Processing Grammatical Gender in German

- An Eye-Tracking Study on Spoken-Word Recognition -

Bachelor’s Thesis

by

Nadine Hartmann

1st examiner: Dr. Graham Katz
2nd examiner: Dr. Sabine Reinhard
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLES</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>0 Abstract</td>
<td>1</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>2</td>
</tr>
<tr>
<td>1.1 Gender in General</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Properties of the German Gender System</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Research on Gender</td>
<td>7</td>
</tr>
<tr>
<td>1.4 Eye-Tracking</td>
<td>9</td>
</tr>
<tr>
<td>2 Experiment</td>
<td>12</td>
</tr>
<tr>
<td>2.1 Hypothesis</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Method</td>
<td>13</td>
</tr>
<tr>
<td>2.2.1 Participants</td>
<td>13</td>
</tr>
<tr>
<td>2.2.2 Stimulus Materials</td>
<td>14</td>
</tr>
<tr>
<td>2.2.3 Procedure</td>
<td>15</td>
</tr>
<tr>
<td>2.2.4 Coding and Data Analysis</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>18</td>
</tr>
<tr>
<td>2.3.1 Condition 1</td>
<td>19</td>
</tr>
<tr>
<td>2.3.2 Condition 2</td>
<td>20</td>
</tr>
<tr>
<td>2.3.3 Condition 3</td>
<td>21</td>
</tr>
<tr>
<td>2.3.4 Condition 4</td>
<td>22</td>
</tr>
<tr>
<td>2.3.5 Analyses of variance</td>
<td>24</td>
</tr>
<tr>
<td>3 Discussion</td>
<td>25</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>XXVIII</td>
</tr>
</tbody>
</table>
TABLE OF FIGURES

Figure 1: The Corneal Reflections Produced by Different Eye-Head Positions ................................................................. 10
Figure 2: Arrangement of Objects on the Screen................................. 14
Figure 3: Fixation Probability over Time of Condition 1 ................. 20
Figure 4: Fixation Probability over Time of Condition 2 ................. 21
Figure 5: Fixation Probability over Time of Condition 3 ................. 22
Figure 6: Fixation Probability over Time of Condition 4 ................. 24

TABLES

Table 1: German Definite Articles by Gender, Number, and Case .......... 7
ACKNOWLEDGEMENTS

I would like to thank all the people that helped in many different ways with the completion of this thesis. Special thanks to Dr. Graham Katz for a lot of good advice, always offering assistance, and a lot of patience in particular. In addition, I would like to thank the people who helped me finish this thesis when I was in the depths of despair: Boris Gutbrodt, Sarah Irwin, and my family, especially my sister Ivonne Hartmann.
Abstract

An eye-tracking experiment examined grammatical gender effects in sentence processing in German. Participants followed spoken instructions to click on one of four pictures on a computer screen (e.g., *Klicken Sie auf die blaue Rakete.* [Click on the blue rocket.]) while their eye-movements were monitored. The name of the target object was preceded by a definite article of accusative case and an adjective, both carrying gender information. When a target picture was accompanied by a gender-matching or color-matching competitor in the visual display, fixations were higher to these objects than to unrelated distractors. Gender-matching competitors were activated earlier during processing than color-matching ones. In addition, prior gender information did not prevent gender-mismatching objects to be activated, if these agreed in color with the target. On the whole the results speak against a post-lexical influence of gender marking, but rather support the notion of an earlier effect as the interactive view.

The results given provide evidence that morphosyntactic gender information is not able to uniquely specify a target object. They do underline the importance of investigating gender in sentence processing, and the advantages of the eye-tracking paradigm, though. Hopefully the problems that occurred here will be considered thoroughly by future experimenters, and coming experiments concerning gender processing can strengthen existing results and also move a step closer towards grasping the essence of grammatical gender.
1 Introduction

The present research focuses on linguistic gender in spoken-word recognition investigated within the eye-tracking-paradigm. Some attention has been paid to this phenomenon, but the role of grammatical gender in linguistic processing is still barely understood and only few truly sophisticated findings exist. This introductory part serves as a short overview of research done in the past on the topic of this paper. First, an overview of grammatical gender and its implications on a language is given and expatiated upon German on this basis. As a fundamental source for this section served the work of Corbett (1991), which is an insightful analysis of over 200 gender systems in the world. Afterwards, some previous research on gender is reconsidered with respect to how and where this information influences the word recognition process. At the end, the eye-tracking-paradigm and some findings with this methodology will be discussed.

1.1 Gender in General

Grammatical gender, henceforth referred to simply as gender, is an intriguing phenomenon in many of the world’s languages. It can be defined as a system of classes of nouns that is reflected in the behavior of associated words. Every noun must thus belong to one of the classes and there should be few which belong to several classes at once (Hockett, 1958; cited in Corbett, 1991, p. 1; Wikipedia). The number of these classes might vary considerably from one language to another: Most Indo-European languages are organized around two or three genders typically labeled masculine/feminine or masculine/feminine/neuter, such as French and Italian, or Russian and German. In contrast to this, many African, especially the Bantu languages, comprise from approximately ten up to over twenty usually numbered so-
called noun classes (Corbett, 1991). Hence, gender distinctions represent different grammatical classifications of nouns, but these classes do not necessarily need to mirror natural (biological) gender.

