

WM load influences the interpretation of referring expressions

Jacolien van Rij

University of Groningen
J.C.van.Rij@rug.nl

Hedderik van Rijn

University of Groningen
D.H.van.Rijn@rug.nl

Petra Hendriks

University of Groningen
P.Hendriks@rug.nl

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 present a computational model (in ACT-R, Anderson, 2007) to explain how referring subjects are used and interpreted. We furthermore report on an experiment that tests predictions that follow from simulations. The results of the experiment support the model predictions that WM load only affects the interpretation of pronouns in stories with a topic shift, but not in stories without a topic shift.

1 Introduction

How do listeners interpret a potentially ambiguous referring expression? To describe a particular event, object or character, often different referring expressions can be used. Some referring expressions are more specific than others. For example, a proper name such as ‘Eric’ is more specific than the personal pronoun ‘he’, which can refer to any of the male characters in a given linguistic context.

Generally, pronouns are used and interpreted as referring to the most salient character or object (the *topic*) in the linguistic context (a.o., Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993). In contrast, full noun phrases or proper names are used to introduce new characters or to refer to less salient characters. Different factors have been found to affect the saliency of characters or objects in the linguistic context (see Arnold, 1998, for a review), among others the grammatical role. The subject of the previous sentence is likely to be the current topic (Grosz, Weinstein, & Joshi, 1995). As a result, listeners will often interpret a pronoun as referring to the previous subject (a.o., McDonald & MacWhinney, 1995; Stevenson, Crawley, & Kleinman, 1994).

However, for children up to the age of 7, the grammatical role seems to be a less important cue than for adults. Manipulating the discourse structure, Koster and colleagues showed that children interpret subject pronouns in a different way than adults do (Koster, Hoeks, & Hendriks, in press). They presented adults and children with prerecorded short stories about two characters of the same gender. Two types of stories were tested, stories with and without a topic shift. In the stories with topic shift the topic shifted halfway by changing the grammatical role of the characters (see Table 1): the second character becomes the subject of next sentences, rather than the first character. In all stories the final sentence started with a potentially ambiguous pronoun (‘he’ or ‘she’).

Story with topic shift (+TS)

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.’
3. Philip/haalt/Eric/na het eten/met de auto op.
‘Philip picks up Eric after dinner by car.’
4. Hij/voetbalt/al twintig jaar.
‘He has played soccer for twenty years.’

Story without topic shift (-TS)

1. Eric/gaat/voetballen/in de sporthal.
‘Eric is going to play soccer in the sports hall.’
2. Eric/vraagt/Philip/om mee te rijden/naar de training.
‘Eric asks Philip to carpool to the training.’
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.’

Q Wie voetbalt al twintig jaar?
‘Who has played soccer for twenty years?’

Table 1: Example of the Dutch sentences (and the English translations) of a story with and without topic shift.

Adult listeners interpreted this pronoun as referring to the second character in stories with a topic shift, and as referring to the first character in stories without a topic shift. Children, on the other hand, did not distinguish between these two types of stories: they showed a small preference for the first character as the referent of the pronoun. Koster et al. found that children with a higher auditory working memory capacity performed more adult-like, which raises the question whether limited WM capacity causes decreased performance.

We have implemented a cognitive model to investigate the effect of WM load on the interpretation of subject pronouns in discourse. To test the prediction following from our model that WM load can decrease adults' comprehension of stories with a topic shift, we also performed an experiment.

2 Modeling pronoun interpretation

We have implemented a cognitive model within the cognitive architecture ACT-R (Anderson, 2007) that can simulate both the use and interpretation of referring subjects (Van Rij, Van Rijn, & Hendriks, submitted). Here we focus on the interpretation of subject pronouns.

2.1 Computational simulation

To simulate the results of Koster et al. (in press), the model is presented with stories of 6 sentences, with or without a topic shift. The stories are provided to the model word by word. During on-line sentence processing the model builds a (simplified) representation of the preceding discourse: every character in the story is represented in the declarative memory. Each representation (referred to as "chunk") has a certain amount of activation that reflects the saliency of the character in the current discourse. The model determines the referent of the potentially ambiguous pronoun in the final sentence, by selecting the chunk with the highest level of activation as the current discourse topic and as the referent of the pronoun.

