

# **A linking hypothesis**

## **connecting eye movements and linguistic processing**

Juliane Steinberg,  
*Lund University Cognitive Science*  
Supervisor: Kenneth Holmqvist

---

### **Abstract**

Eye-tracking methodologies have been used extensively in spoken language research. However, no qualitative account of the relationship between language processing and gaze behavior has been given so far. Such an account is derived from different theories, essentially stating that the observed correspondence between visual and linguistic processing is based on activation spreading between the different sorts of representations included in a concept. An experiment showing that gaze also follows understanding when gaze reactions are not task-relevant supports this linking hypothesis. As a result of the specific gaze behavior observed in the experiment, two parameters are proposed that might, in unison with the linking hypothesis, influence gaze reactions. The functioning of these factors has yet to be fully explored in further research.

---

### **1. Introduction**

Eye-tracking has been used in many studies on language comprehension as well as language production. Researchers value this method because it is very sensitive to language processing without interfering with it, which allows non-disruptive observations in natural settings. Studies usually involve subjects facing one or more pictures and having their eye movements recorded while they either say something themselves or listen to utterances. Whenever a gaze to a certain object is temporally correlated to that object being mentioned in the speech stream conclusions about the understanding or planning of that expression are drawn.

The problem that the present paper takes as its starting point is that most of the studies done in this field so far take the observed connection between eye movement behavior and language processing as a given, using it for their purposes but not asking what it might actually look like.

The eye-mind assumption by Just and Carpenter (Just & Carpenter, 1980) posits that the eye and the mind are linked in a way that the eye fixates an object as long as it is processing it. This assumption is constrained by the fact that it is formulated explicitly as part of a theory on reading, an activity that necessarily requires processing of the visual display. Just and Carpenter state that their eye-mind assumption is also supported by observations made in spatial problem-solving tasks, but it still remains constrained to tasks where an inspection of the visual environment is necessary.

Those eye-tracking studies that are not at all explicit on what the connection between the observed eye movements and the language processing inferred from

them looks like (Arnold, Eisenband, Brown-Schmidt & Trueswell, 2000; Barr & Keysar, 2002; Griffin, 2001; Sedivy, Tanenhaus, Chambers, & Carlson, 1999) build on an eye-mind assumption like Just and Carpenter's. Additionally, the large body of evidence showing a clear linkage between subjects' gaze behavior and spoken language stimuli or subjects' own utterances clearly argues for the usage of eye-tracking paradigms in language research. However, the very fact that a number of researchers feel the need to conclude that eye-tracking is a sensible instrument in language research from their and others' results (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Tanenhaus & Chambers, 2001; Griffin & Bock, 2001) already shows that there is still a need to defend the use of eye-tracking in language research. The research paradigm is thus motivated by the fact that it produces sensible results, but a more or less explicit cognitive theory on why an object is fixated while it is being processed by the brain is still missing<sup>1</sup>.

The present paper presents results that might challenge the eye-mind assumptions held so far, and tries to outline how further research could help understanding the connection between eye and language better.

## 2. Theoretical Background

### 2.1. Eye-tracking studies on spoken language processing

Eye-movement behavior and spoken language processing have been connected by several empirical studies. The two main approaches in this area are provided by language comprehension and language production studies – two different research fields unveiling different phenomena that might still build on the same cognitive processes.

In the following three basic behavioral patterns found in these studies will be described shortly.

The first kind of behavior occurs when individuals hear spoken utterances containing expressions that refer to their visual environments. Most often this means that subjects are instructed to 'pick up' or 'move' objects (Allopenna et al., 1998; Barr & Keysar, 2002; Dahan et al., 2002). Beginning from about 200 ms after the onset of a word that refers to an object in their visual environment, individuals shift their gaze to this object. Under the most favourable conditions reactions occur rapidly and with a very high reliability, up to around 90% (this is only the basic effect that has often been varied to examine lots of different influencing factors: Allopenna et al., 1998; Arnold et al., 2000; Barr & Keysar, 2002; Dahan et al., 2001; Sedivy, 1999; Spivey, 2002; Tanenhaus, 2000). In these cases eye movements and thus visual processing are obviously steered by linguistic processing.

In a speech production context eye behavior is not as predictable as in the speech comprehension contexts examined in the abovementioned studies. Meyer and Dobel (2003) review a number of studies on the issue and describe the following tendencies in behavior: When subjects produce utterances about the visual scenery surrounding them, two general patterns in eye movements can be observed. The first one is an inspection pattern where relevant information is

---

<sup>1</sup> A very recent exception to this pattern is Griffin's discussion on the link between eye movements and language production (Griffin, 2004).

retrieved from the scenery and the utterance's structure is planned accordingly. After this phase a second pattern can be observed during which the retrieval of linguistic representations for the mentioned objects can be observed through eye movement patterns. Subjects usually fixate objects immediately before naming them<sup>2</sup>. A number of experiments has shown that during these fixations a number of linguistic representations are retrieved, so far as to the phonological representation of the object's name.

The steering of the eyes to the objects that are just about to be named, having already inspected the whole scene, is again a clear influence of linguistic on visual processing, because eye movements are determined by the structure of the sentence being uttered. On the other hand, why would subjects fixate on an object while retrieving its name (having already looked at the object earlier) if there were absolutely no effects in the opposite direction as well? As also the construction of utterance structure during the inspection phase shows, there must be an influence of visual processing on linguistic processing as well.

The best proof for this is a study conducted by Zelinsky and Murphy (2000), which provides evidence for unconscious linguistic encoding triggered by visual stimuli. When told to inspect and memorize objects presented to them, subjects fixated the objects for varying periods of time, depending on how long the names of the fixated objects were. This study shows that visual processing of objects can cause linguistic processing of these objects, even when retrieving the objects' names is not required by the task<sup>3</sup>. In that way this study constitutes a counterpart to the present paper which intends to show that visual processing can also be steered by linguistic processing, even when the visual processing is not task-relevant.

This short summary has shown some instances of interaction between visual and linguistic processing that suggest that there actually is a mutual functional relationship between speech and vision that reaches beyond "what we look at is what we think about".

## 2.2. *An Aspect of Relevance Theory*

The relevance theory by Sperber and Wilson (1995) is a theory of communication which draws its explanatory power from describing the general cognitive processes that are involved in producing and especially understanding utterances. There is one aspect of this theoretical framework that is of particular significance to the present paper and will be shortly sketched here.

This core aspect is the notion of relevance. It characterizes an economic processing behavior that determines all aspects of communication. What is more, it can also be applied to cognitive processing in general and perceptual mechanisms and perceptual salience in particular.

Notions such as the relevance of phenomena are of major importance for the issues discussed here. A phenomenon (which is a quite general notion that includes perceived stimuli) can be more or less relevant for an individual in a certain context: The more the processing of the phenomenon changes the

---

<sup>2</sup> This does not happen if the object's name already is highly available – for example because the object has been named shortly before or its name is very easy to retrieve.