But how are nouns allotted to genders in a language then? According to Corbett (1991), semantic and formal principles build the basis for these connections. Semantic rules assign a gender to nouns with respect to their meaning (semantics). In strict semantic systems like Tamil – so called natural gender systems – semantic rules often take the principle of humanity or animacy as the semantic core of gender. For example all animate male nouns receive masculine assignment, all animate females will be feminine and all other (inanimate) nouns will be of neuter gender. But this kind of language is relatively rare and most languages do not follow such a straight-forward system. In these more formal systems gender may be determined by morphological attributes of the noun, e.g. specific affixes, or phonological attributes, e.g. syllables etc. These types are often (inter-)connected and it is not that easy to distinguish between the two of them. Importantly, formal systems always have a semantic basis, i.e. there is no language that operates solely on formal rules, but these rules intervene where semantic ones do not succeed. When these principles contradict, semantics usually takes precedence over syntactic criteria, but in general they show extensive overlap.

Another way in which languages differ that is closely connected to the former is the type and scope of their agreement domain, i.e. the constituents that can show the gender. They might be restricted to the noun phrase itself or extend to various other elements of which the noun phrase is an argument. A noun’s gender may thus have an influence on the modification of constituents like definite/indefinite articles, numerals, demonstratives, possessives, and adjectives as well as on participles, verbs, (relative and personal) pronouns, adverbs, pre-/postpositions, and even complementizers (Corbett; 1991; for review van Berkum, 1996). The realization of gender on these elements is
primarily reflected by inflectional affixes, as they do not have a gender by themselves, e.g. *stylo* [pen] is masculine in French and mainly the article and adjective, but other constituents, too, have to agree with the noun and carry out the agreement (gender) marking (Grosjean, Dommergues, Cornu, Guillelmon, & Besson, 1994): *Le stylo auquel j’écris est vert.* [The (masc.) pen with which I write is green (masc.).] In the next section, this rough basis will be elaborated for gender in German.

### 1.2 Properties of the German Gender System

German, belonging to the Indo-European language family, employs a three gender system, i.e. each noun is either masculine, feminine, or neuter. Some homonymies are of two genders, in which a category shift is meant to resolve ambiguities (e.g. *der Leiter* [the (masc.) chief] and *die Leiter* [the (fem.) ladder]; *der Gehalt* [the (masc.) concentration] and *das Gehalt* [the (neut.) salary]; *die Steuer* [the (fem.) tax] and *das Steuer* [the (neut.) steering-wheel]). Only a couple of nouns may have double-gender, with no difference in meaning (e.g. *der/das Liter* [the (masc./neut.) liter]; *das/der Dotter* [the (neut./masc.) yolk] (Lübke, 1999)), but typically one form is regarded to be more colloquial and the other more formal and less common. Taken word frequency into account, genders are almost equally distributed, with the neuter gender appearing only slightly less (26%) than feminine (35%) or masculine (39%) gender (Schiller & Caramazza, 2003).

A person who has not studied German can form no idea of what a perplexing language it is. [...] Every noun has a gender, and there is no sense or system in the distribution; so the gender of each must be learned separately and by heart. There is no other way. To do this, one has to have a memory like a memorandum book. In German, a young lady has no sex, while a turnip has. Think what overwrought reverence that shows for the turnip, and what callous disrespect for the girl.

(Mark Twain, 1880)
Mark Twain’s citation of “The awful German Language” gives an idea about how complex and unpredictable the German gender system seems to be for somebody learning German as a second language, although it does not appear to be a difficulty for a native speaker. But behind this arbitrariness at first glance, there emerges a complex combination of semantic, morphological, and phonological principles (Corbett, 1991; Köpcke, 1982). A lot of research in the past was carried out to discover the regularities of gender assignment in German, where probably the most widely acknowledged contributions come from Zubin and Köpcke (Köpcke, 1982; Köpcke & Zubin, 1983, 1996). But it is not the purpose of this paper to identify these rules and regularities, only some examples and interesting exceptions should be mentioned to get a small impression. The most obvious semantic principle concerns natural gender: Gender reflects sex almost without exceptions. Nouns referring to males receive masculine assignment and to females feminine respectively (Köpcke & Zubin, 1996). This holds for human beings as well as for animals. But here we also find the most common and cited exception as well. The German word Mädchen takes neuter gender assignment (das Mädchen [the(neut.)/that(neut.) girl]), because the diminutive –chen, as its morphological suffix, is usually neuter. Nevertheless, it may take the feminine personal pronoun sie (e.g. Mein Freund hat ein kleines Mädchen, das/*die ein Meerschweinchen hat. [My friend has a(neut.) girl that(neut.)/that(fem.) has a guinea-pig.] – Mein Freund hat ein kleines Mädchen. Es/Sie hat ein Meerschweinchen. [My friend has a(neut.) girl. It(neut.)/She(fem.) has guinea-pig.]). In this case, where a noun does neither take agreement of one consistent pattern nor belongs to more genders than one, this noun is supposed to be a hybrid noun (Corbett, 1991). Morphological rules, like the one above, are often not clearly distinguishable from semantic ones, because they sometimes have a meaning on their own. Other examples are the very productive suffix –heit (status, nature), that always carries feminine assignment or –tum (demonstrates membership to group), which is always neuter (Weber, 2001). Phonological assignment rules were elaborated extensively by Köpcke
(1982) on the basis of monosyllabic nouns. He came to the conclusion that almost 2/3 of the German monosyllabic nouns are of masculine gender. Exceptions occur, when they end in –ur/ür, which normally implies that they receive feminine assignment.

