Activating Gender Stereotypes During Online Spoken Language Processing
Evidence From Visual World Eye Tracking

Pirita Pyykkönen,1,2 Jukka Hyönä,1 and Roger P. G. van Gompel3

1Department of Psychology, University of Turku, Finland
2Department of Computational Linguistics, Saarland University, Germany
3School of Psychology, University of Dundee, UK

Abstract. This study used the visual world eye-tracking method to investigate activation of general world knowledge related to gender-stereotypical role names in online spoken language comprehension in Finnish. The results showed that listeners activated gender stereotypes elaboratively in story contexts where this information was not needed to build coherence. Furthermore, listeners made additional inferences based on gender stereotypes to revise an already established coherence relation. Both results are consistent with mental models theory (e.g., Garnham, 2001). They are harder to explain by the minimalist account (McKoon & Ratcliff, 1992), which suggests that people limit inferences to those needed to establish coherence in discourse.

Keywords: inference, gender stereotypes, visual world eye tracking, anaphor, spoken language

During language comprehension, people not only use linguistic information to understand texts, but they also use general world knowledge. One type of knowledge, which has received much attention, is gender stereotypes. The question of when and exactly how gender stereotypes are activated during online comprehension is currently under debate (see e.g., Duffy & Keir, 2004; Irmen, 2007; Reynolds, Garnham, & Oakhill, 2006). Stereotype information can be used to establish coherence by relating currently processed information to the previous discourse, in which case it is called a backward or bridging inference (Clark & Haviland, 1977). One widely studied example of bridging inferences is anaphors, which are explicit markers that prompt people to link them to a suitable antecedent in the previous discourse. There is ample evidence that stereotype information becomes activated when making a bridging inference to resolve an anaphor: when processing the pronoun he after surgeon, people infer that he refers to the surgeon because surgeons are stereotypically male (for an overview, see e.g., Duffy & Keir, 2004). However, the question of whether stereotype information is also activated elaboratively during the processing of a stereotype noun (e.g., surgeon) has not yet been resolved. Information is activated elaboratively when it is not literally stated in the text and is not used to establish coherence between currently processed information and the preceding discourse. When people activate information elaboratively, it does not increase text coherence. In the case of activating stereotype information elaboratively, either specific knowledge that a particular person is a surgeon may be activated or generic knowledge that surgeons tend to be male.

Elaborative versus Backward Activation of Gender Stereotypes

Studies of lexical semantics suggest that gender stereotypes can be activated when a stereotype noun is presented in isolation. For example, Banaji and Hardin (1996) showed that when people were asked to judge the gender of pronouns he, she, him, her or to judge whether the target word was a pronoun or not, responses were slightly faster when the words were primed by gender-stereotype words congruent with the pronoun’s gender than by gender-incongruent words. However, the priming effect of stereotypical gender was smaller than that of lexical gender. In addition, Oakhill, Garnham, and Reynolds (2005) showed that when people were asked to judge whether words like surgeon-brother or surgeon-sister could refer to the same person, the effect of gender stereotype was fast and not affected by the type of instruction. Even when participants were encouraged to pay attention to unambiguous lexical gender relations (landlady-sister) and were told that some occupations were not clearly marked for gender but could be held either by men or women (i.e., in the case of gender-stereotype occupations), they were...
less likely to accept stereotypically inconsistent than consistent pairs and the response times for the inconsistent pairs were also prolonged compared to consistent pairs. Oakhill et al. therefore proposed that gender stereotypes may be difficult or even impossible to suppress. In a follow-up study, Cacciari and Padovani (2007) used Italian gender-ambiguous role names associated with male or female stereotypes to prime the gender-marked target pronouns lui (he) and lei (she). In a gender decision task, both pronouns were judged faster when they were primed with stereotypically congruent than incongruent nouns when presenting with a 500-ms prime-target stimulus onset asynchrony.

