

Speech vs. reading comprehension:

An explorative study of gender representations in Norwegian

Ute Gabriel¹, Dawn M. Behne¹ & Pascal M. Gygax²

¹Norwegian University of Science and Technology

Department of Psychology

NO-7491 Trondheim, Norway

²University of Fribourg

Department of Psychology

Rue P.-A.-de-Faucigny 2

CH-1701 Fribourg, Switzerland

POSTPRINT: Gabriel, U., Behne, D. & Gygax, P. (2017). Speech vs. reading comprehension: An explorative study of gender representations in Norwegian. *Journal of Cognitive Psychology*, link: <http://www.tandfonline.com/doi/full/10.1080/20445911.2017.1326923>

Author Note

This research was supported by an internal grant from the Department of Psychology NTNU.

We thank Magnus Alm, Håvard Tveit and Ane Torsdottir for their support in stimulus planning and preparation, experiment programming and data collection.

Correspondence concerning this article should be addressed to Ute Gabriel, Department of Psychology, NTNU, 7491 Trondheim, Norway. Email: ute.gabriel@ntnu.no

Abstract

As research on the construction of a mental representation of referent gender in speech comprehension is scarce, this study examined whether factors identified in reading comprehension exert similar influence in speech comprehension. Conceptually replicating previous research, a sentence continuation evaluation task was set up in two modalities, as a listening task and as a time confined reading task (i.e., to mimic the time constraint when listening). In line with previous findings from self-paced reading paradigms we found gender representations in language comprehension to be grounded in the interaction between textual (grammatical) and background (stereotypical) information. Extending previous research, the effect of stereotypical information was however modulated by presentation modality. In all, although speech and reading comprehension share higher-level processes of comprehension, this study provides first evidence that differences in comprehension might occur due to differences such as orthographic access or attention allocation.

(142 words)

Keywords:

Gender representation in language, Role nouns, Speech comprehension, Norwegian, Grammatical gender, Gender stereotypes

Speech vs. reading comprehension:

An explorative study of gender representations in Norwegian

When referring to a person's social, personal or professional role as scientist, traveller or manager, knowing the person's biological sex is not always crucial for comprehending the discourse. Nevertheless, research suggests that readers of sentences in which gender is not explicitly specified, such as "The traveller had to change trains", tend to spontaneously activate gender, or infer the gender of the person referred to (e.g. Irmen, 2007, in German). Some authors have suggested this process to be *automatic* (e.g., Oakhill, Garnham & Reynolds, 2005, in English), meaning very difficult, if not impossible, to avoid. In terms of social cognition, such an activation may narrow readers' mental representations to the relative disadvantage of one gender or the other (e.g. Stahlberg & Sczesny, 2001) in that readers activate and encode a particular gender that may anchor and bias subsequent processing (such as evaluation of behaviours according to stereotypes or adequacy of behaviours according to social context).

Surprisingly, previous research on these issues has predominantly focused on reading comprehension, leaving spoken language – despite its ontogenetic primacy – out of the equation. A possible reason for the focus on reading comprehension might be found in the assumption that, although listening to speech and reading differ in the basic (i.e. lower-level) processes, they share the same higher-level processes involved in comprehension (e.g. Jobard, Vigneau, Mazoyer & Tzourio-Mazoyer, 2007, Buchweitz, Mason, Leda, & Just, 2009). Consequently, cognitive processes in discourse comprehension (such as inference making) have been described by some scholars as being relatively independent of input modality (e.g. Booth, Perfetti, MacWhinney, & Hunt, 2000).

Despite the latter assumption of comprehension as independent of input modality, we wish to argue that similarities in higher-level processes do not necessarily mean that the

resulting mental representations are the same between spoken and written discourse comprehension. When specifically looking at gender representations, differences in comprehension might derive from differences at the lower levels of information processing (e.g., processing surface grammatical cues), an issue that, to our knowledge, has not yet been clarified. As such, the focus on written language when investigating gender representations might therefore exemplify a case of unfortunate one-sided research, and any conclusion from it that has been applied to language in general could be misleading.

The present study is intended to spark this discussion by exploring the construction of the mental representation of gender in speech comprehension, and to directly compare it to the mental representation of gender in reading comprehension. More specifically, inspired by central findings obtained with reading paradigms, we investigate the role of grammatical and stereotypical gender information in Norwegian, both in reading and speech comprehension. To the best of our knowledge, this is the first study to compare gender representations in reading and speech comprehension, thus contributing to a neglected area of research in this particular type of inference. If the exploration of speech comprehension yields similar results to reading comprehension, it will add to outlining the external validity of previous findings on gender representations. However, if it yields different results, it will inform the theoretical discussion on modality specific and modality independent language processing.

Gender inferences in language: Grammar & Stereotypes

Irrelevant of the differences between written and spoken discourse comprehension, one can assume that comprehenders of discourse always go beyond explicit information to form a mental representation, or mental model, composed both of explicit and implicit elements (Garnham & Oakhill, 1996). The process by which this is done is that of *inferencing* (Graesser, Singer, & Trabasso, 1994). Resulting mental representations enable readers to maintain coherence, both at a local level – to link adjacent sentences for example – and at a

more global level – to understand why a certain character behaves in a certain way throughout the text for example. Although there has been some level of disagreement as to which elements are or are not included in comprehenders' mental models during reading, the assumption that inferences are grounded in the interaction between the textual elements and comprehenders' *general knowledge* is rather undisputed (Marmolejo-Ramos, Elosua de Juan, Gygax, Madden & Roa, 2009).

Gender inferences very nicely illustrate such an interaction. When comprehending role nouns that are not lexically gender marked, such as *travellers*, *psychologists*, *neighbours* or *soldiers*¹, two information sources, depending on the language under investigation, have been shown to be of importance, namely grammatical gender (i.e., textual elements) and gender stereotypical knowledge. For comprehenders of languages that lack a sex-based grammatical gender system (e.g. Finnish; Pyykkönen, Hyönä & van Gompel, 2010) and languages with a pronominal gender system (e.g. English; Banaji & Hardin, 1996; Carreiras, Garnham, Oakhill, & Cain, 1996; Duffy & Keir, 2004; Garnham, Oakhill & Reynolds, 2002; Kennison & Trofe, 2003; Sturt, 2003) a reliable influence of gender stereotypes has been shown, whereas both sources (grammatical and stereotypical information) seem to influence comprehenders of languages with a sex-based grammatical gender systems (e.g. German, Irmen, 2007; Esaulova, Reali, & von Stockhausen, 2014; Italian: Cacciari, Corradini, Padovani, & Carreiras, 2011; French: Garnham, Gabriel, Sarrasin, Gygax, & Oakhill, 2012).

