Neuter is not Common in Dutch:

Eye Movements reveal Asymmetrical Gender Processing

Hanneke Loerts, Martijn Wieling, and Monika S. Schmid

University of Groningen

Author Note

Hanneke Loerts, Department of Applied Linguistics / School for Behavioral and Cognitive Neurosciences (BCN), University of Groningen, The Netherlands;

Martijn Wieling, Department of Humanities Computing, University of Groningen, The Netherlands;

Monika S. Schmid, Department of English Language and Linguistics / Center for Language and Cognition Groningen (CLCG), University of Groningen, The Netherlands.

Correspondence concerning this article should be addressed to Hanneke Loerts, University of Groningen, Department of Applied Linguistics, P.O. Box 716, 9700 AS Groningen, The Netherlands. E-mail: h.loerts@med.umcg.nl

Keywords: visual world paradigm, spoken language comprehension, grammatical gender, lexical anticipation, mixed-effects modeling

Acknowledgements

The authors would like to thank Joyce Haisma for lending her voice to the present experiment. We are grateful to Michael K. Tanenhaus and his colleagues for helpful discussions of this work and useful advice concerning data analysis.

Abstract

Native speakers of languages with transparent gender systems can use gender cues to anticipate upcoming words. To examine whether this also holds true for a non-transparent two-way gender system, i.e. Dutch, eye movements were monitored as participants followed spoken instructions to click on one of four displayed items on a screen (e.g., *Klik op de*$_{COM}$ *rode appel*$_{COM}$, ‘Click on the$_{COM}$ red apple$_{COM}$’). The items contained the target, a colour- and/or gender-matching competitor, and two unrelated distractors. A mixed-effects regression analysis revealed that the presence of a colour-matching and/or gender-matching competitor significantly slowed the process of finding the target. The gender effect, however, was only observed for common nouns, reflecting the fact that neuter gender-marking cannot disambiguate as all Dutch nouns become neuter when used as diminutives. The gender effect for common nouns occurred before noun onset, suggesting that gender information is, at least partially, activated automatically before encountering the noun.

Neuter is not Common in Dutch:

Eye Movements reveal Asymmetrical Gender Processing

While listening to language, people are not passively interpreting each word separately. Instead, our brain is actively trying to predict what the next word will be. People are therefore faster in responding to words that are predictable on the basis of semantic context (e.g., Stanovich & West, 1983) and can decide what word will be pronounced after only hearing the first few phonemes (Grosjean, 1980). The speed with which certain (non-)linguistic cues in the speech stream are used to anticipate and hence facilitate language comprehension has increasingly been investigated by following people's eye movements while they receive auditory input. These eye tracking experiments are based on the initial finding by Cooper (1974), who showed that while people are listening to short narratives, they strongly tend to look at those objects in the visual field that are most closely related to the words they hear in the speech stream. Over two decades after this discovery, Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy (1995) re-introduced and further developed Cooper's (1974) methodology into what is now known as the 'visual world paradigm'. By measuring participants' eye movements while they are following instructions to manipulate objects, the studies by Tanenhaus and his colleagues revealed a real-time interdependence between spoken input and the surrounding visual world. If, for example, subjects were instructed to '*pick up the can...*' they fixated equally often on a picture of a candle and a picture of a candy (Allopenna, Magnuson, & Tanenhaus, 1998). This effect illustrates the incremental nature of language comprehension: the onset of a word activates a cohort, a relevant set of lexical candidates that compete for recognition until the uniqueness point of the word is heard (Marslen-Wilson, 1987).

More recent models argue that competition occurs on the basis of all acoustic input (e.g., the TRACE model (McClelland & Elman, 1986), Shortlist (Norris, 1994), and the neighbourhood activation model (NAM; Luce & Pisoni, 1998)). Ambiguities in the speech stream may therefore not only concern phonological overlap between words, but items may also compete due to ambiguous morphological cues in the input. Accordingly, gender-marked items preceding the noun might facilitate access by reducing the set of possible lexical candidates to gender-matching nouns. This hypothesis was first tested in French by Dahan, Swingley, Tanenhaus, and Magnuson (2000). Their participants were instructed to click on one of four displayed pictures (e.g., *Cliquez sur le bouton*, 'Click on the$_{\text{MASC}}$ button$_{\text{MASC}}$'). Every screen of four pictures contained a target, a cohort competitor (i.e. a word sharing its initial phonemes with the target, e.g., *bouteille*, 'bottle$_{\text{FEM}}$'), and two unrelated distractors. When the auditory input contained no gender-marking cues, as in the baseline condition containing French plural '*les*', the same cohort competitor effect as reported by Allopenna et al. (1998) was found: people looked equally often at a picture of a bottle and a picture of a button upon hearing '*cliquez sur les bou...*'. Interestingly, this cohort effect disappeared when singular gender-marked determiners preceded the noun in the speech stream. Upon hearing '*cliquez sur le*$_{\text{MASC}}$ *bou…*', natives never fixated on the picture of a feminine referent (*la*$_{\text{FEM}}$ *bouteille*) more often than on the unrelated distractors, indicating that gender cues facilitate language processing by minimizing the set of possible lexical candidates. Dahan et al. (2000) also found that, by itself, a gender-marked article does not seem to prime all gender-matching nouns. Upon hearing the feminine article, e.g., in *la*$_{\text{FEM}}$ *louche* ('the$_{\text{FEM}}$ ladle'), participants did not fixate more on a picture of a gender-matching noun that did not overlap in phonemic onset (e.g., a sock: *chaussette*$_{\text{FEM}}$). These additional results suggest that gender information, as present on the definite determiner preceding the noun, might be used post-lexically, i.e.

requiring the output of the lexical processing system for its operation. The gender effect could then be defined as an inhibitory rather than a facilitatory effect.

The absence of a gender effect in Dahan et al. (2000) with regard to phonemically non-overlapping article-noun combinations might be a matter of locality (Paris, Weber, and Crocker, 2006). Subjects may have had too little time to pre-activate and in turn use gender information on the article, because the short French articles are immediately followed by the target nouns. To increase the time people have to possibly pre-activate gender congruent nouns, Paris et al. (2006) studied the phenomenon in German which allows adjectives to intervene between determiner and noun. Confirming their hypothesis, they found that participants began fixating targets and gender-matching pictures significantly more than gender-mismatching pictures immediately after having heard the gender-marked article in a sentence such as e.g., *Wo befindet sich die*$_{FEM}$ *schwere*$_{FEM}$ *Melone*$_{FEM}$ (‘Where is the$_{\mathrm{FEM}}$ heavy$_{\mathrm{FEM}}$ melon$_{\mathrm{FEM}}$?’). This effect only decreased when subjects heard the onset of the target noun (which was unique to the target). The authors do mention that the effect they found is smaller than the effect found with phonemically overlapping nouns (Dahan et al., 2000), but the presence of a gender effect with non-overlapping nouns suggests that gender cues on articles influence referential processes before encountering the target noun. This result provides at least some support for a pre-activation account in which gender cues facilitate language comprehension in a more automatic fashion.

Until now, a gender effect (either facilitatory or inhibitory) has been found in eye tracking studies in German (e.g., Paris, Weber, & Crocker, 2006), French (Dahan et al., 2000), Russian (Sekerina, Brooks, & Kempe, 2006), and Spanish (Lew-Williams and Fernald, 2010). Gendered languages, however, differ greatly with respect to the amount and type of gender categories they have as well as the extent to which a noun's gender can be interpreted on the basis of phonological and/or morphological features (Corbett, 1991). Most researchers believe that the assignment of nouns to a particular gender category is a rather arbitrary process and in most instances, e.g., for inanimate objects, there is no clear relationship between the noun's gender and any male or female characteristics of its referent in the real world. In languages with an overt gender system the noun's gender is reflected in the morphology of its agreement targets as well as in the form of the noun itself. In many Romance languages, gender has a phonological correlate. For example, Spanish nouns ending with the suffix '*–a*' are mostly feminine ('*la senora*': madam), whereas nouns ending in '*–o*' are usually masculine ('*el marido*': husband), although there are exceptions to this principle. Other languages, such as French and German, contain more abstract, morphological cues that help to determine the gender of a noun. For example, French nouns that end with the suffix '*-ion'* generally receive the feminine gender (Foucart, 2008) and German nouns are assigned feminine gender when ending in the feminine suffix '*-keit*' (Zubin & Köpcke, 1984).