Although gender stereotypes have been shown to become activated when gender-stereotype words are presented in isolation, this does not necessarily imply that they are activated during text processing. However, evidence that comprehenders make stereotype inferences during text processing comes from Sanford (1985), who showed that when a text indicates that a previously mentioned surgeon is female, people find it hard to understand the story (see Dunning & Sherman, 1997 for similar findings). Similarly, Reynolds et al. (2006) found that when readers encountered stereotypically inconsistent information following surgeon, they slowed down their reading and sometimes even failed to interpret who the surgeon was. They argued that the increased reading time reflects elaborative activation of gender-stereotype information during the processing of surgeon, that is, before readers reached the gender-stereotypically inconsistent information following it. Furthermore, Carreiras, Garnham, Oakhill, and Cain (1996) observed reading difficulty when gender-stereotype information was inconsistent with the morphological gender of the noun’s article in Spanish. This suggests that stereotype information is activated when encountering a stereotype noun in the text, at least when it mismatches grammatical gender.

In a recent study in German, Irmen (2007) demonstrated a gender-mismatch effect between the plural form of gender-stereotype roles names such as *Installateur* “plumbers” (male stereotype) or *Geburtsfahrer* “obstetricians” (female stereotype) and the noun phrase *diese Männer/Frauen* “these men/women” later in the text: later measures of reading (as evidenced by readers’ eye fixation patterns) for the noun region (*Männer/Frauen*) were reliably affected by the gender-stereotype mismatch.

Because in most reading studies the evidence for stereotype activation is based on text recall or reading times following the point at which readers had to establish coherence between the currently processed text and an earlier presented gender-stereotype noun, these results are likely to reflect backward rather than elaborative inferencing. In addition, although plausibility judgment experiments suggest that gender stereotypes can be activated elaboratively in the absence of a bridging inference (Garnham, Oakhill, & Reynolds, 2002), they do not necessarily represent normal text processing, because the judgments may have encouraged readers to activate information that they do not routinely activate. Therefore, it is important to investigate whether stereotype information is activated before the point at which coherence needs to be established in discourse.

### Activation of Stereotypes According to the Mental Model and the Minimalist Accounts

Elaborative activation of gender stereotypes is in line with the view that during text comprehension people construct a mental model, a representation of the situation in the world described by the text (e.g., Garnham, 2001; Johnson-Laird, 1983; Zwaan & Radinsky, 1998). The representation includes both explicitly stated text information and implicit information activated on the basis of world knowledge. The latter type of knowledge may include information about gender stereotypes. Thus, when people encounter surgeon, they may elaboratively activate information that this person is male even when this inference is not necessary to establish coherence in the text (Garnham, 2001; Garnham et al., 2002; Sanford, 1985).

According to the minimalist account, however, elaborative activation of stereotype information is unnecessary for textual coherence and thus is not routinely made (McKoon & Ratcliff, 1986, 1988, 1992). The account suggests that people do not fully represent the situation that can be inferred on the basis of the text. Only two types of inferences are assumed to be made: inferences that help to build local coherence in the text and inferences that are readily available for the comprehender. The meaning of “readily available” is somewhat open to interpretation, but McKoon and Ratcliff (1992) followed Kintsch and Dijk (1978) in assuming that information that is readily available in short-term memory contains the explicitly stated words in the current or immediately preceding context and the propositions derived from them. Gender stereotypes do not belong to this category and should therefore not be inferred during text processing.

Clearly, these two accounts make opposite predictions with respect to the activation and use of gender stereotypes in online language processing. In order to differentiate between these accounts, a fine-grained measure that can follow listeners’ online processing is needed that does not require any secondary tasks to be carried out. One such method is the visual-world eye-tracking method. In this paradigm, listeners’ eye movements to visually depicted entities are recorded while they listen to stories referring to entities depicted on the screen (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Several studies have shown that linguistic properties of the spoken stimuli have a very rapid influence on listeners’ eye movements, often within 300 ms of the word onset (for an overview, see e.g., Dahan, Magnuson, & Tanenhaus, 2001). Considering that it takes approximately 200 ms to plan and execute a saccade (e.g., Matin, Shao, & Boff, 1993), there appears a close link between spoken language comprehension and eye-movement behavior. Previous studies investigating pronoun comprehension have shown that the method can also be applied to study inferencing during online pronoun resolution (e.g., Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Järvikivi, Van Gompel, Hyyönen, & Bertram, 2005). For example, Arnold et al. (2000) showed that listen-
ers use grammatical gender information during pronoun resolution, that is, they looked more often at the character on the screen whose gender matched than mismatched the pronoun’s gender. Thus, looks to the gender-congruent character can be used as an index of pronoun resolution. This study is the first to adopt this method to investigate the activation and use of gender stereotypes during spoken language processing.