The moderating role played by the grammatical gender system of a language, can be exemplified by a systematic comparative reading study conducted in English, French and German, (Gygax, Gabriel, Sarrasin, Oakhill & Garnham, 2008) and one in Norwegian (Gabriel & Gygax, 2008). In both studies, a sentence evaluation paradigm (based on Tanenhaus & Carlson, 1990) was applied, whereby participants had to decide whether a second sentence containing explicit information about the gender of one or more of the

characters (e.g. ... one of the women/men...) was a sensible continuation of a first sentence introducing a role noun (e.g. The spies came out of the room).

Within each language tested, the role nouns were chosen to differ in *gender stereotypicality* (female stereotyped role nouns, male stereotyped role nouns, non-stereotyped role nouns, as measured by Gabriel, Gygax, Sarrasin, Garnham & Oakhill, 2008), yet some of the languages differed in their grammatical gender systems. French and German are illustrative of *sex-based* gender systems, in which most role nouns are grammatically marked and allow for a systematic grammatical-to-referential gender mapping (e.g. in German: die Lehrerinnen_{fem} [*the female teachers*] vs. die Lehrer_{masc} [*the male teachers*]). In contexts in which referent gender is considered irrelevant, or is unknown, the masculine form can also be used, in which case it carries a generic meaning (e.g., die Lehrer_{masc} [*the teachers*]). Essentially, the decision between a specific or a generic meaning is left to the comprehenders. English has a pronominal gender system, with only few morphologically marked role nouns such as *actor – actress* or *waiter – waitress* still being in use. Finally, Norwegian (*bokmål*) has a grammatical gender system, which is gradually losing the feminine gender-marking in role nouns, with few grammatically marked role nouns such as lærerinnene_{fem} – lærerne_{masc} [*the teachers*], flyvertinnene_{fem} – flyvertene_{masc} [*the flight attendants*], venninnene_{fem} – vennene_{masc} [*the friends*] still being in use.

Gygax et al. (2008) showed that in English, where no formal marking of gender was present, the proportion of yes- and no-responses on whether the second sentence was a sensible continuation of a first sentence, depended on the stereotype of the role nouns being tested. English participants, for instance, were more likely to respond negatively (no-response) when a sentence composed of *women* followed a sentence in which a stereotypically male role noun was presented (e.g. *mechanics*), than when a sentence composed of *men* followed a sentence with a stereotypically male role noun. In French and

German, where role nouns were written in the grammatically masculine form, although interpretable as a generic form, the proportion of negative answers was higher when the second sentence was composed of *women*, independent of the stereotype portrayed by the role nouns. These results illustrate the idea that the specific interpretation of the masculine form (i.e., masculine form = male referent) is activated through a *passive* and *hard-to-control* process, as further developed by Gygax, Gabriel, Lévy, Pool, Grivel, and Pedrazzini (2012). Essentially, readers of French (or other grammatical gender languages) cannot avoid the activation of the specific meaning of the masculine grammatical form (Gygax et al., 2012).

Most interestingly, in the corresponding study on Norwegian (Gabriel & Gygax, 2008), participants' representations were biased by the stereotypicality of the role nouns when reading female (e.g. *nurses*) and male (e.g. *pilots*) stereotyped role nouns (replicating findings from the English sample), but male biased when reading non-stereotyped role nouns (replicating findings from the French and the German samples). While the use of the suffixing to indicate female referents is declining, those prevailing are sufficient reminders of the suffix-referent mapping to impact readers' inferences in the absence of other context information (e.g. gender stereotypical). As raised by Gabriel and Gygax (2008), Norwegian can thus be considered an example in which stereotypical information and grammatical gender truly interact during reading comprehension, making the language particularly suited for the issues at hand in this present paper. Whether this is true or not for speech comprehension is an empirical question, which has yet received no attention.

Reading and speech comprehension

We are aware of only one study that has focused on gender inferences in speech processing of role nouns, namely that of Pyykkönen, et al. (2010). Employing a visual-world paradigm in an eye-tracking study, in which participants listen to sentences of a short discourse while looking at a scene presented on a computer screen with their eye-fixations

being monitored, the authors investigated whether listeners would activate gender stereotypes when hearing generically used nouns (embedded in a story), and whether they would do so even if such an activation (leading to an inference) was not needed for coherence. In their study, whilst participants heard a sentence portraying a female and a male character, followed by a sentence mentioning a job referred to by a gender stereotype noun (e.g., chimney sweep), two of the pictures presented on the screen, represented either the female or the male character. Analysing participants' fixations on the target pictures, the authors mainly showed that, in Finnish, listeners made gender inferences based on stereotypes (e.g., more fixations on the male character when listening to a male stereotyped job), and they did so in an elaborative way, meaning even when the inference was not needed for coherence.

The lack of empirical research on this issue is somehow surprising, as oral communication is prominent in day-to-day activity and as is even exclusive in the first years of language acquisition. More importantly, and as argued in the introduction, although listening to speech and reading might share higher-level processes, differences between reading and speech comprehension might nevertheless derive from differences at the lower levels of language processing. In particular, we suggest that, when focusing on gender representations, differences might arise for three main reasons.

First, when reading, comprehenders have direct access to orthographic information, whereas in listening, this is not necessarily the case. Although in Norwegian, a rather transparent language, word endings (including grammatical gender suffixes) are pronounced the same way as they are written (i.e., consistent phonology-orthography relationship), being visually presented with orthographic information might emphasize its pertinence in forming a mental representation of gender.

Second, and related to the first point, when listening to speech, comprehenders might be influenced by a number of factors associated to oral speech, each distancing them from the

focal aspects of the grammatical gender cues. For example, aspects of prosody, such as intonation or rhythm (i.e., always present), may direct comprehenders to sentence content words rather than grammatical cues. Of course, this would not be the case if speakers stress the endings of content words, making them more salient. However, generally, the inherent property of listening may well lead to differences in comprehension compared to reading.

The third aspect when contrasting listening and reading comprehension pertains to the fleetingness of speech and hence the greater demand on memory in spoken discourse comprehension. While auditory information is temporarily available, visual information is permanent. As a consequence, listeners are assumed to focus on essential aspects of the content of speech, building more *immediate* or *spontaneous* mental representations. By comparing speech and reading comprehension we should thus be able to compare representations of gender in language comprehension that vary in their relative *immediacy*, the results of which could inform the debate on what kind of gender-related information is activated and used during the course of processing (e.g. Gygax et al., 2012; Esaulova, et al., 2014).