In German the associated gender is not overt on the noun itself, but has morphosyntactic repercussions on constituents of the noun phrase instead, which is called noun-inherent gender (noméninhärentes Genus; Weber, 2001) or covert gender (van Berkum, 1996). Constituents showing gender-marking can be found among almost all pronouns (demonstratives, possessives, interrogative-, and indefinite- as well as personal- and relative pronouns), numerals, adjectives and articles. Only the latter will be discussed a bit further, as it is meant to be the one that reflects gender primarily. Gender is not the only nominal category however that determines the inflected alignment of constituents in a sentence. All three nominal categories together, where number and case are the other two, have to be chosen in their respective context (by the speaker) and generate congruency within a sentence, and especially within a noun phrase. In German, as a relatively richly inflected language, the information about these three categories merges together in only one inflectional morpheme which is generally not unambiguous at all. For 16 possible functions in gender, number, and case – gender differences only occur in the singular and disappear in the plural –, there are only six different inflectional elements in the case of the German definite articles. Hence, no form is a unique combination of gender, number, and case. Table 1 is meant to illustrate this situation:
Table 1: German Definite Articles by Gender, Number, and Case

<table>
<thead>
<tr>
<th>Case</th>
<th>Gender</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>masc.</td>
<td>fem.</td>
<td>neut.</td>
</tr>
<tr>
<td>Nominative</td>
<td>der</td>
<td>die</td>
<td>das</td>
</tr>
<tr>
<td>Genitive</td>
<td>des</td>
<td>der</td>
<td>des</td>
</tr>
<tr>
<td>Dative</td>
<td>dem</td>
<td>der</td>
<td>dem</td>
</tr>
<tr>
<td>Accusative</td>
<td>den</td>
<td>die</td>
<td>des</td>
</tr>
</tbody>
</table>

As one might have seen by now, dealing with gender is not a trivial thing. But why exactly is it that many languages stick to such a cognitive enormously demanding category? Grammatical gender must exhibit a certain role, such that it is not just a burden for our memory and grammatical processing – otherwise it would probably have been eliminated over time. Different approaches to this question are discussed in the next section.

1.3 Research on Gender

Many researchers have already focused on lexical gender effects using diverse experimental paradigms to explore the value for processing and investigate the boundaries of where and how gender information affects lexical access of the subsequent noun. Three different points of view were summarized and discussed by Friederici & Jacobson (1999; see Bölte & Connine (in press) for summary): The post-lexical, the (intra-)lexical, and the interactive approach. One obvious function of gender is that it generates congruency among constituents that belong together within and over sentences, hence establishing local and global coherence. By providing pronouns, it permits to keep track of reference in discourse in an elegant and efficient way. Support for this view of gender in linguistic processing was provided by Colé & Segui (1994), who examined grammatical priming in French using a lexical decision task. The observed effect reveals that a word is recognized faster when it is preceded by a valid than by an invalid gender marking on a prior adjective. Gender priming results have been explained to be due to a post-lexical congruency test, i.e. the
influence of gender information is thought to be subsequent to lexical activation. Within this view lexical is supposed to be independent of previous syntactic or semantic information.

Using two methodologies, gating and lexical tasks, Grosjean et al. (1994) studied recognition of French nouns that were either preceded by an appropriate gender marked article or no article at all. Within both paradigms, strong facilitatory effects of conditions with the gender carrying article were obtained (faster reaction times in the lexical decision and earlier recognition points in the gating task). Grosjean et al. (1994) regarded a post-lexical locus as well as a lexical one, where gender information affects lexical activation, as possible.

The third view considers gender marking as an interactive venture predicting gender priming to occur pre-lexically taking semantic and syntactic information into account. This opinion is supported by Bates, Devescovi, Hernandez, & Pizzamiglio (1996) who examined inhibition and facilitation effects by word repetition, gender monitoring and grammaticality judgement tasks. Word repetition showed a clear gender priming effect, gender monitoring provided facilitatory effects, and grammaticality judgements revealed faster reaction times for items, which agree in gender, than for rejection of gender inconsistent phrases.

