

## Eye Tracking as a Tool to Investigate the Comprehension of Referential Expressions

**Anke Karabanov, Peter Bosch, Peter König**  
Institute of Cognitive Science, University of Osnabrück

*In the study reported here we use eye-tracking methods closely related to the “Visual World Paradigm” to examine the processes underlying the comprehension of referential expressions. An investigation of the time course of these processes is of special interest for the controversial question of whether pronouns are understood the same way as full noun phrases. Our results show that both full noun phrases and unambiguous anaphoric pronouns are immediately followed by increased fixations on the corresponding referent in the visual scene and that both reach their fixation peak at about 1000 ms after the onset of the referential expression. This suggests that anaphoric pronouns are referentially interpreted very much like definite full NPs, and that no extra processing time is needed to resolve the anaphoric reference.*

### Introduction

It is crucial for the language comprehender to be able to track to whom or what different expressions refer. Whereas new discourse referents are often introduced by proper names or full lexical noun phrases, all languages also have a wide variety of different anaphors<sup>1</sup> that are used for referring back to referents previously introduced into discourse. Besides the fact that anaphorically used expressions cannot be used entirely "out of the blue", and although there are certain classes of expressions that are indeed primarily used anaphorically, like pronouns in particular, there is a considerable variety of expressions that allow for anaphoric use and there is considerable variation in their lexical specificity. In most languages, anaphors can range from various forms of zero anaphors through pronouns and definite noun phrases to repeated proper names. Due to these differences in lexical specificity anaphors also vary in the degree to which their interpretation is governed by the surrounding discourse. This means that full lexical noun phrases and proper names are typically more limited in their referential options by their lexical content than pronouns, which depend in their interpretation much more on the surrounding context. In this study, we will mainly focus on the interpretation process for full lexical noun phrases (NPs) and for anaphoric pronouns that are unambiguously determined in their reference by the preceding discourse.

In order to understand a sentence containing a pronoun, the listener must be able to pick up the interpretation from the relevant antecedents in the text.

(1) *The shop assistant told the craftsman that she was angry.*

To understand sentence (1) we must know that the word *she* refers to the shop assistant. While computational linguistics has been struggling for a long time with the complex task of

---

<sup>1</sup> Here and in the following text we will use the term *anaphor* in the wide, classical sense ("expression used anaphorically") and not in the more technical sense that is found in Binding Theory.

pronoun resolution, humans encountering pronouns in discourse solve this problem with unexceptional ease and without even being aware of any effort. As speakers of a language we feel that we can immediately relate the pronoun to its referent. This effortless is astonishing, especially since it seems to be necessary to consider quite a number of syntactic, semantic, and pragmatic constraints in order to determine the correct antecedent or referent for a pronoun (cf. the listing of factors in Nicol & Swinney, 2002).

But how do we manage to figure out to whom or what a pronoun refers in a certain context? The mechanisms of this process are not yet clear and have been the issue of controversial discussions in syntactic and semantic theory as well as in psycholinguistics for many years.

In the last two decades two opposing hypotheses about pronoun resolution have been held: Gernsbacher (1986) claimed that pronoun resolution happens in a two step process. First the antecedent of the pronoun is identified and then, in a second step, the connection to the referent of the antecedent is established. Tyler & Marslen-Wilson (1982), however, claimed that pronouns are immediately interpreted referentially, just like full lexical NPs or proper names. However the debate between these two positions as to the cognitive processes underlying reference phenomena, until recently, had to rely on empirical evidence of a largely indirect nature, such a reading times, reaction times, or eye movement during reading (cf. surveys in Nicol & Swinney 2002, Rayner & Clifton 2002).

To decide the issue, we need to look at the time course these processes. Eye tracking during spoken sentence comprehension seems to be a method perfectly suited for this purpose. With a head-mounted eye tracker we are able to record up to four eye fixations per second. Considering that each fixation can be seen as an unconscious decision of the participant about where to direct her attention it becomes clear that eye tracking is a very powerful tool to investigate comprehension processes online. Compared to more conventional psycholinguistic measures such as reading time or probe verification, eye tracking cannot only produce a huge amount of data, it can also monitor cognitive processes, like language comprehension, without any interruption, by directly monitoring the attentional focus of a participant at any time during a task.

Two main experimental setups are used in most eye tracking experiments concerned with language processing. In one setup participants eye movements are recorded during a reading task. Obviously this type of experiment is mainly suited to investigate the reading process, specifically the orientation of participants in the text. The second has become known under the term “Visual World Paradigm”. In this setup, developed by Tanenhaus and colleagues (Tanenhaus e.a. 1995), eye movements are recorded while participants view a visual display that is paired with an accompanying linguistic auditory stimulus. In the last ten years this paradigm has extensively been used to study different aspects of language processing.

In the experiment to be reported here we make use of a slightly modified Visual World setup. In contrast to many other studies within the Visual World paradigm (Runner, Sussman & Tanenhaus 2003, Arnold 2000, Tanenhaus 2000, etc.), we did not use cartoons or line drawings for our visual displays but detailed photographs of scenarios built up with Playmobil™ toy characters. This was motivated by the idea that line drawings may, and usually do, already contain an interpretation of the drawn object. The artist pre-selects certain features, which are depicted in the drawing, highlights some features and discards others. This may of course influence the focusing behaviour of the viewers. By using photographs of scenes composed from prefabricated objects as visual stimuli we thus hope to increase the naturalness and the general validity of our findings.

In the following we will look at two questions, which are essential for the understanding of pronoun resolution. First we want to find out if there are any interesting differences in the fixation probabilities for the referents of different referential expressions. Second, we want to investigate the time course of fixations for both full noun phrases and pronouns. And of special interest is the question whether pronouns require additional time for “resolution” or whether they are interpreted immediately, just like proper names, as was argued by Tyler & Marslen-Wilson (1982) and Garrod, Freudenthal & Boyle (1994). By determining if there is additional time needed for the resolution of pronouns, we hope to be able to provide evidence for or against the assumption of a distinct mechanism that establishes a link between pronoun and antecedent. If no temporal difference in the resolution process of pronouns and definite full NPs is found, we have a strong indication for Tyler & Marslen-Wilson’s assumption that pronouns are interpreted directly with respect to the discourse representation and that no prior linking to antecedent expressions needs to be assumed.