In sum, by focusing on differences between speech and reading comprehension, we explore written- and speech-related factors that may influence the way comprehenders process grammatical and stereotypical information when building mental representations of the gender of human referents.

Our experiment

Our study used the same materials and sentence-evaluation paradigm as Gabriel and Gygax (2008), yet we adapted the paradigm to directly compare written and spoken text comprehension. In Gabriel and Gygax (2008), participants' reading was *self-paced*. In the present study, to better compare reading to listening, we modified the presentation mode to include a *time-confined reading* condition, which we directly compared to a (naturally time-

confined) *listening* condition. Essentially, we tried to minimize the effects of readers being able to go back to the text before moving on, yet not entirely erasing the difference in fleetingness between written and spoken language.

As mentioned earlier, Gabriel and Gygax (2008) found that participants' representations were biased by the stereotypicality of the role nouns when reading female (e.g. *nurses*) and male (e.g. *pilots*) stereotyped role nouns, but male biased when reading non-stereotyped role nouns. Based on these findings, when reading stereotyped role nouns, even under a *time-confined reading* condition, comprehenders should activate stereotyped information. Based on Pyykkönen et al. (2010), this should be the same when the role nouns are presented orally. However, we believe that differences could emerge when considering those role nouns that carry no stereotypes. In effect, the differences between the two representations will depend on the focal attention given to the grammatical gender mark. More specifically, we believe that speech-related factors, such as prosody, stress or fleetingness, will move comprehenders' attention away from the gender grammatical mark, reducing therefore the male bias found in Gabriel and Gygax (2008).

METHOD

Design

To compare how listening and reading influence the processing of grammatical and stereotypical information when building mental representations of gender, our experimental design was a two (Modality: listening vs. confined reading) by three (Stereotypicality of the role noun in sentence A: female vs. non stereotyped vs. male) by two (Continuation in sentence B: women vs. men) mixed factorial design. We also added Task Order (Task order: listening condition first vs. confined reading condition first) as a factor, to ensure that the order of modality was not a confound. As such, Task Order was a between-participant factor and all other factors were within-participant.

Materials

Materials were based on Gabriel and Gygax (2008) in which Norwegian participants were presented text with pairs of sentences (sentence continuation paradigm based on Tanenhaus & Carlson, 1990) using role nouns.

Role noun selection was based on a norming-study (Gabriel *et al.*, 2008) on 126 role nouns that was run in English, French and German. In the norming-study participants indicated the percentage of men and women they thought occupy these roles. To inspect the stereotypicality of those role names in Norwegian, Gabriel and Gygax (2008) had them translated and pre-tested applying the questionnaire format from Gabriel *et al.*, 2008, Study 2. Although stereotypicality of role nouns might not be reliable across time, we decided to stick to the original items used by Gygax *et al.* (2008) and Gabriel and Gygax (2008) to allow for comparability². In the current study the sentence pairs from Gabriel and Gygax (2008) were presented both as text and as speech.

Sentence Pairs

In each sentence pair, sentence A introduces a role noun, which is female, male or non-stereotyped (e.g. *Politikerne gikk inn på kontoret.* [The politicians entered the office.]), and sentence B mentions the gender ("men" or "women") of some of the members of the group (e.g. *Det var tydelig at de fleste av kvinnene/av mennene var skikkelig sinte.* [It was obvious that most of the women/men were angry.]). Sentence B would therefore either match or mismatch the stereotypicality of interpretation of the role noun indicated in sentence A. The non-stereotyped role nouns represent the only role noun type that was intended to trigger an unbiased response.

In total, 36 experimental sentence pairs and 36 filler sentence pairs were included in the study. The experimental sentence pairs included 12 first-sentences (sentence A) with a female stereotyped role noun, 12 with a male stereotyped role noun and 12 with a non-

stereotyped role noun, respectively. For each stereotype condition, six of the role nouns were combined with *women*-continuations and six with *men*-continuations resulting in two sentence pair lists, such that if a role noun was continued by *women* in one sentence pair list, it was continued by *men* in the other sentence pair list.

All experimental sentence pairs were intended to elicit a *yes* response. As a consequence, the thirty-six filler items were included to elicit a clear *no* response. Three versions of filler items were used: referents in sentences A and B did not match (e.g., The football players walked through the park. One could see that several of the trade unionists were quiet.), the definitional gender of the role noun mentioned in sentence A and did not match sentence B (e.g., The bridesmaids waited on the steps. It was obvious that some of the men were very sad.), and finally sentences A and B lacked semantic coherence (e.g., The professors took a break in the sun. Because of the bad weather the majority of the women were holding umbrellas).³

Audio versions of the text materials

The written materials were those used in Gabriel and Gygax (2008) and an audio version of them needed to be prepared. As this is the first study to compare reading with speech comprehension of grammatical and stereotypical information, and as we did not want to add yet another factor into our experiment, the sex of the speakers voice in the listening condition was maintained constant by only using a female voice.

Audio recordings of the experimental and filler sentences were made of a young adult female, native speaker of East Norwegian who was instructed to read the Gabriel and Gygax (2008) sentence pairs with a natural speaking rate and intonation, avoiding unnatural rising or falling intonation at the end of sentences. By having the sentence pairs read, rather than repeated in conversation style, the duration of the sentence pairs approximated the amount of time participants in the experiment would need to read the text materials being used in the

confined reading task.

The speaker was seated in a sound insulated studio in the NTNU Speech Laboratory, Department of Psychology, NTNU where all combinations of sentence A and B were digitally recorded using a Røde NT3 microphone connected via an M-Audio Firewire 1814 sound card to an Apple Macintosh G5 computer. The audio was recorded at a sampling rate of 44.1 kHz using PRAAT, version 5.0.32 (Boersma, 2001). The onset and offset of the sentence pairs were determined by visual inspection of the waveform. Multiple recordings of each sentence pair were evaluated for naturalness, even intonation, and lack of stray sounds picked up during the recordings. The audio versions of sentence pairs selected for use were further edited in PRAAT to the same unweighted intensity. In the final audio versions of the sentence pairs, the average duration of sentence A was 1900 ms (range: 1384 to 2828 ms) and for sentence B the average duration was 2800 ms (range: 1988 to 3702 ms).

Participants

The sample included 72 (36 female participants) NTNU students between the age of 19 and 31 years (mean= 22years, 6 months) who had Norwegian as their native language. All participants self-reported normal hearing and corrected-to-normal vision, further confirmed with a computer-based version of the Snellen test. All participants were right-handed or ambidextrous based on results from a variant of the Edinburgh Handedness Inventory (Oldfield, 1971).⁴

Procedure

Participants were presented the 72 sentence pairs twice, once in a time confined reading task and once in a listening task. To ensure equal cell sizes, female and male participants were quasi randomly assigned to four groups based on sentence pair list (list 1 and 2) and whether they began with the listening or confined reading condition, thus counterbalancing a possible effect of repetition. This resulted in 18 participants (9 female, 9

male) in each group.