<sup>3</sup> Although one might argue that names are retrieved to ease the organization of the memorized stimuli.

knowledge of the individual and the less processing effort this requires, the more relevant the phenomenon is.

For example, when perceptual information is being processed, relevance is the guiding principle: perceptual input is processed in such a way as to yield the most interesting information at the lowest possible effort – the most relevant phenomena are preferred. Sperber and Wilson (1995) even state that humans in general automatically turn their attention to what seems most relevant to them.

This thought is important for the model that will be developed in the next section.

### 2.3. Theories on concepts

An important piece of the puzzle needed to establish the relationship between language processing and eye movements is a theory of concepts.

Many authors use the notion of meaning synonymously with conceptualisation (Langacker, 1991). Conceptualisation is a much wider notion than concept: according to Langacker it includes "novel conceptions as well as fixed concepts; sensory, kinesthaetic, and emotive experience; recognition of the immediate context (social, physical, and linguistic), and so on" (Langacker, 1991, p. 2).

This paper focuses on the level of word processing, since words can be assumed to be the smallest linguistic entities that can trigger eye movements to objects in the visual environment<sup>4</sup> - sentence understanding should be examined only once the mechanisms on a word level are understood. Therefore it should be acceptable to focus on one type of conceptualisations only, namely the fixed concepts that already exist in an individual's mind. After all, experiments exploiting the eye-tracking paradigm need to use words which can have concrete, visualizable counterparts in the real world – eye-tracking research is far from exploring all the other aspects of conceptualisation Langacker names.

These considerations motivate the reduction of meaning to word meaning and concepts in this paper. It will be assumed that the activation of concepts constitutes the primary link between the visual system and linguistic processing.

Zelinsky and Murphy (2000) posit, on the basis of their experiments, that visual and linguistic processing systems synchronize their use of working memory – visual sketchpad and articulatory loop work closely time-locked to each other. With this thesis Zelinsky and Murphy (2000) go deeper into the issue than the authors of the abovementioned eye-tracking studies, but they still do not explain why the two systems work together.

If the activation of concepts is included in this model, the missing link might be found – if both linguistic and visual information together can be located in concepts.

There are in fact theories that afford this. Langacker (1991) assumes concept(ualisation)s to be characterized by different dimensions of information like shape, size, material, use and other associated information. Though he does

---

<sup>3</sup> Studies have been conducted on how spoken stimuli influence gaze on a phonological level (Allopenna et al., 1998). These effects are, however, also based on the subjects' anticipating complete words that refer to the fixated objects.

not state this explicitly, the spoken and written words that belong to a concept (if there exists a word describing the concept) presumably also constitute such dimensions.

Murphy (2003) presents a conceptual view on word meaning in which word meaning is 'represented psychologically by mapping words onto conceptual structures' (Murphy, 2003, p. 388)<sup>5</sup>. What is important for the present purpose is Murphy's review of several studies that showed typicality and basic level effects, both well-documented properties of concepts, in various different experimental settings. The human conceptual system showed the same specific behavior no matter through which kind of stimuli it was accessed: dot patterns, alphanumeric strings, patches of colour, geometric shapes, or linguistic stimuli. This is clear evidence showing that visual as well as linguistic information is included in concepts and furthermore that the same concepts can be accessed by accessing those or other different types of information.

#### *2.4. The coordination of eye movements and linguistic processing – the linking hypothesis*

So far, each of the theoretical and empirical pieces of the puzzle in the preceding sections supports a bit of the rough hypothesis on the coordination of eye movements and linguistic processing that will be described in the following. This linking hypothesis might seem neither very daring nor very sophisticated, but actually there does not yet exist any theory explicating the illustrated processes; and the proposed explanation might be taken as a starting point to direct further empirical studies, which could in turn lay a foundation for a better theory on the issue<sup>6</sup>.

The model proposed here is as follows (see also fig. 1): In the human mind there is a tight coupling between the phonological, graphematic, visual and many other types of information belonging to a single concept. As soon as a concept is activated, by any of its components being perceived or activated through other cognitive processes, activation spreads within the concept, so that the other components of the concept are also activated<sup>7</sup>. Completing Zelinsky & Murphy's (2000) approach: When for example a spoken word is heard it is first processed in working memory's articulatory loop. From there, the concept containing the phonological form of the word is activated and through spreading activation the corresponding visual representation as well. Through this activation the visual representation is transferred to the visual sketchpad. If a similar representation is already present in the visual sketchpad (because the individual has just inspected his visual environment), an eye movement towards the corresponding object is triggered (see fig. 1).

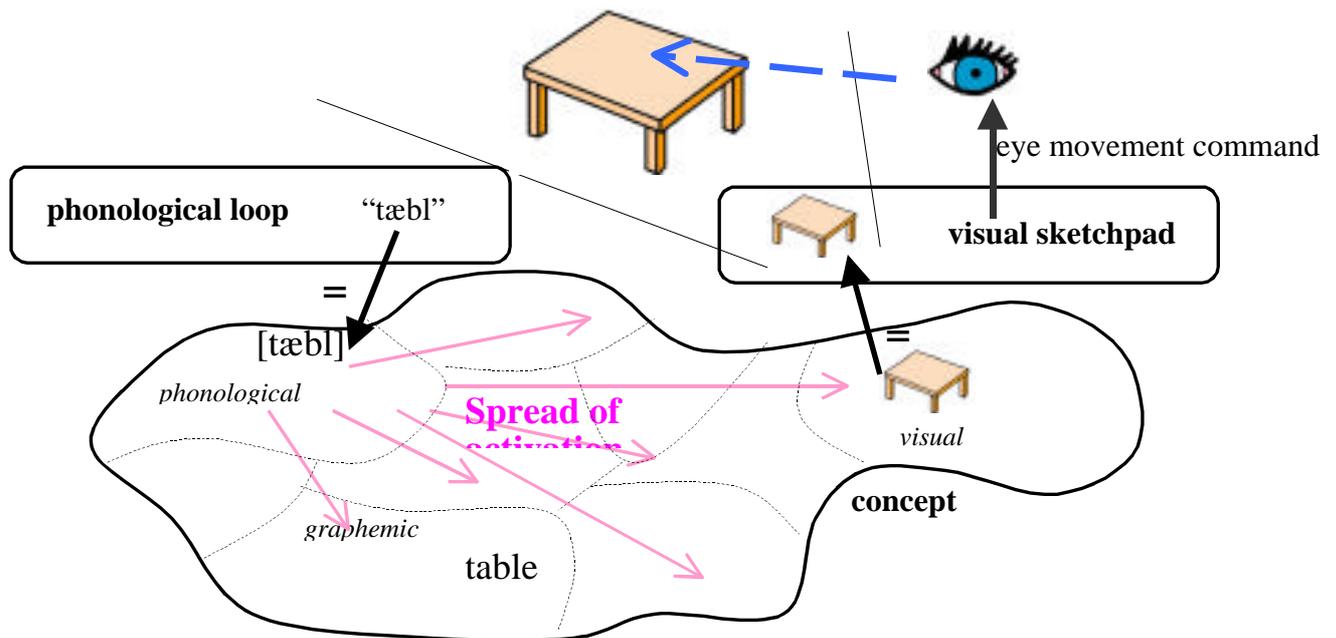
---

<sup>5</sup> It is not clear if this includes separating the word as a phonological or graphemic representation from the concept itself, or if "word" denotes an abstract entity in linguistics.