**Experiment**

Two questions were addressed in this study: (1) Do comprehenders elaboratively activate gender stereotypes when they encounter generically used nouns and (2) do they use gender stereotypes to revise an earlier established coherence relation? These issues were investigated using Finnish stories such as the one presented in Table 1.

First, in order to test whether people activate gender stereotypes elaboratively, a gender-stereotype word such as *chimney sweep* occurred in Sentence 2. *Chimney sweep* is used generically to refer to any chimney sweep rather than to a specific person. Thus, it corefers with neither the female nor the male character. Eye movements to either Sinikka (female) or Mikko (male) following *chimney sweep* would therefore not indicate that listeners assume this particular person to be a chimney sweep, but instead, an increased probability to fixate the male character would suggest that participants elaboratively activate the information that chimney sweeps are usually male. In other words, comprehenders activate stereotype information even though it does not contribute to text coherence at this point, as suggested by the mental models theory (e.g., Garnham, 2001; Garnham et al., 2002; Reynolds et al., 2006). This should result in more looks to the picture of the male than the female character when first encountering the male-biased word in Sentence 2 (context 1a and 1b in Table 1). If the activation of gender stereotypes is hard to suppress (Oakhill et al., 2005), the probability to fixate the male person after hearing *chimney sweep* should increase even if listeners do not consider the visually depicted man to represent that role.

Second, we investigated whether people use stereotype information to revise an already established coherence relation. If comprehenders make only the minimum number of inferences necessary for establishing textual coherence (McKoon & Ratcliff, 1986, 1988, 1992), one might expect that they do not revise already established inferences. Once comprehenders have made a bridging inference that improves the coherence of their text representation, they may not subsequently make an additional inference that requires them to revise their representation. We tested this using Sentence 3, which contained an anaphoric verb (e.g., *kouluttauduttaan*) followed by the same stereotype noun as in Sentence 2. The verb form contained the gender-ambiguous third person anaphoric suffix –*AAN*. The suffix indicates that the verb is related to one of the two characters, but because Finnish lacks gender marking it

---

**Table 1. An example of the two versions of a male biased and a female biased story**

**Male-biased story**

<table>
<thead>
<tr>
<th>Context 1a: female salient</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. On the screen you see Sinikka, a 35-year-old woman from Jyväskylä and Mikko, a 40-year-old man from Tampere.</td>
<td></td>
</tr>
<tr>
<td>2. While doing yard work Sinikka evaluated with Mikko the dangerous situations a chimney sweep gets into on slippery roofs.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Context 1b: male salient</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. On the screen you see Mikko, a 40-year-old man from Tampere and Sinikka, a 35-year-old woman from Jyväskylä.</td>
<td></td>
</tr>
<tr>
<td>2. While doing yard work Mikko evaluated with Sinikka the dangerous situations a chimney sweep gets into on slippery roofs.</td>
<td></td>
</tr>
</tbody>
</table>

Third sentence after both 1a and 1b

3. *Kouluttauduttaan nuohoojaksi hän oli oppinut monia keinooja hoitaa työnsä turvallisesti.*

**Female-biased story**

<table>
<thead>
<tr>
<th>Context 2a: female salient</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. On the screen you see Saara, who comes from Kauniainen and Ville, who comes from Tervakoski.</td>
<td></td>
</tr>
<tr>
<td>2. Yesterday on the phone Saara talked with Ville about a make-up artist’s possibilities to create impressive theater make-ups.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Context 2b: male salient</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. On the screen you see Ville, who comes from Tervakoski and Saara, who comes from Kauniainen.</td>
<td></td>
</tr>
<tr>
<td>2. Yesterday on the phone Ville talked with Saara about a make-up artist’s possibilities to create impressive theater make-ups.</td>
<td></td>
</tr>
</tbody>
</table>

Third sentence after both 2a and 2b

3. *Opiskeltuaan meikkajaksi hän oli oppinut monia keinoja hoitaa työnsä turvallisesti.*

Study [INF. PAST + PER.SUFF.] make-up artist he/she (gender ambiguous pronoun) has learned many ways to work safely.

After having graduated to become a make-up sweep, he/she (gender ambiguous pronoun) has learned many ways to work safely.