Explaining children's and adults' performance

In ACT-R, the activation of chunks reflects the chunk's history, because activation is dependent on the frequency of use (the more frequently used, the higher the activation) and the recency of the last retrieval (the more recent the last retrieval, the

higher the activation). The activation of chunks decays with time, but is increased when the chunk is retrieved. In addition to this base-level activation, spreading activation can temporarily boost the activation of a chunk in a particular context, reflecting the usefulness of that chunk in that context¹. 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 spreads activation to all discourse elements associated with it.

To explain the difference between children's and adults' interpretation of subject pronouns, we manipulated the amount of spreading-activation from the previous subject. 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 the other discourse elements. As a result, the model will retrieve the subject of the previous sentence as the current discourse topic. However, 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 which referent is retrieved.

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 that were provided to the model. The second character, referent 'b', is introduced in the third sentence. The +TS condition starts to differ from the -TS condition in sentence 4, where the second character be-

¹ In ACT-R the activation of chunk i is defined by:

$$A_i = \ln \left(\sum_{k=1}^n t_k^{-0.5} \right) + \sum_{j=1}^m W_j S_{ji} + \varepsilon_i, \text{ with } n \text{ being the}$$

number of presentations of chunk i , and t_k the time since the k th 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 it from declarative memory: $T = e^{-A_i}$.

comes the subject in the +TS stories but not the –TS stories (cf. Koster et al., in press). With a high WM capacity, the model selects the subject of the previous 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, and often chooses the first character. Similarly to children’s performance, the models’ interpretation of pronouns is not affected by grammatical role (Figure 1, left).

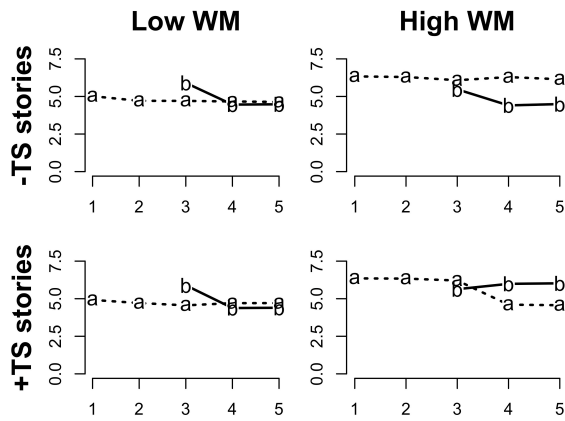


Figure 1. Mean levels of activation of the first character (a) and the second character (b) in stories with (+TS) and without (-TS) a topic shift, measured at the end of each of 5 sentences (x-axis). In sentence 6, the model selects the character with the highest activation as the referent of the pronoun.

Prediction of the model

On the basis of these simulations we propose that an individual’s WM capacity determines how much 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 that is 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 child-like performance in high WM load conditions: they will more often select the first character as the current discourse topic. In addition, the model predicts that it will take more time to retrieve a discourse element in a high WM load condition, because the retrieval time is determined by the level of activation (the lower the activation, the longer it takes to retrieve the information).

3 Experiment

We performed a dual-task experiment to test our prediction that adult listeners will show difficulties with the comprehension of a topic shift if they have less WM capacity available. To manipulate WM load, participants were asked to perform a combined task: memorizing a sequence of digits for later recall and performing a moving-window task (Just, Carpenter, & Woolley, 1982).

3.1 Methods

Digit task

At the start of each trial, participants were presented with a sequence of either three or six digits (low and high WM load conditions) that they had to memorize. Digits were shown for 1 second each in the center of a computer screen. After completing the moving-window task, the participants recalled the memorized digits. The digits were pseudo-randomly chosen from 1 to 9, while ensuring that not all the digits were the same.

Moving-window task

After the presentation of the digits, the moving-window task started. In this task, participants had to read stories 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 dashes 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. All 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 in this way, a question was presented in the center of the screen, and two answer alternatives were presented in the bottom corners of the screen. 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: the reading times per region, accuracy and reaction times for the questions and the number of errors in reproducing the digits.

Design

Stories. In every story two characters of the same gender played a role, of which the first names started with a different letter. All names consisted of 4-8 characters, and two syllables. Importantly, the final sentence started with a subject pronoun *hij* ('he') or *zij* ('she') that was ambiguous: the pronoun could refer to both characters, so that the only clue to the interpretation of the pronouns 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 shift of topic. The topic shift is realized by making the second character ('Philip') the subject of the second sentence. If there was no topic shift, we expected participants to prefer the firstly introduced character as the referent of the ambiguous pronoun, but 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.