In a recent study on auditory word recognition in German, Bölte & Connine (in press) conducted two corpus analyses indicating that an article can constrain the gender of a subsequent word. In two further experiments using lexical decision tasks, they were able to find robust gender priming effects. A phoneme monitoring experiment, however, did not show an influence of gender marking, only showing lexical similarity effects. Altogether, their results suggest that gender is not used early in word recognition processes, but rather increases initial familiarity of a word.
1.4 Eye-Tracking

As a more recent methodology to gain insight into language processing, eye-tracking, i.e. monitoring eye-movements of subjects, while reading or listening to spoken language, has been developed. Positions and latencies of eye-movements on pictures are recorded using a camera usually mounted on a headband. Two types of eye-movements can thus be distinguished – fixations, when the eye rests on an object, and saccades, when the eyeball jerks to another position. The saccades are ballistic in nature and assumed to be programmed during the fixations. Via the reflection of some infrared illumination, typically two optical features are processed hand in hand by the eye-tracker. First, it registers parameters of the pupil by the lack of reflectance (dark pupil tracking) and in combination the corneal reflection, or first Purkinje image, which reflects the striking light in only one small point that appears as a small bright dot within or right next to the pupil (figure 1, taken from Richardson & Spivey, 2004). These characteristics together provide good and stable information about the point of gaze, because their constellation differs when the subject moves his/her eyes or head. There exist several other techniques to monitor participants’ eye-movements, but it would probably be out of the scope of this paper to review them all (for further information the interested reader is kindly referred to the above references).
The first pioneering work has already been three decades ago, however. Cooper (1974) demonstrated in his ground-breaking experiment that people are likely to spontaneously guide their eye-movements to elements in the visual field which are semantically related to the spoken language currently heard. He used a fixed-head eye-tracker to record eye-movements to pictures while subjects listened to a story. Tanenhaus and his colleagues (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Eberhard, Spivey-Knowlton, Sedivy & Tanenhaus, 1995) further elaborated this work and derived the eye-tracking, also known as “visual-world”, -paradigm. In these experiments, participants interact with a display of multiple objects and receive spoken instructions to manipulate these objects, while their eye-movements are measured. The tasks revealed some interesting results, showing that this methodology is eminently suited for investigating linguistic processing. Eberhard et al. (1995) and Tanenhaus et al. (1995) conducted several experiments demonstrating that reference is established incrementally by the listener with eye-movements being closely time-locked to referents in the speech stream. Moreover, they found clear effects of visual context on the spoken input, since it affected the resolution of temporary ambiguities within words, as well as within sentences (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). It was shown that non-
linguistic context affected syntactic ambiguity resolution, when it supported the normally less preferred structure of a prepositional phrase. Predominantly, the activation of a visually mediated lexical competitor has been studied broadly to gain insights into linguistic processing. During recognition of spoken words, multiple candidates are assumed to get simultaneously activated, such that this search space becomes reduced when inconsistencies between candidates and the continuing speech input occur. While subjects hear the name of a target, they are more likely to initially look at pictures with the same onset as the target than to phonological unrelated distractors (e.g., upon hearing the word “beaker”, a picture of a “beetle” is more likely to be fixated than distinct distractors). Such competition effects closely map to activation levels of word candidates as simulated in computational models of spoken-word recognition such as TRACE (Allopenna, Magnuson, & Tanenhaus, 1998). After all, some advantages of this technique are obvious already. In properly designed experiments it is useful as an on-line investigation method to study cognitive processing with non-invasive tasks, and it allows for a relatively realistic environment under practically natural conditions (Tanenhaus & Spivey-Knowlton, 1996).

Underlying this present research is an eye-tracking study by Dahan, Swingley, Tanenhaus, & Magnuson (2000), which examined gender in spoken-word recognition in French. They conducted two experiments to investigate whether a gender-marked article restricts the set of candidates initially considered for recognition of the forthcoming noun. In the absence of gender marking (i.e., using the French plural article *les*) a “cohort” competitor effect for nouns sharing the initial phoneme was found. For another condition, in which they tested the use of gender information carried by an article when no phonological overlap between object names was given, they discovered that this information only is not sufficient to restrict attention to pictures with the same gender as the target word. In the second experiment they examined if a
gender-marked article can eliminate the activation of the phonological similar cohort. Their results provide evidence that the initial set of lexical candidates is constrained by the preceding information of the article: Upon hearing “Cliquez sur le bouton.” [Click on the (masc.) button.], French listeners did not fixate the picture of the feminine cohort bouteille [bottle] more often than unrelated distractors. This present study is meant to examine whether it might be different for the German language and gender information is sufficient to recognize a subsequent noun, i.e. that it is possible to limit participants’ attention to pictures with the explicitly named gender. The idea was raised, because French definite articles are even shorter than the German ones and, in addition, it is possible to insert a gender marked adjective. This might reduce the high co-occurrence of definite articles and nouns and gives gender information more time to be managed.