Participants were tested in groups of up to five at a time in the NTNU Speech Laboratory. Each participant was seated in front of an iMac (11.3, OS10.6.8) computer with a 17" monitor (1440x900 pixels) positioned at a 90 degree angle at a distance of ca. 60cm by which participants were presented instructions and written sentence pairs. Connected to the computer were AKG K271 stereo closed dynamic circumaural studio headphones and a Cedrus response box where a left button was labelled "no" and a right button was labelled "yes".

Instructions displayed on the computer monitor informed participants that their task was to read the sentences displayed on the monitor (Confined reading condition) or to listen to sentences presented over the headphones ca. 68 dBA (Listening condition). The participants' task was to decide as accurately and as fast as possible whether the sentence B was a sensible continuation of sentence A, and respond by pressing either the "no" or "yes" button. Participants were asked to keep their left and right index fingers respectively on the "no" and "yes" buttons at all times during the experiment. The instructions stressed that participants should read or listen as they would normally do and to respond without prolonged contemplation.

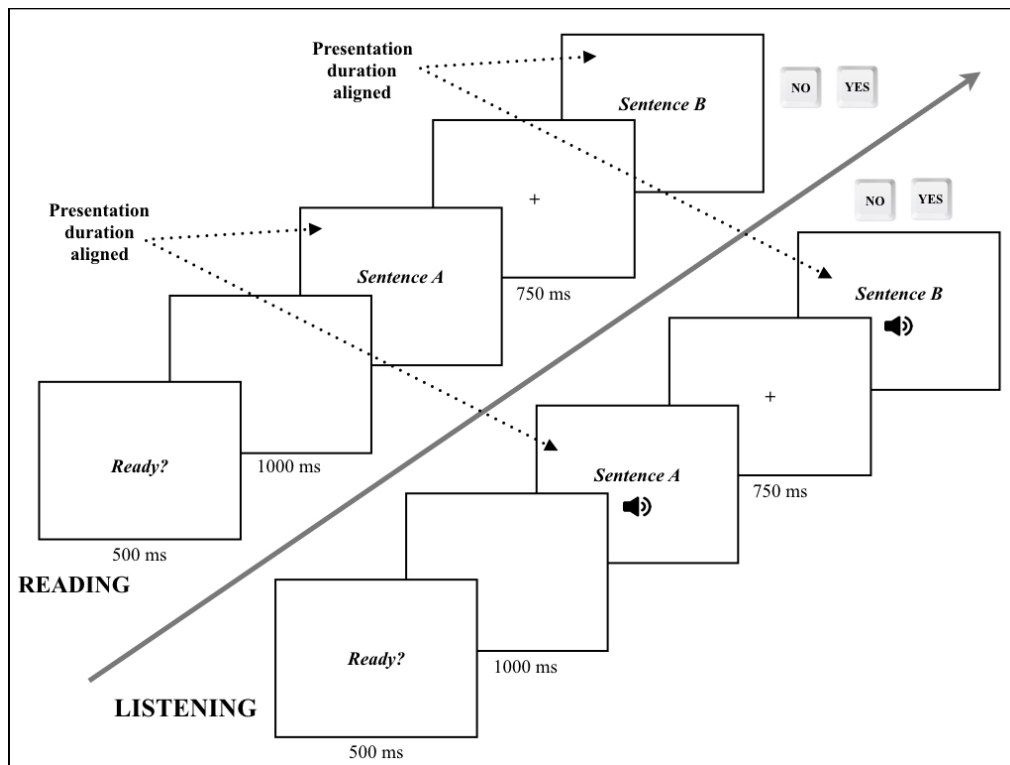


Figure 1. Timeline for trials in the confined reading condition and the listening condition. Sentences in the confined reading condition were presented for the amount of time corresponding to the respective sentence in the listening condition.

The experiment was set up and run using Superlab 4.5 (Cedrus, San Pedro, California, USA). The presentation of sentence pairs was blocked for the listening condition and confined reading condition. The sentence pairs within each modality were presented in random order for each participant.

The structure of the listening trials and reading trials is presented in Figure 1. All trials began with a 500 ms prompt, followed by a 1000ms blank screen. Sentence A was then presented. In the listening trials the audio version of sentence A was presented and in the confined reading trials sentence A was displayed in black text for the same amount of time as the duration of the corresponding audio version of the sentence. Sentence A was followed by a 750ms black fixation cross, displayed at the centre of a blank screen. This was followed by sentence B such that in the listening trials the audio version of sentence B was presented and

in the confined reading trials sentence B was displayed for the same duration as the audio version of the sentence. Sentence B was followed by a blank screen until the participants made their decision. In this way sentences A and B in the confined reading condition were presented for the same amount of time as corresponding sentences in the listening condition. Responses were collected from onset of sentence B.

To familiarize participants with the experiment procedure, six practice trials were given each for the listening condition and for the confined reading condition at the start of the corresponding block. To minimize fatigue, participants had five forced breaks in the course of the experiment: 15s. breaks came after 12, 24, 48 and 60 trials, and a 60s. break came half way through the experiment after 36 trials.

RESULTS

Data Analyses

We conducted separate analyses on participants' responses (yes/no) and their response times for yes-responses (i.e., for accepting sentence B as a reasonable continuation). As shown in Gabriel and Gygax (2008), both dependent variables show similar effects on some of the factors, yet may also differ on others. We nevertheless expected difficulties in integrating the information from sentence A (stereotype) and sentence B (women/men continuation) to mainly show up as fewer yes-responses and slower yes-responses.

Mixed-effects logistic regression (generalized linear mixed-effect regression, *glmer*) was used to model participants' yes/no responses, and linear mixed-effect regression (*lmer*) was used to analyse participants' response times for yes-responses. All analyses were performed using R software (R Studio Team 2015; R Studio for Mac version 0.99.486; R version 3.3.0) and the *lme4* (Bates, Maechler, Bolker & Walker, 2015) and *LMERConvenienceFunctions* packages (Tremblay & Ransijn, 2015; *fitLMER*-function for *glmer*-model and *bfFixefLMER* and *ffRanefLMER* functions for *lmer*-model).