Lists. The presented materials were part of a larger experiment, in which we additionally tested another two variants of every story. In total, 64 test stories were designed in four different variants. Four lists of 64 test stories were constructed, so that every list received a different variant of the test stories and thus contained 16 test stories per condition. Besides the test stories, the lists also contained 128 filler items (the same for all lists), so that the lists consisted of 192 items in total. The filler stories had the same structure as the test stories, so 64 filler stories per condition. 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 experimental manipulations. Half of the filler questions asked about one of the characters, the other half asked about a non-referent (what- or where-question). Note that in contrast to the test questions, that were designed to elicit an interpretational preference, filler questions were not ambiguous and could be unambiguously scored as right or wrong.

Here, we report on 2 times 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.

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 first received instructions, followed by a practice trial suited for the current WM load condition. Between the two blocks participants received instructions for the other digit task.

Participants

Sixty-two first-year psychology students (17 men, 40 women; mean age 20) 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) was used for the statistical analyses.

3.2 Results

In this section we first discuss the performance on the digit tasks, followed by the off-line story comprehension results, i.e., answers on the questions and the related reaction times, and the self-paced reading data.

Digit task results

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%, mean errors per trial: 3-digits=0.343, 6-digits=0.852), indicating that the 6-digit condition was indeed more difficult. We did not find any effect of story condition on the number of errors in the digit task or on the percentages correct trials.

Off-line results

Answers. Figure 2 shows the preference for either the first or second character as the referent of the ambiguous pronoun at the end of the test stories.

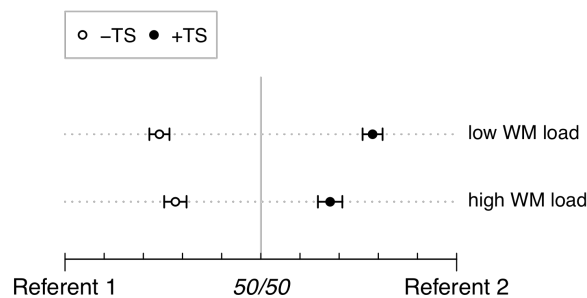


Figure 2: Referent preference for stories without (-TS) and with (+TS) topic shift (\pm SE), plotted separately for both WM load conditions.

Figure 2 shows that participants were sensitive to the topic-shift manipulation. In both WM load conditions, the expected referent was preferred (i.e., Referent 1 in -TS, and Referent 2 in +TS).

We examined the effects of *Topic shift*, *WM load*, and *Trial position*, the position of the trial in the experiment, on the choice for the first character (yes or no) using logistic mixed-effects models (cf. Baayen, 2008). More complex models that included additional predictors did not show qualitatively different effects. In all the presented models, participant and item (i.e., all variants of a story were labeled as the same item) were included as crossed-random effects.

We compared different models using a stepwise variable deletion procedure, starting with the complete interaction model. For every model comparison, we examined whether the difference in -2 log likelihood is significant, given the difference in degrees of freedom using the chi-square distribution. If this difference is significant, the

reduced model has a significantly lower goodness-of-fit, indicating that the deletion of the variable or interaction is not justified. As removing the 3-way interaction did not show a significant difference with the complete interaction model, we selected the model without this three-way interaction as the baseline (or full model). Figure 3 summarizes the model comparisons performed to investigate whether WM load and the type of story affect the choice for the first character (left graph). All two-way interactions (*Topic shift* by *Trial position*, *Trial position* by *WM load*, and importantly *Topic shift* by *WM load*) needed to be included in the statistical model.

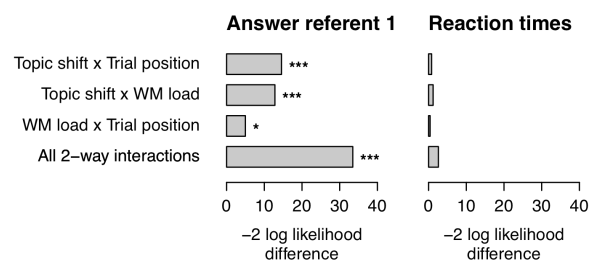


Figure 3. Explained variance in -2 log likelihood of interactions compared to a full model (see text). The statistical significance is calculated with a χ^2 -test (with 1,1,1, or 3 degrees of freedom respectively). (*' $p < .05$, ***' $p < .01$, ****' $p < .001$)

The best model showed that, in stories with topic shift (+TS), the first character was selected more often in the high WM load condition than in the low WM load condition ($\beta=0.844$, $z=3.57$; $p < .001$), in line with the assumption that decreasing working memory capacity reduces pronoun resolution performance. The model showed no general effect of WM load ($\beta=0.308$, $z=1.10$; $p > .1$). Thus, no differential effect for WM load condition was found for the condition without topic shift. In addition, participants were more likely to select the first character in stories with a topic shift as the experiment progressed ($\beta=0.008$, $z=3.81$; $p < .001$), but in the high WM load condition this effect was reduced ($\beta=-0.005$, $z=-2.28$; $p=.023$).

