

## Beyond sentence boundaries: Grammatical gender information in short-term recall of texts

JUDITH SCHWEPPE, RALF RUMMER, AND ANNE FÜRSTENBERG  
*University of Erfurt, Erfurt, Germany*

In the context of text recall it is often stated that surface representations are quickly forgotten. Jarvella (1971) and Sachs (1967) argued that what is retained beyond sentence boundaries is meaning information, whereas lexical and syntactic information is only available for the most recent constituent. We based a text recall experiment on Jarvella's paradigm, in order to demonstrate that both meaning and grammatical gender information contribute to the recall of short text passages. Although it is known that grammatical gender information is used in anaphor resolution, even if noun and pronoun do not belong to adjacent sentences, there is no direct evidence for a gender contribution to text memory so far. The present experiment demonstrates that grammatical gender information, even when of no semantic importance, is retained beyond sentence boundaries and can contribute to the memory representation of subsequent text. All materials and additional statistics may be downloaded from [mc.psychonomic-journals.org/content/supplemental](http://mc.psychonomic-journals.org/content/supplemental).

We often read in psychology textbooks that in memory for discourse, a representation of surface information (such as word form and syntactic structure) is available only for the most recent constituent of a text (e.g., Anderson, 1990, pp. 366ff). It is assumed that, beyond clause and sentence boundaries, the processing system deals with condensed meaning information. However, research on *text comprehension* gives rise to doubts about these assumptions concerning *text memory*. The purpose of the present article is to address the role of surface information in text memory directly by investigating the influence of grammatical gender congruency on lexical substitution errors in short-term recall of text passages.

This article is inspired by the idea that for short-term retention of verbal material, those types of information are available which are activated or generated in the course of language comprehension, and that language processing thus determines what kind of information can potentially contribute to a short-term memory task (e.g., phonological, morphosyntactic, or lexical-semantic information). In addition, we assume that language processing indicates (or even determines) for how long certain types of representations are available. In other words, how long a type of information is retained in short-term memory does not depend on whether or not it is conceptual information, but on how long it is used for understanding or producing verbal utterances. In order to test this assumption, the present article asks whether grammatical gender information is retained in short-term memory beyond sentence boundaries.

Why grammatical gender? In German (our study language) all nouns belong to one of three gender categories:

masculine, feminine, or neuter. Particularly for inanimate objects, the grammatical and conceptual genders of a noun are often *not* related. For instance, to produce German beer, which is neuter (das<sub>[neut]</sub> Bier), one needs masculine hops (der<sub>[masc]</sub> Hopfen), feminine barley (die<sub>[fem]</sub> Gerste), and neuter water (das<sub>[neut]</sub> Wasser). Furthermore, even near-synonyms can belong to different gender categories ("der<sub>[masc]</sub> Ozean," "die<sub>[fem]</sub> See," and "das<sub>[neut]</sub> Meer" all mean "the sea" or "the ocean"). Hence, German grammatical gender can be considered surface rather than conceptual information. Nevertheless, there are studies on text processing that indicate a role for grammatical gender information even beyond sentence boundaries. For instance, Garnham, Oakhill, Ehrlich, and Carreiras (1995, Experiment 3) demonstrated that in French, pronouns may be interpreted more easily when the referent can be identified unambiguously on the basis of grammatical gender, even when antecedent and pronoun are separated by an entire sentence (for similar results concerning pronoun production, see Meyer & Bock, 1999, Experiment 1). This is the case for both animate and inanimate referents. Because we assume that the way in which linguistic representations are used in language processing indicates how long they are accessible in a memory task, we predict an influence of grammatical gender also in tasks that directly address short-term memory beyond sentence boundaries. The (standard) assumption that only conceptual information is retained beyond sentence boundaries, however, would give rise to a contrary prediction, at least as long as inanimate nouns are investigated. We wish to demonstrate the use of grammatical gender information in a memory test by applying the paradigm that provides evidence for

the loss of surface information at sentence boundaries (Jarvella, 1971).

In Jarvella's (1971) paradigmatic experiments, participants had to listen to texts. Whenever the presentation was interrupted, they were to recall as much text as they could, as accurately as possible. Jarvella's manipulation concerned the final three clauses preceding each interruption: The middle clause could be connected either to the previous or to the final clause. In the former case it was thus not part of the most recent sentence, whereas in the latter case it was. The wording of the middle clauses did not differ between these two conditions (see Examples 1A and 1B below, from Jarvella, 1971, p. 411; the middle clause is italicized).

1A. Taylor did not reach a decision until *after he had returned to Manhattan*. He explained the offer to his wife.

1B. With this possibility, Taylor left the capital. *After he had returned to Manhattan*, he explained the offer to his wife.