<sup>6</sup> Actually the proposed hypothesis shares some similarities with Just and Carpenter's theory on reading (Just and Carpenter, 1980): The spreading activation between associated concepts and their transfer to working memory as basic mechanisms in reading are also of crucial importance here.

<sup>7</sup> This could presumably be supported by neuropsychological theories on the spread of activation (causing such effects as priming and associations, for example). Conceptual information being closely linked in the brain simply means that activation spreads easily between it.

This is the point where Sperber and Wilson's (1995) notion of the relevance of phenomena comes in. The activation of a visual representation that shares many similarities with an object already present in the visual sketchpad makes that object a relevant phenomenon. This has two reasons: On the one hand there is a high probability that processing a visual stimulus that is an instance of a concept already activated by other cognitive processes will yield interesting results (a high effect can be expected). On the other hand, the object will not require much processing effort, since its concept is already active (low effort is needed). So the tendency to turn attention to relevant stimuli will lead to a gaze shift to the object corresponding to the activated concept.



**Figure 1:** An illustration of the linking hypothesis between eye movements and linguistic processing proposed in this paper.

At the core of this linking hypothesis lies the assumption that concepts, as the psychological counterparts to word meaning, can be accessed through each of their components, causing activation to spread to other components. This principle links the processing of visual information to the processing of linguistic information. It does not make strong assumptions on concepts' structure and thus should be compatible to quite a number of different theories on conceptualisation. It is robust to effects of polysemy or word onset similarity etc. because activation can of course spread to all kinds of related information and does not stop at the 'boundaries'<sup>8</sup> of one single concept.

To lay a basis for a possible confirmation of the linking hypothesis this experiment will try to show that Just and Carpenter's eye-mind assumption can be applied more generally than it was meant to. If it can be shown that gaze follows language understanding even when the task does not require this, an underlying connection must be assumed that at least partly builds on an automatic link between visual and phonological representations as proposed in the linking hypothesis above.

<sup>8</sup> Which do not exist anyways, as for example typicality effects show (Murphy, 2002).

The experiment conducted for this paper included subjects being faced with two pictures of simple everyday objects while listening to instructions to write a certain word on one of them. The word they were to write always corresponded to one of the two pictures. Thus a situation was created in which subjects heard a word and were only required to retrieve its graphemic representation, but not to look at a picture corresponding to it. However, a gaze reaction to the picture of the word that was to be written was also possible and expected. This possible reaction could be compared to the 'control condition' of the second noun occurring in each instruction, denoting the picture subjects were to write on. Here eye movements to the picture were predicted by previous studies, because looking at the picture was task relevant.

With this rough description of the experiment in mind, the concrete hypothesis to be tested can be formulated, derived from the above considerations.

### 2.5. Hypothesis

Upon hearing the word they are supposed to write, subjects will react with an eye movement to the picture corresponding to that word. This should be visible in significantly more gaze samples falling on this than on the other picture, in the time interval between the onset of the word subjects are to write and the onset of the word denoting the picture to write on.

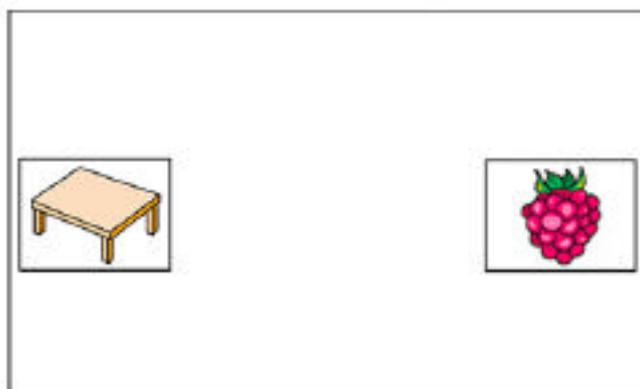
## 3. Method

### 3.1. Subjects

Sixteen students of Lund University took part in the study in exchange for a chocolate bar. All subjects were native Swedish speakers and all except for one had normal or corrected-to-normal vision. One subject was cross-eyed and produced strongly deviating eye-tracking data, which is why these data were not included in the analysis.

### 3.2. Stimulus

The subjects were presented with the stimuli in the form of a computer program while sitting at a table in front of a 19" TFT monitor, with a keyboard and a mouse for them to use. Their task was to follow instructions telling them to write a word on one of two pictures they saw on the screen. This could be done by clicking with the mouse on the picture they intended to write on and then using the keyboard to type the word they were to write (for an example see fig. 2).



"Write **raspberry** on the **table**. /  
 Write **table** on the **raspberry**. /  
 Write **raspberry** on the **raspberry**. /  
 /  
 Write **table** on the **table**. "



**Figure 2:** An example of a visual stimulus subjects were presented with, together with all four possible variants of an auditory stimulus that could have occurred with these two pictures.

Visual stimuli consisted of two pictures displayed simultaneously. The pictures were chosen to represent basic and everyday objects that should be easy to name. Which pictures were presented together on one trial was determined by the names of the objects they showed: these names should start with different phonemes, so that inference effects (see Allopenna et al., 1998) could be avoided. Most of the pictures were obtained from clipart-galleries on the internet, some were drawn by hand and scanned in. All pictures were coloured and measured 7.5 x 6.5 cm on the screen. The distance of the pictures from each other was large enough to avoid that both pictures could be processed foveally at the same time.

Auditory stimuli always had the structure "Write ... on the ...." (in Swedish), for example "Write raspberry on the table". The first noun in these sentences will from now on be called the 'write word', the second one the 'where word'<sup>9</sup>. Both of the nouns, the write as well as the where word, always had a corresponding picture among the two pictures on the screen, so for the example "Write raspberry on the table." the pictures on the screen would be one of a raspberry and one of a table. The picture corresponding to the write word will from now on be called the write picture, the picture corresponding to the where word is the where picture. Write and where words could be different or the same on one trial, so that subjects could not build up expectations on what the where word, coming second in the sentence, would be. Both write and where pictures were positioned equally often on the left and right sides of the screen to avoid any predictability (see fig. 2 for possible trials).