In a largely covert gender system, such as Dutch, gender has to be marked on the agreement targets, but is usually not obvious from the shape or meaning of the noun (Haeseryn et al., 1997). Dutch has two genders, common and neuter. Common nouns receive the definite article '*de*' while neuter nouns (as well as diminutives) take '*het*'. Attributive adjectives always receive a schwa-ending, except when modifying a neuter noun in an indefinite noun phrase (NP; see also Table 1). The special status of the neuter gender in indefinite NPs is accompanied by an asymmetrical distribution of the two gender categories. Common gender evolved from the collapse of the original masculine and feminine genders and consequently comprises around 75% of all Dutch nouns (Van Berkum, 1996). This overrepresentation of nouns taking '*de*' is further reinforced by the fact that '*de*' is also the definite plural article, irrespective of the gender of the noun. The relatively scarce evidence in the input for neuter gender causes monolinguals to overgeneralize the definite, common determiner '*de*' until the age of 6 (Blom, Polisenskà, & Weerman, 2008; Van der Velde, 2004). This is rather late in comparison to French native children, who have been reported to have acquired their gender system at the age of 4 (Van der Velde, 2004) as well as German monolinguals, who appear to master their system by age 3 (Müller, 1990; Szagun, Stumper, Sondag, & Franik, 2007).

[Insert Table 1 here]

The Dutch gender system with its asymmetrical distribution of gender classes and agreement markers thus largely has to be learned on an item-by-item basis. An interesting question, therefore, is if gender-marking in such an asymmetrical and non-transparent gender system is also used to facilitate and/or inhibit spoken word recognition?

Brouwer, Unsworth, and Mak (2010) recently investigated this question using the eye tracking methodology. They recorded eye movements from monolingual Dutch natives (both children and adults) while they looked at two pictures on a computer screen. The two nouns on the screen always shared the same colour, but the competitor either had a different gender or shared the gender with the target. Participants listened to speech stimuli containing gender-marked articles such as '*Can you see the*$_{NEU}$ *yellow house*$_{NEU}$*?*'. As a baseline, the non-gender-marked possessive adjective '*mijn*' (my) preceded the adjective and the noun. The authors reported that adults are marginally sensitive to gender-marking, as revealed by earlier detection of the target picture when the item was preceded by a gender-marked article as opposed to the possessive '*mijn*'. It should be pointed out that a potentially confounding factor in this study is the fact that the non-gender-marked adjective '*mijn*' might be a problematic baseline as it is somewhat longer than both gender-marked determiners (*'de'* or *'het'*). Consequently, the earlier detection of the target picture in the gender-marked conditions might, at least to some extent, be caused by the shorter duration of these determiners as opposed to the possessive '*mijn*'. The present study improves on Brouwer's approach by using similar durations throughout the auditory stimuli and therefore attempts to investigate more clearly whether, despite the non-transparency of the system, Dutch natives are able to use prenominal gender cues during language comprehension to pre-select upcoming nouns. The setup used was similar to the visual world paradigm of Tanenhaus et al. (1995), asking subjects to manipulate the objects in their visual scene. The nouns depicted varied along two dimensions: syntactic gender and colour. Reasoning in line with Paris et al. (2006), the Dutch language also allows for the placement of adjectives in between articles and nouns, possibly reducing the high co-occurrence of definite articles and nouns and allowing more time for gender information to be processed.

If gender-marking on an item immediately preceding the noun does not allow the listener enough time to process and use the information, then a possible gender effect is only expected in the definite condition, i.e. the condition in which gender-marking occurs on the determiner that is followed by an intervening adjective allowing for more processing time. This definite condition was directly compared to the indefinite 'baseline' condition, in which gender information only becomes apparent at the offset of the adjective and might thus not allow the listener enough time to process and use the information. The present study used only phonemically non-overlapping nouns in order to examine the gender effect in more detail.

METHODS

Participants

Twenty-eight adult native speakers of Dutch (mean age: 48, range: 34-60; 16 female) were recruited around the city of Groningen by means of advertisements in a local newspaper. All participants filled out a questionnaire to ensure that they had normal or corrected-to-normal vision, no hearing deficits and no form of colour blindness.

Visual Stimuli

A set of 48 nouns (24 common and 24 neuter nouns) was selected based on several criteria. First of all, the nouns chosen had to refer to objects that could be depicted using images. Only nouns that could be extended by an adjective referring to the colour of the object depicted were used. As Dahan, Magnuson, & Tanenhaus (2001) have pointed out, referents with higher frequency names are more likely to be fixated on, though the fixations tend to be shorter. Therefore, only high frequency nouns were selected from the frequency list of the *Corpus Gesproken Nederlands* (corpus of spoken Dutch), containing nouns that belong to the 5000 most frequently occurring words (CGN 5000). Despite this, there were still relatively large differences in lemma frequency of the nouns in our stimuli set, ranging from 7 to 3203 (mean lemma frequency of 323.2). To correct for the skewness of this distribution, lemma frequency was log-transformed and included in the mixed-effects regression model (see below).

The visual stimuli were derived from the standardized picture set of black and white line drawings (Snodgrass and Vanderwart, 1980), but rendered in one of five different colours: red, blue, green, yellow, and brown. Initially, the goal was to use only primary and/or notable colours. Due to the large number of criteria, however, images that could not be depicted in a primary or secondary colour (such as e.g., a dog) were used and hence coloured brown.

Auditory Stimuli

As gender information in Dutch NPs either appears on the determiner in definite NPs (sentence 1a, below) or on the adjective in indefinite NPs (sentence 1b, below), both definite and indefinite constructions were used. Colour adjectives were placed between the determiner and the noun to reduce possible effects of high co-occurrence probabilities of determiners and nouns as well as to allow more time for gender information to be processed in the definite constructions, see 1a below (Dahan et al., 2000).

1a ‘Klik op de$_{COM}$ rode appel’

Click on the$_{COM}$ red apple

1b ‘Klik op het plaatje met een rode$_{COM}$ appel’

Click on the picture with a red$_{COM}$ apple

The auditory instructions were digitally recorded in a soundproof booth with a sampling rate of 44.1 kHz by a female native speaker of Dutch with normal intonation. A speech editor (Adobe Audition 3.0) was used to slightly adjust the auditory stimuli in such a way that all definite determiners started around 800 ms and all indefinite determiners around 2000 ms after the onset of the sentence. The mean duration of the determiners was 214 ms in the definite condition and 188 ms in the indefinite condition. On average, adjectives following a definite determiner lasted for about 532 ms and adjectives in the indefinite condition lasted about 493 ms. The nouns also lasted approximately equally long across conditions with 669 ms and 681 ms for definite and indefinite sentences, respectively. The relatively constant timing delay of 1200 ms for the indefinite NPs as compared to the definite constructions allowed for a direct comparison of the responses to these different types of sentences. Possible effects of these slight timing differences between the two conditions, however, were controlled for by including these timing variables in the mixed-effect regression analysis.

[Insert Table 2 here]

Table 2 shows the four conditions in the present experiment. Each trial consisted of a screen divided into four quadrants in which four pictures appeared: the target, a competitor matching in colour and/or gender with the target, and two distractors that never matched in gender or in colour with the target. The initial segments of the nouns of the four pictures never overlapped to exclude any phonological competition effects. Positions of the target, competitor and distractor objects on the screen were counterbalanced. In addition, each item occurred on the screen once as a target, once as a competitor, and twice in the distractor role. Following a Latin Square design, four different lists were created in such a way that each participant was presented only one version of each item in a certain condition.

Two blocks were created: one block containing instructions with definite NPs, and one block containing instructions with indefinite NPs. Each block consisted of 48 experimental plus 22 filler trials.