## **Methods**

The participants, who volunteered for the experiment, were 12 native German speakers (5 male). All participants were students of Cognitive Science at the University of Osnabrück. They were aged between 20 and 25 (mean 21.9), had normal or corrected-to-normal vision, and none reported any speech or hearing deficits which could have influenced their performance. One participant had to be excluded from the experiment because this participant's gaze remained almost static during the whole experiment, and we had to abort the experiment with another participant due to poor calibration (mean error  $>0.5^\circ$ ). In total, we were able to include ten participants in the analysis. All participants were naïve about the purpose of the experiment and received either course credits or payment for their participation. All participants were informed of the purpose of the experiment only after it had been completed.

Ten photographs were paired with pieces of narrative discourse. The photographs showed pseudo-natural everyday situations built up with Playmobil<sup>TM</sup> toy characters (See Figure 1). Each of the pictures comprised three objects that were named in the narrative discourse. Two of these objects were human characters, whereas the third one was either an inanimate object or an animal. The three referents were named in the corresponding discourse both by a lexical NP and by pronouns. The distractor objects that were present in each photograph were either inanimate objects or animals that fit the general context of the scene. For each photograph, a corresponding piece of narrative discourse was presented via loudspeakers.

The discourses consisted of three German sentences. The first sentence always described the general scene that was visible in the corresponding photograph, without referring to any specific object in the scene. The second sentence introduced the two human referents. The third sentence referred to each of the human referents at least once with a pronoun and introduced the non-human referent with a full lexical NP as in Table 1.



**Figure 1:** Shows an example of the visual stimuli presented to the participants; in this case for a stimulus text like the following: *"Heute ist Markt im Dorf. Die Marktfrau streitet mit dem Arbeiter. Sie sagt jetzt gerade, daß er kein' Ärger machen und das neue Fahrrad zurückgeben soll, das er sich geliehen hat."* [It's market day in the village. The market woman is quibbling with the worker. She's just saying that he should not make any trouble and should give the new bike back that he borrowed.]

The sentences were pre-recorded and were spoken by a female native German speaker (the first author). All pieces of discourse had the same number of syllables and the duration of the discourse pieces ranged between 13.5 and 13.9 seconds. Each of the discourses had four variants. The first two sentences, introducing the whole scene and the two human referents, did not change. In the third sentence, however, the pronouns were permuted as in Table 1

All conditions start with a pronoun (Pro1) that has NP1 as its antecedent. Pro2 has as its antecedent NP2 in the conditions A and C and NP1 in conditions B and D. Pro3 in conditions C and D has the antecedent NP1 in Condition C and NP2 in condition D. All four conditions end with Pro4, which has NP2 as its antecedent.

The idea behind the permutation is to see whether there are differences in fixation probability that might occur due to the relationship between pronoun and antecedent. Linguistic theory distinguishes between pronouns that have their antecedents within the same sentence and those that do not. Whereas pronouns with antecedents in another sentence are referential anaphoric pronouns, pronouns with their antecedent within the same sentence may either also be regular anaphoric pronouns or may be c-commanded and bound by their antecedents. In our material, all pronouns with their antecedent in the same sentence belong to the class of c-commanded pronouns that are bound by their antecedents and, as argued in Bosch (1983), occur non-referentially. In contrast, all pronouns in our material with antecedents in the preceding sentence are ordinary referential pronouns. According to this distinction, we would expect higher fixation probabilities on the matching referents for the referential pronouns. From the arrangement of our four conditions we hope to be able to test whether the

differentiation in bound versus referential pronouns is just a theoretical one, or if it can also be observed in human language understanding.

1. Heute ist Markt im Dorf. [It's market day in the village.]
2. **Die Marktfrau** (NP1) streitet mit **dem Arbeiter** (NP2). [The market woman is quibbling with the worker.]
3. **Sie** (Pro1) sagt jetzt gerade, [She's just saying]
  - A daß **er** (Pro2) kein' Ärger machen und **das neue Fahrrad** (NP3) zurückgeben soll, das **er** (Pro4) sich geliehen hat.  
[that he should not make any trouble and should give the new bike back that he borrowed.]
  - B daß **sie** (Pro2) kein' Ärger machen und nur **das neue Fahrrad** (NP3) zurückhaben will, das **er** (Pro4) sich geliehen hat.  
[that she does not want any trouble and only wants to have the new bike back that he borrowed.]
  - C daß **er** (Pro2) **ihr** (Pro3) jetzt **das neue Fahrrad** (NP3) zurückgeben soll, das **er** (Pro4) sich geliehen hat.  
[that he should now give her the new bike back that he borrowed.]
  - D daß **sie** (Pro2) **ihm** (Pro3) jetzt **das neue Fahrrad** (NP3) zurückgeben will das **er** (Pro4) sich geliehen hat.  
[that she now wants to give him the new bike back that he borrowed.]

**Table 1:** Sample narrative

Due to the fact that each narrative had four different conditions, we have a total amount of 40 stimulus sentences (10 pieces of discourse times 4 conditions). Each of the participants heard either conditions A and C or conditions B and D of each story plus 30 additional stories that served as fillers. In total 20 experimental and 30 filler scenarios were presented to each participant. The auditory presentation of the discourses varied between 13.5s and 13.9s for the experimental discourses and between 13s and 17s for the filler discourses. Even though the auditory stimulus was often shorter, each picture was presented to the participants for 17 seconds.

Eye movements were recorded using a binocular eye-tracker (Eye Link II, SR Research, Mississauga, Ontario, Canada, 2003). Three infrared cameras recorded the position of the participant's head and the movements of both eyes. The two cameras that record the eye fixations were placed under the participant's eyes. The eye-tracker was controlled by a Pentium 4 PC (Dell Inc., Round Rock, TX, USA) that sampled the eye position signal at a

rate of 250 Hz. Besides video-based pupil tracking, the eye-tracker included infrared cornea reflection which reduces susceptibility to headband slips and motion.