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). However, we did not find any significant interaction (see Figure 3). The best fitting model, which

included the main effects, but no interactions, only showed a significant contribution of *Trial position*: Participants became faster in answering during the experiment ($\beta=-0.002$, $t=-5.97$; $p<.001$).

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 more often select the first character 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 if they experience more WM load.

3.3 Reading time data

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 compared linear mixed-effects models in the same way as before to test the effects of *Topic shift*, *WM load*, and *Trial position* for all moving-window regions on the log-transformed reading times. More complex models that included additional predictors did not show qualitatively different effects. We found no significant 3-way or 2-way interactions in the analyzed regions that needed to be included in the statistical model. We therefore only report on the main-effects model.

Sentence 1. The first sentence of the story is identical in both variants of the stories (-TS and +TS). Figure 4 displays reading times of the first three regions for the two working memory conditions, collapsed over the two story types. The main-effects model showed an effect of *Trial position* (participants read faster as the experiment progressed, $\beta=-0.004$, $t=-13.00$; $p<.001$), an effect of *WM load* (increased reading time in high WM condition, $\beta=0.245$, $t=9.44$; $p<.001$), but no effect of *Topic shift* ($\beta=0.012$, $t=0.48$; $p>.1$). Similar results were found for region 2.

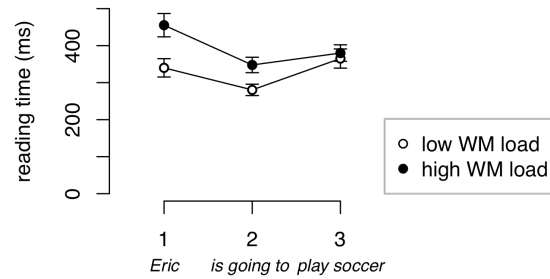


Figure 4. Reading times (raw data) of the first three regions of sentence 1 (\pm SE). (English translation of an example sentence from Table 1)

Sentence 2. In the stories with a topic shift (+TS), the topic shift is initiated in the second sentence, by introducing a new character in subject position. Therefore, we would expect to see differences in reading times between the two story types. In addition, we expected to find an interaction between *WM load* and *Topic shift*, as an early measure of the effect of *WM load* on the off-line data: in the high WM load condition, the previous subject has less influence, therefore we would expect the difference in reading times to be reduced. However, we did not find any significant interaction.

Figure 5 shows the normalized effects of regions 1-4 of the second sentence (normalized by the first region). For region 1, the main effects model revealed that participants became faster over the course of the experiment ($\beta=-0.004$, $t=-12.21$; $p<.001$). However, *Topic shift* ($\beta=0.024$, $t=0.83$; $p>.1$) and *WM load* ($\beta=-0.010$, $t=-0.33$; $p>.1$) did not contribute to the fit of the data.

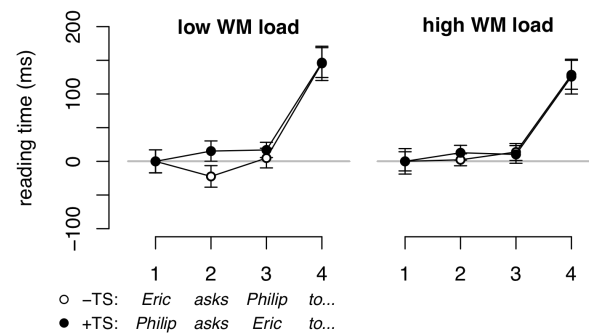


Figure 5. Normalized reading times (difference with region 1) of the first four regions of sentence 2 (\pm SE). (English translations of example sentences from Table 1)

For analyzing the reading times of region 2 we removed possible confounding effects at the beginning of the sentence, such as the effect of *Trial position*, by taking the difference in reading time between the second and first region.² The main-effects model for analyzing region 2 without *Trial position* showed a significant increase in reading time for the stories with a topic shift in comparison with the stories without a topic shift ($\beta=0.085$, $t=2.79$; $p=0.004$), indicating that participants expected to see the subject of the previous sentence instead of a new referent. However, there was no significant contribution of *WM load*. Analyzing the remaining regions of sentence 2 did not show an effect of *Topic shift* or *WM load*. In sentences 3 and 4, we did not find significant effects of *Topic shift* or *WM load*, nor an interaction between these two factors.

