TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 5 (2013) 564–580

ISSN:1756-8757 print / 1756-8765 online
DOI: 10.1111/tops.12029

How WM Load Influences Linguistic Processing in Adults: A Computational Model of Pronoun Interpretation in Discourse

Jacolien van Rij, Hedderik van Rijn, Petra Hendriks

Department of Psychology, University of Groningen

Received 15 February 2012; received in revised form 13 December 2012; accepted 11 January 2013

Abstract

This paper presents a study of the effect of working memory load on the interpretation of pronouns in different discourse contexts: stories with and without a topic shift. We discuss a computational model (in ACT-R, Anderson, 2007) to explain how referring expressions are acquired and used. On the basis of simulations of this model, it is predicted that WM constraints only affect adults' pronoun resolution in stories with a topic shift, but not in stories without a topic shift. This latter prediction was tested in an experiment. The results of this experiment confirm that WM load reduces adults' sensitivity to discourse cues signaling a topic shift, thus influencing their interpretation of subsequent pronouns.

Keywords: Working memory load; Language processing; Cognitive modeling; Pronouns

1. Interpretation of subject pronouns

Listeners interpret referring expressions such as pronouns using information from the preceding linguistic discourse. However, representing and accessing this discourse information may come with certain costs. In this study, we investigate the influence of working memory load on adults' interpretation of pronouns in discourse.

Table 1 shows two short stories about two characters of the same gender. In the stories, the final sentence starts with a potentially ambiguous pronoun (*he*), which can refer to both male characters in the given linguistic context (*Eric* and *Philip*).

Generally, pronouns are used to refer to the most salient character or entity (i.e., the *topic*) in the linguistic discourse (a.o., Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993).

Correspondence should be sent to Jacolien van Rij, University of Groningen, Groningen, The Netherlands. E-mail: vanrij.jacolien@gmail.com

Table 1
Example of a story in Dutch with (A) and without (B) topic shift

A. Story with Topic Shift	B. Story without Topic Shift
1. Eric/gaat/voetballen/in de sporthal. "Eric is going to play soccer in the sports hall."	1. Eric/gaat/voetballen/in de sporthal. "Eric is going to play soccer in the sports hall."
2. Philip/vraagt/Eric/om mee te rijden/naar de training. "Philip asks Eric to carpool to the training."	2. Eric/vraagt/Philip/om mee te rijden/naar de training. "Eric asks Philip to carpool to the training."
3. Philip/haalt/Eric/na het eten/met de auto op. "Philip picks up Eric after dinner by car."	3. Eric/haalt/Philip/na het eten/met de auto op. "Eric picks up Philip after dinner by car."
4. Hij/voetbalt/al twintig jaar. "He has played soccer for twenty years."	4. Hij/voetbalt/al twintig jaar. "He has played soccer for twenty years."

Note. Question: Wie voetbalt al twintig jaar? / "Who has played soccer for twenty years?"

In contrast, full noun phrases or proper names are used to refer to less salient characters or to introduce new characters. The interpretation of pronouns is influenced by various factors, including the grammatical roles of potential referents (see Arnold, 1998; for a review). As the subject of the previous sentence is likely to be the current topic (Grosz, Weinstein, & Joshi, 1995), listeners will often interpret a pronoun as referring to the previous subject (a.o., McDonald & MacWhinney, 1995; Stevenson, Crawley, & Kleinman, 1994).

Hendriks, Koster, and Hoeks (in press) presented children (age 4–6) years and adults with prerecorded stories, of which the last sentence started with a potentially ambiguous pronoun. In half of the stories, the topic shifted halfway by changing the grammatical role of the characters (similar to the stories in Table 1). In agreement with the literature, they found that adult listeners prefer to interpret the ambiguous subject pronoun as referring to the subject of the previous sentence (*Philip* in Table 1a, and *Eric* in Table 1b). Children, however, showed a general preference for the first character of the story as the referent of the pronoun (*Eric* in Table 1a and 1b) and did not distinguish between stories with and without topic shift. Interestingly, Hendriks et al. report that children with a higher auditory working memory capacity performed more like adults in their pronoun use.

These results indicate that pronoun interpretation is not only influenced by discourse information but also by cognitive resources such as WM capacity. This is in line with earlier findings that people with a high WM capacity may be more sensitive to sentence ambiguity (a.o., Just & Carpenter, 1992; Miyake, Just, & Carpenter, 1994; Nieuwland & van Berkum, 2006), which suggests that sentence processing is influenced by WM capacity. However, most linguistic accounts of pronoun processing do not specify how pronoun processing is influenced by individual differences or differences in cognitive resources. To investigate these issues, we implemented a computational model within the cognitive architecture ACT-R (Anderson, 2007). This architecture places restrictions on its models based on what is currently known about human information processing. Our computational model simulates the acquisition of referring expressions in discourse (van Rij, van

Rijn, & Hendriks, unpublished data), which allows us to generate precise and testable predictions about pronoun processing.