Procedure

Eye movements were recorded in the Eye Lab of the University of Groningen using a remote Tobii T120 eye tracker with a sampling rate of 60 Hz. Auditory and visual stimuli were presented using E-Prime (Schneider, Eschman, & Zuccolotto, 2002) and reaction-times as well as eye movement data were recorded using the E-Prime Extensions for Tobii software.

Participants were seated in a comfortable chair at approximately 60 cm from the eye tracking monitor. After a standard 5-point calibration procedure, participants were instructed to click (using the mouse) on the correct picture as quickly as possible. The experiment began after a short practice session containing 6 filler items.

Each trial started with a fixation cross at the center of the screen. To ascertain that the participants' eyes were focused at the center of the screen at the beginning of each trial, subjects had to fixate on the fixation cross for at least 1000 ms in order for the trial to start. Auditory instructions started as soon as the display of four pictures was presented on the screen. Immediately after clicking on a picture, the set of pictures disappeared and the next trial started.

Each subject completed two blocks: one containing instructions with definite and one with indefinite NPs. Half of the subjects received the definite instructions first and the other half received the indefinite instructions first. Each block lasted around 12 to 13 minutes and subjects were allowed a short break between the two blocks.

Statistical Analyses

In most eye-tracking studies, the dependent variable is the proportion of fixations towards the target. At present, the vast majority of studies analyses this data using subject and item ANOVAs or t-tests in order to obtain generalizable results. This approach *incorrectly* follows Clark (1973) who proposed a min-F measure based on F1 and F2 to generalize over subjects and items. As min-F is not equal to conducting a separate F1 and F2 analysis, many studies may report spurious significance (type-I errors). Additionally, it is not always necessary to take F2 into account, especially when the research design is counterbalanced (Raaijmakers, Schrijnemakers, & Gremmen, 1999). Using ANOVA to analyze categorical data (or proportional data, which is bounded by 0 and 1) is even more problematic as it violates the assumptions necessary to use an ANOVA and can lead to spurious significance (type-I errors) and null results (type-II errors; Jaeger, 2008). More suitable analyses have therefore been proposed, including growth curve analysis (Magnuson, Dixon, Tanenhaus, & Aslin, 2007) and mixed-effects (logistic) regression (Barr, 2008; Huettig, Rommers, & Meyer, 2011).

In this study we will apply mixed-effects regression modeling (for introductions, see, e.g., Baayen, 2008, Ch. 7 and Baayen, Davidson and Bates, 2008). In mixed-effects regression modeling fixed-effect and random-effect factors are distinguished. Fixed-effect factors have a small number of levels that exhaust all possible levels (e.g., the gender of a participant is either male or female). In contrast, random-effect factors have levels sampled from a much larger population of possible levels.

In our data, there are two random-effect factors that are likely to introduce systematic variation, namely participant and item (which has a one-to-one correspondence with a certain scene). Our observations (i.e. measurements) are specific to each participant. As these participants are a sample of a much larger set of possible participants which might have participated in our study, participant is a random effect factor. Similarly, since we could have selected different items, item is also a random effect factor.

By including random-effect factors, the structural variability associated with these factors can be taken into account. For example, some participants might focus faster on the target item than others. These adjustments to the population intercepts (‘random intercepts’) can be used to make the regression formula (for each participant and item) as precise as possible.

Similarly, it is possible that there is variability (associated with the random-effect factors) in the effect a certain predictor has. For example, the effect (i.e. slope) of participant age might vary per item, indicating that some items might be more easily recognized by older participants than younger participants, while other items may show the opposite pattern. Together with the random intercepts, these random slopes make the regression formula as precise as possible (for each participant and item). The necessity of including these random intercepts and random slopes is verified with likelihood ratio tests evaluating whether the increase in the number of parameters is justified given the increase in goodness of fit.

The analyses used in the present paper considered such random-effect factors, as well as a contrast to distinguish the presence or absence of a colour competitor and a gender competitor in the scene (i.e. the contrast we are most interested in). In addition, we included item-related characteristics such as the gender and word frequency of the item, as well as participant-related characteristics such as gender and age. To control for a possibly imbalanced design, factors such as the position and colour of a specific item were also included in the analyses. Potential effects of fatigue and/or learning during the course of the experiment were accounted for by including predictors such as trial-number as well as dependent measures corresponding to the previous trial and testing their significance in the model.

We evaluated the significance of fixed-effect factors by means of the usual *t*-test for the coefficients. Since our data set contains a large number of observations (about one thousand items), the *t*-distribution approximates the standard normal distribution and factors are significant ($p < 0.05$) when they have an absolute value of the *t*-statistic exceeding 2 (Baayen et al., 2008).

As we had two dependent variables, reaction time (henceforth RT) and eye movement data (henceforth EYE), we created two separate mixed-effects regression models. The models were fitted using a stepwise variable deletion procedure: predictors that did not contribute significantly to the model were removed. Potential nonlinear effects of each of the continuous predictor variables were also assessed. As collinearity can be a problem even in rather balanced designs, predictors were centred before analyses (Jaeger, 2008).

RESULTS

Of all 2688 item responses (28 subjects x 48 items x 2 blocks), 2677 were answered correctly (99.6%) and only 11 items (0.4%) were answered incorrectly and therefore removed from further analysis. For 6 additional items, responses had only been given later than 1.5 seconds after the onset of the noun. These 6 extreme outliers (0.2%) were also removed. In addition to minimal a-priori data trimming, potentially influential outliers were removed after fitting a first version of the model (i.e. model criticism, see Baayen & Milin, 2010). In this study, linear mixed-effects models were fit using the *lmer* function of the lme4 package (Bates, 2005) implemented in R (version 2.11.1: The R Foundation for Statistical Computing, 2010).

The mean timing delay of 1200 ms in the indefinite constructions (compared to the definite constructions) was subtracted from the timing data in order to be able to directly compare responses from both conditions.

Behavioural Data (Reaction Times)

Out of the 2671 responses of the initial data set, absolute standardized residuals exceeding 3 standard deviations were removed (0.5%). A linear mixed-effects model was fitted to the log-transformed RT data using a stepwise variable deletion procedure. The best fitting model including all necessary fixed and random effect factors (shown in Tables 3 and 4) explained approximately 67% of the variance in the RT data and the residuals of the model followed a normal distribution.

Table 3 shows the coefficients and the associated *t*-values of the fixed-effect factors. The random-effects structure is outlined in Table 4. Table 3 reveals an intercept of about 957 ms (6.8640 in logarithmic scale), which can be seen as the base reaction time for a given item. The coefficients of the remaining predictors or fixed factors reveal whether this base value increases or decreases when a specific variable is present.

The main question addressed in the present experiment was whether gender-marking influences the comprehension process by facilitation of gender-matching and/or inhibition of gender-mismatching nouns. As shown in Table 3, the presence of a competitor that shared gender with the target (a *SameGenderCompetitor*) did not by itself significantly impact reaction times ($\beta = -0.0459$, $t = -1.78$). However, when looking at the interaction with age, we observed that older people had a stronger negative impact of the presence of a *SameGenderCompetitor* than younger people ($\beta = 0.0012$, $t = 2.31$, $p < 0.05$, two-tailed test). In other words, a *SameGenderCompetitor* only increased RTs in the older participants. As consistently found across studies (see. e.g., Der & Deary, 2006), older participants were also generally slower in choosing the correct target picture ($\beta = 0.0050$, $t = 3.51$, $p < 0.01$).

The second interaction (*SameColourComp:SameGenderComp*) revealed that the effect of colour, i.e. significantly slower RTs when the competitor shared colour with the target (β = 0.1536, $t = 19.48$, $p < 0.001$), was less strong when the competitor also shared gender with the target (β = 0.1536 - 0.0152 = -.1384, $t = -2.17$, $p < 0.05$). This counterintuitive result suggests that a gender competitor speeds up the visual decision making process.