---

1 The structure is called a converb construction and is relatively common in Finnish – especially in sentence initial position (Hakulinen, Karlsson, & Vilkuna, 1980).
can refer to either of them. In the preceding context, one of the two characters was made more salient by mentioning him/her before the other as the subject of the second sentence, thus making it the preferred antecedent of the anaphoric verb (e.g., Gernsbacher & Hargreaves, 1988; Järvičivi et al., 2005; Kaiser & Trueswell, in press). Thus, comprehenders are assumed to interpret the anaphoric suffix to refer to the salient character, which should result in more looks to this character relative to the less salient character. More interesting is what will happen when comprehenders subsequently hear the stereotype noun (now used in a nongeneric, specific sense). If people limit the number of inferences, they should not use stereotype information to revise the already established anaphoric relationship, so stereotype information should not affect looks to the characters at this point. In contrast, if people do revise their initial interpretation, we should observe an increased probability to fixate the stereotypically consistent character.

Method

Participants

Thirty undergraduates (13 male) from the University of Turku participated in the experiment for payment. All were native Finnish speakers with normal or corrected-to-normal vision.

Apparatus

Eye movements were recorded using an EyeLink II eyetracker manufactured by SR Research Ltd. The eyetracker is a near-infrared video-based tracking system combined with hyperacuity image processing. It samples pupil location at a rate of 500 Hz and has a spatial accuracy of less than 0.5°. Viewing was binocular, but only one eye was tracked.

Materials

Participants heard three-sentence stories (Table 1) that contained a male and a female character. Sentence 1 introduced the characters. In Sentence 2, they discussed a job referred to by a gender-stereotype noun (e.g., chimney sweep). Here the stereotype noun did not refer to either character, but was used generically. In half of the stories the stereotype noun was male biased and in the other half female biased. The final word of Sentence 2 referred to a noun (e.g., roof) that was depicted on the computer screen in order to attract listeners’ gaze away from the pictures of the characters prior to Sentence 3. Importantly, the words following the stereotype noun were not stereotypically biased and thus any preference to fixate a stereotypically consistent character following the stereotype noun must be due to this noun’s gender bias.

In Sentences 1 and 2, one of the characters was made more salient than the other by mentioning him/her earlier in the two sentences and by having the character appear as the subject and linguistic topic of Sentence 2. Thus, it was a strong candidate for the antecedent of the anaphoric verb (presented at the beginning of Sentence 3), which contained a gender-ambiguous third person anaphoric suffix – AAN (e.g., in kouluttauduttaaan [graduate INFINITIVE-PAST TENSE + gender ambiguous third person suffix]). All verbs were gender neutral. In contrast to English, Finnish does not differentiate between masculine and feminine for anaphoric suffixes and pronouns, so the anaphoric marking was gender ambiguous. Because Finnish is practically “genderless”, it allowed us to investigate how people use stereotype information following the anaphor.

To test whether listeners used gender-stereotype information in Sentence 3 even though they may already have established coreference between the anaphoric verb and the previous context using salience information, the stereotype noun was repeated immediately after the anaphoric verb (note that the verb’s anaphoric suffix must refer to the same person as the stereotype noun). In half of the conditions the gender of the salient character and the stereotype noun matched; in the other half they mismatched.

The experiment contained 36 experimental stories, 18 stories containing a male-biased and 18 stories a female-biased stereotype noun. Each male and female stereotype story had two versions, one in which the male character was salient and another in which the female character was salient. These versions were counterbalanced across two lists, and the items in both lists were presented in a randomized order. Participants were randomly assigned to the lists. Fifty-nine filler stories were presented in each list. Fillers were two- or three-sentence long stories and were similar in structure to the experimental items: The stories introduced two interacting characters by giving their names and some additional information about them. In some stories the two characters were talking about occupations; only occupations rated as gender neutral in a rating test were used in the filler sentences. We used as many fillers as possible (keeping the length of the experiment within a reasonable limit) in order to hide the purpose of the experiment.

A rating test with 20 undergraduates was conducted to find strongly biased male and female stereotype nouns. In total, 124 nouns (occupations and role names) were rated for gender stereotypicality using a 7-point scale, with 1 indicating an extremely masculine and 7 an extremely feminine occupation. The 18 most male-biased and 18 most female-biased nouns were chosen for the experiment. The mean ratings of the chosen nouns were below 2.04 and above 5.48, respectively.