The computational simulations of our ACT-R model suggest that limited WM capacity may cause decreased performance on children's choice and interpretation of referring expressions in discourse, thus providing a detailed, mechanistic explanation of the empirical correlation reported by Hendriks et al. (in press). Our ACT-R account is in line with previously proposed computational models in different frameworks that explain the relation between WM capacity and language processing, such as CC READER (Just & Carpenter, 1992) or 4CAPS (Just & Varma, 2007). In these models, WM capacity is implemented as a limited amount of activation, which is different for different individuals. If more capacity is required for processing or storage than is available, this will result in longer processing times or retrieval errors (e.g., Daneman & Carpenter, 1980). However, alternative accounts have been put forward to explain the relation between WM capacity and language processing. For example, MacDonald and Christiansen (2002) propose that differences in WM capacity are differences in skill that arise from variations in exposure to the language and biological differences. Alternatively, Ericsson and Kintsch (1995) and Kintsch (Construction-Integration model; 1998) argue that WM capacity reflects the efficiency of retrieval cues present in the active portion of working memory. In their view, more efficient cues result in more accurate and faster retrieval of information from declarative memory.

In this paper we show that our ACT-R model makes specific predictions about adults' linguistic performance under different conditions of WM load. Based on simulations of our ACT-R model, we expect adults' interpretation of subject pronouns in discourse to be influenced by the amount of WM capacity available for interpretation. We present empirical evidence supporting this novel prediction.

2. Cognitive model

In a previous study, we implemented a computational model within the cognitive architecture ACT-R (Anderson, 2007) that simulates children's acquisition of referring expressions in discourse (van Rij et al., unpublished data). In our implementation, we focused on the differences between potentially ambiguous pronouns (*he*) and more explicit referring expressions such as full noun phrases (*the soccer player*) or proper names (*Eric*). The cognitive model was shown to simulate the acquisition data reported in Hendriks et al. (in press). This paper presents new predictions, following from simulations of the cognitive model, for *adults'* interpretation of referring expressions in discourse. Before describing the simulations, we will give a short overview of the model.[1]

2.1. Overview of the model

The model consists of two parts, namely the memory principles that determine the representation of the linguistic discourse, and the linguistic principles that underlie the

choice and interpretation of referring expressions. The model uses general memory principles to build a representation of the discourse during sentence comprehension. The saliency of discourse referents is represented by the activation of elements in memory. The discourse referent with the highest activation is taken to be the current discourse topic. In addition to this memory-based process that automatically constructs a discourse representation, the model uses a rule-based process to perform its task of producing or interpreting a referring expression. To *produce* a referring expression, the model evaluates possible forms, such as a pronoun or a full noun phrase, on the basis of linguistic principles. It then takes the perspective of the listener to select the expression that is not only economical to use but also understandable for the listener in the current discourse. To *interpret* a referring expression, the model evaluates possible discourse referents, such as the current topic, based on linguistic principles. Then it takes the perspective of a speaker to determine whether a speaker would indeed have used the expression for the selected interpretation. The model thus integrates linguistic theories based on the accessibility of referents (e.g., Ariel, 1990) and those that assume a pragmatic reasoning process (e.g., Gundel et al., 1993). Embedding insights from these linguistic theories in a cognitive architecture allows for the generation of testable and cognitively plausible predictions with respect to linguistic performance.

Here, we focus on the interpretation of subject pronouns, which is mainly driven by the linguistic discourse and not by sentence-internal factors, in contrast to object pronouns (van Rij et al., unpublished data). The linguistic discourse must be represented in the listener's memory. By modeling the memory principles determining the representation of the linguistic discourse, we can generate model-based predictions about children's and adults' interpretation of referring expressions in discourse.

2.2. *Simulations: The effect of WM capacity on pronoun interpretation*

Simulating the task of Hendriks et al. (in press), the model is presented with stories with and without topic shift. The six-sentence stories are presented to the model word by word. During online processing, the model builds a (simplified) representation of the preceding discourse: Every time a character is encountered in the story that character is represented in the declarative memory. Each memory representation (referred to as *chunk*) has a certain amount of activation that reflects the saliency of that character in the current discourse (consistent with Foraker & McElree, 2007; but see Grosz et al., 1995; Gundel et al., 1993; who propose discrete cognitive states to represent a character's saliency). The final sentence of each story starts with a potentially ambiguous pronoun. As the current discourse topic, and hence the referent of the pronoun, the model selects the chunk with the highest level of activation.

In ACT-R, the activation of a chunk reflects the chunk's history and the chunk's usefulness in the current context (see Appendix 1). The activation of chunks decays with time but is increased when the chunk is retrieved. This base-level activation is dependent on the frequency of use (the more frequently used, the higher the activation) and the recency of the last retrieval from memory (the more recent the last retrieval, the higher

the activation). In addition, spreading activation from other chunks can temporarily boost a chunk's activation: Chunks that are currently being processed spread activation to other, connected chunks in declarative memory. As the amount of spreading activation determines the ability to maintain goal-relevant information, differences in spreading activation account for individual differences in working memory capacity (Daily, Lovett, & Reder, 2001).