2 Experiment

This experiment was designed to assess the impact of the German gender system on sentence processing. While their eye-movements were monitored, participants received auditory instructions to mouse-click on one of four pictures presented on a computer screen (e.g., Klicken Sie auf die blaue Rakete. [Click on the (fem.) blue (fem.) rocket.]). The objects varied among two dimensions: their syntactic gender and their coloring. The experimental manipulation was realized in variations of the objects offered together with the target. Depending on the type of trial, the distractor objects included either a “gender-matching competitor” (GMC) with the same gender as the target, a “color-matching competitor” (CMC) with the same color as the target, a gender- and color-matching competitor (GCMC), or an unrelated distractor (UD) neither related in gender nor in color, in various combinations.
There were four different conditions in total: In the first one, all pictures were of the same color and among the distractors there was one picture also sharing the gender of the target, thus one GCMC and two CMCs. In the second type of trial, the color of the objects was the same, too, but no distractor was of the same gender as the target, only CMCs appeared. The third condition was a bit different, as always two pictures were of the same color and the gender of the target was unique, like in condition 2 (one CMC and two UDs). In the last condition, colors were distributed as in the previous one, but a gender-matching competitor belonged to the target’s opposite color (one GMC, one CMC, and one UD).

2.1 Hypothesis

As this study was meant to contradict in some sense to the one of Dahan et al., results pointing to a more obvious usage of gender marking were desirable. The expectations were that it would be easier for the participants to correctly fixate the target, when no gender-matching competitor is given. Hence, it should be simpler to identify the goal object in condition 2 than in condition 1. Furthermore, it was hypothesized that the gender information given should be able to decrease the attention to incongruent color-matching competitors like in condition 4.

2.2 Method

2.2.1 Participants

The participants volunteering for the experiment were twenty native speakers of German, mainly students at the University of Osnabrück from diverse faculties. They were aged between 17 and 31 with the average age lying at 23.8 and with an equal distribution among men and women. All of them had
normal or corrected-to-normal vision and none reported any speech or hearing
deficits which could have influenced their performance.

2.2.2 Stimulus Materials

A set of ninety-one relatively common German nouns was selected – 31 of
masculine, 30 of feminine, and 30 of neuter gender (see Appendix 1). The only
restrictions on the choice were that they had to be unambiguous in gender, to
refer to picturable objects, and could be extended by an adjective. All the
adjectives used were color adjectives or referred to hair-dyes. The pictures
chosen were normal clip art line drawings whose color could be modified. In
groups of four, they were displayed to the subjects, always with a normal, white
background. The screen was divided by an invisible 2x2 grid, with each
individual object situated in the center of one of the four resulting fields.
Figure 2 represents the arrangement of the objects on the screen.

![Figure 2: Arrangement of Objects on the Screen](image)

The auditory instructions were spoken by a female German native speaker (the
author) and recorded in a soundproof room at a sampling rate of 32 kHz with
a sampling size of 16 bits using the CoolEdit 2000 program. All of them were of the same structure: “Klicken Sie auf ...” [click on ...] followed by a noun phrase (definite article, adjective (both with accusative case), and noun). After editing the audio files, some basic durations with respect to the trials were measured: In condition 1 “Klicken Sie auf" [click on] was on average 707 msec long, the determiners 191 msec, the adjectives 454 msec, and nouns 559 msec, the whole trial was thus 1911 msec long. The means in the other conditions did not differ much from these results. Condition 2: beginning: 710 msec, det: 200 msec, adj: 398 msec and noun 557 msec, 1864 msec long in total; Condition 3: beginning: 714 msec, det: 188 msec, adj: 474 msec, and noun 661 msec, 1890 msec long altogether; Condition 4: beginning: 716 msec, det: 184 msec, adj: 400 msec, and noun 574 msec, 1874 msec long overall.

Twelve trials were implemented for each of four conditions, so that there were 48 different targets to be clicked on in total (see appendix 2). Care was taken that the three genders appeared within each condition equally often – as target, as well as among the particular distractor objects. Every object was shown at most three times during a whole session.

2.2.3 Procedure

All participants were tested individually in the usability laboratory of the University of Osnabrück. Prior to the eye-tracking-experiment subjects received a handout containing the pictures they were to see in the following. In order to establish naming-norms, the names of these images were read out aloud from the experimenter and in the next step repeated by the participants. If the subjects named an object incorrectly, they were corrected by the experimenter and asked to say the right word again. The repeated words and intended names in the first step matched in 95.1% of the cases.
For the following, participants were seated at a comfortable distance from the subject’s computer screen. A lightweight SMI EyeLink I eye-tracker was used to monitor their eye-movements relative to their head. This was accomplished by two special high-speed infrared cameras mounted on a headband which recorded the observer’s pupils at a sampling rate of 250 Hz and a so-called head-camera, which communicated with four infrared markers in the corners of the screen to balance out head-movements. Before the experiment could begin, a calibration procedure had to be carried out and some basic instructions were given.

The course of each trial always took place in the same manner: Participants were allowed to scan the display containing four objects freely until the mouse cursor emerged in the middle of the screen. It appeared approximately 1600 msec after stimulus onset as a central fixation target equidistant to any other object on the screen to control eye-position at the start of each trial (Tanenhaus & Spivey-Knowlton, 1996). Subsequently, participants were supposed to look at that arrow until the presentation of the auditory stimulus which was about 2000 msec after picture onset and 400 msec after the mouse cursor appeared. These time intervals seemed reasonable and are very similar to the ones used by Allopenna et al. (1998), as the participants had just enough time to scan the four objects and then center their fixations to eliminate baseline differences on the pictures when the auditory instructions began. These were presented via two loudspeakers next to the monitor, such that the subjects listened to them and clicked on the appropriate object as soon as possible. After each trial the display was cleared and a dot appeared in the middle of the screen that permitted the experimenter to check whether the calibration was still adequate enough to proceed.