Both words' onsets varied relative to the sound files' starts on the different trials and were measured by hand in the video editing program used to create the single sound files (Peak 7.0). Because it is difficult to determine the exact onset of a single word in a speech stream these measurements can be expected to have a mistake of around  $\pm 20$  ms. Onsets of the where words relative to the write words' onsets ranged from 567 ms to 1033 ms and were at 799 ms on average.

All auditory stimuli were recorded from a female native speaker, using a Sony DCR-TRV 900E mini dv - video camera. The video tape was read into a computer, converted and edited into single .wav-files with Peak 7.0.

The program that presented the stimuli was created with e-studio, a program from the e-prime suite. With this program it is possible to define every single state of the computer screen during an experiment, play sound files, define how subjects can interact with the stimulus program, and send signals to the eye-tracking equipment to mark the beginnings of critical phases for measurement. Also provided by e-prime was accurate time logging of every event during the experiment, a feature necessary in the analysis when stimulus onsets must be located in the eye-tracking data.

A problem caused by the American origin of e-prime was that some Swedish letters (å, ä and ö), once typed in with the keyboard, did not appear correctly on the screen. This problem could not be fixed, but it was considered not to be too

---

<sup>9</sup> Both words are sometimes called target words, the corresponding pictures target pictures.

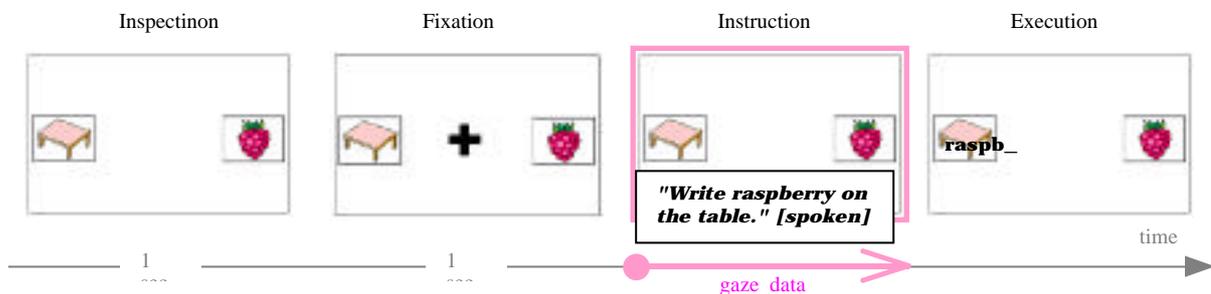
disruptive, because only the very beginning of each trial was of interest for the analysis – the point in time when subjects were actually writing did not occur until much later.

On the whole subjects were presented with 64 trials. No picture appeared more than once, and no filler trials were used. As already mentioned above, stimuli were balanced for position (left or right) of write as well as where pictures and for whether write and where words were the same ("Write raspberry on the raspberry") or different ("Write raspberry on the table").

The two conditions compared in this experiment occurred after one another on each trial: the time following the onset of the write word was the experimental condition, as opposed to the time following the where word which constituted the control condition.

Trials were run in random order for each subject: This was taken care of by e-prime after parameters had been set accordingly during the creation of the stimulus program.

A whole trial (see fig. 3) consisted of a one-second inspection phase during which only the two pictures were on the screen, so that subjects would have time to understand what the pictures showed. After this second, a fixation cross appeared in the middle of the screen for one second, while the two pictures remained unchanged. The moment the fixation cross disappeared the audio file with that trial's instruction (auditive stimulus) was played, the pictures still remaining unchanged. Then subjects were to click on one picture (depicted by the where word) and write the write word on it.



**Figure 3:** An example trial: one second inspection time is followed by one second where subjects were to fixate a cross in the middle of the screen; the cross disappeared together with the onset of the spoken instruction and then subjects wrote a word on one of the two pictures (execution of the task). The Inspection screen is the one during which eye gaze data was collected.

### 3.2. Apparatus

A head-mounted eye-tracking system (SensoMotoric Instruments) and Polhelmus head-tracker were used to track subjects' eye movements. Both systems were fastened on a bicycle helmet. Subjects' right eyes were filmed and pupil positions were determined at a rate of 50 Hz by an SMI-program measuring the corneal reflex.

### 3.3. Procedure

Upon their arrival subjects were welcomed and then made familiar with the head-mounted eye-tracking equipment. When they had been seated with the eye-

tracker on their heads they were told that their pupils would be filmed during the experiment and that a calibration procedure had to be run in preparation for that. They would be allowed to look wherever they wanted to during the whole experiment except for the calibration phase, during which they were asked to move as little as possible.

The calibration for eye- as well as head-tracker was done and after that subjects were told that they would have to follow the instructions given to them by a computer program started by the experimenter. The program would begin by explaining what the task was, followed by some practice and then subjects would have the chance to ask questions about the task if they had any.

Also subjects were informed about the problem with the display of å, ä and ö and were asked to ignore it as well as they could – they were told that the program would know which keys they had hit, no matter which characters would appear on the screen in their place.

The experimenter then started the stimulus program. After an introduction sequence demonstrating what subjects were to do, reminding them to look at the fixation cross as long as it was on the screen and telling them to try to solve the task as quickly as possible<sup>10</sup>, five practice trials were run, the program paused and the subjects were asked if they had any questions about what they were expected to do. Having cleared up any doubts, the subjects were told to start the real experiment by pressing a key on the keyboard.

The 64 trials were run, thereafter subjects were freed from the eye-tracker, got their chocolate bar and were seen off.

## 4. Results

### 4.1. Data Coding

The raw data from the experiment were obtained in the form of text files with coordinates of gaze positions on the computer screen at every 20 milliseconds from the beginning of the experiment. All analyses were done using Microsoft Excel.

Using the absolute gaze coordinates and the positions of the two pictures on the screen it was determined for every instance of gaze data which picture gaze had fallen on, if any. These data were calculated relative to the onset of the spoken word that was assumed to steer eye movements at that moment in time: from the moment the write word had started, gaze hits on the corresponding picture were counted as target hits, from the onset of the where word the other picture could become the target, according to the structure of that trial (see fig. 2).