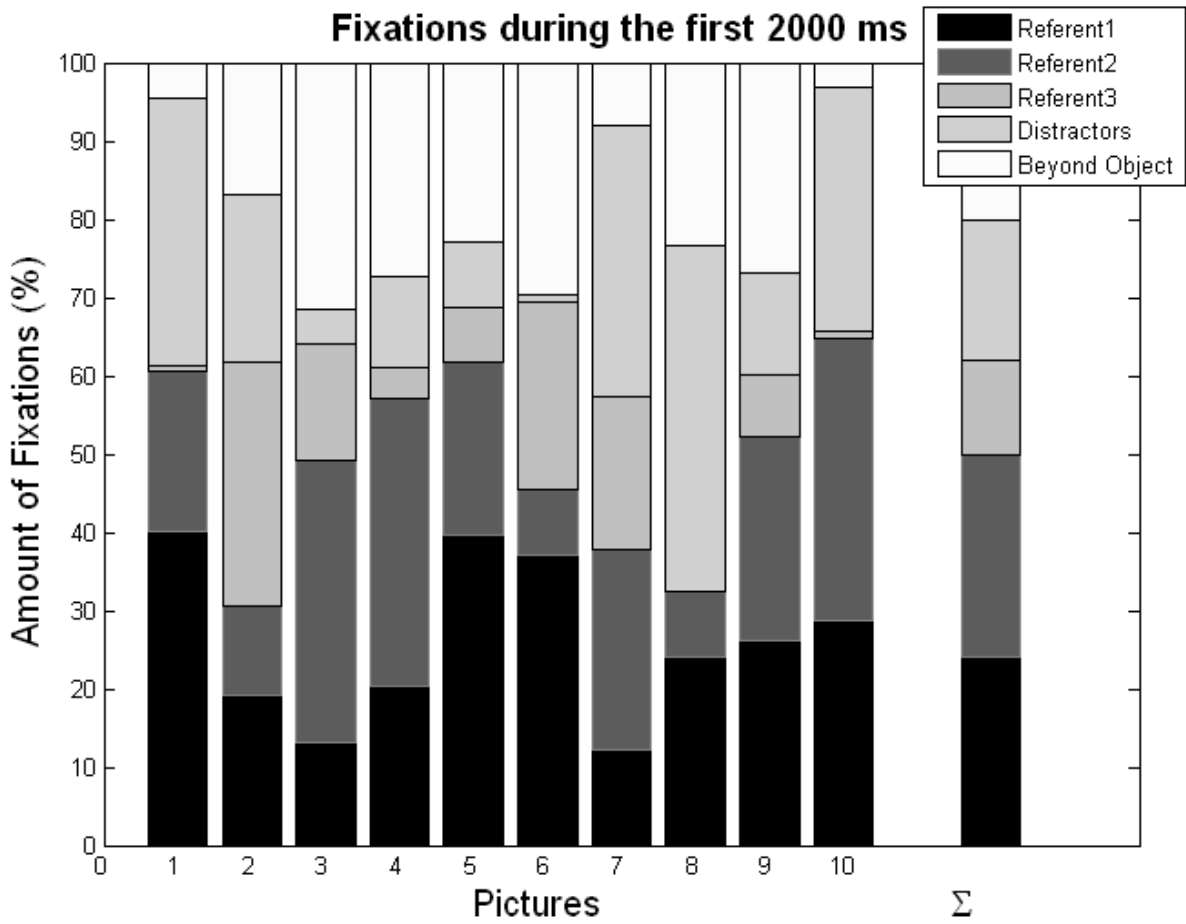
At the beginning of the experiment no information about the purpose was given. Both eyes were calibrated using the nine-point grid procedure. During this procedure, participants were asked to fixate on a small point, which appeared randomly at one of nine locations on the monitor. Only calibration values with a mean error  $<0.5^\circ$  were accepted during the validation procedure. Using a standard setting of the EyeLink II, the better eye was selected. Before each stimulus presentation a fixation point was presented. Stimulus presentation was triggered by the experimenter after the participant had stably fixated on the fixation point. This fixation point was used to perform a correction for drifts and slips of the eye-tracker and allowed participants to take a short break between trials. A total amount of 50 stimuli (20 experimental and 30 filler) was presented to each participant and the order of presentation was randomized. The experiment lasted about 30 minutes. Participants were instructed to “study the images carefully”.

## **Results**

As our visual stimuli differed considerably from the stimuli used in earlier visual world studies we first had to check for big variance in fixation probabilities between the different stimulus pictures and for the different objects within single scenes. By doing this, we could make sure that our stimulus material was valid and worthy of further examination. By looking at the first 2000 ms of our stimulus presentation – in which the scene was introduced without direct reference to any of the objects in the scene – we were able to ascertain that our scenes were perceived as meaningful stimuli with the two human referents in the centre of the scene. The analysis showed that the participants had a preference for the human characters in the scene already before they were mentioned explicitly in the narrative (see Figure 2).

This analysis furthermore revealed a relatively high fixation probability for distractor objects and thereby demonstrated that the distractor objects were also perceived and that the participants' focus did not exclusively lie on the three referents relevant in the stimulus sentences. The relatively high fixation probabilities that could not be assigned to any object in the scene (16.1 % Beyond Object Fixations) can be explained by the quite narrow definitions of regions of interest and by the fact that objects often stood so close to each other that it might have been possible to fixate both by looking in the empty space between them.

To check for the general influence of full NPs and pronouns on the fixation probabilities we summed up the fixation probabilities for all participants and all pictures. By this method we were able to obtain a graph that shows the change in fixation probability over the whole time course of the story (see Figure 3). Since we also summed over all the conditions this first evaluation of general fixation probabilities caused by full NPs and pronouns we can only include the first and the last of the pronouns into our analysis since only they were the same in all conditions. Figure 3 shows that both full noun phrases and pronouns elicit an increase in fixation probabilities for the matching referent. However, the fixation probabilities for full noun phrases (mean 43%) are higher than the fixation probability for pronouns (mean 35%). The time frames in which the fixation probabilities on a referent were significantly higher ( $p < 0.005$ ) than the fixation on other referents are depicted in the horizontal bars above Figure 3.