Verbatim recall of the final clause was always more accurate than that of the previous ones and, critically, recall performance for the middle clause strongly depended on which clause it was attached to: The middle clause was recalled as poorly as the previous clause was when it belonged to the same sentence. Yet recall of the middle clause improved significantly when it was connected to the final one and was thus part of the most recently processed sentence. Jarvella (1971) concludes that it "appears that the propositional meaning of sentences was remembered shortly after they were heard, although, as measured by verbatim recall, the form of the sentences was quickly forgotten" (p. 415).

As indicated above, we investigated whether grammatical gender information (as a type of surface information) is retained beyond sentence boundaries. To this end, we have combined the Jarvella (1971) text recall paradigm with the *intrusion paradigm* (Potter & Lombardi, 1990), a modified version of which was used by Schweppe and Rummer (2007).

Schweppe and Rummer (2007) presented sentences for immediate recall. Each sentence contained a target word that was not an optimal fit for the sentence content, for example:

Die Wissenschaftler wollten mit *einem Test*<sub>[masc]</sub> mehr über das Gedächtnis erfahren. (By means of *a test*, the scientists attempted to learn more about memory.)

Prior to the sentence, a list of five unrelated words was presented for subsequent recognition. Whether this distractor list contained an unrelated control word (in one third of the trials) or a *lure word*, a (near-)synonym for the target word and a better fit for the content of the sentence, was subject to manipulation. In one third of the trials, the lure word was congruent with the target word in grammatical gender (in the example above: "Versuch"<sub>[masc]</sub>/

study) and in the final third it was gender incongruent ("Experiment"<sub>[neut]</sub>/experiment).<sup>1</sup> The dependent variable was the proportion of cases in which participants erroneously reproduced the sentence with either lure word. These intrusion errors occurred spontaneously (when neither lure word was presented) or were induced (when either the gender-congruent or the gender-incongruent lure word was part of the distractor list). The main finding was a *gender congruency effect*; that is, the gender-congruent lure word induced more intrusions than the gender-incongruent one did. Crucially, such a gender difference was observed even when there was no discrepancy between the form of the target determiner and the determiner that had to be used in the case of a gender-incongruent intrusion (e.g., in the example above the dative determiner "einem" goes along with "Test"<sub>[masc]</sub> and "Versuch"<sub>[masc]</sub> as well as with "Experiment"<sub>[neut]</sub>; Schweppe & Rummer, 2007, Experiment 3). These results demonstrate that grammatical gender information contributes to short-term retention of sentences.

In the present study, participants had to recall short texts that ended in three-clause passages comparable with those in Jarvella's (1971) experiments. Each text existed in two versions; the penultimate clause formed a sentence with either the previous or the final clause. A further manipulation was added: In the style of the modified intrusion paradigm, the critical middle clause contained a (nonanimate) target word for which either a gender-congruent or a gender-incongruent lure word was presented in a prior context sentence (see Table 1 for examples).

The experiment was based on a two-factorial design, with lure condition and middle clause attachment as the two within-subjects factors. With respect to lure condition, we manipulated the context passage so that an unrelated control word (without lure), a gender-incongruent lure word, or a gender-congruent lure word was presented prior to the final three clauses. The middle clause of the critical three clauses was connected to either the final or the previous clause.

We expected a gender congruency effect, which would indicate the contribution of grammatical gender information to text recall. Whereas both types of lure words are activated because of their conceptual appropriateness and lexical priming, the activation level of the gender-incongruent lure word ought to be reduced due to the conflicting gender information from the target word. This is not the case for a gender-congruent lure word. Since we assume that grammatical gender information is retained beyond sentence boundaries, a gender congruency effect is expected for both middle clause attachment conditions. In contrast, if one assumes that a representation of surface information—and thus of grammatical gender information, for instance—is only available for the most recent sentence, the gender congruency effect should be restricted to the condition in which the critical middle clause is part of it (see Noordman & Vonk, 1999, for a review of this position; for a different point of view, see Von Eckardt & Potter, 1985).

**Table 1**  
**An Example Item As Used in the Experiment, Consisting of Context Passage and Target Region (German Original, Literal Translation, and Glosses)**