Participants were not told anything about the purpose of the experiment. In addition, they were misinformed about the uniqueness of the targets. It was mentioned that if an object appeared as target, it was still possible to be the
intended item another time. This was necessary, because the order of the 48 trials was absolutely randomized and participants should not be able to adopt a strategy with regard to the prediction of the goal object.

2.2.4 Coding and Data Analysis

The spatial coordinates of participants’ fixations were recorded together with the position of the mouse click. This data collected from the eye-tracker was edited and time frames of 20 msec duration were established. For each trial and each participant, the frame slot was filled with the object (including the cross) the subject was looking at. This was accomplished by using the invisible grid to determine the eye-position. Fixations were coded on each trial from the onset of the determiner until the end of that trial, when the subject clicked on the target picture. To ascertain this onset of the critical auditory input in the tracker data, a marker was placed during recording indicating the onset of the audio file. This point in time plus the time used for the instruction “Klicken Sie auf” [click on] represented the onset of the determiner.

For this experiment, the subjects’ gaze was reckoned as a fixation, if it remained on the picture for at least two time frames. The normal duration of a saccade is regarded to be 10 – 80 msec (EyeLink II User Manual – Version 1.05, 2002). In the rare cases where it took longer than these four time frames to reach a new object, the additional time was attached to the previous fixated object. Blinking time, too, was always added to the prior fixation.

The average latency of a saccade, i.e. the time passing until the eye-movement is executed, is estimated to be around 200 msec (Tanenhaus & Spivey-Knowlton, 1996). Therefore, an eye-movement caused by acoustic information should be detected only a few hundred milliseconds after it is launched. For this reason, a time window for comparison of fixation proportions extending from 300 to 1500 msec seemed reasonable to establish, over which fixations to
each picture or the cross were summed. To account for the influence of the constituents of the noun phrase, this interval was divided into smaller time windows: A 300 – 500 msec window to relate it to the determiner only, 500 – 900 msec to address determiner and adjectival information together, and 900 – 1500 msec for the whole noun phrase.

In order to test whether each of the competitors or the distractor(s) were fixated longer, planned comparisons were conducted between the fixation proportion to the particular type of objects in condition 2, 3 and 4 [one-tailed t-tests by items (t₁) and by subjects (t₂)]. These comparisons were computed for each of the time windows mentioned earlier. To assess the time course of lexical activation as speech unfolded, the fixation probability over time for each picture was computed and the mean for each of the four conditions was depicted graphically. A factorial analysis of variance (ANOVA) with repeat measures was conducted on the means of the first fixation to the target. For each participant in each trial the first target-fixation was determined. The means over the twelve trials in each condition was computed and served as a basis for the ANOVA. By this, it should be tested whether a difference in the speed of identifying the target can be observed, when variations in its definiteness are given. The same ANOVA was conducted on the final fixations to the goal picture just before clicking occurred to check, if an object unique in gender is recognized faster than in presence of a GMC or a GCMC in an overall trial.

2.3 Results

For some subjects a few trials were unusable due to technical failure. In addition, it happened five times that distinct subjects clicked on an object other than the target. In both situations plus the cases in which a subject’s gaze remained on the same object the whole trial long, the results were excluded
from further analysis. Altogether, 23 of 960 trials were missing thereby, which are almost equally distributed among conditions (6 in condition 1; 6 in condition 2; 4 in condition 3; and 7 in condition 4). One concern was that participants might have become aware of the purpose of the experiment, because only a small set of pictures and no filler trials were used. But only one of the subjects reported that he came to an idea what the experiment is all about by the end of his session, and these results did not differ much from the ones before.

2.3.1 Condition 1

Over the 300 – 1500 msec time window, the proportion of fixations to the target was 45.75%; to the GCMC 14.9%; to each averaged CMC 8.03%; and to the cross 19.18% in condition 1. The results do not sum up to 100%, because of the time taken up by the saccades where no object was fixated. Planned comparisons (one-tailed $t$-tests) between fixations to the GCMC and the averaged CMCs revealed significant results for each of the above mentioned time windows. The GCMC was fixated longer than the CMCs during the overall time window ($t_1(22) = 4.76, p < .0001; t_2(19) = 4.02, p < .0005$), the interval from 300 – 500 msec ($t_1(22) = 3.09, p < .005; t_2(19) = 1.98, p < .05$), from 500 – 900 msec ($t_1(22) = 4.03, p < .0005; t_2(19) = 5.58, p < .01$), and from 900 – 1500 msec ($t_1(22) = 2.97, p < .005; t_2(19) = 4, p < .0005$). This result demonstrates that the GCMC was activated during the speech stream, probably because of the same gender of the object.