In our model, the subject of the previous sentence is temporarily stored as goal-relevant information before it is overwritten with the information from the next sentence. Therefore, the subject of the previous sentence spreads activation to all discourse elements associated with it (similar to the implementation of short-term syntactic priming of Reitter, Keller, & Moore, 2011). We manipulated the amount of spreading activation to explain the difference between children's and adult's interpretation of subject pronouns reported by Hendriks et al. (see Appendix 2). If the amount of spreading activation is high, the chunk representing the subject spreads a large amount of activation and discourse elements that are associated with the subject become more activated in comparison with other discourse elements. As a result, the model will retrieve the referent that was the subject of the previous sentence as the current discourse topic. On the other hand, if the subject spreads a small amount of activation, reflecting a low WM capacity, then there will be no effect on the discourse elements associated with the subject. In that case, the effects of frequency and recency will be the main determinants of referent retrieval.

Figure 1 shows the effect of WM capacity (i.e., the amount of spreading activation) on the activation of the two referents in the stories with topic shift presented to the model. The second character is introduced in Sentence 3. The topic shift condition starts to differ from the no topic shift condition in Sentence 4, where the second character becomes the subject in the topic shift stories, but not in the stories without topic shift (cf. Hendriks et al., in press). With a high WM capacity, the model selects the subject of the previous

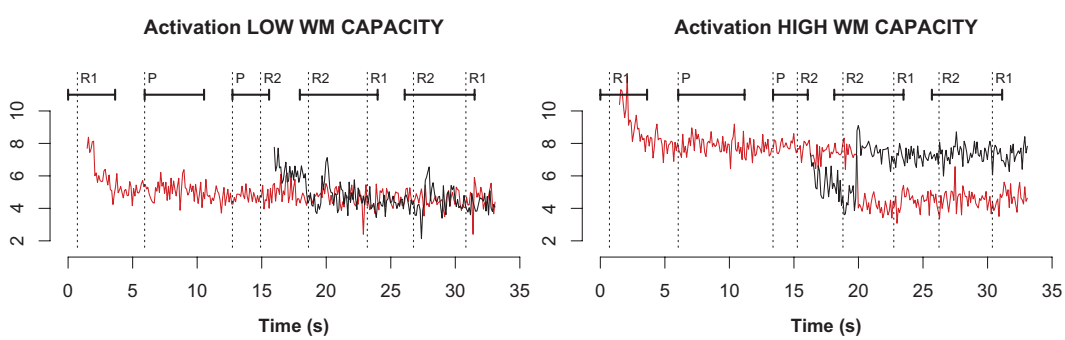

Fig. 1. Activation of the two referents over time (s) in a story with topic shift in a low WM capacity and a high WM capacity simulation. The horizontal lines represent the duration of the first five sentences of the story, whereas the vertical lines depict the start of a referring expression (R1 and R2 represent expressions unambiguously referring to referent 1 and 2, and P is a pronoun). The red activation line represents the activation of the first introduced referent; the black activation line represents the activation of the second referent (adapted from van Rij et al., unpublished data).

sentence as the referent of the pronoun in Sentence 6, because this discourse element clearly has the highest activation (Figure 1, right). However, with a low WM capacity, the model will show a much-reduced preference for the second character as the referent of the pronoun. Similarly to children's performance, the model's interpretation of pronouns is not affected by grammatical role (Figure 1, left). Averaged over multiple simulations, the model shows a small preference for the first character because this referent is more frequently mentioned in the story.

2.3. New prediction: Effect of WM load on adults' interpretation of pronouns

On the basis of our model's simulations, we propose that an individual's WM capacity influences the extent to which the grammatical structure of the previous sentence plays a role in resolving a potentially ambiguous subject pronoun. If this hypothesis is correct, we expect that adults show difficulties in detecting a topic shift when their WM is taxed by having to perform a memory task in parallel. This prediction follows directly from the ACT-R model: Goal-relevant information spreads a proportion of the total spreading activation to other chunks in the declarative memory. If the number of sources from which activation is spread increases, the amount of spreading activation received by individual chunks decreases. In a high WM load situation, more information needs to be maintained in an activated state and as a result, the subject of the previous sentence spreads less activation to the discourse elements associated with the subject. Therefore, the model predicts that adult listeners or readers show more childlike performance in high WM load conditions: They will more often select the first character as the current discourse topic. In addition, as the level of activation determines the retrieval time, the model predicts that it will take more time to retrieve a discourse referent in a high WM load situation: Due to spreading activation, the activation of discourse referents in a high WM load condition is lower.

3. Dual-task experiment

We performed a dual-task experiment to test our prediction that adult listeners will show difficulties in their comprehension of a topic shift if they have less WM capacity available (see for a similar dual-task manipulation in the domain of the processing of referring expression; Campana, Tanenhaus, Allen, & Remington, 2011). To manipulate WM load, participants were asked to perform a combined task: memorizing a sequence of digits for later recall while performing a moving-window task (Just, Carpenter, & Woolley, 1982).