	German Original		Literal Translation		Glosses	
Context Passage	Denise hatte sich für den Samstag fest vorgenommen, einen Stadtbummel zu machen. Am frühen Morgen schon ließ sie sich mit ihrem <i>Auto</i> <sub>(gender-congruent lure word, neut)/<i>Wagen</i><sub>(gender-incongruent lure word, masc)/<i>Hund</i><sub>(control word)</sub> abholen und nahe der Fußgängerzone absetzen.</sub></sub>		Denise had herself for the Saturday definitely planned a stroll through the town to make. At early morning already let she herself with her <i>automobile</i> <sub>(gender-congruent lure word, neut)/<i>car</i><sub>(gender-incongruent lure word, masc)/<i>dog</i><sub>(control word)</sub> pick up and near the pedestrian mall drop.</sub></sub>		For Saturday, Denise had planned to take a stroll through the town. Early in the morning she was picked up with her <i>automobile</i> <sub>(gender-congruent lure word)/<i>car</i><sub>(gender-incongruent lure word)/<i>dog</i><sub>(control word)</sub> and was dropped near the pedestrian mall.</sub></sub>	
Target Region	(a) Der Parkplatz füllte sich langsam, während der Chauffeur das Fahrzeug <sub>(target word, neut)</sub> bewachte. Denise ging zum Einkaufen in die Innenstadt.	(b) Der Parkplatz füllte sich langsam. Während der Chauffeur das Fahrzeug <sub>(target word, neut)</sub> bewachte, ging Denise zum Einkaufen in die Innenstadt.	(a) The car park filled itself slowly while the chauffeur the vehicle <sub>(target word, neut)</sub> watched. Denise went for shopping into the city.	(b) The car park filled up slowly. While the chauffeur the vehicle <sub>(target word, neut)</sub> watched, Denise went for shopping into the city.	(a) The car park filled up slowly while the chauffeur was watching the vehicle <sub>(target word)</sub> . Denise went downtown for shopping.	(b) The car park filled up slowly. While the chauffeur was watching the vehicle <sub>(target word)</sub> Denise went downtown for shopping.

## METHOD

### Participants

Fifty-four students of Saarland University (all of them native speakers of German) participated in the experiment. They either received course credit or were paid.

### Materials

Thirty short texts were constructed. Each of them ended in a three-clause structure, with two versions of this target region: Either the first and the second clause formed a sentence and the final clause stood alone, or the first clause stood alone and the second and the third clause formed a sentence. Regardless of this structure manipulation, except for obligatory word order differences in the final clause, the wording of the sentences was the same in both versions. The middle clause contained a target word for which two more suitable lure words existed (one gender congruent and one gender incongruent).

Each of the three-clause passages was preceded by a context passage consisting of two to five sentences. The length of these passages varied, so participants could not guess which part of the text they would have to recall, but each three-clause passage was presented with the same context passage in each condition (except for the lure/control words). The use of text materials allowed either a lure word or a control word to be placed unobtrusively within the text. It always appeared in the sentence preceding the target region (see Table 1). All three-clause passages and lure words are available at [mc.psychonomic-journals.org/content/supplemental](http://mc.psychonomic-journals.org/content/supplemental).

### Stimulus Norming

Several pretests were conducted to ensure that grammatical gender was the only critical difference between the two types of lure words. Twenty native speakers of German (none of whom participated in the experiment) rated the semantic similarity between lure and target words (in isolation) on a scale ranging from 1 (*no semantic similarity*) to 10 (*identical meaning*). In addition, both types of lure words had to be an equally good fit for the sentence content, which was measured by a sentence completion task (cloze probability). The two versions of the three-clause passages with target determiners and nouns omitted were completed by 40 more participants (so that 20 participants completed each of the two middle clause attachment versions).

Furthermore, the two lure types were matched with respect to word length (in characters and in syllables) and word frequency (according to the CELEX Lexical Database; Baayen, Piepenbrock, & van Rijn, 1993). Since grammatical gender and phonological struc-

ture are sometimes related (e.g., the suffix “-keit” always indicates a feminine noun), we also matched phonemic overlap between the final syllables of lure and target words. None of these tests yielded significant mean differences between the gender-congruent and the gender-incongruent lure words. All results are given in Table 2.

### Procedure

For each participant, one third of the texts (i.e., 10) included a gender-congruent lure word, one third a gender-incongruent lure word, and one third a control word semantically unrelated to the target word. Across participants, each text appeared equally often with a gender-congruent lure word, a gender-incongruent lure word, and a control word. Each participant recalled 15 texts in which the critical clause was connected to its preceding clause, and 15 texts in which it was connected to its subsequent clause. Text assignment to one of the middle clause attachment conditions was counterbalanced across participants.

Each trial was started by pressing the space bar of the computer keyboard. Three asterisks then appeared in the center of the screen for 300 msec, followed by a blank screen for 350 msec. Then the text was presented auditorily at normal speech rate (about six syllables per second). After a 500-msec break, in which a row of percentage signs was displayed, an auditory cue indicated the starting point for text recall. The cue always consisted of the beginning (up to the first noun) of the target region. Finally, participants had to recall the last two sentences of the text as accurately as possible. Three practice trials were presented before the experiment started.

The experiment lasted about half an hour. Participants' utterances were recorded and transcribed. The number of correct reproductions of the target word, of gender-incongruent lure intrusions, and of gender-congruent lure intrusions served as dependent variables. In addition, the number of target omissions<sup>2</sup> was recorded. Recall accuracy of the final two clauses will be reported as a further dependent measure.

## RESULTS

We will first report the results with respect to lure intrusions and, second, present analyses referring to general recall performance.