### 4.2. Reactions to write word vs. where word

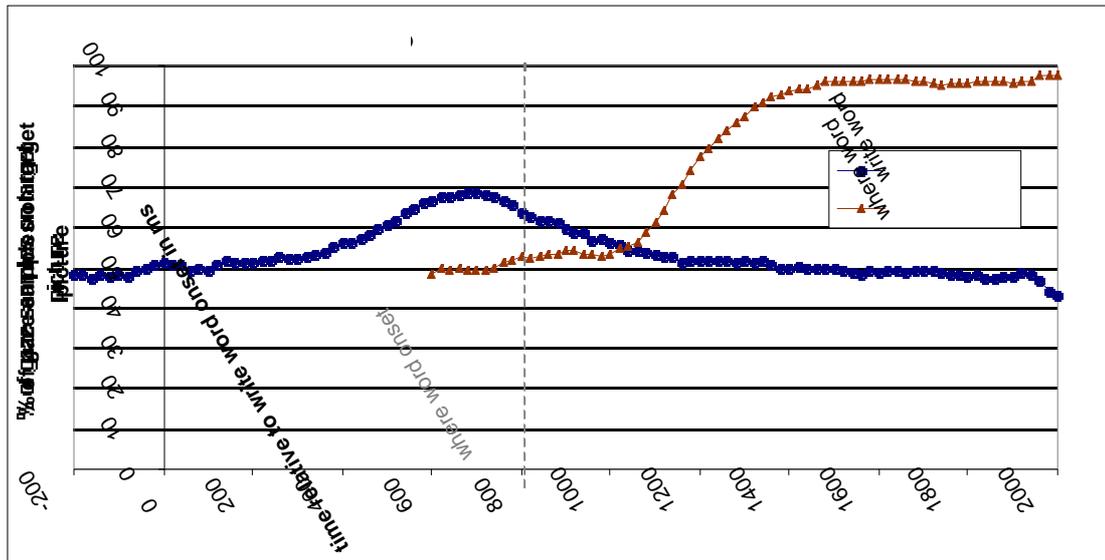
Averaging gaze measures over all trials of all subjects and calculating the percentages of the gaze samples hitting the target pictures produced the results in fig. 4.

Three aspects of this figure need some further explanation:

---

<sup>10</sup> This instruction was given to keep subjects from hypothesizing about the real intention of the experiment by making them believe their reaction speed or correctness mattered.

Firstly, the graphs of the write and where word conditions appear after each other in the figure, to illustrate the order in which write and where word appear on each trial. The onset of the where word at 800 ms in this figure is an average value over all trials relative to the write word onset. It is important to note that this averaging affects only the position of the where word curve in the figure – the curve itself results from calculations based on the exact onsets of the where words on each trial.



**Figure 4:** Average percentages of gaze samples that fall on the pictures corresponding to the write and where words. The write word starts at 0 ms, the where word on average (see explanation below) 800 ms after that. Percentages are calculated as those gazes hitting the target picture out of all gazes hitting one of the two pictures, so that a gaze percentage of 50% on the target picture means that both pictures are looked at equally much.

The second issue about figure 4 that needs to be explained is the decline of the write word curve from about 700 ms. This decline is caused by the structure of the experiment: on half of the trials write and where pictures are the same ("Write table on the table"), on the other half write and where pictures differ ("Write raspberry on the table."). This means that the target picture of the write word changes on half of the trials with the onset of the where word. In those cases gazes hitting the target picture are from then on counted as non-target gazes and gazes hitting the non-target picture become target hits. That way the reaction data is reversed on half of the trials, becoming the (almost) exact opposite to the reaction data from the other half of the trials<sup>11</sup>. So for every target gaze in one half of the data there will be a non-target hit in the other half, which makes the write word curve decline to 50% of the gaze samples hitting the target picture. Had the where word set on at the same time on every trial this decline would have been abrupt; but as where word onsets vary the write word curve descends smoothly.

In other words, the analysis of the gaze data as a reaction to the write word is rendered meaningless with the onset of the where word, because from then on a

<sup>11</sup> Assuming that reactions to the write word do not differ depending on whether or not the where word is the same as the write word. This assumption is justified, since subjects do not know the write word beforehand because of the randomised trial order. See also section 4.5. on this question.

new instruction steers just the critical behaviour that constituted the reaction to the write word before. This shows in the write word curve declining to a level of 50%, meaning random viewing behaviour.

The third point about fig. 4 that needs to be clarified is the fact that the percentages of gazes on write and where pictures sum up to more than 100%. This has to be attributed to the target picture staying the same for write and where words on half of the trials, so that in those cases hits on the target picture are counted as gazes on the target pictures on both conditions and are thus counted twice in the sum of the write and where word curves.

One can clearly see in figure 4 that subjects react to the write and where words by shifting gaze to the respective target pictures. The decline of the write word curve after approximately 700ms can be attributed to the onset of the where word when the target picture changes on half of the trials. Otherwise the curves for both conditions look quite similar, only the write word curve does not reach as high a peak as the where word curve does.

#### 4.3. Reliability of reactions

A one-sided t-test for paired samples was performed on the absolute numbers of gaze samples hitting the target pictures and non-target pictures in the time interval from the write word onset at 0 ms to the average where word onset at 800 ms. It showed that there were significantly more gaze samples falling on the target picture than to the non-target picture following the write word. The same analysis performed on the 800 seconds of data following the onset of the where word also yielded significant results, as can be seen in table 1. There is, however, a significant difference between the percentages of gaze samples that land on the target pictures following the onset of the write word and the onset of the where word. The results of these tests are to be found in table 2.

	average number of gazes on target picture	average number of gazes on non-target picture	significance of the target advantage
write word	328	234	$p = 7.88 \cdot 10^{-9}$
where word	548	168	$p = 2.33 \cdot 10^{-10}$

**Table 1:** The results of a comparison of the absolute numbers of gaze samples on write and where word trials for the interval from 0 to 800 ms relative to the onset of the respective target words. On both write and where word trials subjects look at the target pictures significantly more often than at the non-target pictures.

average percentage of gaze on write picture	average percentage of gaze on where picture	significance of the difference between reactions
57.9	74.6	$p=9.3 \cdot 10^{-12}$

**Table 2:** The results of a comparison between the target advantages as a reaction to write vs. where words. On the where word conditions target advantages are significantly stronger than on the write word condition.

So subjects reacted to the write and where words by almost immediately shifting their gaze to the corresponding picture, but they did so more reliably in response to the where word than to the write word.