3.1. Methods

3.1.1. Digit task

At the start of each trial, participants had to memorize a sequence of either three or six digits (low and high WM load conditions). Digits were shown for 1 s each in the

center of a computer screen. The digits were pseudorandomly chosen from 1 to 9, while ensuring that not all the digits were the same. After completing the moving-window task, the participants had to recall the memorized digits.

3.1.2. Moving-window task

After the presentation of the digits, participants had to read stories consisting of four sentences each (see Table 1), followed by a comprehension question. The sentences were presented one by one and were subdivided into smaller word clusters (indicated by slashes in Table 1). Using a typical moving-window paradigm (Just et al., 1982), only the letters of one single word cluster were visible at a time, and the other letters were replaced by a dot. By pressing a button, the participant could move the window to the next word cluster. After reading the four-sentence story, a question was presented with two answer alternatives. Participants had to press the corresponding button to answer the question. After answering the question, they had to type in the digits that were presented at the beginning of the trial. At the end of each trial, participants only received feedback on the digit task to ensure sufficient focus on the WM task. We collected different measures per trial: reading times per region, accuracy, and reaction times for the questions and accuracy in reproducing the digits.

3.1.3. Design

3.1.3.1. Stories: Every story featured two characters of the same gender. The final sentence of each story started with an ambiguous subject pronoun *hij* ("he") or *zij* ("she"): The pronoun could refer to both characters, so that the only clue to the interpretation of the pronoun was the structure of the story. We designed two variants of every test story (see Table 1), in which we manipulated whether there was a topic shift or not. The topic shift is realized by placing the second character (*Philip*) in subject position in Sentence 2. If there was no topic shift, we expected participants to prefer the first introduced character as the referent of the ambiguous pronoun. If there was a topic shift, we expected participants to prefer the second character. At the end of every test story, a question was presented to elicit the preferred interpretation of the ambiguous pronoun.

3.1.3.2. Lists: The presented materials were part of a larger experiment, in which we additionally tested two other variants of the stories. In total, 64 test stories were designed in four different variants. Four lists of 64 test stories (16 test stories per condition) were constructed to separate the different variants of the test stories. The lists also contained 128 filler items with the same structure as the test stories (32 filler stories per condition, the same for all lists). The filler stories were followed by a question about the first or second sentence of the story, to avoid reading strategies and to mask the goal of the experiment. Half of the filler questions asked about a character; the other half were what- or where-questions. Note that in contrast to the test questions, filler questions were not ambiguous and could be scored as right or wrong.

Here, we report on 2 × 32 test items, and the 64 filler items with the same two discourse structures. One test story (both variants) was removed from the data because of a technical problem during presentation.

3.1.4. Procedure

Participants were randomly assigned to one of the four lists. The experiment consisted of two blocks: a low WM load block (3 digits) and a high WM load block (6 digits). The order of blocks was counterbalanced; within blocks, the items were randomly distributed. Participants received instructions for each block. The experiment started with a practice trial suited for the current WM load condition.

3.1.5. Participants

Sixty-two first-year psychology students (18 men, 44 women; M_{age} = 19) participated in the experiment in exchange for course credits. Five participants could not complete the experiment because of technical problems. Another 5 participants were excluded from data analysis because they answered less than 75% of the filler questions correctly in the low WM load condition and/or performed at chance level in one of the two types of filler questions. Data of 52 participants (15 men, 37 women) were used for the statistical analyses.

3.2. Results

In this section, we first discuss the performance on the digit task. We then present the offline results on the linguistic task (i.e., the answers on the questions and the reaction times), followed by the online results on this task (i.e., the reading times).

3.2.1. Results on digit task

Participants made more errors on the digit task in the high WM load condition than in the low WM load condition (percentage correct trials: 3-digits = 77.2%, 6-digits = 52.0%), indicating that the 6-digit condition was indeed more difficult.

3.2.2. Offline results on linguistic task

3.2.2.1. Answers: Figure 2 (left) shows the percentages of times that the subject of the previous sentence was chosen as the referent of the ambiguous pronoun. In addition, the percentages of correct answers on the filler questions (right) are given for comparison. Figure 2 shows that participants were sensitive to the topic-shift manipulation. In both WM load conditions, the subject of the previous sentence was preferred.