### Lure Intrusions

Table 3 shows absolute frequencies and percentages of the four different response types (i.e., correct target

**Table 2**  
**Means and Standard Deviations for the Target and Lure Words in the Pretests**

	Target Word		Gender-Congruent Lure Word		Gender-Incongruent Lure Word	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
	Semantic similarity (1–10)	—	—	5.9	1.7	5.7
Cloze probabilities (%): Target in penultimate sentence	0.8	1.9	22.0	22.7	30.3	21.7
Cloze probabilities (%): Target in ultimate sentence	1.0	2.4	22.0	20.7	28.3	23.5
Word frequency (per million)	561	1,296	431	559	389	440
Word length (in characters)	6.5	2.3	6.5	3.1	6.5	2.1
Word length (in syllables)	2.1	0.9	2.0	0.8	2.1	0.8
Shared phonemes between target and lure (final syllable)	—	—	0.6	0.8	0.6	0.7

reproductions, gender-incongruent intrusions, gender-congruent intrusions, and target omissions) per lure condition (i.e., without lure, with gender-incongruent lure, and with gender-congruent lure) and middle clause attachment condition (i.e., target word part of the penultimate sentence vs. part of the ultimate sentence). For both types of lure words, induced intrusions were more frequent than spontaneous ones. Crucially, induced gender-congruent intrusions occurred more frequently than induced gender-incongruent ones did, even when the target word was part of the penultimate sentence.

Statistical analyses were based on hierarchical log-linear models (e.g., Howell, 2002) using SPSS LOGLINEAR (SPSS for Windows, Version 12.0). Log-linear analyses are appropriate for determining the fit between observed and expected cell counts (as are standard cross-tabulation chi-square tests) and provide the opportunity to test for main effects and interactions in a multifactorial design (e.g., Scheepers, 2003). Even more important, log-linear analyses allow the analysis of categorical frequencies such as those obtained in the present experiment. They do not rely on parametric assumptions concerning the dependent variable and do not require factor levels to be independent of one another, which is important for the comparison of correct target reproductions and intrusion errors (though not for the clause recall data).

Partial associations likelihood-ratio chi-squares ( $\chi^2_{LR}$ ) were computed. They derive from tests in which expected frequencies are calculated from all effects competing with the respective effect at the same level of the effect hierarchy. For each analysis, we will present results with participants included as a factor ( $\chi^2_{LR(1)}$ ) and results with items included ( $\chi^2_{LR(2)}$ ) in order to assess the generality of effects across individuals and materials. The inclusion of either participants or items implies that expected fre-

quencies for the interactions of interest are calculated in competition with interactions involving the participant or item factor, at the same hierarchical level. Although this procedure is not equivalent to a random effects procedure in an ANOVA, it allows to some degree for the adjustment of interindividual (or interitem) variation in log-linear analysis (see also Scheepers, 2003).

To test the hypotheses, three types of analyses are required (1A–1C). As in standard chi-square tests, each of these analyses compares intrusion errors and correct reproductions of the target word as a function of lure condition and middle clause attachment. In addition, we will report three more comparisons. The first one directly tests whether there is a gender congruency effect when the target word is part of the penultimate sentence (2A). The other two analyses investigate the influence of middle clause attachment on the intrusion effect for gender-congruent lure words (2B) and for gender-incongruent lure words (2C).

#### 1A. Intrusion effect for gender-congruent lure words.

The first analysis compares gender-congruent lure intrusions and correct target reproductions in the condition in which neither lure word appeared in the text and in the condition in which the gender-congruent lure word appeared in the text. An intrusion effect for gender-congruent lure words would be reflected in a higher proportion of induced than of spontaneous gender-congruent intrusions. As the relative frequency of intrusions and correct target reproductions is analyzed, this means that we expect a two-way interaction between lure condition and response type. In these analyses, response type is treated as a factor. Thus, an interaction between a manipulated factor and the response type factor relates to a main effect of the respective factor in an ANOVA. (Analogously, a log-linear three-way interaction is similar to an ANOVA two-way

**Table 3**  
**Absolute Frequencies and Percentages of Gender-Congruent and Gender-Incongruent Lure Intrusions, Correctly Reproduced Target Words, and Target Omissions As a Function of Lure Condition and Middle Clause Attachment**