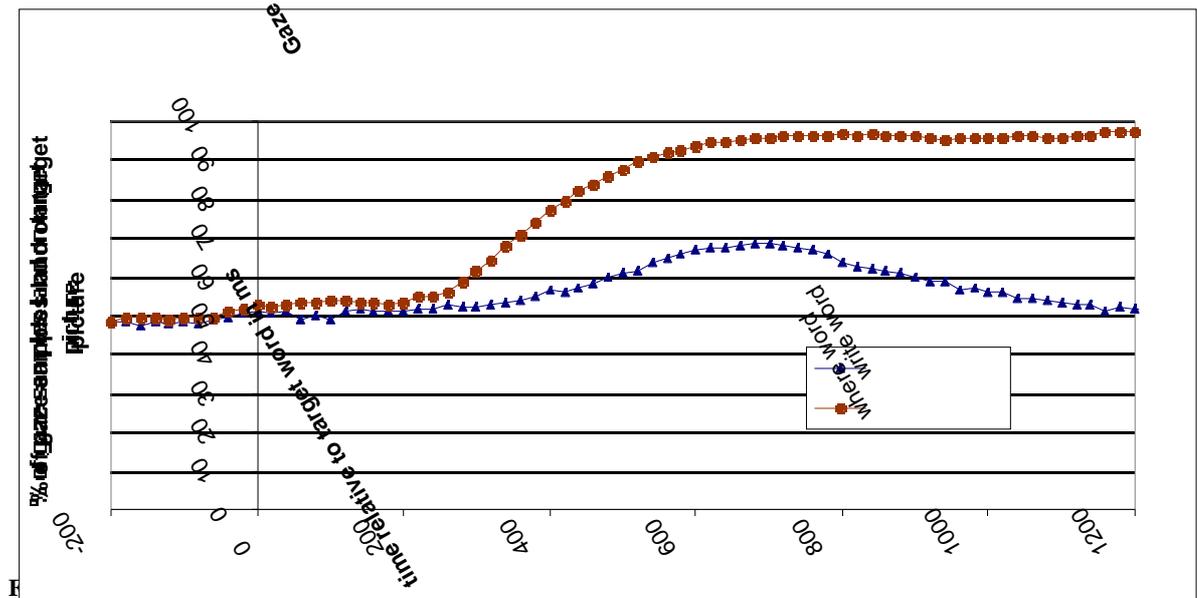
The curve of reactions to the where word reaches its peak at about 800 ms after the where word onset and then remains at that level of target advantage. In contrast to this the write word-gaze curve drops after having reached its peak at 700 ms. The facts that the onset of the where word is responsible for bringing the write word curve back down to a level of 50% and that the earliest where word

onset already occurs at 567 ms impose a severe methodological problem: It is not possible to say if the percentages of gaze samples hitting the write picture had gone up to the same level as those hitting the where picture or not. From the point where the slope of the write word curve begins to decrease (at about 570 ms) influence of the where word has already begun and it is not possible to separate this influence from the 'real' effect hidden in the data by mathematical calculations.

4. Speed of reactions

To give the reader a first impression of the speed with which eye movements towards write and where words take place, the curves from fig. 4 were matched in fig. 5.

A mathematical analysis used by many authors in related studies to compare reaction times is the analysis of 100 ms- or 200 ms-intervals of data, checking from which interval on target advantages get significant (Allopenna et al., 1998; Arnold et al., 2000).



does not represent the order of events on a trial any more, but takes both target words' onsets as an orientation point and shows how gaze reactions develop relative to this.

Data from the write word and where word condition were partitioned into small time intervals and these were examined more closely. 100ms-samples of the absolute gaze data were tested on whether or not there were significantly more gaze samples hitting the respective target pictures than the non-target pictures. The samples were taken from the first 500 ms after the onsets of the respective target words. The results of this analysis can be seen in table 3.

**Write word**

analysed interval	average number of gazes on target picture	average number of gazes on non-target picture	significance of the target advantage
1 <sup>st</sup> 100 ms	235	232	p= 0.48
2 <sup>nd</sup> 100 ms	274	266	p= $4.03 \cdot 10^{-4}$
3 <sup>rd</sup> 100 ms	290	269	p= $1.45 \cdot 10^{-4}$
4 <sup>th</sup> 100 ms	301	263	p= $1.82 \cdot 10^{-4}$
5 <sup>th</sup> 100 ms	323	239	p= $1.87 \cdot 10^{-4}$

**Where word**

analysed interval	average number of gazes on target picture	average number of gazes on non-target picture	significance of the target advantage
1 <sup>st</sup> 100 ms	326	290	p= $2.82 \cdot 10^{-5}$
2 <sup>nd</sup> 100 ms	336	292	p= $4.81 \cdot 10^{-4}$
3 <sup>rd</sup> 100 ms	354	284	p= $2.76 \cdot 10^{-3}$
4 <sup>th</sup> 100 ms	455	217	p= $2.46 \cdot 10^{-4}$
5 <sup>th</sup> 100 ms	611	137	p= $3.44 \cdot 10^{-11}$

**Table 3:** Results of one-sided paired t-tests on 100 ms intervals of the write word resp. where word data. Significant target advantages can be seen from 100 ms after write word onset and from immediately after the where word onset.

This means that subjects' gaze moved to the picture corresponding to the where word immediately after the where word's onset, while reactions to the write word were significant from 100 ms after the write word's onset.

These results obviously do not have a great value besides from the fact that they unveil significant target advantages on both conditions that will be discussed in the next section. They do not provide insight into whether or not gaze reactions to the write word occur equally quickly as the reactions to the where word.

The problem that foils reaction speed analysis is that it is unclear if the write word curve could have reached a higher peak had the where word not occurred. Knowledge of both curves' peaks is necessary to compare how quickly that peak is reached and thus to make a conclusion on how fast gaze reactions occur. It is possible that the write word curve could have risen further after 700 ms (which is where it peaks in the present data) and reached about the same level as the where word curve, if the where word had not occurred and steered gazes to the where picture. Had the write word curve risen further to about the where word curve's peak with a slope similar to the one between 200 and 600 ms that reaction would have to be called much slower than the one to the where word. Had the write word curve not risen above the level reached at 700 ms it would have reached its peak equally fast as the where word curve which would make reactions equally quick but less reliable on the write word condition.

The question on reaction speeds is again a question that cannot be solved by purely analysing the data.

#### 4.5. Early target advantage

The very early 'reactions' in the form of gaze landing on the target pictures immediately after the target words' onsets (see table 3) need to be examined more closely, because they deviate from the expected gaze behaviour: Since even the

programming of a saccade takes at least 150 ms (Allopenna et al., 1998; Barr & Keysar, 2002; Dahan et al., 2001), the earliest gaze reactions to a word cannot be expected before these 150 ms after word onset, and indeed other experiments only observed preference for the target from about 200 ms after word onset (Allopenna, 1998; Arnold, 2000).

The significant target picture advantage during the first 100 ms and maybe even the first 200 ms cannot be attributed to understanding of the write or where words, but must be attributed to lucky guessing. Subjects could not build up reliable strategies for determining the structure of the actual trial before the instruction, because all four possible trial structures (see fig. 2) occurred equally often in the experiment and were run in random order for each subject.

It is thinkable, though, that subjects built up expectations about what the next trial would look like and used these strategies successfully a number of times. Statistically seen any strategy building on the structure of previous trials should only be successful on half of the trials (because of the random trial order they would also fail on half of the trials), but since only 17 random trial orders were picked out and used in the experiment it cannot be completely assured that those trial orders did not favour such strategies above the level of chance. Actually on an average of 33.7 trials subjects could guess the write word target picture correctly if they just looked at the last trial's write picture (with the number of those trials amounting to up to 43 with one subject) and on an average of 31.2 trials they could guess the write word correctly if they looked at the last trial's where picture (in a perfectly balanced experiment the guessing would always have worked on 32 trials).