We examined the effects of *Topic shift*, *WM load*, *Trial position* (the trial position in the experiment), and *Session* (the first or the second block of the experiment) on the choice for the previous subject (yes or no) in the test items using logistic mixed-effects models (cf. Baayen, 2008). More complex models that included additional predictors such as list, the position of the subject answer on the screen (left or right), or the accuracy of repeating the memorized digits, did not show qualitatively different effects. In all the

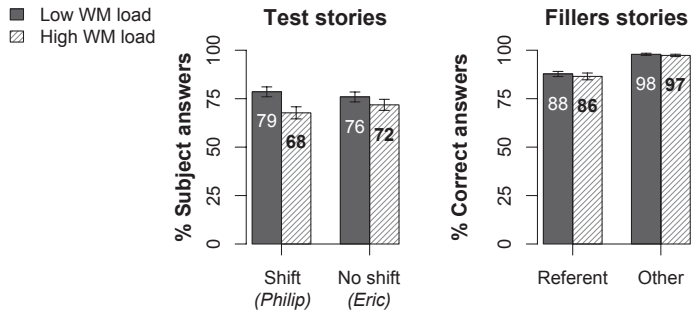

Fig. 2. Left: Preference for the subject of the previous sentence as the referent of the pronoun for stories with and without topic shift (± SE); this would be *Philip* in the story with topic shift in Table 1 and *Eric* in the story without topic shift. Right: Percentage correct answers for filler questions asking about a referent or a non-referent (± SE).

presented models, participant and item (all variants of a story were labeled as the same item) were included as crossed-random effects, with the maximum random effect structure that was justified by the data. We compared different models using a stepwise variable deletion procedure, starting with the complete interaction model (Baayen & Milin, 2010).

To explain the choice for the subject of the previous sentence (yes or no), the four-way interaction between *Topic shift*, *WM load*, *Trial position,* and *Session* needed to be included in the model ($\chi^2(1) = 7.950$, $p = .005$). The full-interaction model (Appendix 2, Table A1) showed that in Session 1, the subject was selected less often in the high WM load condition than in the low WM load condition for stories with a topic shift ($\beta = -0.676$, $z = -2.71$; $p = .007$), in line with the assumption that decreasing WM capacity reduces the performance on pronoun resolution. However, *Trial position* attenuates this effect: near the end of the session, the difference between the WM load conditions became smaller ($\beta = -0.336$, $z = -3.15$; $p < .001$). A negative main effect of *Trial position* ($\beta = -0.009$, $z = -2.98$; $p = .003$) indicates that participants were less likely to choose the subject as referent near the end of the experiment. Note that these interactions between *WM load*, *Topic shift*, *Trial position,* and *Session* are mainly driven by the topic shift condition: A mixed-effects model to explain the stories with a topic shift shows a significant three-way interaction between these predictors ($\chi^2(1) = 5.707$, $p = .017$). Crucially, no effects of *WM load*, *Trial position*, or *Session* were found using the same analysis for the stories without a topic shift.

3.2.2.2. Reaction times: In the same way as we analyzed the choice of referent, we analyzed the log-transformed reaction times after excluding the short outliers (≤ 50 ms; less than 1% of the data). We found a significant interaction between *Trial position* and *Session* ($\chi^2(1) = 7.783$, $p = .005$), indicating that participants became faster in answering during the experiment, but no significant effects of *WM load* and *Topic shift.* The best-fitting model additionally included a predictor *Subject* ($\chi^2(1) = 11.597$, $p < .001$): Participants were faster when selecting the subject than when selecting the non-subject referent.

To summarize, we found that WM load affects the comprehension of stories with a topic shift, but not the stories without a topic shift: Participants less often select the subject of the previous sentence as the referent of the ambiguous pronoun in the high WM load condition. However, we did not find a difference in reaction times between the two types of stories, suggesting that the questions after stories with a topic shift are not more difficult to answer. These findings support our prediction that adults will show difficulties in processing a topic shift under higher WM load.

3.3. Online results on linguistic task

3.3.1. Reading times

Before analyzing the reading time data, we removed missing data (2%), short outliers (smaller than 50 ms, 19%), and used a log-transform to reduce the effect of the long outliers (cf. Baayen & Milin, 2010). The relatively large amount of short outliers was caused by a technical problem. As the outliers were equally distributed over the story conditions and the WM load conditions ($\chi^2(1) = 0.925$, $p > .1$), it is unlikely that this influences our results in qualitative ways.

We analyzed the log-transformed reading times of Sentence 2 and Sentence 4 to determine whether WM load influenced the processing of the topic shift and the processing of the ambiguous pronoun.

3.3.2. Sentence 2

Figure 3 (left) shows the reading times of Sentence 2 on a logarithmic scale. We expected an interaction in reading times between the topic shift conditions and the WM load conditions at the start of the sentence: The introduction of a new referent is expected to cause an increase in reading times, in comparison with repeated reference to the previous subject. However, this difference between the topic shift conditions is not expected for the high WM load condition because the model predicts that a high WM load decreases the activation of the subject referent. To test these hypotheses, we analyzed the