	Target Word Part of the Penultimate Sentence			Target Word Part of the Ultimate Sentence		
	Without Lure	With Gender-Congruent Lure	With Gender-Incongruent Lure	Without Lure	With Gender-Congruent Lure	With Gender-Incongruent Lure
Gender-congruent intrusions	20 (7.4%)	58 (21.5%)	16 (5.9%)	20 (7.4%)	49 (18.2%)	18 (6.7%)
Gender-incongruent intrusions	11 (4.1%)	5 (1.9%)	38 (14.1%)	8 (3.0%)	3 (1.1%)	22 (8.2%)
Correct target reproductions	163 (60.4%)	153 (56.7%)	158 (58.5%)	196 (72.6%)	176 (65.2%)	181 (67.0%)
Target omissions	76 (28.2%)	54 (20.0%)	58 (21.5%)	46 (17.0%)	42 (15.6%)	49 (18.2%)

interaction.) The expected two-way interaction between lure condition and response type turned out to be significant [ $\chi^2_{LR(1)}(1) = 33.55, p < .001$ ;  $\chi^2_{LR(2)}(1) = 35.52, p < .001$ ]. Participants produced more gender-congruent intrusions when the lure word had been presented in the context passage (19.8%) than when it had not been presented (7.4%). The target was reproduced more often in the “without lure” condition (66.5%) than in the “gender-congruent lure” condition (60.9%). In total, this reflects an intrusion effect for gender-congruent lure words. The three-way interaction between lure condition, response type, and middle clause attachment was not significant [all  $\chi^2_{LRS}(1) < 1.3$ , all  $ps > .25$ ], which indicates that the intrusion effect was not modulated by the syntactic status of the target clause. The three-way interaction between lure condition, response type, and the participant factor was not significant in this analysis either ( $p > .1$ ), but a significant three-way interaction with items [ $\chi^2_{LR(2)}(1) = 47.89, p = .015$ ] suggests that the size of the intrusion effect varied between items.

**1B. Intrusion effect for gender-incongruent lure words.** In the second analysis, gender-incongruent lure intrusions and correct reproductions of the target word in the control condition and in the condition in which the gender-incongruent lure word was presented are included. An intrusion effect for gender-incongruent lure words may be demonstrated if the proportion of induced gender-incongruent intrusions is higher than the proportion of spontaneous gender-incongruent intrusions. Indeed, the analyses revealed a significant two-way interaction between lure condition and response type [ $\chi^2_{LR(1)}(1) = 25.6, p < .001$ ;  $\chi^2_{LR(2)}(1) = 24.79, p < .001$ ]. There were more gender-incongruent intrusions when the incongruent lure word was part of the context passage (11.1%) than when an unrelated word was presented instead (3.5%). Correct target reproductions were more frequent in the “without lure” (66.5%) than in the “incongruent lure” condition (62.8%). This interaction was modulated neither by the attachment of the target clause (both  $ps > .1$ ) nor by the participant or item factor (both  $ps > .3$ ).

**1C. Induced gender-incongruent versus induced gender-congruent intrusions.** Most important is the third analysis. It contrasts gender-incongruent intrusions and correct target reproductions in the condition in which the gender-incongruent lure word was presented and gender-congruent intrusions and correct reproductions of the target word when the gender-congruent lure word was presented. We expect a gender congruency effect (i.e., more induced gender-congruent intrusions than induced gender-incongruent intrusions). There was indeed a significant interaction between lure condition and response type [ $\chi^2_{LR(1)}(1) = 11.76, p = .001$ ;  $\chi^2_{LR(2)}(1) = 13.96, p < .001$ ]. As expected, the presentation of the congruent lure word induced more intrusions (19.8%) than did the presentation of the incongruent lure word (11.1%). Correct responses in the incongruent (62.8%) and in the congruent (60.9%) lure condition did not differ. Again, the three-way interaction between lure condition, response type, and middle clause attachment did not reach significance [ $\chi^2_{LR(1)}(1) < 1, p > .4$ ;  $\chi^2_{LR(2)}(1) = 2.66, p > .1$ ]; neither

did the three-way interaction between lure condition, response type, and participants ( $p > .2$ ). Yet there was a significant three-way interaction between lure condition, response type, and items [ $\chi^2_{LR(2)}(1) = 62.52, p < .001$ ].

**2A. Induced gender-congruent versus induced gender-incongruent intrusions when the target belongs to the penultimate sentence.** These analyses tested for a gender congruency effect in the condition in which an additional sentence boundary separated the presentation of the target word and its recall. There was a significant two-way interaction between lure condition and response type [ $\chi^2_{LR(1)}(1) = 4.21, p = .04$ ;  $\chi^2_{LR(2)}(1) = 4.03, p = .045$ ]. Participants produced 14.1% induced incongruent intrusions and 21.5% induced congruent intrusions. They correctly reproduced the target word in 58.5% and 56.7% of the cases, respectively. The three-way interaction with participants was not significant ( $p > .2$ ), but again, the interaction with items was [ $\chi^2_{LR(2)}(1) = 67.33, p < .001$ ].

**2B. Effect of sentence boundaries on the induction of gender-congruent intrusions.** We compared gender-congruent intrusions and correct target reproductions in the gender-congruent lure condition in the two middle clause attachment conditions. Middle clause attachment and response type did not interact significantly [ $\chi^2_{LR(1)}(1) = 2.14, p > .5$ ;  $\chi^2_{LR(2)}(1) = 1.6, p > .2$ ]. None of the higher order interactions was significant either (all  $ps > .3$ ). The proportion of induced congruent intrusions and correct target reproductions did not differ depending on which clause the target clause was connected to.