The sentence pairs used were originally developed for use in a self-paced reading study (Gygax et al., 2008) and consequently vary in orthographical length and thus in presentation duration. In Gygax et al. (2008) and Gabriel and Gygax (2008) this was less of an issue as individual reading times of A sentences were utilized to estimate the (baseline) reading time of B sentences, and as the difference between estimated and actual response time to B sentences was used as a dependent variable. As this is not an available option for the current data, instead, we control for variation in presentation duration of the B sentences by including presentation duration as an additional predictor in our analyses, both as a main effect and as an interaction effect by modality (confined reading vs. listening).

For both DVs an *initial model* was fitted that included all experimental variables (Modality, Task Order, Stereotypicality of Role Noun, Continuation) and their interactions as well as Presentation Duration and Presentation Duration by Modality as fixed effects, and the random intercepts for both subject and item as random effects. Next, following Baayen (2008) and Baayen and Milin (2010), the fixed effect structure was back-fitted, random effects (by subject random slopes for all experimental variables and by item random slopes for modality) were forward-fitted, and finally the fixed effect structure was re-back-fitted.

Presentation duration ($M = 2780$ ms, $SD = 442$ ms, $Min = 1990$ ms, $Max = 3700$ ms) was z-transformed. All categorical predictors were sum coded; for Stereotypicality of Role Nouns, the first comparison contrasted stereotypically female with male role nouns, and the second comparison non-stereotyped with stereotypically male role nouns.

Response times were log transformed, data were screened for outliers, no outliers were removed. The final model was refitted with data points with absolute standardized residuals exceeding 2.5 standard deviations removed (88 data points, 2.3% of data). After this trimming, the residuals approximated normality more closely. Trimming did not impact the fixed effect structure.⁵

Note that since *sex of respondents* never showed any effects in previous related experiments, we did not have any hypothesis concerning this specific factor and hence did not include it in our analyses.

YES/NO Responses

The final model ($N = 5135$) contained random intercepts by subject and by item, and Task Order slope by subject. There were significant main effects of Stereotypicality, Wald $\text{Chi}^2(2) = 14.87, p < .001$, and Continuation, Wald $\text{Chi}^2(1) = 5.2, p = .02$, which were qualified by a significant disordinal Stereotypicality by Continuation Interaction effect (cf. Figure 2), Wald $\text{Chi}^2(2) = 82.79, p < .001$; after sentences containing female stereotyped role nouns the likelihood for yes-responses was higher (+12%) when the continuation mentioned *women* than when it mentioned *men*, whereas the opposite hold true for sentences containing male stereotyped role nouns (-13%, $z = 8.1, p < .001$). The pattern for non-stereotyped role nouns (-4%) was not significantly different from that for male stereotyped role nouns ($z = .3, p = .76$).

In addition, the likelihood for yes-responses was significantly higher in the Listening (83%) than in the Time Confined Reading modality (79%, Wald $\text{Chi}^2(1) = 12.57, p < .01$) and in the Second presentation (85%) than in the First presentation (77%, Wald $\text{Chi}^2(1) = 24.9, p < .001$).

Finally, the final model contained an ordinal Stereotypicality by Modality interaction effect (cf. Figure 2), Wald $\text{Chi}^2(2) = 5.28, p = .07$, revealing that the main effect of Modality (i.e. higher likelihood for yes-responses in Listening modality) was similarly strong for stereotypically female (+3%) and male (+2.4%) role nouns ($z = -1.62, p = .11$) but significantly stronger for the non-stereotyped role nouns (+6.8%, $z = 2.18, p = .03$).

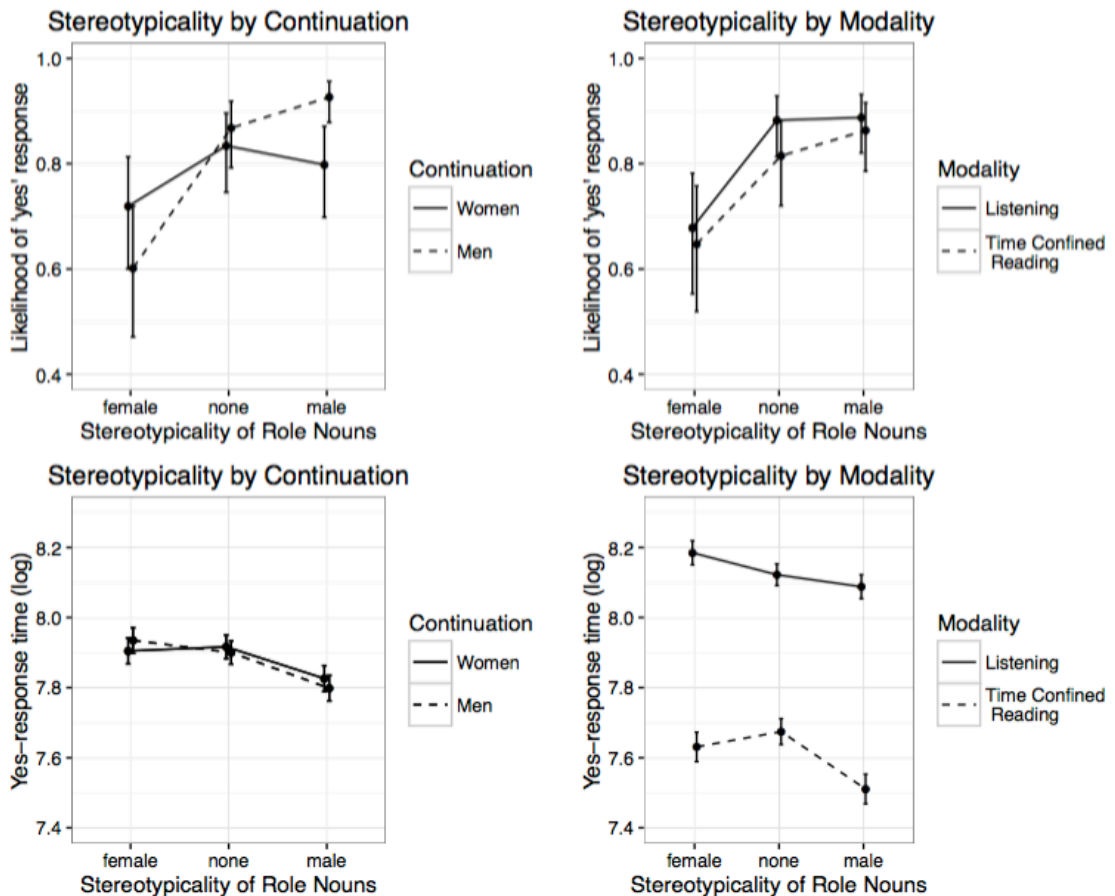


Figure 2: Means with 95% CI for Stereotypicality by Continuation (left column) and Stereotypicality by Modality (right column) interaction effects for Yes/No Responses (upper row) and Yes-response times (lower row)

Response times for yes-responses

The final model ($N = 3700$) contained random intercepts by subject and item, and modality slope by item. Unsurprisingly, Presentation Duration was a significant predictor, $F(1, 3688) = 531.5, p < .001$, and there was a significant Presentation Duration by Modality interaction, $F(1, 3688) = 26.9, p < .001$, revealing that Yes-response times depended more strongly on Presentation Duration in the listening than in the time confined reading condition (estimated differences in slopes = .025).