It still does not seem convincing that these strategies have been followed by many subjects – after all, the trial orders *were* randomized and subjects also came across numerous occasions where such strategies could not be used successfully.

Some subjects were even asked explicitly after the experiment if they had developed any strategies to fulfil the task as well as possible. However, none of them could answer this question, which is why it was not asked to more than five subjects. One could suspect though that subjects could have named a strategy building on trial structure had they used one and been conscious about it. So the reactions to this question on explicit strategies can also be taken as support for the assumption that the early target picture advantage was not caused by the trial orders.

One can say that the early target picture advantage remains a mystery. This does not endanger the findings depicted above, though. What is important for the conclusions to be drawn here is that both write and where word curves rise rapidly from 200 – 300 ms after target word onset, which is in line with other research on speech comprehension and eye movements. Since the behavior of the data changes drastically around 200 ms one can assume that a new factor, comprehension, has taken over there and that whatever factors influenced gaze before lost their impact at that point in time.

To be able to examine the time course of gaze reactions in a more fine-grained manner it would be necessary to eliminate the 'lucky guessing' effects, because with the present data and its noise it cannot be determined when a target preference starts to be caused by comprehension of the target word.

#### 4.6. Task specific behavior?

It is important to exclude the possibility that the behavior that is investigated (in this case the reactions to the write word) was caused by the experimental task itself. To do this the time course of gaze behavior over the 64 experimental trials was investigated, assuming that task specific behavior would betray itself through either increasing the target advantage towards the end of the experiment (if gaze on target pictures was the result of a viewing behavior acquired during the experiment) or enhancing it at the experiment's beginning (if gaze on target pictures was only a result of subjects' confusion about which picture to pick as the write picture).

To investigate possible learning/task specificity effects, the time course of gaze on the first 20 trials of all subjects' experimental runs was compared to the corresponding last 20 trials.

First the percentages of gaze samples falling on the write and where pictures were calculated for all subjects for the first as well as the last 20 trials of each experimental run. The results can be seen in fig. 6.

One can see in this graph that the reactions to where and write words are about 70 ms delayed on the last compared to the first 20 trials. The curves of the last 20 trials reach about the same heights as on the first 20 trials. So subjects look at the target pictures equally reliably during the whole experiment, but their reactions get slower towards the end.

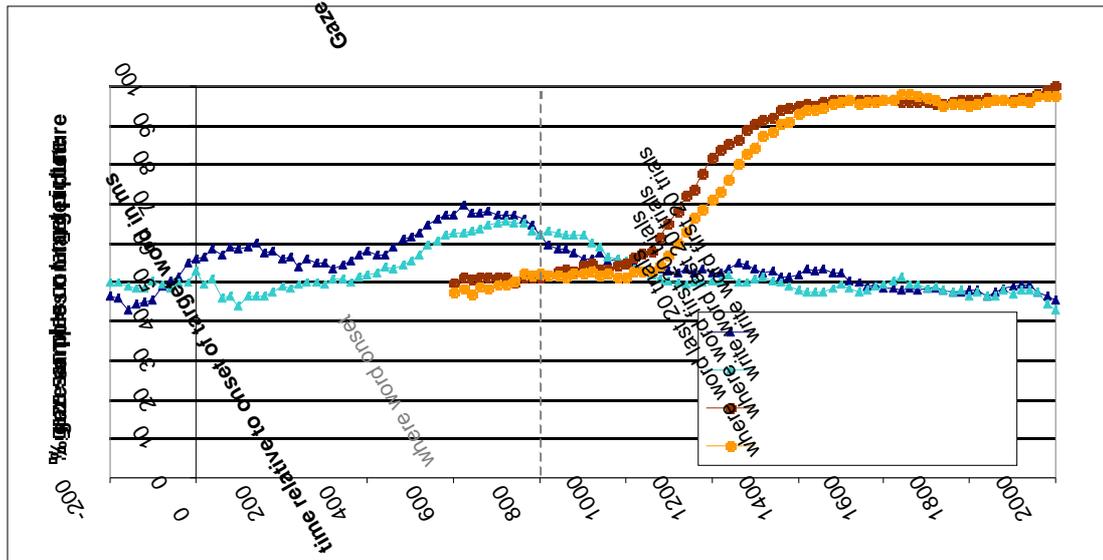
In addition to this one can observe a strong tendency for subjects to look at the write picture during the first 300 ms after the onset of the write word on the first 20 trials of each experimental run. Attempts to explain this with the subjects using strategies to guess the write word fail just as the abovementioned attempts to explain gaze at target pictures on all trials during the first 200 ms after word onset.

These results can be taken to exclude the possibility that the general trends observed in the experiment have to be ascribed to learning effects, because reactions to the target pictures get slower, but not weaker or stronger.

The fact that reactions get slower with both write and where words by about the same amount of time (approximately 70 ms) can presumably be attributed to the subjects getting tired. It is important to notice that both curves are shifted equally much to the right and keep their shape<sup>12</sup>, because it is this observation that defends the main result of this experiment against the claim that it could depend on task-specific strategies: The essential outcomes of this experiment are that reactions to write words occur, reliably and quickly, but are weaker than the reactions to where words. Towards the end of the experiment reactions to the write words are still strong and quick, and the relation between write and where word reactions does not change, because both reactions are delayed by the same amount of time. This is why it can be assumed that general factors like fatigue cause the change of behavior towards the end of the experiment, not a qualitative change of the processes underlying word recognition and the steering of eye movements.

---

<sup>12</sup> ignoring the guessing effect with the write word on the first 20 trials



**Figure 6:** Percentages of gaze samples falling on the target pictures of write and where words, on the experiment's first and last 20 trials. This figure has to be understood in a similar way as fig. 4 (see clarifications beginning on p. 11).

Another possible factor causing the slower reactions on the last trials could be that subjects do not care about responding as quickly as possible anymore (as they were instructed in the beginning of the experiment), maybe because they try to respond more correctly instead, or just because they become bored and thus less engaged.

## 5. Discussion

### 5.1. What steers eye movements during speech comprehension?

The hypothesis formulated in this paper, saying that it would be possible to observe a shift of gaze to the write picture following the onset of the write word, could be confirmed.

This result can be taken as proof for the basic assumption of the linking hypothesis presented in this paper. There was a significant tendency to shift gaze to the picture corresponding to a spoken word in a context where looking at that picture was not task-relevant. This observation can be explained well referring to spreading of activation between the different components of a concept, which is an automatic process that takes place every time one component of the concept is activated – the essential idea of the linking hypothesis proposed in the beginning of this paper.