**2C. Effect of sentence boundaries on the induction of gender-incongruent intrusions.** For gender-incongruent intrusions, the two-way interaction between middle clause attachment and response type was significant [ $\chi^2_{LR(1)}(1) = 6.55, p = .038$ ;  $\chi^2_{LR(2)}(1) = 6.76, p = .009$ ]. There were more induced incongruent intrusions (14.1%) and fewer correct target reproductions (58.5%) when the target word was part of the penultimate sentence than when it belonged to the final clause (8.2% and 67%, respectively). The three-way interactions with participants ( $p > .99$ ) or items ( $p > .1$ ) did not reach significance.

**Clause recall.** In addition, we recorded the proportion of correctly reproduced words per clause (irrespective of order and excluding the target noun phrase). Since the recall cue constituted a substantial part of the first clause, this clause was not analyzed.

An ANOVA, including the factors middle clause attachment and clause position, revealed significant main effects both for middle clause attachment [ $F_1(1,53) = 19.14, p < .001$ ;  $F_2(1,29) = 12.48, p = .001$ ] and for clause position [ $F_1(1,53) = 8.19, p = .006$ ;  $F_2(1,29) = 4.24, p = .048$ ]. Yet these effects were modulated by a significant two-way interaction [ $F_1(1,53) = 5.77, p = .02$ ;  $F_2(1,29) = 6.56, p = .016$ ].

Planned comparisons showed that recall of the middle clause was significantly better when it formed a sentence with the final clause (79.9%) than when it was connected to the previous clause (73.6%) [ $F_1(1,53) = 13.77, p < .001$ ;  $F_2(1,29) = 17.56, p < .001$ ; CI =  $\pm 3.37$  for a mean difference of 6.24]. In contrast, recall performance for the final clause was not affected by the middle clause at-

tachment factor (both  $F_s < 1$ ). Unexpectedly (and in an unexpected direction), the difference between the middle and the final clause reached significance when the two clauses belonged to the same sentence (79.9% vs. 71.2%) [ $F_1(1,53) = 41.64, p < .001$ ;  $F_2(1,29) = 8.61, p = .006$ ;  $CI = \pm 2.7$  for a mean difference of 8.67], but not in the other middle clause attachment condition (73.6% vs. 70.5%) [ $F_1(1,53) = 2.26, p = .14, F_2 < 1$ ].

## DISCUSSION

The purpose of this experiment was to find out whether the contribution of grammatical gender information to short-term recall would be influenced by the syntactic status of the critical information. We suggested that linguistic information is available as long as it is potentially used in language processing; since grammatical gender information is used in pronoun interpretation beyond sentence boundaries, this implies that it should be retained across sentence boundaries. Thus, a gender congruency effect ought to be observable not only when the target word is part of the most recent sentence, but also when it is part of a prior sentence. The findings support this assumption: In the overall analysis (1C), the gender congruency effect was not modulated by the middle clause attachment factor, and (as a more direct test) the effect could also be observed when only the subset of cases was analyzed in which the target word appeared in the previous sentence (2A). Consequently, the present results call into question the strict assumption that only conceptual information is represented as soon as a sentence boundary is reached.<sup>3</sup>

An interesting result of our study is that the attachment of the middle clause influenced the number of induced gender-incongruent intrusions (2C), but did not influence the number of induced gender-congruent intrusions (2B). This indicates that the influence of grammatical gender decreases to some degree when crossing sentence boundaries. In a way, these results can be regarded as reflecting the actual use of grammatical gender information in language processing. Whether or not a grammatical gender cue is sufficient for anaphor interpretation depends on the circumstances. The more discourse topics intervene between antecedent and pronoun, the higher the probability will be that more than one possible antecedent matches the pronoun in grammatical gender, with the most recent gender-congruent noun phrase being the most probable antecedent for the pronoun. This suggests that although there is still a need for gender information to be accessible even after an intervening sentence, the influence decreases (see Clark & Sengul, 1979).