In addition to the variables of interest, we tried to account for potential fatigue and/or learning effects and tested for both the effect of trial (*Sequence*) and effects of RTs on the previous trial (*PreviousRT*). An overall effect of *Sequence* was found (β = -0.0006, $t = -3.85$, $p < 0.001$), showing that participants' RT generally decreased during the course of the experiment. In addition to this learning effect, participants responded more slowly when they had a slower RT in response to the previous trial (β = 0.0744, $t = 4.92$, $p < 0.001$). By adding this control variable, temporal dependencies are not only controlled for, but inclusion also allows for a more precise estimation of the effects of the other predictors (Baayen & Milin, 2010). A second control variable, *LemmaFrequency,* also reached significance, showing that higher frequency items were responded to significantly faster (β = -0.0071, $t = -2.46$, $p < 0.05$).

The analyses also included item as well as trial-related characteristics such as the colour of the item, its competitor and the position of these pictures on the screen. Inclusion of these predictors showed that participants were significantly slower when the target picture was coloured brown (β = 0.0297, $t = 2.80$, $p < 0.05$). In addition to the saliency of different colours, positioning of the pictures also influenced RTs: people were slower in clicking on the target when it was shown on the left side of the screen as opposed to the right (β = 0.0103, $t = 2.81$, $p < 0.05$). Although this latter result is not straight-forward to interpret, it might be related to the right-handedness of our subjects, and consequently their preference moving their hand from left to right. This might cause a delay in manual responses to objects at the left of the screen.

In order to match the targets, competitors, and distractors, the present design consisted of a set of 48 images of which each image appeared on the screen once as a target, once as a competitor, and twice as a distractor. The analyses revealed that a higher number of times a competitor had previously been on the screen in a different role, e.g., the role of distractor, significantly increased the time it took participants to select the correct referent ($\beta = 0.0046$, $t = 2.25$, $p < 0.05$, two-tailed test). This might constitute a recognition effect for the competitor, but the absence of such an effect for the amount of times the target had been on the screen reveals that this effect probably results from the tendency to pre-select the target picture, which is more likely to be a picture that has previously been on the screen in a different role, such as the role of distractor.

Table 4 shows the two random-effect factors (labelled as Groups) of the present experiment: *Item* and *Subject*, where *Item* refers to a specific picture or noun and *Subject* refers to the participant. The table lists the standard deviations for the adjustments to the intercept for both random effect factors and, for subject standard deviations are also listed for the adjustments to coefficients (i.e. slope) of 4 covariates: *SameColourCompetitor*, *TargetIsBrown*, *IsIndefinite*, and *Sequence*.

The population slope for *SameColourCompetitor* was 0.1536 for different gender trials and 0.1384 (0.1536 - 0.0152) for same gender trials. Figure 1 shows, however, that subjects not only differ with respect to the effect that a specific variable has on their RTs, but that these different variables also correlate with one another. Subjects with a more positive slope for

SameColourCompetitor tend to be the fast responders, while slow subjects have a more negative slope for *SameColourCompetitor*. In other words, faster participants suffered more from a competitor that shared colour with the target resulting in a relatively large increase in RTs as compared to a relatively small increase in RTs for the overall slower participants.

[Insert Figure 1 here]

Similarly, Figure 1 shows that the slower subjects suffered less from a brown target as compared to fast subjects. The coefficient of *TargetIsBrown* for the population is estimated to be 0.0297. Fast subjects, i.e. subjects with negative adjustments to the intercept, are characterized by upward adjustments for *TargetIsBrown* slopes. Furthermore, there is a positive correlation of the adjustments for *SameColourCompetitor* and *TargetIsBrown* (0.466), showing that subjects who suffer more from the presence of a competitor that shares colour with the target also suffer more when the target depicted is coloured brown.

In addition, by-subject random slopes for the effects of *IsIndefinite* and *Sequence* significantly improved model fit. In other words, some subjects were slower in response to indefinite constructions, while others were faster. Similarly, the learning effect that was found differed significantly among subjects, with some of them showing a large effect of *Sequence* while others only show minor or no learning effects across the duration of the experiment.

Eye Movement Data

Since all relevant and distinctive information that might be used to pre-select the target picture (i.e. gender-marking and colour information) occurs between the onset of the determiner and the onset of the noun, this time period was used to analyze eye movements (taking into account the 200 ms it takes to plan and launch an eye movement).

[Insert Figure 2 here]

For each item and each time point, the gaze location was calculated as being either on the Target, on the Competitor, on the Distractors, or not on an Area of Interest. For 8% of the data, the Tobii system reported to be fairly confident that the data was incorrect or corrupted (validity codes > 2). In order to retain as much data as possible, the gaze location for these specific time points was set to NA. In addition, for each individual item and subject combination an error percentage was calculated. Items for which the validity codes of the eyes were higher than 2 for more than 30% of the time were removed (9.1%). For the resulting data set, mean proportions of fixations were calculated across the entire time period (i.e. from determiner onset until noun onset plus 200 ms) per subject and per item.

It would be expected that, if participants use gender-marking cues and colour information to pre-select the correct picture, this would result in more fixations towards the target as compared to the competitor when the information is disambiguating. Therefore, the dependent variable *TargetRatio* was calculated as the percentage of looks towards the target divided by the total percentage of looks towards the target and its competitor. To avoid problems associated with performing statistical analyses on proportion data, mixed-effects models were fit to the empirical logit transformations of the variable *TargetRatio* (Barr, 2008; Jaeger, 2008).

A linear mixed-effects regression model was fitted to the data using a stepwise variable deletion procedure. Absolute standardized residuals exceeding 3 standard deviations were removed (1.3%) after which the model was refitted. The best model including all fixed and random effect factors (shown in Tables 5 and 6) explains approximately 32% of the variance in looking behaviour and the residuals of the model followed a normal distribution. Before explaining the random effect structure of this model, the fixed-effects will be explained in detail.

As opposed to the reaction time data, the eye movement data do reveal a difficulty in distinguishing target and competitor when both referents shared gender. The negative coefficient for *SameGenderCompetitor* in Table 5 reveals that participants looked relatively more towards the competitor when it shared gender with the target ($\beta = -0.0484$, $t = -2.22$, $p < 0.05$). In addition to this general effect of gender, the analysis also revealed a difference between common and neuter nouns: when the *TargetIsNeuter*, i.e. when it is preceded by '*het*' in the definite or a bare adjective in the indefinite condition, people took significantly longer to fixate on the target ($\beta = -0.0856$, $t = -3.37$, $p < 0.01$). Interestingly, these two predictors interacted in that the observed gender effect was present in common nouns, but reversed when both the target and the competitor were neuter ($\beta = 0.0839$, $t = 2.67$, $p < 0.05$). This difference between common and neuter nouns is also visible in Figure 2, where the upper panel clearly reveals an earlier selection of the target picture when it has unique gender (upper-right picture) as compared to same-gender trials. The lower panel, depicting eye movements in response to neuter nouns, shows that the opposite happened in response to neuter nouns: people tended to pre-select the target when target and competitor shared gender, while relatively more looks towards the competitor occur when the gender of the target is unique. The absence of a gender effect for neuter nouns can be directly linked to the fact that all Dutch nouns, regardless of the gender they carry, are preceded by the neuter determiner '*het*' when used as a diminutive. Consequently, hearing neuter gender-marking preceding the noun does not automatically mean that all referents to common nouns can be ignored as possible targets. The fact that the gender effect appears to be reversed for neuter nouns, i.e. the fact that a competitor sharing both colour and gender with the target makes it easier as compared to a competitor that only shares colour with the target, is very difficult to explain. Although Figure 2 does suggest that looks towards the competitor decrease earlier in the different-gender as compared to the same-gender trials, there is an obvious preference for the competitor in the different-gender and a preference for the target in the same-gender trials.