The stories were read with a normal sentence intonation by a male speaker and recorded at a sampling frequency of 22,050 Hz using Praat (Boersma, 2001). For the analyses, the acoustic onsets of each word in each trial were marked.

2 The names were the most commonly used male and female names in Finnish. The pictures were collected from the Internet, but the names and other personal information were imaginary. The persons looked clearly male or female according to our intuitions.
in the sound files using visual spectrogram and auditory information.

While the participants listened to the stories, they were shown four colored photographs on the computer screen: A picture of the male and female characters, the place or instrument mentioned in Sentence 2 (snow shovel depicting yard work), and the noun mentioned last in Sentence 2 (roof). The pictures were 150 × 150 pixels in size and located in the four corners of an imaginary square; the positions were counterbalanced between items (The pictures are available from the first author upon request).

Procedure

Participants sat approximately 70 cm from a 21 inch color monitor. They were asked to carefully listen to the stories and at the same time follow with their gaze the pictures depicted on the screen, in the same order they appear in the story. Prior to the experiment, a 9-point calibration grid was used to calibrate the eye-tracker. The pictures appeared on the screen 50 ms before the sound file onset and remained there until the sound file ended. After about every tenth story the participant had to continue the story with one sentence using the persons and places mentioned therein. This task ensured that participants listened to the stories for comprehension.

Results

Looks to the pictures of the male and female character were analyzed for three regions: (1) From the stereotype noun onset in Sentence 2 (nuohooja “chimney sweep”) until the onset of the sentence-final word, to investigate whether listeners elaboratively activate stereotype information after hearing a generically used role name; (2) from the anaphoric verb onset in Sentence 3 (kouluttauduttuaan “after having graduated”) until the stereotype word onset, to test whether people use salience to link the anaphoric verb to the previous context; and (3) from the stereotype noun onset in Sentence 3 (nuohoojaksi “chimney sweep”) until the next word (a pronoun) onset in order to determine whether stereotype information was used to revise the coherence relation between the anaphoric verb and the preceding context. The acoustic onset of each target word was determined on a trial-by-trial basis.

First, we divided each analysis region in 30-ms time frames starting from the region onset. Next, for each 30-ms frame we calculated the probability that each of the two characters was fixated (out of all trials). Finally, we averaged these proportions across each of the three analysis regions defined above. Thus, this yielded a proportions-of-looks measure for the male and female character.

For all the three analysis regions, we conducted 2 × 2 ANOVAs with salience (salient vs. nonsalient character) and stereotype (stereotypically consistent vs. inconsistent character) as within-participants and within-items variables. As it takes around 200 ms to plan and execute a saccade (Matin et al., 1993), we started the analyses from the 210 ms time frame after the target word onset.

The Stereotype Noun in Sentence 2

Across the stories, the shortest interval from the stereotype noun onset until the final word onset was 2700 ms; thus, the time window of analysis was defined as 210–2700 ms. Trials in which participants fixated one of the characters at the stereotype noun onset were included in the analyses, because one of the characters was just mentioned and listeners were expected to fixate his/her picture. Because of this, we expected to find a “preference” for the nonsalient character; a saliency effect here would simply indicate that people were still fixating the last mentioned character following its mention, but this is not of theoretical interest.

We found a marginal main effect of salience in the participant analysis and a significant effect in the item analysis, $F_1(1, 29) = 3.81, p = .06, \eta^2 = .116$; $F_2(1, 35) = 39.54, p < .001, \eta^2 = .530$, denoting more looks to the recently mentioned character (20.4% vs. 30.6%, respectively). More importantly, we also found a significant main effect of stereotype, $F_1(1, 29) = 7.44, p < .05, \eta^2 = .204$; $F_2(1, 35) = 8.73, p < .01, \eta^2 = .200$, denoting more looks to the stereotypically consistent than inconsistent character and suggesting elaborative activation of gender-stereotype information. The percentages of looks to the stereotypically consistent and inconsistent characters during the stereotype words both in Sentences 2 and 3 are presented in Figure 1. The interaction did not reach significance, $F_1(1, 29) = 1.51, p > .2, \eta^2 = .050$; $F_2(1, 35) < 1$.3

The Anaphoric Verb in Sentence 3

The shortest interval from the anaphoric verb onset until the stereotype noun onset was 440 ms; given that the time frames for which we calculated the looks were 30 ms, the analysis region was defined as 210–420 ms. Because the noun that was mentioned at the end of Sentence 2 was depicted on the screen, listeners were expected to fixate the picture of this noun at the verb onset rather than either of the characters. If they did fixate one of the characters, such trials (7.9%) were excluded.