Such a "yes but less" influence of grammatical gender information after an intervening sentence is also reflected in the aforementioned anaphor comprehension study by Garnham et al. (1995). Unlike most studies that investigate the influence of gender information on pronoun interpretation, Garnham et al. dissociated conceptual from grammatical gender. In study languages that provide grammatical gender information for genderless entities (French, Experiments 1 and 3; Spanish, Experiment 2), they compared

pronouns referring to people and pronouns referring to things. When the pronoun was presented in a sentence that directly followed the sentence containing the antecedent, the use of a gender cue was not influenced by the antecedent type (object vs. person; Experiments 1 and 2). When there was an intervening sentence between the antecedent and the referring pronoun (Experiment 3), participants were still able to make use of the fact that the antecedent could be identified simply on the basis of a gender cue. This advantage was more pronounced when the antecedent was a person (and thus marked with respect to grammatical and conceptual gender) than when it was a thing (and thus marked solely with respect to grammatical gender). Since the present experiment aimed to minimize the influence of conceptual gender, it is comparable to Garnham et al.'s object condition. Their results indicate that even beyond sentence boundaries, there is an influence of morphosyntactic information with respect to text comprehension. Our study demonstrates that this is also the case for short-term memory for text. What is more, we have shown the contribution of grammatical gender information to text recall using the experimental paradigm by means of which compelling evidence was found for the decline of short-term memory for surface information past sentence boundaries.

In the following, we will turn to possible objections to the method used here and to certain aspects of our data. The present experimental procedure applied a modified version of the sentence recall intrusion paradigm, but it differs from this paradigm in that the lure words were discourse antecedents for the target, whereas lure and target words were structurally unrelated in the sentence experiments. Since it is plausible to assume that the lure word is reactivated when the target is processed in the present experiment, this variation might bring about a change in the memory processes involved.<sup>4</sup> One might thus question whether our findings can be generalized to situations in which gender information for unrelated processing units has to be retained. However, there is evidence within the standard intrusion paradigm suggesting a gender influence beyond immediate sentence recall. Schweppe (2006, Experiment 3) demonstrated a gender congruency effect in a delayed recall version of the intrusion paradigm, in which the sentence is presented first, followed by the distractor list, the probe recognition task, and finally sentence recall.

Another issue worth considering is that interitem variation was comparatively strong in the present study. We assume that this is due to the high proportion of target omissions (see Table 3), in particular when the target word was part of the penultimate sentence. Moreover, certain items caused several target omissions, whereas others caused hardly any. When the target is frequently omitted, both the number of correct target reproductions and the number of lure intrusions are reduced unsystematically. Since all relevant analyses were concerned with the relative proportion of intrusion errors and correct reproductions of the target word, this induces a large amount of noise.

A more serious problem might be the discrepancy between our clause recall data and Jarvella's (1971). His findings were replicated insofar as recall performance

of the critical penultimate clause was influenced by how this was attached (i.e., the clause was recalled better when it belonged to the most recent sentence than when it belonged to the previous one). Our results were different for the comparison between recall accuracy for the penultimate and for the ultimate clause, however. In contrast to Jarvella's findings, there was no general advantage for the most recent clause. Instead, the penultimate clause was recalled more accurately than the ultimate one when they belonged to the same sentence.

How can these differences be accounted for? We think that characteristics of the materials are a plausible reason for the absence of a clause recency effect in our data. As was the case in Jarvella's (1971) experiments, there is a potential confound in the comparison of the second and the third clause. One compares different clauses with different wording and different syntactic structures. Since the materials were not designed to achieve the highest possible likeness between the two clauses, any difference in recall accuracy might be caused by differences in the complexity of the clauses. We have reanalyzed the materials to detect complexity differences between the final and the middle clauses (the detailed analyses are provided along with the materials). The final clauses turned out to contain a higher proportion of content words, whereas the middle clauses contained more function than content words. Function (or closed class) words can be inferred or reconstructed more easily than can content (or open class) words (e.g., Potter & Lombardi, 1990); this should facilitate recall of clauses with a high proportion of closed class words. Indeed, function word recall was more accurate than content word recall in the present study. Consequently, the—unintended—complexity difference between the second and the third clauses may have caused the unexpected recall advantage for the middle over the final clause.

The present article was based on the theoretical assumption that the requirements of language processing influence verbal short-term memory, and vice versa. Following from this, we have concluded that information important for anaphor resolution beyond sentence boundaries should also exert an influence on a short-term memory task beyond sentence boundaries. With respect to grammatical gender information, our data suggest that this is indeed the case.

What about other types of information? A further implication is that retention times ought to be shorter for linguistic representations *not* required for subsequent discourse processing (although they are relevant for initial sentence processing). This applies to phonological information, for instance, which plays a dominant role in recall of unrelated word lists and contributes to immediate sentence recall (e.g., Alloway, 2007). Yet its influence should be restricted to the most recent sentence, since, after initial processing steps, maintenance of a lexical representation is sufficient for further discourse processing. In other words, unlike grammatical gender information, phonological information is expected not to be retained beyond sentence boundaries. An early indicator of faster forgetting rates for (lexical) phonological than for syntactic information comes from Sachs (1974). She

had participants decide whether or not a probe sentence had been part of a previously presented text. The probe sentences were changed in content, syntactic structure, word order, or word form. After the briefest delay between presentation and test, detection of word order and lexical changes were at chance level already, whereas syntactic changes were still detected reliably. Recognition of content changes was accurate even after the longest delay. Sachs's (1974) results are thus in line with our assumption that syntactic information is not forgotten as fast as other types of surface information, though faster than conceptual information.