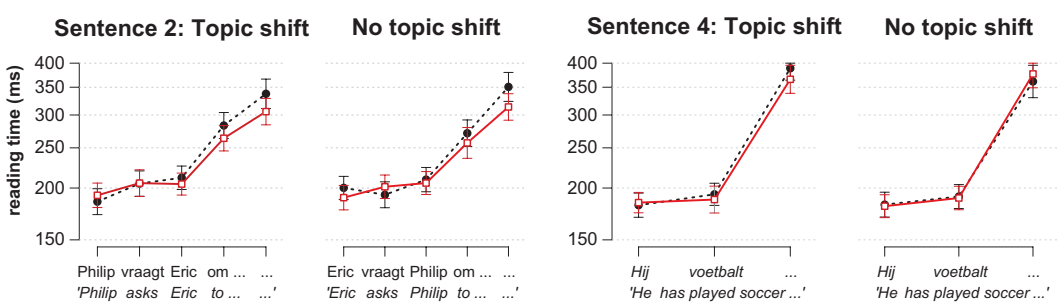

Fig. 3. Reading times (± SE) of Sentence 2 (left) and Sentence 4 (right) in the low WM load condition (dashed line) and the high WM load condition (solid line). Note the logarithmic scale on the *y*-axis.

effects of *WM load* and *Topic shift* in the first and second region, which may reflect residual effects of retrieving the discourse referent after processing the word.

In the analysis of the first region, we included *Trial position*, *Session*, and *Letters* (the number of letters of the word) in the linear mixed-effects model. The best-fitting model (Appendix 2, Table A2) contained significant interactions between *Trial position* and *Session* ($\chi^2(1) = 6.371$, $p = .012$), and between *WM load* and *Trial position* ($\chi^2(1) = 18.829$, $p < .001$). The latter interaction indicates that reading times became faster toward the end of the experiment in the high WM load condition ($\beta = 0.103$, $t = -2.42$), although reading times were generally slower in the high WM load condition ($\beta = 0.001$, $t = 2.09$). There was no significant effect of *Topic shift* ($\beta = -0.011$, $t < 1$).

In the analysis of the second region, we included *Trial position*, *Session*, and *Log-RT1* (the log-transformed reading times on the first region), to control for autocorrelation effects. However, we did not find any significant interaction. The main-effects model (Appendix 2, Table A3) did show a main effect of *Topic shift* ($\beta = 0.028$, $t = 2.15$, $p = .032$), indicating that readers needed more time to read the second region in the stories with topic shift, but no significant effect of *WM load* ($\beta = 0.015$, $t = 1.13$; $p > .1$).

Additional analyses that evaluate all reading times of the sentence to investigate the effects of *WM load* and *Topic shift* over time showed significant effects of these predictors at the end of the sentence but no interaction. These effects may point to a reasoning process or a task-specific strategy, which is beyond the scope of this study.

3.3.3. Sentence 4

Figure 3 (right) shows the reading times for the final sentence of the story, which started with an ambiguous pronoun. We analyzed this sentence using the analyses described before to determine whether WM load influenced the retrieval of the referent of the pronoun. However, we did not find significant effects of *Topic shift* or *WM load*, nor an interaction between these two factors in the first two regions of Sentence 4.

To summarize, in Sentence 2, we found an effect of *WM load* on the first word and an effect of *Topic shift* on the second word, but no interaction between *WM load* and *Topic shift*. The prediction that more salient discourse referents give rise to shorter retrieval times was not reflected in the reading time data.

4. Discussion

We predicted, on the basis of our cognitive model, that adults would show more difficulties in processing a topic shift in the higher WM load condition because higher WM load makes it more difficult to use information about the structure of the previous sentence. As a consequence, participants would rely on alternative factors such as frequency and recency instead. We performed a dual-task experiment to investigate this prediction. We hypothesized that as WM load increased, adult readers would show a significant

decrease in their preference for the second character as the referent of a pronoun in the stories with a topic shift. In addition, we expected an increase in reading times in stories with a topic shift as a result of the topic shift, but we expected that this increase would diminish in the high WM load condition.

The offline data support the model's prediction: Participants selected the first character as the referent of the ambiguous pronoun significantly more often in the high WM load condition. No differences in reaction times were found, suggesting that the comprehension questions were equally difficult to answer for the two types of stories.

With respect to the reading times, we found an increase in reading times immediately after presenting a new referent in subject position. This indicates that readers expect to see the subject of the previous sentence again, instead of a new referent. However, we did not measure a significant interaction between WM load and story type. Different explanations are possible for why this interaction did not reach significance, contrary to our expectations. It could be that WM load does not affect the processing of the sentence, but only affects the updating of the discourse representation with new sentence information. In that case, sentence wrap-up effects could have masked an interaction between topic shift and WM load. An alternative explanation is that the moving-window task is not suited for detecting the effect of WM load. It is reasonable to assume that the effect of WM load on topic shift is spread out over different regions and is thus more difficult to detect. ERP studies provide support for this explanation because for unexpected noun phrases readers show an ERP effect 300–600 ms after the determiner of the unexpected noun phrase (Otten & van Berkum, 2009), which is much longer than it took participants in our experiment to read one region (~200 ms).