Table 5 reveals the same substantial effect of *SameColourCompetitor* that was visible in the reaction time data ($\beta = -0.3138$, $t = -15.61$, $p < 0.001$). In this case, the presence of a competitor with the same colour as the target picture causes a significant decrease in *TargetRatio*, i.e. people fixated less on the target and relatively more on the competitor when there is a S*ameColourCompetitor* present. Reaction times already showed that people also suffered when the target picture was coloured brown: participants were significantly slower in clicking on such target pictures. The negative effect of *TargetIsBrown* is replicated in the eye data with people looking significantly less towards a brown target as compared to the competitor ($\beta = -0.0836$, $t = -3.97$, $p < 0.001$). Furthermore, looks to the target significantly increased when the target was presented in one of the two upper quadrants of the screen ($\beta = 0.0666$, $t = 2.65$, $p < 0.05$, two-tailed test). When the competitor was placed in one of the right corners of the screen, (i.e. *CompIsRight),* looks towards the Competitor also increased relative to the looks towards the Target ($\beta = -0.0396$, $t = -2.49$, $p < 0.05$, two-tailed test). Similar to the RT data, *Age* significantly affected the dependent variable with older people taking significantly longer in choosing the correct target picture resulting in a lower *TargetRatio* ($\beta = -0.0045$, $t = -3.23$, $p < 0.01$, two-tailed test). The negative coefficient for *IsMale* shows that male participants generally showed a lower score of the dependent variable *TargetRatio*, reflecting a more equal distribution of looks towards target competitor as compared to female participants (β = -0.0558, t = -2.86, $p < 0.05$, two-tailed test).

The auditory stimuli were recorded and manipulated in such a way that the determiner always started around 800 ms after the onset. Minor differences in timing between the items, however, could not be further reduced and the timing details were included in the model. Table 5 reveals that an increase in *TimeOnsetNoun*, i.e. the time (in ms at which the noun started in the speech stream), significantly reduced the *TargetRatio* (β = -0.0010, t = -6.67, $p < 0.001$). A later *TimeOnsetDeterminer*, however, increased *TargetRatio* (β = -0.0016, t = 6.36, $p < 0.001$). These results suggest that the later the participant heard and could process the determiner, the more the target was focused on relative to the competitor before the onset of the noun. It should be noted that the variable *TargetRatio* was calculated per item and per subject. The slight difference between the timing of the onset of the determiner and the noun thus caused a difference in the length of the time-window over which *TargetRatio* was calculated. This length was taken into account in the calculation of the empirical logit transformation of *TargetRatio.* The lower *TargetRatio* values in response to a later *TimeOnsetNoun* thus most likely reflect the fact that the information on the determiner and/or adjective had not yet been adequately processed for these items. The higher *TargetRatio* in response to items with a later *TimeOnsetDeterminer* is more difficult to explain, but might reflect that people have had more time to scan the pictures, which might help in detecting the target picture more efficiently.

Potential effects of fatigue and/or learning were again accounted for by testing both the effect of *Sequence* and the effects of the *TargetRatio* for the previous trial. *Sequence* did not affect the output measure, but subjects were found to have a higher *TargetRatio* when they had a higher *TargetRatio* on the previous trial (β = 0.0361, t = 1.83, $p < 0.05$, one-tailed test).

In contrast to the RT data, the eye movement data did not require a random intercept for item, which was confirmed by likelihood ratio tests. This is probably related to the time-window that was analysed, from determiner onset to noun onset, in which the noun's referent has not yet been mentioned in the speech stream. In relation to this, *LemmaFrequency* had no effect on the mean proportions of fixation in the time-window analysed. As shown in Table 6, our model did require a random intercept for *Subject*, revealing again that some people have higher *TargetRatio* scores while other people have lower *TargetRatio* scores. Similar to the RT data, people tend to be differentially affected by the presence of a *SameColourCompetitor*, with some people suffering more than others from this predictor. Whether the target appeared in the upper or in the lower panel of the screen also impacted people's eye movements differently and inclusion of by-subject random slopes for these effects significantly improved model fit.

DISCUSSION

Previous eye-tracking experiments have shown that competition between words during language comprehension does not only occur on the basis of phonological information (e.g., Allopenna et al., 1998), but also on the basis of morphological cues such as gender-marking (e.g. Dahan et al., 2000; Paris et al., 2006). The types and amount of gender categories as well as their impact on agreement targets, however, varies between languages. The Dutch gender system, as opposed to many Romance languages for example, is relatively covert in that a noun's gender is rarely predictable on the basis of phonological and/or morphological cues in the input (Haeseryn et al., 1997). The two-way Dutch gender system is characterized by its asymmetrical distribution with about 75% of all Dutch nouns having common gender (Van

Berkum, 1996). The distinct status of neuter gender is enhanced by the fact that all words use the common determiner ‘*de*’ in the plural. On the other hand, every Dutch noun can become neuter when used in the diminutive form. The present study investigated whether gender-marking in such an asymmetrical and non-transparent gender system is used by Dutch natives during language processing to anticipate gender-congruent and/or inhibit gender-incongruent nouns.

Eye movements were recorded while people listened to sentences such as *Klik op de*$_{COM}$ *rode appel*$_{COM}$ (‘Click on the$_{COM}$ red apple$_{COM}$’) while looking at a screen containing the target, a competitor matching in colour and/or gender with the target, and two distractors that never matched in gender nor colour with the target. We hypothesized that, if people use prenominal gender-marking to pre-select the correct referent, this should become visible before hearing the onset of the noun (which was unique to the target), resulting in a relatively higher amount of fixations towards the target as compared to the competitor when the information in the relevant time-window contained disambiguating colour and/or gender information. Mixed-effect models were fitted to both the (log-transformed) reaction time (RT) data and the relative difference in fixation proportions between target and competitor in the relevant time-window, i.e. the time-window from determiner onset to noun onset in which both gender and colour information was presented.

A gender effect was found only in the eye data and, more importantly, only common gender-marking appears to affect processing of the subsequent noun by inhibiting gender-incongruent and/or anticipating gender-congruent nouns. This result is an immediate reflection of the special status of neuter gender in Dutch: all Dutch nouns can become neuter when used in the diminutive form. A common noun can thus still become a neuter noun and consequently, neuter gender-marking cues cannot disambiguate between common and neuter nouns. While some studies suggest that gender-marking on the article does not, by itself, seem to restrict the possible set of lexical candidates to gender-congruent nouns (Dahan et al., 2000), others suggest that gender cues facilitate processing immediately after hearing the gender-marked article (Paris et al., 2006). Reasoning in line with Paris et al. (2006), we used intervening adjectives in the present study to allow the participants more time to process gender cues. More importantly, since Dutch gender-marking either appears only on the determiner (definite NPs) or on the adjective (indefinite NPs), the present experiment investigated both constructions in order to disentangle this issue of locality. The fact that no difference was found between the two conditions suggests that the effect of gender is present in and similar for both constructions. In other words, gender-marking on the adjective (as in indefinite NPs) occurs immediately before the noun and the presence of a gender effect in these conditions suggests that Dutch natives do have enough time to process gender information and hence pre-select the correct referent on the screen based on the gender-marking cues present in the auditory input.

The results reported by Paris et al. (2006) were replicated in that a gender effect was found before the onset of the noun, which was unique to the target. The fixation proportions also reveal, however, that this effect occurs at the offset of the adjective as opposed to immediately after hearing the determiner. This might be the result of colour being an extremely salient feature. Both the RT data as well as the eye movement data revealed a large effect of colour: participants waited until they had heard the (onset of the) adjective before choosing the correct target picture and the fixation proportions suggest that, even if there is a disambiguating determiner present in the input, people tend to look at both the competitor and the target equally often until right before hearing the unique initial phonemes of the noun.