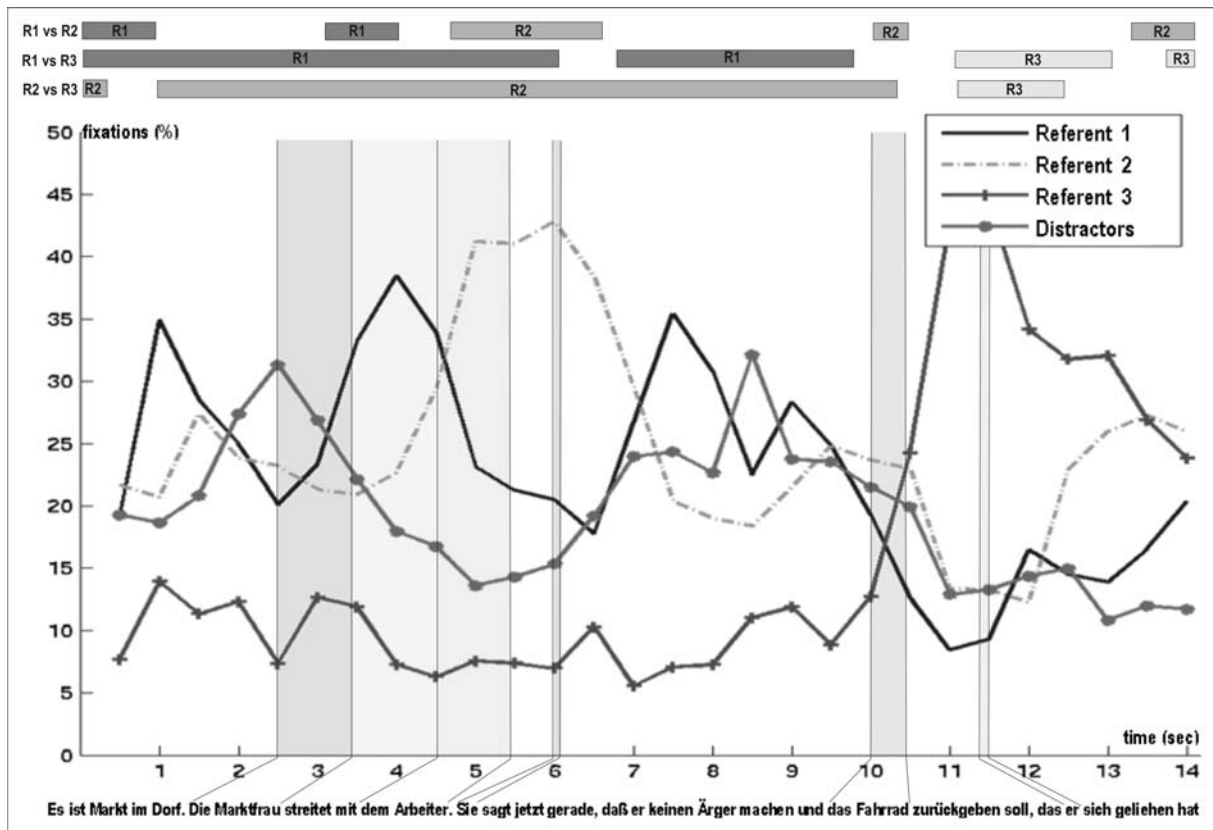


**Figure 2:** Shows the accumulated number of fixations on different objects in the pictures over the first two seconds before the onset of Sentence 2, which introduces the referents. As indicated on the x-axis, the columns stand for the different visual stimuli, with the last column showing the summed fixation probabilities over all pictures. On the y-axis, the fixation probability is depicted. The grey-shading indicates the fixated object.

In order to compare the peaks in fixation probabilities that were caused by full NP reference with those caused by pronoun reference, we looked at the fixation curve of each referent during both explicit and pronoun naming in a time window beginning 500ms before the onset of the referential expression and lasting until 2000 ms after the onset of the expression (Figure 4A for pronoun naming and Figure 4B for explicit naming). In Figure 4A we can see the fixation probabilities caused by the two pronouns. Pronoun 1 is the first pronoun in our narrative, standing very close to its antecedent, whereas Pronoun 2 is the last pronoun of our story, and has the longest distance to its antecedent. It can be seen in Figure 4A that Pronoun 1 reaches its probability peak at 1000 ms after the onset of the referential expressions and decreases directly after that. Pronoun 2 reaches its peak only at about 1500 ms.