To summarize, we found an effect of *WM load* in the first sentence and an effect of *Topic shift* in the second sentence. The longer reading times on the first sentence in the high WM load condition probably reflect some final rehearsing of the digits. However, after this first sentence, no effect of WM load is found.

4 Discussion

We predicted, on the basis of our cognitive model, that adults will show more difficulties in processing a topic shift in higher WM load conditions. 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 preferences 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 off-line data support the prediction of the model: 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

similarly 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, which indicates that readers expected to see the subject of the previous sentence instead of a new referent. However, we did not measure a significant interaction between WM load and type of story. 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 the effects of WM load. An alternative explanation is that the moving-window task is not suited to detect the effect of WM load. It is reasonable to assume that the effect of WM load on the 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.

The link between WM capacity and language processing is not new. For example, 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. The fluctuating activation of chunks also plays a role in our account of the interpretation of pronouns in discourse. This implementation is consistent with the memory account of Foraker and McElree (2007) that characterizes the prominence of discourse elements as differences in strength of their representations in memory (in contrast with a.o. Grosz et al., 1995; Gundel et al., 1993). Our implementation is similar to the account of Reitter, Keller and Moore (2011), who use ACT-R's spreading activation mechanism to explain short-term priming of syntactical structures. In addition, to explain the difference between children's and adults' performance, we implemented the WM theory of Daily, Lovett, and Reder (2001), who manipulated the amount of spreading activation to account for

² Analysis of the absolute reading times revealed the same effects. The reading times of region 1 were included in the analysis of the absolute reading times.

individual differences in working memory capacity on digit span tasks.

Our account is also 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 that is used for storage of intermediate results and for computation. The amount of activation is different for individuals. If more capacity is required for processing or storage than is available, this will result in longer processing times or retrieval errors. On the basis of this theory, Daneman and Carpenter (1980) predicted longer reading times on discourse pronouns for readers with a low WM capacity. In contrast, MacDonald and Christiansen (2002) have argued against the limited capacity theory of Just and Carpenter: they propose instead that differences in WM capacity are differences in skill that arise from variations in exposure to the language, and biological differences. However, our data that shows that WM load can affect the interpretation of stories with a topic shift, is difficult to explain in terms of language skills.

To conclude, on the basis of earlier research (Koster et al., in press) we hypothesized that limited WM capacity might cause decreased comprehension of stories with a topic shift. To investigate how WM capacity affects the comprehension of the discourse structure, we implemented a cognitive model. Our model implied that sufficient WM capacity is necessary for an adult-like interpretation of a potentially ambiguous subject pronoun. With sufficient WM capacity, information about the grammatical roles of the referents in the previous sentence determines the interpretation of the ambiguous pronoun, but readers or listeners without sufficient WM capacity rely more on the base level activation of discourse elements. To test whether adults' performance would decrease when their WM is taxed, we performed a dual-task experiment in which we manipulated the WM load. The results confirmed that with higher WM load, adults are less likely to distinguish between stories with and without a topic shift, similarly to children. Thus WM load can affect the interpretation of ambiguous subject pronouns.

References

- Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press, USA.
- Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.
- Arnold, J. E. (1998). *Reference Form and Discourse Patterns*. Unpublished Ph.D. thesis, Stanford University.
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press.
- Baayen, R. H., & Milin, P. (2010). Analyzing Reaction Times. *International Journal of Psychological Research*, 3(2), 12-28.
- 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. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, 19(4), 450-466.
- 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.
- 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.

- Koster, C., Hoeks, J., & Hendriks, P. (in press). Comprehension and production of subject pronouns: Evidence for the asymmetry of grammar. In A. Grimm, A. Müller, C. Hamann & E. Ruigendijk (Eds.), *Production-comprehension asymmetries in child language*. Berlin: De Gruyter.
- 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.
- 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.
- 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.
- Van Rij, J., Van Rijn, H., & Hendriks, P. (submitted). Production and comprehension of referring subjects: A computational model of the interaction between linguistic and cognitive constraints.