While the eye movement data in response to common nouns show a gender effect, i.e. a same-gender competitor slows down the process of locating the target; the opposite pattern was found for neuter nouns, i.e. a same-gender competitor appears to speed up the process of locating the target as compared to a different-gender competitor. Although the data, as presented in Figure 2, do suggest that looks towards the competitor decrease earlier in the different-gender as compared to the same-gender trials, there is an obvious preference for the competitor in the different-gender and a preference for the target in the same-gender trials. Similarly, the RT data suggested that the effect of colour decreased when the competitor also shared gender with the target. This result initially seems rather counterintuitive, but another recent study on gender processing yielded similar results. Cubelli, Paolieri, Lotto, and Job (2011) asked their subjects to semantically categorize two objects that either matched or did not match in gender (in Italian). The time to categorize pictures decreased when the two objects were from the same grammatical gender category. The authors hypothesized that, during the categorization task, lexical representation associated with the objects are activated and nouns within the same gender category activate each other, facilitating processing and speeding up response times. Similar results have recently also been reported in 30-months old children when confronted with two pictures from different semantic categories. These children were reported to look more towards the target picture when the competitor shared gender with the target as compared to the different-gender trials (Bobb & Mani, 2012). These results, in combination with the gender effect for common nouns occurring before noun onset, implies that gender information of the objects on the screen is made available before encountering the gender-carrying noun itself and argues in favour of a pre-activation account of gender. The fact that this facilitation of same gendered items is especially visible in the eye movement data for neuter nouns is in line with the fact that neuter Dutch nouns are relatively scarce and make up about 25% of all nouns (Van Berkum, 1996). It is thus easier to activate neuter gender nouns simply because there are less to activate and hence facilitation within this gender category speeds up the recognition process.

The RT data also reveal that the effect of gender increases with age, which corresponds to the overall age effect that was found in both the RT and the eye data. As expected, item-related characteristics such as frequency only affected the RT data with faster reaction times in response to higher frequency items (Dahan et al., 2001). Similar frequency effects were not replicated in the eye fixations, which is not surprising when considering that we analysed eye movements in the time-window before noun onset.

By using mixed-effects regression analyses, the present experiment allowed to include variables in the model that are normally considered to be rather balanced. Although the present experiment also used a counter-balanced design, usage, placement and colouring of the pictures appeared to impact both reaction times and eye fixations. To create a balanced design, the present study presented every item picture once as a target, once as a competitor, and twice as a distractor. The analyses revealed that the number of times that a competitor had previously been on the screen in a different role significantly increased the relative looks towards this competitor. This effect was only found for pictures in the competitor role and not in the target role, suggesting that participants focus was primarily on their task to select the target picture as soon as possible. The target is more likely to be a picture that had previously been on the screen in a different role, such as the role of distractor, and hence there is a greater chance that this should be the picture in the competitor role.

Not only the number of times a picture was shown affected participants' behaviour, but also the colouring and positioning of these items on the screen. When the target was brown, people were slower in locating it with their eyes, but they were also slower in clicking on the target with the mouse. It has been known for a long time that colour plays an important role in capturing a person's attention and some colours appear to attract more attention than other colours, simply because they are more salient (Osberger & Rohaly, 2001). As a colour of low intensity, brown is a tertiary colour, which means that it is a mix of three subtractive primary colours. This causes brown objects to be less noticeable and hence more difficult to find and respond to. The RT data also revealed that subjects were slower in clicking on a target when it was positioned on the left side of the screen. We hypothesized that, since all subjects were right-handed, they tended to prefer moving their hand from left to right and hence cause a delay in manual responses to objects on the left of the screen. Similarly, the difference in fixations towards the target and competitor positively increased when the target was positioned in one of the upper quadrants of the screen. Contrary to this, the target was more difficult to find when its competitor occurred in one of the right quadrants on the screen. The latter effect is hard to explain, but the fact that upper pictures are more easily detected most probably arises from the fact that these are the first pictures the subjects fixate on. Most, participants tend to scan the screen from left to right and from the upper to the lower pane. Consequently, the concepts corresponding to the upper pictures have already been activated and hence eye movements are more easily directed towards these referents as compared to the referents in the lower pane. The fact that, as visible in the eye data, people tend to suffer from a competitor on the right side of the screen might be caused by them not having activated the referent of these competitors yet. This, however, remains speculative and it should be stressed that these predictors are not of main interest to the present study, but their inclusion in the model proves to improve our understanding of other, more important variables. The same holds for the timing variables that significantly affected the eye movement data.

Minor differences in timing between the onset of determiners and nouns in the auditory input differentially affected participants' eye movements. A later onset of the noun was related to more equally distributed fixations on target and competitor, while a later onset of the determiner tended to increase relative fixations on the target as compared to the competitor. Although these conflicting results are difficult to explain, we hypothesized that a later onset of the noun is more likely to relate to a later onset of the gender information and this information probably had not yet adequately been processed for these items. Relatively more fixations on the target in response to later onsets of the determiner might well reflect the increase in the amount of time participants had to scan the pictures prior to receiving disambiguating information, which might help in detecting the target picture more efficiently.

Mixed-effects modeling not only allows including multiple predictors that could not have been added in an ANOVA analysis, but it also allows for the inclusion of random intercepts for both subject and item and to vary the slopes of specific predictors across subjects or items. Both the RT data as well as the eye movement data showed that a competitor sharing colour with the target differentially affected participants, with some of them suffering more from a same coloured competitor than others. The RT data additionally showed that fast responders suffered more from a competitor with the same colour than slower subjects and these subjects, likewise, also suffered more from a target being brown. Mixed-effects regression models allow the researcher to acknowledge that people respond differently to certain variables.

The present experiment found a gender effect in eye movement data, but not in reaction time data, confirming the advantages of studying language processing by investigating eye movements. Given that the gender effect is present in a non-transparent language such as Dutch, argues against the idea that previously reported gender effects, such as in French (Dahan et al., 2000), were dependent on the onset of the noun reflecting a phonological or lexical effect. Instead, the present study suggests that gender information is, at least partially, activated automatically before encountering the noun. We therefore have provided at least some evidence for a more automatic, pre-activation account on gender processing.

REFERENCES

Allopenna, P. D., 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.

Baayen, R.H. (2008). Analyzing linguistic data: A practical introduction to statistics using R. Cambridge University Press.

Baayen, R.H., Davidson D., & Bates, D. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59,* 390-412.

Baayen, R.H. & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3,* 12-28.

Barr, D.J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, Special Issue: Emerging Data Analysis*, *59,* 457-474.

Bates, D. (2005). Fitting linear mixed models in R. *R News, 5*, 27–30.

Blom, E., Polisenska, D. & Weerman, F. (2008). Articles, adjectives and age of onset: the acquisition of Dutch grammatical gender. *Second Language Research*, *24*(3), 297-331.

Bobb, S.C. & Mani, N. (2012). *Categorizing with Gender: Does Implicit Grammatical Gender Affect Lexical Access in Young Children?* Talk presented at International Conference on Infant Studies (ICIS XVIII), Minneapolis, Minnesota, June 7-9.

Brouwer, S., Unsworth, S., & Mak, P. (2010). *Processing grammatical gender in Dutch.* Talk presented at Generative Approaches to Language Acquisition North America (GALANA 4), Toronto, Canada, Sep 3.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84−107.

Corbett, G. (1991). *Gender.* Cambridge: CUP.

Cubelli, R., Paolieri, D., Lotto, L., Job, R. (2011). The Effect Of Grammatical Gender on Object Categorisation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 449-460.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42,* 317−367.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory & Language, 42*, 465-480.

Der, G. & Deary, I.J. (2006). Age and Sex Differences in Reaction Time in Adulthood: Results From the United Kingdom Health and Lifestyle Survey. Psychology and Ageing, 21, 62-73.

Foucart, A. (2008). Grammatical Gender Processing in French as a first and second language. PhD Dissertation, University of Edinburgh and University of Provence, France.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*, 267–283.

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J. & Van den Toorn, M.C. (1997). *Algemene Nederlandse Spraakkunst (ANS)*. Groningen: Martinus Nijhoff.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137,* 151−171.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Lew-Williams, C. & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language, 63,* 447-464.

Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing, 19,* 1-36.

Magnuson, J. S., Dixon, J., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31, 133-156.

Marslen-Wilson, W.D. (1987). *Functional parallelism in spoken word recognition*. In U.H. Frauenfelder & L.K. Tyler (Eds.), Spoken word recognition. Cambridge, MA: MITPress.

McClelland, J., & Elman, J. (1986). The Trace model of speech perception. *Cognitive Psychology, 18,* 1–86.

Müller, N. (1990). *Developing two Gender Assignment Systems simultaneously.* In: Meisel, J. (Eds.), Two first languages. Early grammatical development in Bilingual Children. Dordrecht: Foris.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.