More importantly, there was a main effect of Stereotypicality, $F(2, 3688) = 17.0, p <$

.001, which was qualified by a significant disordinal Stereotypicality by Continuation interaction effect (cf. Figure 2), $F(2, 3688) = 9.3, p < .001$; yes-responses were faster to Women continuations when following female stereotyped role nouns (- 82 ms) but slower when following male stereotyped role nouns (+ 68 ms, female vs. male stereotyped role nouns: $t = -4.21, p < .001$) and non-stereotyped role nouns (+ 44 ms, non-stereotyped vs. male stereotyped role nouns: $t = 1.52, p = .13$).

There was also a strong effect of Modality, $F(1, 3688) = 6937.8, p < .001$, with slower yes-responses in the listening ($M_{\text{listen}} = 3385$ ms) than in the confined reading task ($M_{\text{read}} = 2003$ ms), and a main effect of Task Order, $F(1, 3688) = 11, p < .001$; yes-responses were faster in the second task ($M_{\text{second}} = 2560$ ms) than in first task ($M_{\text{first}} = 2715$ ms).

In addition, the model contained a semi-disordinal Stereotypicality by Modality interaction effect (cf. Figure 2), $F(2, 3688) = 37.8, p < .001$, revealing that the Modality effect (slower responses in listening task) was stronger for female stereotyped (+1525ms) than for male stereotyped role nouns (+1427ms, $t = 2.16, p = .03$) and stronger for male stereotyped role nouns than for non-stereotyped role nouns (+1215 ms, $t = -8.65, p < .001$).

Additional analysis: No-responses to filler items

Contrary to the background of our (general) rationale for the dependent variables (i.e. *fewer* yes-responses and *longer* yes-response times indicate difficulties in integrating the information from sentence B) the main effects for modality (i.e. *more but slower* yes-responses to experimental items in the listening than in the confined reading condition) seem puzzling at first glance. However, being a main effect, these findings might instead generally reflect more elaborate or more difficult information processing in the listening condition than in the confined reading condition. If the modality manipulation in our experimental set-up lead to such differences in information processing, this should also show in how participants responded to the filler items. To check for this assumption, we ran an additional analysis on

the filler items. More specifically, we checked whether the number of (correct) no-responses to fillers could be predicted by presentation duration, task order and modality.

Filler items.

The final model contained random intercepts by subject and by item, and Modality slope by subject. There were significant main effects of Presentation Duration Wald $\chi^2(1) = 8.93, p < .01$, the likelihood for (correct) no-responses to filler items increased by Presentation Duration; and Modality, Wald $\chi^2(1) = 6.72, p < .01$, the likelihood for (correct) no-responses to filler items was significantly higher in listening (86%) than in time confined reading (82%). In essence, these data suggest some extra processing in the listening task, whether because of difficulty or because of extended attention remains open.

DISCUSSION

By comparing gender inferences elicited in a sentence-evaluation paradigm based on listening with those based on time confined reading, we explored written- and speech-related factors that may influence the way listeners and readers process grammatical and stereotypical information when building mental representations of the gender of human referents. Both for the yes-responses and the yes-response times, we first found an interaction effect of stereotype by continuation (similarly to Gabriel & Gygax, 2008), documenting general gender representation mechanisms. Second, we found main effects for modality and task order, and an additional interaction effect of stereotype by modality, these documenting modality-specific mechanisms.

Corroborating previous findings (e.g., Gabriel & Gygax, 2008, in Norwegian; Garnham et al., 2002, in English), we found gender stereotypes to have a strong impact on comprehenders' gender inferences in that participants responded "yes" more often and did so faster when the gender that was specified in sentence B matched the stereotypical gender of

the role noun introduced in sentence A, than when it did not match. For non-stereotyped role nouns, the pattern of results was not markedly different from that for male stereotyped role nouns (response: $p > .76$; response times: $p > .10$), suggesting that participants found it easier to link non-stereotyped role nouns to male referents than to female referents. We believe this effect mirrors a gender bias introduced by the use of the grammatical masculine form. More specifically, when no gender-specific stereotypical knowledge is available to comprehenders, they rely on the masculine form – and its specific meaning –, to assign a gender to the role nouns. We will come back to this result when directly comparing our results to those in Gabriel and Gygax (2008).

We believe that the main effects of Task Order – more yes-responses and faster yes-responses in the second task than the first task – merely suggest a training effect for the experimental sentences. However, the main effects of modality for the experimental sentences (more, but slower, yes-responses to experimental items in the listening modality) as well as for the filler sentences (more correct rejections in the listening modality) might indicate that the presentation mode forced a longer processing time in the listening condition than in the confined reading condition. Consequently, compared to the first task, there were more *correct* answers in the second task, both for the filler items (more no-responses) and the experimental items (more yes responses). When only looking at yes-responses, the modality effect (higher likelihood for yes-responses in the listening task) was accentuated for non-stereotyped role nouns. When only looking at yes-response times, the modality effect – slower responses in the listening modality – was stronger for female stereotyped role nouns than for male stereotyped role nouns, and stronger for male stereotyped role nouns than for non-stereotyped stereotyped role nouns. Put differently, compared to time confined reading, listening especially slowed down responses to stereotyped role nouns. This, along with the modality effect associated to yes-responses, might be interpreted as a hint that the absence of a morpho-

orthographical marker (i.e. written suffix –erne) facilitates assessing the semantic gender meaning of the non-stereotyped role nouns.