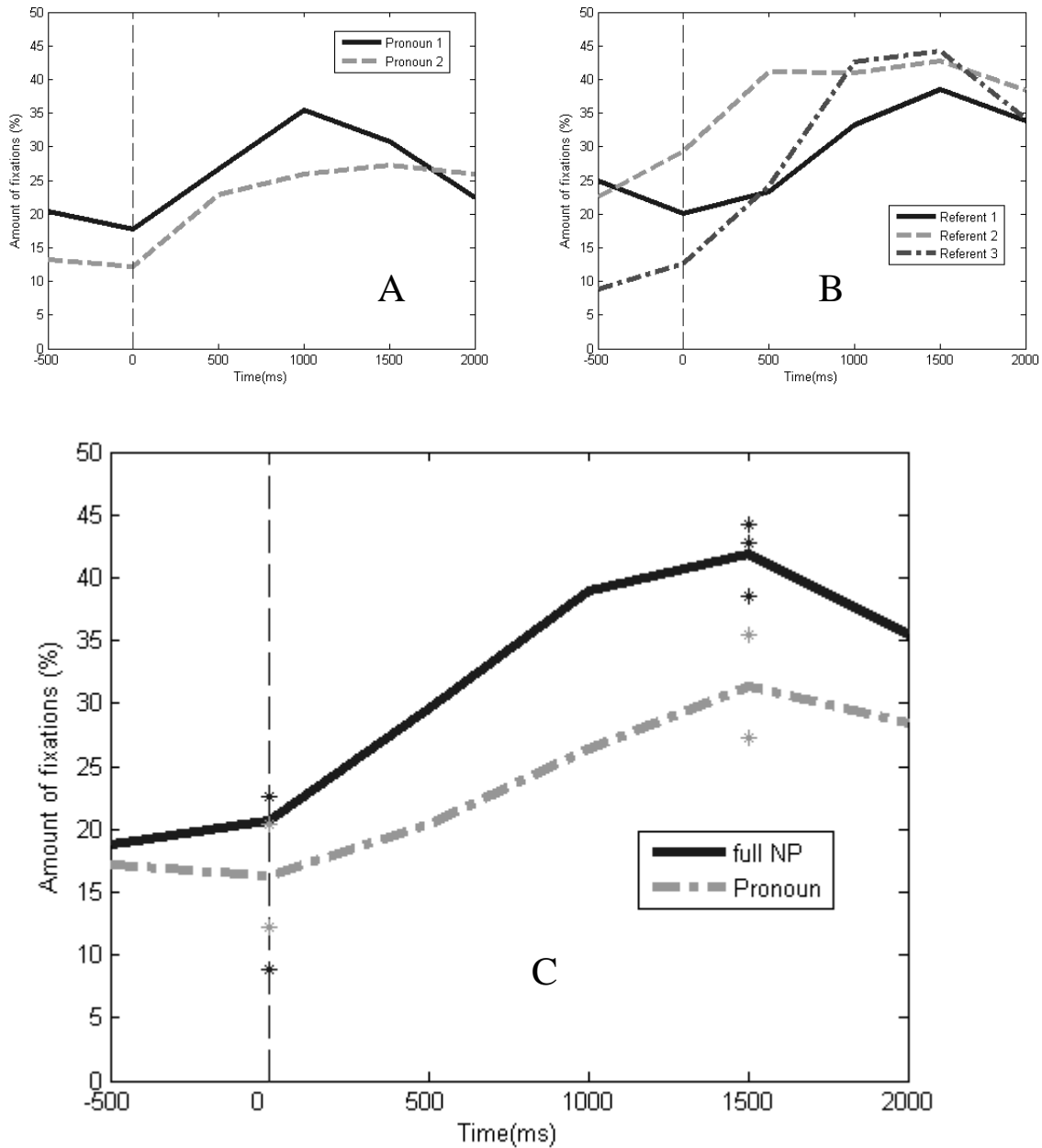
However, the difference in the two fixation curves did not become significant. In Figure 4B the different fixation curves are shown for the three referents that are referred to by a full NP. It is interesting to note that Referent 2 has already reached its peak 500 ms after the onset of the referential expression and stays at this plateau until 1500 ms after the onset of the expression, whereas Referent 1 only reaches its fixation peak at 1500 ms after the onset of the referential expression. Referent 3 has a fixation plateau from 1000 ms to 1500 ms. The fixation pattern of Referent 2, with its early peak, is significantly different from both other fixation curves ( $t = 0.0148$  for the comparison between Referent 1 and Referent 2 and  $t =$

0.128 for the comparison of Referent 3 vs. Referent 2). The difference between Referent 1 and Referent 3 did not become significant.



**Figure 3:** Shows the fixation probabilities for the different referents summed up over all conditions, subjects, and stimulus sentences. The x-axis shows the time course of the stimulus presentation and the y-axis shows the fixation probability. The shaded vertical bars in the background represent the duration of the referential expressions and interaction verbs. The horizontal bars at the top of the figure depict the time slots in which the t-tests between the different referents became significant.

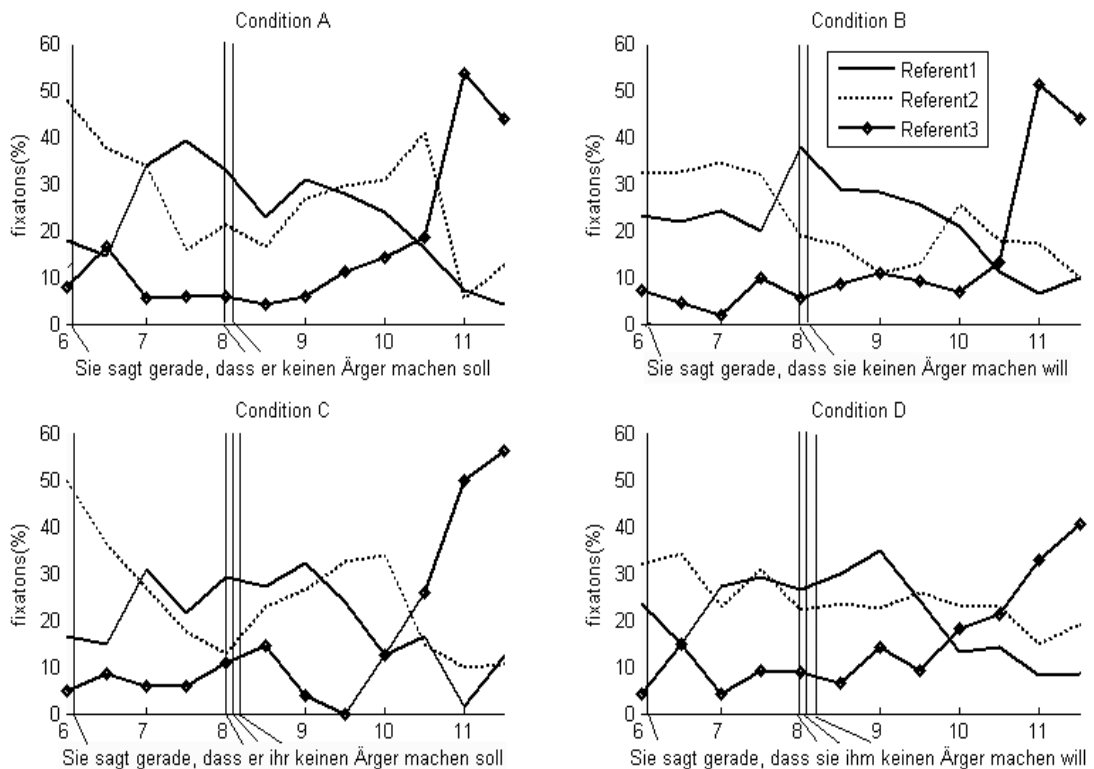
We merged the explicit fixation curves and the pronoun fixation curves for all referents. In doing so, we obtained one fixation curve for pronoun and one for full NP reference (Figure 4C). Both pronouns and full NPs lead to an increase of fixations on the visual referent. The fixation probability for pronouns reached its peak at 1000 ms after the onset of the pronoun. Until 1500 ms after the pronoun onset the fixation probability decreased only slightly, forming a plateau of highest fixation probability between 1000 and 1500 ms. Also in the case of full NPs, the peak of fixation probabilities was nearly reached within 1000 ms after the onset of the expression. However, fixations following full NP reference kept on increasing until 1500 ms after the onset of the expression forming a slightly increasing plateau from 1000 to 1500 ms. The peak in fixation probabilities caused by the full NPs reaches up to 43% and is significantly higher ( $p=0.00005$ ) than the fixation peak caused by pronouns, which reaches up to 32%.