Summarizing, our data show that sufficient WM capacity is necessary for adult-like pronoun interpretation. Furthermore, the amount of WM capacity available for language processing may not only vary between individuals but also within individuals. The link between WM capacity and language processing is not new. Within the context of ACT-R, Lewis and Vasishth (2005) have explained difficulties in sentence processing, which have been attributed to WM load, by ACT-R's fluctuating activation and similarity-based interference in the retrieval of chunks. Our account is also in line with previously proposed accounts that consider WM capacity as a limited source of activation that is different for different individuals (a.o., Daneman & Carpenter, 1980; Just & Carpenter, 1992; Just & Varma, 2007). Alternatively, Ericsson and Kintsch (1995) argue that WM capacity reflects the efficiency of constructed representations in memory (Kintsch, 1998; Chapter 7), which is only consistent with our data under the assumption that higher WM load changes the representations that are constructed in memory. In other words, higher WM load may reduce the probability that the grammatical role is encoded in the discourse representation of a referent. This is subtly different from our implementation, as instead of changing the discourse representation itself, our account suggests that WM load affects the contribution of the previous subject to the activation of referents but does not change the internal representation. Furthermore, our data show that WM load only affects the interpretation of stories with a topic shift, which is difficult to explain in terms of

individual variation in exposure to the language and biological differences (as MacDonald & Christiansen, 2002, propose).

This study may provide new insights into the underlying mechanisms of sentence processing. In our model, the interpretation of pronouns is not fixed, but rather requires the integration of different sources of information. Our results thus contribute to the wider discussion of how people combine all necessary information in the short time available during online sentence processing. Based on simulations of our cognitive model, we argue that the linguistic discourse activates relevant information before the pronoun is encountered, which aids the resolution of pronouns and other ambiguous expressions and speeds up the interpretation process. However, structural information from the preceding sentence is only available with sufficient WM capacity.

To conclude, we found that WM load can affect adults' interpretation of ambiguous pronouns by reducing the influence of the discourse context. This specific effect of WM load on pronoun interpretation was predicted by our cognitive model. Our model implies that readers or listeners without sufficient WM capacity rely more on the base-level activation of discourse elements, instead of using information about the grammatical roles of the referents in the previous sentence. This study thus shows that adults' interpretation of pronouns is not constant, but rather depends on the amount of WM capacity available to them. As a consequence, in certain situations (for example, if they carry out a demanding task while listening), adults may become more child-like in their interpretation of pronouns.

Acknowledgments

The authors thank Charlotte Koster, Gisi Cannizzaro, and Sanne Kuijper for their help with creating the experimental stimuli, and the Cognitive Modeling Group and the Experimental Linguistics Group at the University of Groningen for helpful discussions regarding this study. Petra Hendriks gratefully acknowledges NWO (Vici grant no. 277-70-005) for financial support.

Note

1. The code of the discussed ACT-R model can be retrieved from http://www.let.rug.nl/jacolienvanrij/modelcode.html

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Arnold, J. E. (1998). *Reference form and discourse patterns*. Unpublished Ph.D. thesis, Stanford, CA: Stanford University.

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

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Campana, E., Tanenhaus, M. K., Allen, J. F., & Remington, R. (2011). Natural discourse reference generation reduces cognitive load in spoken systems. *Natural Language Engineering*, *17*(3), 311−329.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science*, *25*(3), 315.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, *56*(3), 357–383.

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, *21*(2), 203–225.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307.

Hendriks, P., Koster, C., & Hoeks, J. C. J. (in press). Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language and Cognitive Processes*.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology*, *111*(2), 228–238.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(3), 153–191.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35–54.

McDonald, J. L., & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, *34*(4), 543–566.

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, *33* (2), 175–202.

Nieuwland, M. S., & van Berkum, J. J. A. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, *1118*(1), 155–167.

Otten, M., & van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291*, 92–101.

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, *35*(4), 587–637.

van Rij, J., van Rijn, H., & Hendriks, P. (submitted). A cognitive model of the acquisition and use of referring expressions: Effects of working memory and perspective taking.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, *9*(4), 519–548.

Appendix

Activation formula of ACT-R

In ACT-R (Anderson, 2007), the activation of chunk i is defined by $A_i = \mathrm{In}(\Sigma_{k=1}^{n} t_k^{-0.5}) + \Sigma_{j=1}^{m} W_j S_{ji} + \varepsilon_i$, with n being the number of presentations of chunk i, and t_k the time since the kth presentation, m the number of chunks that are connected with chunk i, W_j the amount of activation that is spread from chunk j, S_{ji} the strength of association between j and i, and ε_i noise. The activation of a chunk determines the time it takes to retrieve this information from declarative memory: $T = e^{-A_i}$ These formulas are based on a range of psychological experiments (see Anderson, 2007, Chapter 3, for discussion and references).