Additionally the result of the present experiment can be taken to extend Just & Carpenter's eye-mind assumption (Just & Carpenter, 1980) to situations in which the processing of the visual environment is not necessary. As it was shown linguistic and visual processing are also linked in a situation in which the visual display was not task-relevant (on the write word condition).

In addition to this, the results of this experiment give rise to further speculations on factors that influence gaze behavior as a reaction to language understanding, in connection to the linking hypothesis.

Subjects reacted to the where word with quick and reliable gaze shifts to the corresponding pictures. This reaction is consistent with a lot of other studies' results on spoken language comprehension using the visual world paradigm (Allopenna et al. 1998, Dahan et al. 2002, Hanna et al. 2003, Sedivy et al. 1999). The fact that these results could be replicated is valuable - it makes the data of the present experiment comparable to the abovementioned studies. Specifically it suggests that differences in behavior between this experiment's write and where word conditions are genuine and can be attributed to factors that also distinguish the write word condition from the abovementioned previous studies.

As can be seen in section four, gaze reactions to the write word were weaker than gaze reactions to the where word. Whether 'weaker' means slower, less reliable or both at the same time could not be determined. To avoid such problems possible follow-up experiments would have to be designed in a way that different nouns triggering eye movements occur with a much longer time difference between them than in the present experiment.

Two of the possible explanations for this weaker reaction will now be discussed: differing degrees of pictures' task relevance and differing activation centres.

### *5.2. Degree of the visual stimuli's task relevance*

The two experimental conditions of the present experiment differ in how task relevant it is to look at the picture corresponding to the where word or the write word. In previous studies (Allopenna et al., 1998; Arnold et al., 2000; Barr & Keysar, 2002; Dahan et al., 2001; Sedivy, 1999; Spivey, 2002; Tanenhaus, 2000) as well as on the where word condition in the present experiment, the motivation for looking at objects mentioned in the auditory stimulus can be found in the task the subjects got. Tasks required subjects to 'pick up' objects or to 'write on' them – actions that require visual planning and controlling, by looking at the objects in question. If subjects wanted to fulfil the task correctly they were forced to react to where words by looking at the corresponding pictures. Because of this, activating a visual representation of the object they were to find on the display was absolutely necessary.

On the write word condition people did not at all have to look at the write picture to fulfil the task. This was clear to them, since instructions always had the same structure and five training trials had made subjects accustomed to the task. So theoretically it was not at all necessary to activate a visual representation corresponding to the write word and to look at the write picture. It is still thinkable that some subjects considered the write pictures task relevant, just because they were there on every trial. It can be assumed, though, that even in those cases write pictures were at least considered less task-relevant than where pictures.

The weaker reaction to the write word can supposedly at least partly be attributed to this lower task relevance of the write picture. In the terms of the model introduced in section two, the relevance of the visual phenomenon (the write picture in this case) is, though increased through the activation of a corresponding visual representation, not judged to be very large. This judgement is made on the basis of context information (in this case the knowledge that the write word only needs to be written, acquired through learning to know the structure of the instructions).

To generalize the idea of the task-relevance of visual stimuli, one could propose that gaze shifts to objects corresponding to words just processed become less probable the less the actual task requires visually processing the object.

All kinds of different degrees of the task relevance of different visual stimuli can be imagined and could be created in different experimental setups. Follow-up studies to the present paper could aim at creating different experimental conditions on which the relevance of visual stimuli is varied through different situational contexts. With such a series of experiments it could be tested how the task relevance of visual stimuli influences their visual processing as triggered by corresponding linguistic expressions.

### 5.3. Activation centre

Another dimension along which the write and where word conditions differ is which kind of information the write and where words referred to. Where words always explicitly (as subjects had learned during the training trials and through repetition during the experiment itself) referred to one of the two pictures. In contrast to this, write words only referred to their written representations which were to be produced by the subject.

This difference between the two conditions could also be a cause for subjects' different reactions to write and where words. One could assume that different types of information have to be primarily retrieved from the write and where words' concepts to fulfil the task at hand – the activation centres of the concepts differ. In the case of the write word it would be a graphemic representation, in the case of the where word it would be a visual representation. One can predict from this that the activation of the visual representation of the concept is probably stronger on the where word condition than on the write word condition, triggering a stronger gaze reaction to the where word.

Support for this assumption can be found in Langacker (1991): According to Langacker, the different dimensions belonging to a concept differ in the degree to which they are activated on different uses of the concept. Assuming that all other dimensions are activated more weakly than the activation centre, it should be possible to conclude that eye movements are triggered most strongly when the visual representation is a concept's activation centre. This generalization is supported by the results from the experiment, as described above.

However, the present experimental results are far from proving the hypothesis on activation centres – this could only be accomplished by follow-up experiments in which the effect of different activation centres (created by varying situational contexts) on eye movement reactions to spoken words are examined.

The results of the present experiment made it possible to propose lines along which further exploration of the linkage between eye movements and language processing could be carried out. A strong descriptive model of this linkage controlling for important factors that influence eye movement behavior could be a big help in interpreting eye-tracking data from linguistic research as well as open the door for new uses of the eye-tracking paradigm – not only in spoken language research.

## 6. References

- Allopenna, P. S., Magnuson, J. S. & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, 38, 419 – 439. Academic Press
- Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S & Trueswell, J. C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76, B13-B26.
- Barr, D. J. & Keysar, B. (2002). Anchoring Comprehension in Linguistic Precedents. *Journal of Memory and Language*, 46, 391-418.
- Dahan, D., Tanenhaus, M. K. & Chambers, C. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47, 292-314.
- Glenstrup, A. J. & Egnell-Nielsen, T. (1995). Eye Controlled Media: Present and Future State, <http://www.diku.dk/~panic/eyegaze/article.html>.
- Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 81, 1-12.
- Griffin, Z. M. (2004). Why look? Reasons for eye movements related to language production. In: Henderson, J. M. & Ferreira, F (Eds.). *The integration of language, vision and action: Eye movements and the visual world*. New York: Psychology Press.
- Griffin, Z. M. & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274-279.
- Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354.
- Langacker, R. W. (1991). *Concept, Image and Symbol. The Cognitive Basis of Grammar*. Berlin: Mouton de Gruyter.
- Meyer, A. S. & Dobel, C. (2003). Application of eye tracking in speech production research. In: J. Hyöna, J. R. Radach & H. Deubel (Eds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp.253-272). Oxford: Elsevier Science.
- Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, Massachusetts: Bradford Books.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71, 109-147.
- Sperber, D. & Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford UK and Cambridge USA: Blackwell Publishers.
- Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M. & Sedivy, J. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45, 447-481.
- Tanenhaus, M. K., Magnuson, J. S., Dahan, D. & Chambers, C. (2000). Eye Movements and Lexical Access in Spoken-Language Comprehension: Evaluating a Linking Hypothesis between Fixations and Linguistic Processing. *Journal of Psycholinguistic Research*, 29, 557-580.