As the slowest response times were present in the listening modality and were associated with female stereotypes, one might speculate whether the female voice (constant across all the experiment) somehow interfered with the processing of the female stereotyped role nouns, more specifically whether the concurrence of a female voice presenting stereotypical female content hindered assessing the semantic gender meaning of the role noun. At first glance, such an interpretation is inconsistent with previous research that mainly reported inhibition effects when processing gender *incongruent* information. Employing an auditory Stroop task, for example, Most, Sorber, and Cunningham (2007) found that participants were slower to categorize the sex of a speaker's voice when the voice's sex was stereotypically incongruent with the spoken word (e.g., a man saying "lipstick" is more difficult to recognize as a man). Similarly, research on how listeners perceive gender stereotypes and gender of speakers has shown that listeners respond to incongruity between speakers' voice characteristics and the semantic content of their gender-stereotypical (self-referent) utterances (e.g., Lattner & Friederici, 2003; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008; for syntactic processing, Hanulikova & Carreiras, 2015). Finally, recent research suggests a congruency effect of grammatical gender and the speaker's sex, with grammatically feminine words being processed faster when spoken in a female voice than when spoken in a male voice (Spanish: Vitevich, Sereno, Jongman & Goldstein, 2013; Bulgarian: Andonova, 2013)

However, as neither of the cited research actually tested whether properties of a speaker's voice influence the listener's *actual* representation of gender, those findings might not readily be transferrable to our results. Our data also differ in that we register explicit responses. Prolonged response times to sentences following sentences that introduced female

stereotyped role nouns could hence reflect source monitoring issues, i.e. uncertainty whether femaleness was activated by the voice or the content. Future research on this topic may want to focus on how listeners integrate gender-related voice characteristics during speech comprehension.

When looking back at the initial study by Gabriel and Gygax (2008), the current study revealed both similarities and differences. First, although our interaction pattern between Stereotypicality of role nouns and Continuations is perfectly in line with the findings based on the self-paced reading task by Gabriel and Gygax (2008), in the present study, this interaction effect was significant both for yes-responses *and* for yes-response times (which was not the case for Gabriel & Gygax, 2008). This could be a consequence of differences in *self-paced* vs. *fixed presentation* durations, in that confined processing time of priming sentences may increase difficulty of understanding and may compel comprehenders into heightened attention processes. Consequently, comprehenders may allocate relatively more attention to the semantic features (i.e., stereotypes) of discourse constituents, leading them to respond “yes” more slowly for incongruent stereotypical matches. It could, however, also be a consequence of the data handling and analyses, as residual response times were analysed in the self-paced study vs. presentation duration was included as covariate in the present study.

Second, although there was a main effect of continuation, it was only present in the yes-responses, yet not in the yes-response times, as in Gabriel and Gygax (2008). As discussed in the argument above, participants in the present experiment seem to have allocated more attention to the semantic features associated to role nouns, therefore slightly attenuating the male bias found in Gabriel and Gygax’ self-paced reading experiment.

Third and finally, the present experiment revealed a main effect of stereotype that was absent in the self-paced reading study and which is mainly produced by an overall lower likelihood of yes-answers to female stereotyped role nouns and overall faster yes-responses to

male stereotyped role nouns. If the change in the fleetingness of the presentation between the present study and the previous self-paced reading study did generate differences in attention, this main effect of stereotype could illustrate an interesting *match vs. mismatch* focus mechanism. In other terms, in the present experiment, grammar-stereotype matches as well as mismatches might have gained more attention. On the one hand, mismatches between female stereotypical role nouns and the grammatical masculine information generated fewer yes-responses (than expected). On the other hand, matches between male stereotypical role nouns and the grammatical masculine information facilitated yes-responses (i.e., faster response times).

Taken together, we suggest the allocation of additional attention resources as a tentative explanation of the differences between our present experiment and that of Gabriel and Gygax (2008). These resources have to be allocated, as the fleetingness of spoken discourse (and time confined written discourse) does not allow comprehenders to go back, check and/or complete information. As a consequence, semantic features gain more attention and the relative contribution of semantic (here: stereotype) information to the discourse representation is increased.

This study has some limitations that should be addressed in future work. First, the sentence pairs used for this experiment had originally been designed for a self-paced reading task and variation in length was not considered an issue then, but resulted in considerable variation in presentation duration when using the material for the listening and the time confined reading task. Although this variation in presentation duration was not a predictor of participants' likelihood of yes-responses, it was of their yes-response times, especially in the listening task. These effects might be interfering with, or masking other effects. Second, in our listening task, only a female voice was used, which limits any explanations of results based on the sex of the speaker's voice and renders future research that varies speaker's sex

essential. Third and finally, differences in the results between the self-paced reading study and the current study might alternatively be interpreted as signalling changes that happened in the years between the data collections. As such, female stereotyped role nouns may now be more difficult to map onto *women* or *men* continuations due to their possible less frequent use. Also, the continued non-use of feminine gender markings may have (further) weakened language users' associations between the (formerly) masculine suffix and male referents. It is however arguable that such changes should have happened within the seven years between the two data collections (without having been noticed otherwise).

Conclusion

On the one hand, our findings further confirm the external validity of previous findings on gender inferences in language comprehension being grounded in the interaction between grammatical and stereotypical information. More generally, gender inferences are grounded in the interaction between textual elements and the language processor's background knowledge. On the other hand, our findings go beyond previous research in that they provide first evidence of how semantic processing can vary between speech and reading comprehension. They suggest stereotypical information to be more influential in listening (and in time confined reading) than in self-paced reading, and grammatical gender cues (i.e. suffixes) to be less influential in speech than in reading comprehension.

Being the first study to our knowledge to compare gender representations in speech and reading comprehension, our findings raise several new questions, such as about the role of the speaker's voice in speech comprehension and about attention focus across discourse modalities. Future research may want to focus on these issues, as well as on the possibility of varying presentation parameters of both spoken and written utterances.

By confining the presentation of written text to the time it took to orally present it, we kept the length of what was presented similar across the two modalities. In doing so, we

reduced, yet not prevented, readers' possibilities to go back and forth in the text. At the same time, we kept the procedure more "natural" than if we had forced the participants into a specific word-by-word reading pace, as in a moving window paradigm (Reder, 1973).

Although we consider our procedure a good approximation, future research should seek to design tasks that allow for comparing self-paced reading with listening in a more elegant way, for example by combining a stop-making-sense task, i.e. a task in which participants read or listen to text until they decide it no longer makes sense, with gaze tracking for the reading modality.

In all, comparing speech comprehension to reading comprehension is a complex endeavour, but it has the potential to better capture the way stereotypical and grammatical gender information interact when comprehenders process nouns that explicitly refer to people.