#### AUTHOR NOTE

This research was supported by DFG Grant Ru 891/1-3 to the second author. We thank Franca Buge and Jasmin Schneider for their help in running the experiments and analyzing the data. We also thank Christoph Scheepers for helpful discussions concerning the data analysis, and Kay Bock, Rob Hartsuiker, and three anonymous reviewers for valuable comments. Correspondence concerning this article should be addressed to J. Schewpe, University of Erfurt, Department of Psychology, P. O. Box 900 221, 99105 Erfurt, Germany (e-mail: judith.schewpe@uni-erfurt.de).

#### REFERENCES

- ALLOWAY, T. P. (2007). Investigating the roles of phonological and semantic memory in sentence recall. *Memory*, *15*, 605-615.
- ANDERSON, J. R. (1990). *Cognitive psychology and its implications* (3rd ed.). New York: Freeman.
- BAAYEN, R. H., PIEPENBROCK, R., & VAN RIJN, H. (1993). *The CELEX Lexical Database* (Release 1) [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- BOCK, K., & GRIFFIN, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*, 177-192.
- CLARK, H. H., & SENGUL, C. J. (1979). In search of referents for nouns and pronouns. *Memory & Cognition*, *7*, 35-41.
- GARNHAM, A., OAKHILL, J., EHRLICH, M.-F., & CARREIRAS, M. (1995). Representations and processes in the interpretation of pronouns: New evidence from Spanish and French. *Journal of Memory & Language*, *34*, 41-62.
- HOWELL, D. C. (2002). *Statistical methods for psychology* (5th ed.). Belmont, CA: Duxbury.
- JARVELLA, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning & Verbal Behavior*, *10*, 409-416.
- MEYER, A. S., & BOCK, K. (1999). Representations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory & Language*, *41*, 281-301.
- MYERS, J. L., & O'BRIEN, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*, 131-157.
- NOORDMAN, L. G. M., & VONK, W. (1999). Discourse comprehension. In A. D. Friederici (Ed.), *Language comprehension: A biological perspective* (2nd ed., pp. 230-263). Berlin: Springer.
- POTTER, M. C., & LOMBARDI, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory & Language*, *29*, 633-654.
- SACHS, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*, 437-442.
- SACHS, J. S. (1974). Memory in reading and listening to discourse. *Memory & Cognition*, *2*, 95-100.
- SCHEEPERS, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, *89*, 179-205.
- SCHWEPPE, J. (2006). *Shared representations in language processing and verbal short-term memory: The case of grammatical gender*. Saarland University, Saarbrücken: PhD thesis, online publication. Retrieved October 6, 2008, from psydok.sulb.uni-saarland.de/volltexte/2006/852/pdf/DissSchewpe.pdf.
- SCHWEPPE, J., & RUMMER, R. (2007). Shared representations in lan-

guage processing and verbal short-term memory: The case of grammatical gender. *Journal of Memory & Language*, **56**, 336-356.  
VON ECKARDT, B., & POTTER, M. C. (1985). Clauses and the semantic representation of words. *Memory & Cognition*, **13**, 371-376.

#### NOTES

1. The lure and target words never described animate objects; an involvement of conceptual gender information was, therefore, reduced.
2. For the sake of simplicity, we also treated as target omissions the rare cases in which the target was replaced by a word that was neither the gender-congruent nor the gender-incongruent lure word.
3. In the present study, we have addressed this issue regarding morphosyntactic information. Interestingly, research using more implicit memory measures, such as syntactic priming, suggests that syntactic information is rather long lasting as well. It has been found that the syntactic structure of a recently produced sentence influences subsequent sentence production with up to 10 sentences intervening between prime and target (Bock & Griffin, 2000).
4. A potential mechanism involved is resonance. In the resonance model, "concepts and propositions in the discourse representation and

in the reader's knowledge base resonate as a function of the degree of match to the input" (Myers & O'Brien, 1998, p. 132). The effect of gender congruency might thus be viewed as an effect of stronger resonance due to overlap in semantic features *and* grammatical gender for the gender-congruent lure words. However, according to Myers and O'Brien (1998, p. 132), this "match depends on the overlap of semantic and contextual features among concepts and on the argument overlap of propositions." Since gender-congruent and -incongruent lure words were matched with respect to semantic similarity to the target word in the present study, a conceptually based resonance process cannot account for our results.

#### SUPPLEMENTAL MATERIALS

Detailed information about the materials and analyses for this study may be downloaded as supplemental materials from [mc.psychonomic-journals.org/content/supplemental](http://mc.psychonomic-journals.org/content/supplemental).

(Manuscript received August 1, 2007;  
revision accepted for publication September 11, 2008.)