We found a significant main effect of salience, $F_1(1, 29) = 4.83, p < .05, \eta^2 = .143$; $F_2(1, 35) = 4.78, p < .05, \eta^2 = .120$, denoting more looks to the salient than nonsalient character (8.3% vs. 4.7%, respectively). The result shows that participants used salience to link the anaphoric verb to the previous context: As expected, salience

---

3 Follow-up analyses in which the gender bias of the stereotype noun was included as a between-items variable showed that gender bias did not interact with stereotypicality in any time window, indicating that the effect of gender stereotype was similar for both stereotypically female and male occupations.
was used to establish a coherence relationship in a backward fashion. Interestingly, this effect occurred during the anaphoric verb, suggesting that people made this link even before hearing the anaphoric marker that occurred at the end of the word. No stereotype effect occurred in this region, because the stereotype noun had not yet been mentioned. The results showed that neither the main effect of stereotype nor the interaction was significant \( F_s < 1 \).

**The Second Mention of the Stereotype Noun in Sentence 3**

Because the shortest interval between the stereotype noun and pronoun onset was 620 ms, the analysis was carried out for the 210–600 ms region. As we wanted to include trials in which listeners had established coherence rapidly after the verb onset, trials in which participants fixated one of the characters at the stereotype noun onset were included.

We found a significant main effect of salience, \( F_1(1, 29) = 46.78, p < .001, \eta^2 = .617; F_2(1, 35) = 148.20, p < .001, \eta^2 = .809 \), denoting more looks to the salient than less salient character (59.4% vs. 12.2%, respectively). More importantly, we also found a significant main effect of stereotype, \( F_1(1, 29) = 4.64, p < .05, \eta^2 = .138; F_2(1, 35) = 6.12, p < .05, \eta^2 = .149 \), denoting more looks to the stereotypically consistent character (see Figure 1). Interestingly, the stereotype activation occurred even before the stereotype word’s acoustic offset. These effects suggest that even though salience was initially used in building a coherent text representation while participants heard the anaphoric verbs, they readily used gender-stereotype information to revise their initial interpretation. The nonsignificant interaction \( F_8 < 1 \) implies that the effect of gender stereotype was equally strong for the salient and the nonsalient character.

**Discussion**

Our visual world eye-tracking experiment showed that when listeners heard a generically used gender-stereotype noun (chimney sweep), they were more likely to look at a picture of a male than a female character, indicating that they activated stereotype information associated with this word elaboratively, even though it did not refer to any particular male. Similar suggestions have been made earlier (e.g., Garnham, 2001, Garnham et al., 2002, Oakhill et al., 2005), but previous studies did not clearly distinguish between elaborative activation and the use of stereotype information for bridging inferences.

Our finding that comprehenders elaboratively activate stereotype information fits well with the mental models theory (e.g., Garnham, 2001), which assumes that people make elaborative inferences in order to form a detailed representation of the situation that is described in the text. It fits less well with the minimalist account (McKoon & Ratcliff, 1986; 1992), which suggests that inferences not needed for local coherence or based on not easily available information are not made online. Assuming that easily available information is limited to information that is explicitly stated in the text and the propositions derived from them (McKoon & Ratcliff, 1992), the minimalist account does not predict activation of gender stereotypes unless it contributes to textual coherence. Instead, our results from the first mention of the stereotype noun suggest that the processor activates stereotype information in order to update and modify the situation model in an incremental manner during discourse processing (see also Garnham, 2001; Johnson-Laird, 1983; Zwaan & Radvansky, 1998).

The experiment also showed that comprehenders made a bridging inference by looking more at the salient character than nonsalient character during the anaphoric verb in Sentence 3, as was expected based on the earlier studies that suggested that people tend to interpret the anaphoric expressions as referring to the most salient character of the previous discourse (Gemsbacher & Hargreaves, 1988; Järvisvää et al., 2005; Kaiser & Trueswell, in press). More interestingly, comprehenders subsequently revised this coherence relationship using gender stereotype information. The minimalist account suggests that people make only the minimum number of inferences in order to resolve local incoherence, so one might expect that once coherence has been established using saliency information, comprehenders do not.