Figure 3 displays the probability of fixating the target, the averaged CMCs, and the GCMC from the onset of the article on to the end of the time window. It shows that the fixation-probability of the GCMC diverged from the one of the CMCs, and even the one of the target, right from the beginning. This obliged to conduct another $t$-test between the GCMC and CMC, to see if there is a significant difference after article onset that might be due to some
undetermined factors. But this analysis did not show a significant result (two-tailed matched-pairs t-test with $t_1(22) = 1.79; p > .05; t_2(19) = 1.5, p > .1$) and thus allowed to evaluate this condition at all. The probability of fixating the GCMC remained higher than of the one of the CMCs the whole trial long, although the difference declined from 1100 msec on. In relation to the GCMC, the fixation-probability to the target was slightly less until about 600 msec after article onset and began to deviate considerably after approximately 800 msec. (The probabilities do not add up to 100%, especially at the beginning, because fixating the cross is not included to avoid confusion. However, one can imagine what that would look like for all of the conditions: Starting around 0.6 and decreasing with time until 0 at the end.)

![Figure 3: Fixation Probability over Time of Condition 1](image)

**2.3.2 Condition 2**

In condition 2, the proportion of fixations to the target was 46.7%; to each of the three CMCs 9.13%; and to the cross 20.18%. Figure 4 presents the
probability of fixating the target and the averaged CMCs over time, starting with article onset up to 1500 msec.

Figure 4: Fixation Probability over Time of Condition 2

2.3.3 Condition 3

The proportion of fixations over the 300 – 1500 msec time window in condition 3 were distributed as follows: to the target 46.49%; to the CMC 12.75%; to the averaged UDs 7.14%; and to the cross 18.1%. Planned comparisons between the CMC and UD demonstrate that the fixations to the CMC were longer than to the UDs during the time windows for 300 – 1500 msec \( (t_1(22) = 2.97, p < .005; t_2(19) = 4.43, p < .0005) \), 500 – 900 msec \( (t_1(22) = 2.63, p < .005; t_2(19) = 5, p < .0001) \), and 900 – 1500 msec \( (t_1(22) = 2.1, p < .05; t_2(19) = 3.85, p < .001) \). The fixations did not differ for the time interval from 300 to 500 msec as predicted (two-tailed matched-pairs \( t \)-test yielded no significance, \( t_1(22) = 1.79 =; p > .05; t_2(19) = 0.55, p > .5 \)).
The probability of fixating the target, the CMC and the two UDs over time is illustrated in figure 5. Until approximately 400 msec after article onset the probabilities of fixating either object were about equal. Then, the fixation-probability of the target increased rapidly, but stayed about the same for the CMC and UDs until shortly before 600 msec. Fixating one of the UDs became less likely, fixating the CMC however did not until 1000 msec, so that both probabilities were close to 0 by the offset of the noun. Thus, the CMC was activated during the presentation of the auditory stimuli, even though it conflicted with the target in gender.

2.3.4 Condition 4

In this condition, the fixation-proportion to the target was 56.91%; to the GMC 7.34%; to the CMC 8.46%; to the UD 4.81%; and to the cross 17.96%. For the overall time window, planned comparisons confirmed that the GMC and CMC were fixated longer than the UD (one-tailed t-tests between GMC
and UD: $t_1(22) = 2.07, p < .05$; $t_2(19) = 2.74, p < .01$; and CMC and UD: $t_1(22) = 2.66, p < .01$; $t_2(19) = 2.07, p < .05$). A difference between GMC and CMC could not be detected (two-tailed matched-pairs $t$-tests with $t_1(22) = 0.64, p > .5$; $t_2(19) = 1.35, p > .1$). Looking at the time interval a little bit closer, revealed a significant difference between fixations to the GMC and the UD and between the GMC and CMC for the time from 300 to 500 msec (one-tailed $t$-tests between GMC and UD: $t_1(22) = 2.91, p < .005$; $t_2(19) = 2.79, p < .01$; and GMC and CMC: $t_1(22) = 2.93, p < .005$; $t_2(19) = 1.78, p < .05$; two-tailed matched-pairs $t$-tests between CMC and UD $t_1(22) = 0.27, p > .5$; $t_2(19) = 1.13, p > .1$). For the time window extending from 500 to 900 msec, planned comparisons between the objects suggest that the CMC was fixated longer than the other two, but no significant result was found between the GMC and UD anymore (one-tailed $t$-tests between GMC and UD: $t_1(22) = 0.9, p > .1$; $t_2(19) = 1.66, p > .05$; CMC and GMC: $t_1(22) = 1.84, p < .05$; $t_2(19) = 2.55, p < .01$; and CMC and UD: $t_1(22) = 2.85, p < .005$; $t_2(19) = 3.57, p < .005$). This also holds for the last time interval from 900 to 1500 msec (one-tailed $t$-tests between GMC and UD: $t_1(22) = 1.66, p > .05$; $t_2(19) = 1.18, p > .1$; CMC and GMC: $t_1(22) = 2.15, p < .05$; $t_2(19) = 2.14, p < .05$; and CMC and UD: $t_1(22) = 3.7, p < .001$; $t_2(19) = 3.38, p < .005$). These results together propose that the GMC was activated early in time, because of its syntactic similarity to the target, but that this prior activation is diminished as soon as it turns out to be incongruent with the semantic information given by the adjective. In contrast to this, the CMC was activated despite of the incongruent gender information given earlier in the speech stream.