**Figure 4:** Figure 4B shows the fixation probabilities for the three different referents when they were named explicitly. The bold line indicates Referent 1, the dark dashed-dotted line Referent 2, and the light dashed line Referent 3. The vertical dashed line indicates the onset of the referential expression. The time that is depicted on the x-axis ranges from 500 ms before the onset of the referential expression to 2000 ms after its onset. The y-axis shows the fixation probability in percent. Figure 4A shows the fixation probabilities for pronoun naming. In Figure 4C we see the mean fixation probabilities for explicit and pronoun naming. The bold line shows the mean fixation probability for explicit naming and the dashed-dotted line the probability for pronoun naming. The stars located around the word onset and around the fixation peak show the variance.

As already mentioned, each story had four different conditions. To account for differences between the four conditions we calculated the average of fixation probabilities over pictures and participants for each condition. The time frame between 7000 and 10000 ms was of special interest for the comparison between the four conditions, since it was in this time segment that the differences between the four conditions occurred (see Figure 5). To find significant differences between the four conditions, we conducted a series of t-tests ( $t = 0.05$ ), again for each 500 ms time slot. We tested each referent in Condition A against its counter-

part in Condition B, and each referent in Condition C against its counterpart in Condition D. The only comparison that became significant was the comparison of Referent 2 in conditions A and B in the time slot between 10000 and 10500 ms. This time slot is approximately two seconds after the onset of the crucial pronoun in this condition. During this time interval, the probability to fixate Referent 2 in Condition A was significantly greater than in Condition B. No other time frame and no other referent showed significant differences. The comparison between conditions C and D did not yield any significant results.



**Figure 5:** Shows the time frame (6 - 11 seconds) in which the differences between the four conditions occur. The only time slot in which there is a significantly different fixation probability between the four conditions is between 10000 and 10500 sec in conditions A and B. At that time the probability of fixating Referent 2 is significantly higher in Condition A.

## Discussion

In this study we looked at the comprehension of spoken text with a special focus on how referential expressions, in particular full NPs and pronouns, are understood. Our results can be summarized by three main findings:

1. Both full noun phrases and pronouns cause increased fixations on the matching referent.
2. The resolution of pronouns in unambiguous texts happens just as fast as the resolution of full NPs
3. There seem to be some pronouns that do not elicit higher fixation probabilities on the matching referent. These differences in referentiality between different pronouns may be due to the syntactic relation between the pronoun and its antecedent.

In the following discussion, we will first try to interpret each of these three findings, followed by an outlook on how the study presented could be improved in future work. Finally, we will try to place our findings into the existing framework of previous research.

Our results show that both full NPs and pronouns caused increased fixations on the matching visual referents. However, fixation probabilities for full NPs were significantly greater than the fixation probabilities for pronouns. This means that even though both kinds of referential expressions do elicit higher fixation probabilities on their matching referents, we are still able to detect differences between full NPs and pronouns in the amount of fixations they cause. Since we only used unambiguous pronouns<sup>2</sup> in our stimulus material, the differences in fixation probability between NPs and pronouns cannot be attributed to the fact that participants had problems in finding the right antecedents for the pronouns. However, even though we were able to detect this difference in fixation probability, this does not have to mean that full noun phrases generally create a "stronger" connection to the matching referent in the visual scene than pronouns do. We suggest, rather, that the difference in the fixation probability that we found would be due to the fact that, at least in our materials, the full NPs are all cases of "first mention" uses, whereas the pronouns only resume a referent that was already explicitly referred to previously by a full NP. In other words: While the participant's attentional focus gets directed by a full NP to an object not previously mentioned in the discourse, the pronouns only re-direct the focus back to an object that was mentioned and focussed before. It seems reasonable that a newly introduced object causes a higher amount of fixations than one that was already introduced and examined before. We also found that each full NP produces a fixation probability for the matching referent that is significantly higher than the fixation probabilities for all other referents and objects at this time point. The fixation probabilities caused by pronouns were also tested for significance. It is interesting to note that of the two pronouns tested<sup>3</sup>, only the one with the greater distance to its antecedents caused a significant difference in fixation probability compared to the other human referent. The pronoun with the shorter distance to its antecedents did not produce a significant difference compared to the fixation curve for the competing human referent, even though the magnitude of the fixation peak in percent was higher than for the second significant pronoun. This could be explained by the fact that the attentional focus was directed to both of the human referents at the time when the first pronoun occurred, since this pronoun directly followed the explicit introduction of both human referents by a full NP. Another interesting aspect is the temporal duration of the significant intervals for each referential expression. While the significant interval following the full NP referring to Referent 1 starts *at the offset* of the word and the NP referring to Referent 3 starts 500 ms after the offset of the word, the significant interval following the NP referring to Referent 2 already starts 200 ms *after the onset* of the referential expression. The fact that the significant interval for Referent 2 already starts very shortly after the onset of the expression could be due to the fact that Referent 2 is always preceded by an interaction verb like *talk, fight, speak* that requires a second human character as an object. This might enable the listener to anticipate the continuation of the story, since – apart from Referent 1, already referred to by the subject expression – Referent 2 is the only human object present in the visual scene and therefore the only one that allows a plausible unfolding of the story. The significant interval following the second pronoun starts 500 ms after the

---

<sup>2</sup> Gender agreement allowed only for one possible antecedent for each of the two pronouns.

<sup>3</sup> Only two of the four pronouns in each stimulus text were included in this analysis since only these two were the same in all four conditions. A detailed analysis of the pronoun differences in the different conditions will be discussed later.

offset of the referential expression and is thus similar to the pattern of explicit naming for Referents 1 and 3.