Osberger, W. & Rohaly, A. (2001). Automatic Detection of Regions of Interest in Complex Video Sequences. In *Human Vision and Electronic Imaging VI*, 4299, SPIE - The International Society for Optical Engineering, Bellingham, USA, 361-372.

Paris, G., Weber, A. & Crocker. M. W. (2006). German morphosyntactic gender and lexical access. Poster presented at 12th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP 2006), Nijmegen, The Netherlands.

Raaijmakers, J.G.W., Schrijnemakers, J.M.C. & Gremmen, F. (1999). How to deal with "The language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, *41*, 416-426.

Schneider, W., Eschman, A and Zuccolotto, A. (2002). *E-Prime reference guide*, Psychology Software Tools Inc., Pittsburgh.

Sekerina, I. A., Brooks, P. J., & Kempe, V. Gender Transparency Facilitates Noun Selection in Russian. In Levine, J., and Babby, L. (Eds.), *The Proceedings of the 14th Annual Workshop on Formal Approaches to Slavic Linguistics. The Princeton Meeting*. Michigan Slavic Publications.

Snodgrass, J.G. & Vanderwart, M. (1980). A Standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174-215.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. Journal of Experimental Psychology: General, 112, 1-36.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632−1634.

Van Berkum, J.J.A. (1996). *The psycholinguistics of grammatical gender: Studies in language comprehension and production.* Doctoral dissertation, Max Planck Institute for psycholinguistics. Nijmegen, Netherlands: Nijmegen University press.

Van der Velde, M. (2004). L'acquisition des determinants en L1: une etude comparative entre le Francais et le Neerlandais. *Acquisition et interaction en langue etrangere, 21*, 9-46.

Zubin, D.A. & Köpcke, K. M. (1984). Affect classification in the German gender system. *Lingua, 63*, 41-96.

TABLES

Table 1: Overview of the main agreement targets in Dutch.

Adapted from Loerts, Stowe, & Schmid (Under Review). English equivalents of the phrases are printed in italic below each phrase. Note that the indefinite article is always *'een'* in the single case and no equivalent indefinite article exists for the plural case.

	Gender	**Definite Article**	**Adjectives in Definite NPs**	**Adjectives in Indefinite NPs**
Single	common	***de*** appel	***de*** rode appel	een ***rode*** appel
	neuter	***het*** huis	***het*** rode huis	een ***rood*** huis
English equivalent		*the apple/house*	*The red apple/house*	*a red apple/house*
Plural	common	de appels	de rode appels	--
	neuter	de huizen	de rode huizen	--
English equivalent		*the apples/houses*	*the red apples/houses*	--

Table 2: Overview of the conditions in the experiment. Example stimuli used in the present experiment while hearing the instruction to *Klik op de*$_{[COM]}$ *rode appel*$_{[COM]}$ ('Click on the$_{[COM]}$ red apple$_{[COM]}$ '). Note that in the indefinite conditions, gender-marking is on the adjective (suffix '*–e'* when modifying a common noun) while gender-marking only appears on the adjective in indefinite NPs (*schwa*-ending when modifying a common noun).

Target	**Competitor**	**Gender Competitor**	**Colour Competitor**
De$_{[COM]}$ rode appel$_{[COM]}$ *The*$_{[COM]}$ *red apple*$_{[COM]}$	Het$_{[NEU]}$ groene bureau$_{[NEU]}$ *The*$_{[NEU]}$ *green desk*$_{[NEU]}$	different	different
	De$_{[COM]}$ gele zon$_{[COM]}$ *The*$_{[COM]}$ *yellow sun*$_{[COM]}$	same	different
	Het$_{[NEU]}$ rode hart$_{[NEU]}$ *The*$_{[NEU]}$ *red heart*$_{[NEU]}$	different	same
	De$_{[COM]}$ rode taart$_{[COM]}$ *The*$_{[COM]}$ *red cake*$_{[COM]}$	same	same

Table 3: Fixed-effect factors in the model fitted to the reaction time (RT) data.

Estimated Coefficients, standard errors, and *t* values for the mixed-effects regression model fitted to the log-transformed reaction times in the eye tracking experiment. Positive coefficients reflect an increase in response time and negative coefficients a decrease. Factors are significant ($p < 0.05$) when they have an absolute value of the *t*-statistic exceeding 2 (Baayen et al., 2008).

	Estimate	***Std. Error***	***t value***
(Intercept)	6.8640	0.1325	51.80
SameColourCompetitor	0.1536	0.0079	19.48
PreviousRT (log)	0.0744	0.0151	4.92
Age	0.0050	0.0014	3.51
Sequence	-0.0006	0.0002	-3.85
TargetIsLeft	0.0103	0.0037	2.81
TargetIsBrown	0.0297	0.0106	2.80
LemmaFrequency (log)	-0.0071	0.0029	-2.46
*n*SeenCompetitor	0.0046	0.0020	2.25
SameGenderCompetitor	-0.0459	0.0258	-1.78
Age:SameGenderCompetitor	0.0012	0.0005	2.31
SameColourComp:SameGenderComp	-0.0152	0.0070	-2.17

Table 4: Random-effect parameters in the model fitted to the reaction time (RT) data. The numbers in the column entitled 'Correlations with Intercept' represent correlations between the by-subject random intercept and the random slope for *SameColourCompetitor* and *TargetIsBrown*, respectively. The number in the second column represents the correlation between the by-subject random slopes for *SameColourCompetitor* and *TargetIsBrown*.

Groups	***Name***	***Standard Deviation***	***Correlations with Intercept***	***Correlations between Slopes***
Item	Intercept	0.0231		
Subject	Intercept	0.0658		
	SameColourComp	0.0324	-0.719	
	TargetIsBrown	0.0164	-0.950	0.466
Subject	IsIndefinite	0.0313		
Subject	Sequence	0.0005		
Residual		0.0903		

Table 5: Estimated Coefficients, standard errors, and *t* values for the mixed-effects regression model fitted to the output measure *TargetRatio* (the empirical logit transformation of the relative difference between the proportion of looks towards target and competitor) in the eye tracking experiment. Positive coefficients reflect an increase in *TargetRatio* and negative coefficients a decrease. In this case, an increase in *TargetRatio* reflects relatively more fixations on the target as compared to the competitor in the relevant time-window. A decrease in *TargetRatio* signifies a difficulty in distinguishing target and competitor resulting in a more balanced distribution of fixations on the target and the competitor.

	Estimate	***Std. Error***	***t value***
(Intercept)	-2.9161	0.3526	-8.270
SameColourCompetitor	-0.3138	0.0201	-15.61
PreviousFocusDiff	0.0361	0.0197	1.83
TimeOnsetNoun	-0.0010	0.0001	-6.67
TimeOnsetDeterminer	0.0016	0.0003	6.36
Age	-0.0045	0.0014	-3.23
IsMale	-0.0558	0.0195	-2.86
TargetIsBrown	-0.0836	0.0211	-3.97
TargetIsUp	0.0666	0.0251	2.65
CompIsRight	-0.0396	0.0159	-2.49
TargetIsNeuter	-0.0856	0.0254	-3.37
SameGenderCompetitor	-0.0485	0.0219	-2.22
SameGenderComp:TargetIsNeuter	0.0839	0.0314	2.67

Table 6: Random-effect parameters in the model fitted to the eye movement data.