**Figure 1.** The percentage of fixations to the stereotypically consistent and inconsistent character following the onset of the stereotype nouns in Sentences 2 and 3. The error bars show the standard error of mean.
make an additional inference using stereotype information in order to revise their initial interpretation. Thus, these results fit better with the mental models theory, which assumes that comprehenders make inferences continuously during discourse processing.

Both salience and gender-stereotype information were used very quickly in the third sentence of the experimental texts: The salience effect was found at the anaphoric verb and the gender-stereotype effect at the stereotype word. The effect of salience at the anaphoric verb implies that people did not wait until they heard the anaphoric suffix before starting to build a link between the most salient character in the preceding context and the verb phrase. Similarly, the gender-stereotype effect at the stereotype noun was observed before the noun’s acoustic offset. These rapid effects demonstrate that language processing is highly incremental: At each word, people update their interpretation of the text in order to achieve maximal coherence. We propose that this supports a principle that we refer to as the Principle of Maximal Coherence: People try to maximize coherence in the text by establishing and updating coherence relations as soon as they can. This explains why salience and stereotype information had such rapid effects in Sentence 3.

The current experiment demonstrates that the visual-world paradigm is an informative method for investigating (gender-stereotype) inferences. In particular, it allowed us to investigate whether stereotype information is activated when it does not contribute to textual coherence. Investigating elaborative gender-stereotype inferences during discourse processing has been difficult with other experimental settings, because they usually require a mismatch between a stereotype noun and another expression (e.g., a gender-marked pronoun) appearing later in the discourse. The activation of gender-stereotype information at this later point may be the result of a bridging inference rather than elaborative activation of stereotype information.

Participants’ task in the current experiment was to carefully listen to the stories and at the same time follow with their gaze the pictures on the screen, in the same order they appeared in the story (for a similar type of instructions, see e.g., Altmann & Kamide, 1999; Arnold et al., 2000; Järviö et al., 2005). Our reason for using these instructions was that our pictures were static, their positioning with respect to one another was pseudorandom, and they did not visually depict the story plot in any way, so it may not have been clear to the participants that the pictures were related to the stories. Thus, without the instructions participants’ gaze behavior may not necessarily have reflected their story processing. Note, however, that only about half of the noun phrases (i.e., a clear minority of all words) presented in the spoken stories was depicted on the screen. Our task mimics what comprehenders spontaneously do when they listen to stories about persons and objects that are either physically present or presented as an animation (e.g., on TV), where they also identify the persons and objects present in their visual environment and relate them to the spoken discourse. It is nevertheless possible that our instructions make the relation between the story and the pictures very explicit and may thus have encouraged participants to make inferences they would not usually make. However, we consider it unlikely for four reasons. First, research has shown that effects measured by the visual-world paradigm occur regardless of the type of instructions used (Altmann & Kamide, 1999). Second, research by Dunning and Sherman (1997) and Oakhill et al. (2005) suggests that the activation of gender-stereotype information may not be sensitive to task instructions. Third, as noted above, only a clear minority of all words mentioned in the stories were visually depicted on the screen; thus, the task did not encourage participants to search for a picture equivalent of each word. Fourth, the filler stories further ensured that gender-stereotypical occupations appeared relatively rarely in the stories.

In sum, the current study shows that listeners activate stereotypes during online discourse processing even if this does not result in greater discourse coherence. In addition, they also use gender stereotypes to update previously established coherence relations. It also shows that the visual-world eye-tracking paradigm is a very useful method to study both elaborative activation of semantic features and backward inferences during discourse comprehension and to disentangle these two processes from each other.

Acknowledgments

This study was supported by grants awarded by the Academy of Finland (Grant No. 213731), the Graduate School of Psychology in Finland, the Finnish Cultural Foundation, and the Jenny and Antti Wihuri Foundation, all to the first author.

References

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition:


Received August 6, 2008
Revision received March 6, 2009
Accepted March 6, 2009
Published online: November 6, 2009

Pirita Pyykko¨nen

FR 4.7 Psycholinguistics
Department of Computational Linguistics
Saarland University
Building C7.1
66041 Saarbrücken
Germany
Tel. +49 681 302 6557
Fax +49 681 302 6561
E-mail pirita@CoLi.Uni-SB.DE