References

- Andonova, E. (2013). Gender and Sex: The Experiential Basis of Grammar. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society, CogSci 2013, Berlin, Germany, July 31 - August 3, 2013*. Retrieved from <https://mindmodeling.org/cogsci2013/papers/0325/index.html>
- Banaji, M., & Hardin, C. (1996). Automatic stereotyping. *Psychological Science, 7*, 136-141.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48.
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*, 12–28.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International, 5*, 341-345.
- Booth, J. R., Perfetti, C. A., MacWhinney, B., & Hunt, S. B. (2000). The Association of Rapid Temporal Perception With Orthographic and Phonological Processing in Children and Adults With Reading Impairment. *Scientific Studies of Reading, 4*, 101–132. https://doi.org/10.1207/S1532799XSSR0402_02
- Buchweitz, A., Mason, R. A., Leda, T., & Just, M. A. (2009). Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. *Psychology & Neuroscience, 2*. Retrieved from https://works.bepress.com/marcel_just_cmu/23/
- Cacciari, C., Corradini, P., Padovani, R., & Carreiras, M. (2011). Pronoun resolution in Italian: The role of grammatical gender and context. *Journal of Cognitive Psychology, 23*, 416–434. <https://doi.org/10.1080/20445911.2011.526599>

- Carreiras, M., Garnham, A., Oakhill, J. & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology*, *49*, 639-663.
- Duffy, S. A. & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, *32*, 551-559.
- Esaulova, Y., Reali, C., & von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution. *Language, Cognition and Neuroscience*, *29*, 781–803.
<https://doi.org/10.1080/01690965.2013.794295>
- Gabriel, U., & Gygax, P. (2008). Can language amendments change gender representation? The case of Norway. *Scandinavian Journal of Psychology*, *49*, 451–457.
- Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., & Oakhill, J. (2008). Au pairs are rarely male: Norms on the gender perception of role names across English, French, and German. *Behavior Research Methods*, *40*, 206–212.
- Garnham, A., Gabriel, U., Sarrasin, O., Gygax, P., & Oakhill, J. (2012). Gender Representation in Different Languages and Grammatical Marking on Pronouns: When Beauticians, Musicians, and Mechanics Remain Men. *Discourse Processes*, *49*, 481–500. <https://doi.org/10.1080/0163853X.2012.688184>
- Garnham, A., & Oakhill, J. V. (1996). The mental models theory of language comprehension. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 313–339). Mahwah, NJ: Lawrence Erlbaum Associates.
- Garnham, A., Oakhill, J. & Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition*, *30*, 439-446.

- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.
- Gygax, P., Gabriel, U., Lévy, A., Pool, E., Grivel, M., & Pedrazzini, E. (2012). The masculine form and its competing interpretations in French: When linking grammatically masculine role names to female referents is difficult. *Journal of Cognitive Psychology*, *24*, 395–408. <https://doi.org/10.1080/20445911.2011.642858>
- Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes*, *23*, 464–485.
- Hanulíková, A., & Carreiras, M. (2015). Electrophysiology of subject-verb agreement mediated by speakers' gender. *Frontiers in Psychology*, *6*. <https://doi.org/10.3389/fpsyg.2015.01396>
- Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research*, *36*, 431-456.
- Jobard, G., Vigneau, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). Impact of modality and linguistic complexity during reading and listening tasks. *NeuroImage*, *34*, 784–800. <https://doi.org/10.1016/j.neuroimage.2006.06.067>
- Kennison, S. M. & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, *32*, 355-378.
- Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing--evidence from event-related brain potentials. *Neuroscience Letters*, *339*, 191–194.
- Marmolejo-Ramos, F., Juan, M.R.E, Gygax, P., Madden, C. J., & Roa, S. M. (2009). Reading between the lines. The activation of background knowledge during text comprehension. *Pragmatics & Cognition*, *17*, 77–107.

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., ... Sczesny, S.

(2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods*, *46*, 841–871.

<https://doi.org/10.3758/s13428-013-0409-z>

Most, S. B., Sorber, A. V., & Cunningham, J. G. (2007). Auditory Stroop reveals implicit gender associations in adults and children. *Journal of Experimental Social Psychology*, *43*, 287–294.

Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*, 97-113.

Oakhill, J., Garnham, A., & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory & Cognition*, *33*, 972-983.

Pyykkönen, P., Hyönä, J., & van Gompel, R. P. G. (2010). Activating gender stereotypes during online spoken language processing: Evidence from visual world eye tracking. *Experimental Psychology*, *57*, 126-133.

Reder, S. M. (1973). On-line monitoring of eye-position signals in contingent and noncontingent paradigms. *Behavior Research Methods & Instrumentation*, *5*, 218–228.
<https://doi.org/10.3758/BF03200168>

RStudio Team (2015). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA
URL <http://www.rstudio.com/>.

Stahlberg, D., & Sczesny, S. (2001). Effekte des generischen Maskulinums und alternativer Sprachformen auf den gedanklichen Einbezug von Frauen. [Effects of the generic use of the masculine pronoun and alternative forms of speech on the cognitive visibility of women] *Psychologische Rundschau*, *52*, 131–140.

<https://doi.org/info:doi:10.1026//0033-3042.52.3.131>

- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562.
[https://doi.org/10.1016/S0749-596X\(02\)00536-3](https://doi.org/10.1016/S0749-596X(02)00536-3)
- Tanenhaus, M. K., & Carlson, G. N. (1990). Comprehension of Deep and Surface Verbphrase Anaphors. *Language and Cognitive Processes*, 5, 257–280.
<https://doi.org/10.1080/01690969008407064>
- Tremblay A., Ransijn J. (2015). *LMERConvenienceFunctions: Model Selection and Post-hoc Analysis for (G)LMER Models*. R Package Version 2.10 Available at: <http://cran.r-project.org/package=LMERConvenienceFunctions>
- Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The Neural Integration of Speaker and Message. *Journal of Cognitive Neuroscience*, 20, 580–591. <http://doi.org/10.1162/jocn.2008.20054>
- Vitevitch, M. S., Sereno, J., Jongman, A., & Goldstein, R. (2013). Speaker Sex Influences Processing of Grammatical Gender. *PLOS ONE*, 8, e79701.
<https://doi.org/10.1371/journal.pone.0079701>

Footnotes

¹ In contrast to lexically marked role nouns such as *catwomen*, or *chairmen* or nouns with definitional gender such as *ladies* or *knights*.

² As in Gabriel & Gygax (2008), we also ran the analyses considering differences, either from the Pilot in Gabriel & Gygax (2008), or from recently published norms (Misersky et al., 2015). Since none of these analyses signalled any substantial difference, we kept the original materials.

³ Critical information has been underlined for expository purpose.

⁴ Originally 73 participants, but data from one left-hander was removed, as the yes-response was not on their dominant hand.

⁵ In Gabriel and Gygax (2008) yes/no-responses had been analysed with by-item and by-participant ANOVAs, using the share of yes-responses as DV for the yes-no responses. To improve comparability, the data were reanalysed fitting mixed-effect models as described above. Results of their analyses did not change.