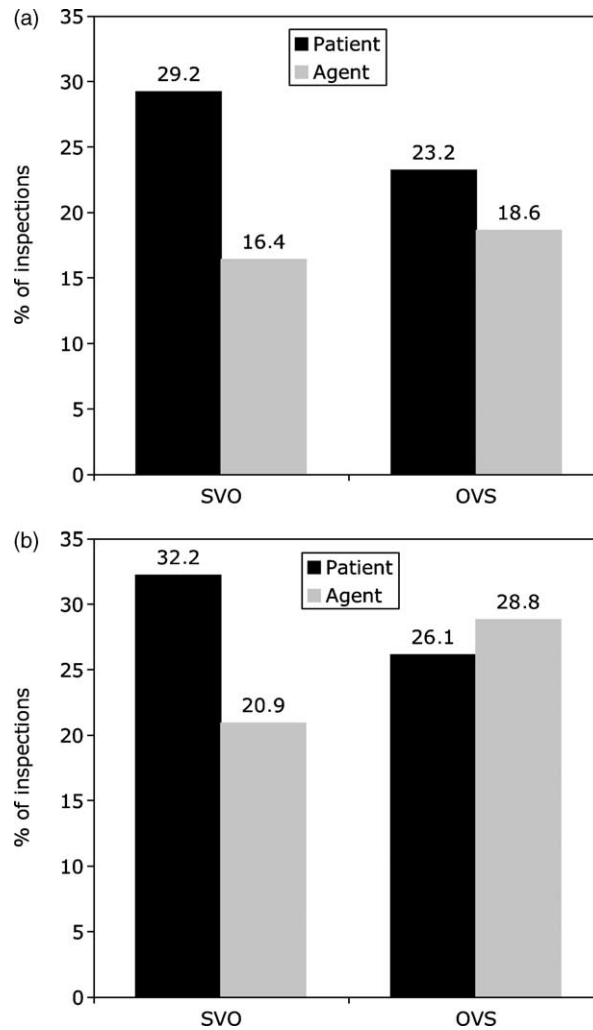


Fig. 3. Percentages of inspections for Patients and Agents during the verb (a) and during the adverb (b).

in pitch will have indicated that the accent on NP1 was nuclear (OVS), or a further accent on the verb will rule out the nuclear accent on NP1 (SVO).

The situation is different, however, in the region following the verb, as is shown in Fig. 3b which depicts average proportions of inspections launched during the adverb (e.g. *womöglich*, 'possibly'). In SVO sentences, suitable Patients (*Vogel*, 'bird') were still more often inspected than Agents (*Hund*, 'dog'); in OVS sentences, however, more inspections of Agents than Patients were found. Crucially, the interaction between sentence type and character type was significant ($LR\chi^2_{(part.)} = 5.14, P < 0.05$; $LR\chi^2_{(item)} = 6.52, P < 0.05$). By the time listeners heard the adverb, prosodic information had influenced the interpretation of grammatical functions: In sentences beginning with the OVS-type intonation, listeners

were less likely to anticipate an Object as second argument. Even though prosody did not fully reverse the preference for the canonical Subject interpretation of an ambiguous NP1—the difference between inspections of Agents and Patients was not significant for OVS sentences ($LR\chi^2_{(part.)}$ & $LR\chi^2_{(item)} < 1$)—it was available to the parsing process early on and modulated the grammatical function assignment as the significant interaction between character type and sentence type showed.⁴

4. General discussion

In the present study we have shown that in the absence of clear morphological and configurational information, prosody can influence the assignment of grammatical function in German. When presented with temporarily ambiguous SVO and OVS sentences, prosodic cues provided by our speaker modulated listeners' Subject and Object assignment: Listeners interpreted case-ambiguous NP1s more often as Subject and expected an Object as upcoming argument only when sentence beginnings carried what we refer to as an SVO-type intonation. This was revealed by more anticipatory eye movements to suitable Patients than Agents in visual scenes. No such preference was found when sentence beginnings had an OVS-type intonation.

The influence of prosodic structure on anticipatory eye-movements reveals a human sentence processor that rapidly integrates prosodic as well as syntactic, and verb-specific role information during the interpretation of temporarily ambiguous sentences. While our findings speak against parsing accounts which strictly prefer the pursuit of canonical (e.g. SVO) structures until some point of lexical disambiguation, it is important to note that the prosodic manipulation in our studies only eliminates the SVO preference; it does not make OVS preferable. This contrasts, for example, with the studies of Knoeferle et al. (2005) in which richly depicted event scenes were able to make the OVS structure preferred. Moreover, Kaiser and Trueswell (2004) report on a discourse-related effect on anticipatory eye movements, in that they are preferentially directed towards discourse-new referents. Taken together, these recent studies paint a picture of a spoken sentence comprehension system that rapidly and adaptively exploits diverse syntactic, semantic, pragmatic, visual, and intonational information for resolving local ambiguity, and for interpretation more generally.