Table A1
Fixed effects of best-fitting logistic mixed-effects model to fit the choice for the subject as the referent of the ambiguous pronoun in the test stories

Predictor	Estimate	SE	*z*-value	*p*-value
Intercept***	2.123	0.348	6.11	.000
WM load[a]	−0.207	0.250	−0.83	.406
Topic shift[b]	0.238	0.330	0.72	.472
Session[c]	−0.170	0.334	−0.51	.610
Trial position[d,]**	−0.009	0.003	−2.98	.003
WM load:Topic shift	−0.138	0.250	−0.55	.580
WM load:Session	−0.274	0.268	−1.02	.307
WM load:Trial position	0.000	0.002	0.12	.901
Topic shift:Session	0.269	0.330	0.82	.414
Topic shift:Trial position	−0.003	0.003	−0.99	.322
Session:Trial position	−0.001	0.003	−0.24	.810
WM load:Topic shift:Session**	−0.676	0.250	−2.71	.007
WM load:Topic shift:Trial position	0.003	0.002	1.36	.173
WM load:Session:Trial position	0.002	0.002	0.72	.474
Topic shift:Session:Trial position	−0.001	0.003	−0.44	.662
WM load:Topic shift:Session:Trial position**	0.007	0.002	2.84	.005

Notes. Model: choiceSubj ~ (WMload + TopicShift + Session + TrialPosition)^4 + (1|Subject) + (1 + TopicShift|Item)
[a]Contrast used for WM load: Low WM load (−1) versus High WM load (1).
[b]Contrast used for Topic shift: No topic shift (−1) versus Topic shift (1).
[c]Contrast used for Session: Session 2 (−1) versus Session 1 (1).
[d]Trial position: range 1–192; linear fit, because log-transformed predictor or more complex transformations did not improve the model.
Significance codes: ***$p < .001$; **$p < .01$; *$p < .05$.

Manipulation of spreading activation in ACT-R

The amount of spreading activation is calculated by the formula: $\Sigma_{j=1}^{m} W_j S_{ji} + \varepsilon_i$, as described in Appendix 1. According to this formula, the activation that is spread from the subject of the previous sentence to the associated discourse elements is dependent on the strength of association between the two chunks (S_{ji}) and the amount of activation that is spread from the subject (W_j). In our model, the subject of the sentence is temporarily stored as goal-relevant information and may spread activation to associated information. The lexical concept of subject of the previous sentence is also associated with the discourse element that represents this lexical concept in the current discourse. Thus, the information about the subject of the previous sentence is connected with one of the discourse elements. The strengths of associations between chunks (S_{ji}'s) are not specified explicitly in our model but are calculated using ACT-R's default formula (Anderson, 2007). In our simulations, the strengths of associations between chunks do not change, because the declarative memory is reset to its original state between different simulations. The amount of activation that is spread from chunks (W_j) is manipulated: In the low WM capacity simulations, the amount of spreading activation (W_j) from the previous subject is set to 0, whereas in the high WM capacity manipulations its value is set to 1.

Table A2
Fixed effects of the mixed-effects model (including all two-way interactions) to fit the reading times on the first region of Sentence 2

Predictor	Estimate	SE	*t*-value[e]
Intercept***	5.458	0.110	49.63
WM load[a]*	0.118	0.053	2.23
Topic shift[b]	−0.046	0.051	−0.89
Trial position[c]***	−0.005	5e−4	−9.18
Session[d]*	0.161	0.061	2.63
Letters	0.022	0.014	1.61
WM load:Topic shift	0.015	0.014	1.08
WM load:Trial position*	−0.001	5e−4	−2.52
WM load:Session	−0.047	0.064	−0.74
Topic shift:Trial position	4e−4	5e−4	0.69
Topic shift:Session	0.012	0.028	0.41
Trial position:Session***	−0.002	5e−4	−4.5

Notes. Model: LogRT ~ (WMload + TopicShift + TrialPosition + Session)^2 + Letters + (1 + WMload|Subject) + (1 |Item).
[a]Contrast used for WM load: Low WM load (−1) versus High WM load (1).
[b]Contrast used for Topic shift: No topic shift (−1) versus Topic shift (1).
[c]Contrast used for Session: Session 2 (−1) versus Session 1 (1).
[d]Trial position: range 1–192; linear fit, because log-transformed predictor or more complex transformations did not improve the model.
[e]Exact *p*-values are not estimated because a random slope for WM load needed to be included for Subjects ($\chi^2(2) = 14.929$, $p < .001$).
Significance codes: ***$p < .001$; **$p < .01$; *$p < .05$.

Statistical analyses

Table A3
Fixed effects of the mixed-effects model (including only main effects) to fit the reading times on the second region of Sentence 2

Predictor	Estimate	SE	*t*-value	*p*-value
Intercept***	3.391	0.189	17.916	.000
WM load	0.015	0.013	1.125	>.1
Topic shift*	0.028	0.013	2.151	.032
Trial position***	−0.003	6e−4	−4.659	.000
Session	−0.040	0.030	−1.331	>.1
Letters	−0.030	0.017	−1.762	.078
LogRT1***	0.442	0.026	16.711	.000

Note. Model: LogRT ~ WMload + TopicShift + TrialPosition + Session + Letters + LogRT1 + (1|Subject) + (1 |Item)
Significance codes: ***$p < .001$; **$p < .01$; *$p < .05$.