Figure 6 shows the probability of fixating each of the four objects graphically. The fixation-probability of the GMC increases, as the article is processed and then decreases very fast, when the object becomes inconsistent with the adjective. Fixating the CMC seems to be rather likely with the time until shortly after adjective offset. (Just before there seems to be a change in activation.)
2.3.5 Analyses of variance

The ANOVA on the first fixation to the target revealed slightly significant differences in the means of the conditions ($F(3, 57) = 7.17, p < .1$). Therefore, one-tailed $t$-tests on this data were conducted to clarify in which conditions the discrepancies appeared. They provided evidence for the assumption that it took more time to fixate the target in condition 1 than in any other condition (comparing condition 1 and 2: $t(19) = 2.33, p > .05$; condition 1 and 3: $t(19) = 5.22, p < .0001$; and condition 1 and 4: $t(19) = 3.13, p < .005$). In addition, the goal object was fixated faster in condition 3 than in condition 2 ($t(19) = 2.93, p < .005$), but other $t$-tests were not significant. The second ANOVA on the final target fixation did not expose important results, too ($F(3, 57) = 4.33, p > .1$).
3 Discussion

First of all, it must be mentioned that the equipment and materials used were not optimally compatible. Working with real-time mode under Windows 2000, a function provided for the EyeLink I eye-tracker to prevent unpredictable delays in the experiment, has a negative effect on playing sounds (Programming EyeLink Experiments in Windows, Version 2.1, 2002). Therefore, the time critical click-measurement was rather excluded and the results presented have to be regarded cautiously, but nevertheless, the direction they point to may be of major importance. In addition, no corpus analysis with respect to the names used in the experiment was done in advance. As Dahan, Magnuson, & Tanenhaus (2001) pointed out, names of higher frequency are more likely to be fixated, though fixations are shorter. Consequently, the choice of pictures might unwillingly have had an influence on the eye-movements of the participants. Besides, the number of participants appears to be very small for most experiments, but is quite common for eye-tracking tests. Despite these problems, the results were thoroughly presented above and will be discussed here. They must be understood more as results of a test experiment and could not be used scientifically on their own. They can be useful, though, as a guideline and prediction of what results an experiment with more participants, a better elaborated design etc. would be likely to produce.

Previous eye-tracking studies have shown that competitor objects sharing initial phonemes or rhyming with the target are fixated more than unrelated distractors (e.g., Tanenhaus et al., 1995; Allopenna et al., 1998). Eye-movements to a certain picture were usually interpreted as evidence for activation of the word matching this object in this respect. Dahan et al. (2000) found that preceding gender information in French can constrain this lexical activation, though this by itself does not restrict the set of candidate words initially considered for recognition. The present eye-tracking study examined
whether this morphosyntactic context might be sufficient to reduce the candidates to gender matching objects. The results show that GMC, CMC, and GCMC compete for lexical activation, because participants were more likely to initiate an eye-movement to one of them than to unrelated distractors in each of the four conditions. Consequently, potential lexical candidates do not necessarily need to be similar to a spoken word in phonological and semantic features, but can also be found among words sharing syntactic information. Moreover, differences among the three competitors are discovered. Gender-matching competitors are the earliest activated and stay so, if their adjectival information does not disagree with the spoken input. This contradicts with the findings of Dahan et al. (2000), which raises the idea that grammatical gender might not fulfill exactly the same functions in the same goodness in every language. Color-matching competitors on the other hand are activated later on, when this rather semantic information is processed. Even if the latter is not consistent with the syntactic gender marking up to that point, the consistent objects are considered to be potential referents. For this reason, it might have been advantageous to a certain extent to use fairly vague adjectives, such as *schön* [pretty] and *groß* [tall], etc. But all in all, these results show nicely that people establish reference incrementally, as previously publicized by Eberhard et al. (1995). Additionally, the results mirror the closely time-locked character of eye-movements to pictures related to referential expressions in the spoken language very well (as first shown by Cooper, 1974). Unfortunately, the results do not provide evidence for faster recognition of the target noun, when gender marking, in principal, is able to uniquely specify the target, or at least further adjectival information becomes incongruent with the goal object. Only comparisons of the first fixation to the target revealed that it is hardest to identify a target accompanied by an object sharing the syntactic and semantic features and that it becomes easier the more distinct the pictures are. Thus, grammatical information seems to be too vague and people wait for reliable, in this case even obvious details.
This experiment suggests a potential role for gender in narrowing down the search space even in early points during the spoken word recognition process, e.g. pre-lexical or lexical access. Facilitatory effects were not detected, however, but this is probably due to the investigation paradigm and the experimental design used. Inhibition is observed, though, as gender inconsistent objects were not activated as much as gender-matching ones.

Altogether, the results obtained here confirm the advantages of the eye-tracking paradigm when investigating linguistic processing and show that this method is well suited for this kind of research.
REFERENCES

Literature


Internet