Besides the significant fixation probabilities caused by referential expressions, we were also interested in the temporal resolution of different referential expressions in general. As a first step, we compared the temporal pattern of the fixation curves caused by the three NPs. This comparison gives results very similar to those obtained by the comparison of the temporal duration of the significant intervals caused by full NPs. It shows that the NP referring to Referent 2 reaches its fixation peak already 500 ms after the *onset* of the referential expression. This is of special interest since it means that the peak is reached even before the end of this NP. As already mentioned, we assume that the interaction verb triggered an anticipation effect that may have caused this early fixation peak. The non-human Referent 3 reached its fixation peak 1000 ms after the onset of the referential phrase, whereas Referent 1 took 1500 ms until the highest fixation peak was reached. That Referent 3 reached its fixation peak 500 ms earlier than Referent 1 may partly be caused by the fact that the expressions referring to Referent 3 were on average 600 ms shorter than those referring to Referent 1.

We further compared the temporal pattern of the fixation curves caused by the pronouns. As already mentioned, this part of the analysis included only those pronouns that are the same in all four conditions. Whereas the first pronoun that we were looking at occurred in the sentence directly following the sentence containing its antecedent, the second pronoun had a full sentence distance to its matching antecedent. Comparing these two pronouns, we found that Pronoun 1, having a short antecedent distance, reaches its fixation peak already after 1000 ms, whereas Pronoun 2 needs 1500 ms until it reaches its fixation peak. However, the difference in percentage between the fixation probabilities caused by the two pronouns is not statistically significant. Comparing the mean temporal fixation pattern of full NPs with the mean fixation pattern for pronouns, we saw that the fixation curves look very similar. Both full NPs and pronouns reach the highest fixation probability between 1000 and 1500 ms after the onset of the referential expression. The fact that both have their highest fixation probability in the temporal interval between 1000 and 1500 ms after the onset of the expression indicates that there is no temporal delay for the resolution of pronouns in unambiguous texts.

One possible objection to this interpretation of our findings is that we compared the fixation curves from the onset of the referential expressions and not from their offset. Due to the longer duration of full NPs, their offset is much later than for pronouns. This means that even though the temporal resolution for both kinds of expressions is equal with respect to their onset, the temporal delay measured from the offset is much bigger for the pronouns. However, we decided to take the onset of the expression as a fixed point since we assume that even as the referential expression unfolds, participants may well already anticipate the matching referent. This assumption is supported by the eye-tracking experiment of Hartmann (2004) which examined gender effects in sentence processing in German. Her results show very convincingly that the gender information carried by the determiner of a full NP has an early effect on fixation probability for a matching referent, and supports an interactive view on language understanding, which claims that the comprehension process already starts during word recognition and not only after the offset of the expression. (Gernsbacher, 1989, MacDonald & MacWhinney, 1990, Tyler & Marslen-Wilson, 1982).

We should add, perhaps, that although we believe that our results contribute to a better understanding of the differences in the comprehension of full lexical NPs as opposed to pronouns, we did obviously not systematically vary all potentially relevant parameters. That

would require a far more comprehensive study. The current experiment should rather be seen as a more modest preliminary to such a systematic comparison, because first of all we had to establish that in the comprehension of spoken text vis-a-vis a relevant visual scene referential expressions do in fact reliably cause a focussing behaviour that is correlated to the hypothetical comprehension processes.

The third main finding of this study was that there seem to be some pronouns that do not elicit higher fixation probabilities on the matching referent. We obtained this finding by comparing the permutations in pronoun order in the four different conditions. The four conditions differed with respect to the syntactic relation that the pronoun has to its antecedent (see Table 1 above). Comparing all four conditions, just one single time slot is shown to be significant: The probability for fixation on Referent 2 is significantly greater in Condition A than in Condition B in the time slot between 1500-2000 ms after the onset of PronounA<sup>4</sup> and PronounB respectively. According to our first claim that both full NPs and pronouns elicit higher fixation probabilities on the matching referent, it was expected that fixation probability in Condition A would become greater for Referent 2 than in Condition B, since PronounA refers to Referent 2 whereas PronounB refers to Referent 1 instead. So, the greater fixation probability in Condition A is not surprising. However, since PronounB refers to Referent 1 we would expect an increase in fixations on Referent 1 in Condition B compared to Condition A. In our data, however, we find no indication for any such increase. On the contrary, after PronounB the fixation curve for Referent 1 keeps on decreasing.

What could be the reason for this strange asymmetry in referential strength between the pronouns in conditions A and B? There are several possible answers to this question. The first one takes the attentional focus of the participants as a possible explanation for the missing fixation increase in Condition B. As already mentioned, the differences between pronouns start after the first pronoun, which is the same in all four conditions. This first pronoun refers to Referent 1 in all of the conditions. This means that that Referent 1 is in the centre of attention when the variation in the different conditions start. Since the subsequent PronounA refers to Referent 2, listeners have to change their centre of attention to Referent 2 when the pronoun occurs. In Condition B however, the focus of attention stays with Referent 1 since PronounB just re-refers to Referent 1. In this condition no new information is added and the attention stays the same over a longer time period. The fact that Condition B does not require a shift in attention to another referent can account for the missing fixation increase.

Another possible explanation for the asymmetry between the pronouns in conditions A and B comes from linguistic theory. It could be that the differences in fixation probability are caused by a difference in the referential properties of the pronouns. Bosch (1983) proposed that pronouns can be divided in two main groups: The regular anaphoric pronouns, which function referentially, and "syntactic" pronouns that do not function referentially, but are c-commanded and syntactically bound by their antecedents. The difference as far as German personal and possessive pronouns are concerned is not a difference in form: The same forms occur in either use. Whereas the anaphoric pronouns occur referentially and the relation to their antecedents is mediated literally by co-reference, the c-commanded pronouns just link up to their antecedents by syntactic agreement. Their relation to their antecedents is free of reference and purely syntactic in nature. According to this theory, the use of the pronoun in Condition A would be anaphoric; PronounA. is in a new sentence and thus cannot be c-

---

<sup>4</sup> Indices behind the pronouns refer to the conditions in which the pronouns occur, e.g. PronounA= The pronoun in Condition A

commanded by its antecedent. The pronoun in Condition B, however, is c-commanded by its antecedent, the subject of the sentence. PronounB would therefore be interpreted syntactically rather than referentially. If this interpretation of our results is correct, it would mean that the distinction between referential and bound pronouns is not just a theoretical one but a difference that is implemented in human language understanding. However, on the basis of the current data we are not able to decide which of the two alternatives is the correct account for the differences in fixation probability between conditions A and B. Further investigation will be required<sup>5</sup>. The comparison between conditions C and D did not yield any significant results. This may be mainly due to a mistake in the construction of the stimuli. As can be seen in Table 1, we tried to per permute the two pronouns in conditions C and D. However the temporal distance between these two pronouns was so small that a comparison was just not possible with an approach based on saccadic movements<sup>6</sup>.