Groups	***Name***	***Standard Deviation***	***Correlations with Intercept***	***Correlations between Slopes***
Subject	Intercept	0.0262		
Subject	SameColourComp	0.0631		
Subject	TargetIsUp	0.0996		
Residual		0.3578		

FIGURES

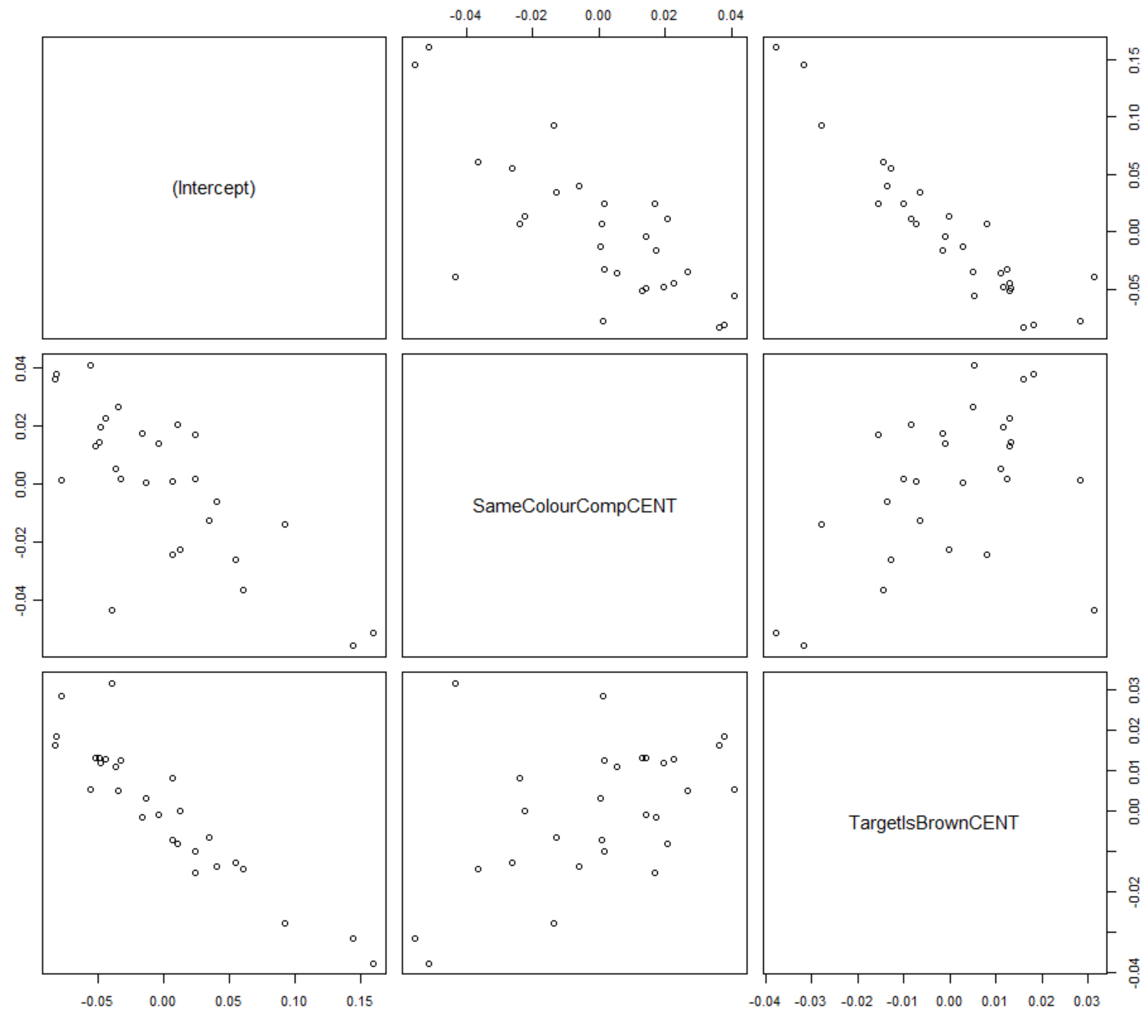

Figure 1: Scatterplots visualizing the correlational structures of the random intercepts and random slopes for *Subject* in the model for the reaction time (RT) data. The adjustments to the intercept position subjects with respect to the average response time. Subjects with large positive BLUPS (best linear unbiased predictions) are slow subjects; those with large negative BLUPS are fast subjects. This graph illustrates that fast subjects tend to suffer more from a *SameColourCompetitor* than slower subjects. The correlational structure between *SameColourCompetitor* and *TargetIsBrown* reveals that those subjects who suffer more from a *SameColourCompetitor* also suffer more when the *TargetIsBrown*.

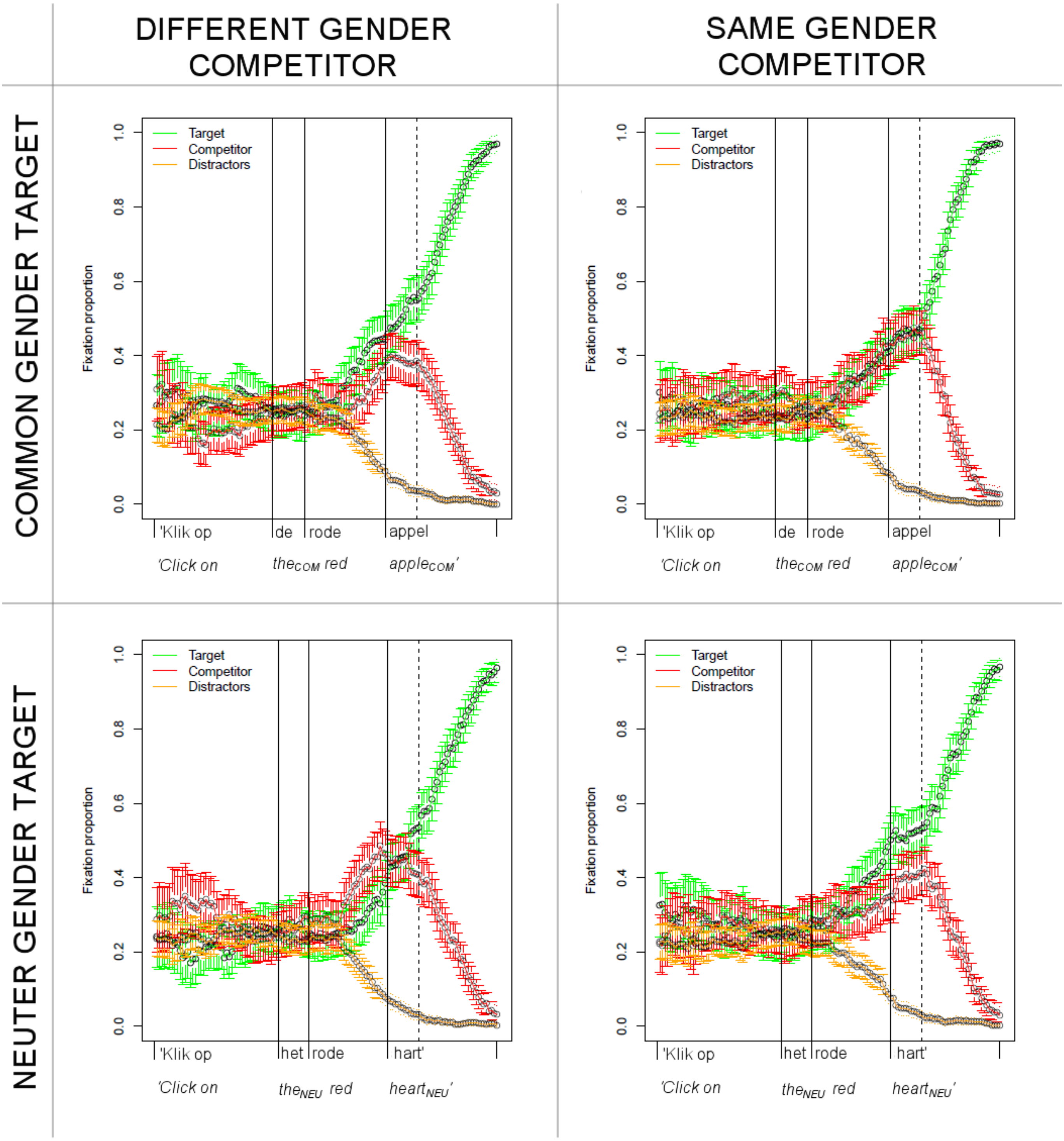

Figure 2: Proportions of fixations towards target, competitor, and distractors are depicted only for the conditions in which target and competitor shared colour. Trials in which the competitor did not share gender with the target are depicted in the left panel and same-gender trials are depicted on the right. The above panel shows the fixation proportions in conditions where the target was common and the lower panel reveals eye movement data in response to neuter gender targets.