A closer look at the intonation patterns and their use in the context of the experiment reveals how prosody was able to affect the interpretation of grammatical function: By default, West-Germanic languages place the nuclear accent on the last argument of the verb (Ladd, 1996). In our study this would be NP2. Default prosody is typically produced on broad focus sentences, for example, in an answer to the question “What happened?”. Thus, in both SVO and OVS sentences the placement of nuclear accents by our speaker, on the verb and on NP1 respectively, was non-default. One might expect that canonical SVO

⁴ Prosodic effects were found prior to the disambiguating second NP even in the first half of the experiment. It is therefore highly unlikely that our findings can be purely attributed to having learned an arbitrary association between a particular prosodic pattern and a sentence structure.

sentences spoken in isolation carry default prosody. However, visual cues can affect the situational givenness of referents, which in turn can influence the accent patterns in a sentence: Actions were not depicted in our scenes, the verb therefore referred to information which was new, or, cognitively speaking, inactive. By contrast, the referents of the verb's two arguments were situationally given (semi-active), since they both appeared in the scene (Baumann & Grice, 2004, in press; Chafe, 1984; Lambrecht, 1994). Prosodic focus on the verb in SVO sentences - the only new piece of information - was therefore appropriate and also perceived as such, as the pattern of anticipatory eye movements showed. Prosodic focus on NP1 in OVS sentences, on the other hand, placed a higher degree of prominence on NP1 than would have been necessary, given that the referent of the argument was displayed in the scene. An interpretation of a marked syntactic structure (OVS) was thus facilitated.

Note, that we do not make the claim that a nuclear pitch accent on NP1 is the only way for OVS sentences to be intoned. In fact, speakers in pilot recordings produced a number of different intonation contours for OVS structures including intonation phrase breaks and silent intervals after NP1. We did not incorporate these contours into the present experiment, to avoid greater durational differences between SVO and OVS sentences. Neither do we claim that the chosen OVS intonation would lead exclusively to the interpretation of Object-first structures. Given, for instance, the focusing question *Wer jagt den Vogel* ('Who is chasing the bird') a nuclear accent on NP1 in the SVO answer *Die Katze jagt den Vogel* ('The cat is chasing the bird') is perfectly acceptable. However, in the absence of further linguistic context, marked prosody on case-ambiguous NP1s, in combination with the visual scene in our experiment, guided its incremental interpretation towards Object.

Acknowledgements

This research was funded by SFB 378 'ALPHA' to the first and third author, awarded by the German Research Foundation DFG. We are grateful to Berthold Crysmann for discussion of syntactic issues and for comments on an earlier draft.

References

- Altmann, G., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347–386). New York: Psychology Press.
- Baumann, S., and Grice, M. (in press). The intonation of accessibility. Special issue on Intonation and Pragmatics. *Journal of Pragmatics*.
- Baumann, S., & Grice, M. (2004). *Accenting accessible Information In Proceedings of Speech Prosody*. Japan: Nara pp. 21–24.
- Chafe, W. (1984). *Discourse, Consciousness, and time*. Chicago, London: University of Chicago Press.
- Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 31–45). New York: Springer.
- Grice, M., & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, 191, 267–298.

- Grice, M., Baumann, S., & Benz Müller, R. (2004). German intonation in autosegmental-metrical phonology. In S.-A. Jun (Ed.), *Prosodic typology—the phonology of intonation and phrasing*. Oxford: Oxford University Press.
- Hemforth, B. (1993). *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens*. Berlin: Infix.
- Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*, 113–147.
- Kamide, Y., Scheepers, C., & Altmann, G. (2003). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37–55.
- Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry*, *8*, 63–99.
- Kjeelgaard, M., & Speer, S. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*, 153–194.
- Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye movements in depicted events. *Cognition*, *95*, 95–127.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: CUP.
- IMSI MasterClips (1990). Premium image collection 303,000 [Image database]. (<http://www.imsisoft.com>).
- Martin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: information processing time with and without saccades. *Perceptual Psychophysics*, *53*, 372–380.
- Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, *11*, 165–192.
- Schafer, A., Speer, S., Warren, P., & White, S. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*, 169–182.
- Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*, 103–130.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, *268*, 1632–1634.