## **Conclusions**

What conclusions can we draw from the results obtained in this study, and how can we relate them to the existing body of knowledge about the resolution of referential expressions?

We were able to confirm that both full NPs and pronouns elicit eye movement but that the in-peak fixation probability is generally smaller for pronouns than for nouns. As far as we know, the literature does not contain an explicit comparison between the fixation probabilities of full noun phrases and pronouns. While Cooper (1975) includes pronouns in the same class as full NPs in his experiment, the experiments of Runner (2003) and Arnold (2000) focus on the fixation probabilities for pronouns and do not comment on the fixation probabilities for the full NPs in their stimuli. Concerning the fixation probabilities for full NPs that we obtained in our experiment, we were also able to detect strong anticipation effects caused either by the word preceding the actual NP (such as an interaction verb indicating that another human referent would follow) or caused by the unfolding NP itself. These results confirm the findings of Hartmann (2004) and Dahan, Swingley, Tanenhaus, & Magnuson (2001), suggesting that anticipation plays an important role in understanding and that fixation probabilities are influenced already by anticipation effects as the stimulus expression unfolds. Concerning the temporal resolution of the comprehension process for different referential expressions, we were able to contribute to a long ongoing discussion in which two main conflicting hypotheses have been around since the middle of the 1980s. As far as we know, eye tracking had not until now been used to investigate this issue. We hope to have shown that eye tracking methods provide a highly suitable online measure for the online investigation of the corresponding comprehension processes. As our results show no difference in the comprehension of nouns and pronouns in unambiguous sentences, they can be interpreted as supporting the hypothesis of Tyler & Marslen-Wilson (1982), who claimed that pronouns are interpreted immediately referentially in the same way as proper names or full NPs. Tyler & Marslen-Wilson's view seems to be much more coherent with our findings than the hypothesis of Gernsbacher (1989), who claimed that pronouns are resolved in two stages, first a search for antecedents and then, via the antecedent, linking up to the referent.

---

<sup>5</sup> A follow-up experiment, conducted by Sarah Jessen at our Institute in July/August 2006 seems to support the interpretation that postulates different referential properties for regular anaphoric pronouns and for bound pronouns.

<sup>6</sup> While the two pronouns in Condition C refer first to Referent 2 and then to Referent 1, the arrangement of the pronouns was exactly the opposite in condition D. However, in both conditions the two pronouns fell into the same time slot of 500 ms, which made it impossible to analyse any differences caused by the arrangement of the pronouns.

According to this theory, the resolution of pronouns has to take longer time than the resolution of full NPs since it requires the additional process of identification of an antecedent. Finally, we found a first indication that the distinction between referential and non-referential pronouns as proposed by Bosch (1983) might be not just theoretical but may have empirical consequences in language comprehension. However, as mentioned, the current data are not conclusive in this respect and certainly here more work is required.

## References

- Arnold, J., Eisenband, J., Brown-Schmidt, & S., Trueswell, J (2000): The Rapid Use of Gender Information: Evidence of the Time Course of Pronoun Resolution from Eye Tracking. *Cognition* 76, B13-B26 (2000)
- Bosch, P. (1983): Agreement and Anaphora - A Study of the Roles of Pronouns in Discourse and Syntax. Academic Press. London & New York. 1983
- Cooper, R. (1974): The Control of Eye Fixation by the Meaning of Spoken Language. *Cognitive Psychology* 6, 84-107
- Dahan, D., Swingle, D., Tanenhaus, M.K, & Magnuson, J.S, (2001): Time Course of Frequency Effects in Spoken Word Recognition: Evidence from Eye Tracking. *Cognitive Psychology* 34, 317-367
- Garrod, S., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the online resolution of sentences in a discourse. *Journal of Memory and Language* 33. 39-68.
- Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, 32, 99-156.
- Gernsbacher, M. A. & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory & Language*, 27, 699-717.
- Hartmann, N. (2005): Processing Grammatical Gender in German. Bachelor Thesis, unpubl. Univ. of Osnabrück, Cognitive Science, [http://www.cogsci.uni-osnabrueck.de/~CL/download/Hartmann\\_GramGender.pdf](http://www.cogsci.uni-osnabrueck.de/~CL/download/Hartmann_GramGender.pdf).
- MacDonald, M. C. & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory & Language*, 29, 469-492.
- Marslen-Wilson, W. & Tyler, L.K. (1987): Against Modularity. In: Garfield, J.L. (ed): *Modularity in Knowledge Representation and Natural Language Understanding*. MIT Press. 1987. pp. 37-62
- Nicol, J. & Swinney, D. (2002): The psycholinguistics of anaphora. In A. Barsz (ed.) *Anaphora*. Oxford. Blackwell. pp. 72-104
- Rayner, K., & Clifton, C., Jr. (2002): Language processing. In D. Medin (Volume Editor) *Stevens Handbook of Experimental Psychology, Third Edition: Volume 2, Memory and Cognitive Processes* (pp 261-316). New York: John Wiley and Sons,
- Runner, J., Sussman R. S. & Tanenhaus, M. K.(2003): Assignment of Reference to Reflexives and Pronouns in Picture Noun Phrases: Evidence from Eye Movements. *Cognition* 81.1, B1-B13 (2003).
- 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.

Tanenhaus, M. K., Magnusson, J. S., Dahan, D., & Chambers, C., (2000): Eye Movements and Lexical Access in Spoken Language Comprehension: Evaluating a Linking Hypotheses between Fixations and Linguistic Processing. *Journal of Psycholinguistic Research*, Vol.29 No.6 (2000)

Tyler, L.K. & Marslen-Wilson, W. (1982): Processing utterances in discourse context: Online resolution of anaphors. *Journal of Semantics* 1: 297-314.