Quantifiers and Working Memory

Jakub Szymanik[1] and Marcin Zajenkowski[2]

[1] Department of Philosophy, Utrecht University
j.szymanik@uva.nl
[2] Faculty of Psychology, University of Warsaw
zajenkowski@psych.uw.edu.pl

Abstract. The paper presents a study examining the role of working memory in quantifier verification. We created situations similar to the span task to compare numerical quantifiers of low and high rank, parity quantifiers and proportional quantifiers. The results enrich and support the data obtained previously in [1–3] and predictions drawn from a computational model [4, 5].
Keywords: working memory; generalized quantifiers; computational semantics; span test

1 Introduction

The role of working memory in language comprehension has been extensively studied (see e.g. [6]). The theory of the specific aspects of memory has been developed by Baddeley and colleagues [7, 8]. They proposed to extend the concept of a short-term memory, suggesting that it could be divided into three separable components. It has been assumed that working memory consists not only from temporary storage units (phonological or visual) but also from a controlling system (central executive). Working together, these components form a unified memory system that is responsible for the performance in complex tasks.

Danemamn and Carpenter (1980) developed span test to asses the working memory construct proposed in [7]. In the task subjects read series of sentences and are asked to remember the final word of each sentence. Data suggest that the result of the span test (the number of correctly memorized words) is a good predictor of language comprehension and other language-processing tasks [9–12]. The main idea of the span test is that solving it requires engagement of both processing and storage functions. In an experimental study a trade-off between them is usually observed. There are two possible explanations of this phenomenon. One is a computational theory according to which storage and processing use the same cognitive resource and compete for a limited capacity [13, 11]. The second is 'multiple resource' theory, where the working memory is viewed as a group of cognitive subsystems each having a specialized function [8, 14]. According to that account performance in a particular task relies on one or more subsystems acting together.

The paper presents a study examining the role of working memory in quantifier verification. We created situations similar to the span task [15]. The aim of our research is to verify the contribution of working memory for a few specific natural language quantifiers.

1.1 Quantifier Verification Model

In [1] the pattern of neuroanatomical recruitment while subjects were judging the truth-value of statements containing natural language quantifiers have been examined using neuroimaging methods. The authors were considering two standard types of quantifiers: first-order (e.g., 'all', 'some', 'at least 3'), and higher-order quantifiers (e.g., 'more than half', 'an even number of'). They presented the data showing that all quantifiers recruit the right inferior parietal cortex, which is associated with numerosity, but only higher-order quantifiers recruit the prefrontal cortex, which is associated with executive resources, like working memory.

The distinction between first-order and higher-order quantifiers does not coincide with the computational resources, like working memory, required to compute the meaning of quantifiers. Cognitive difficulty of quantifier processing might be better assessed on the basis of complexity of the minimal corresponding automata [4, 5]. Taking this perspective, in [3] an analogical reaction time experiment carefully differentiating between the following classes of quantifiers has been conducted (see Table 1). The study has shown that the increase in reaction

Quantifiers	Examples	Minimal automata
logical	'all cars',	acyclic 2-state FA
numerical	'at least k'	acyclic FA with number of states depending on k
parity	'an even number of balls'	2-state FA with loops
proportional	'most lawyers'	PDA

Table 1. Quantifiers and complexity of minimal automata.

time is determined by the minimal automata corresponding to the quantifier. Among others, the results indicate that the numerical and parity quantifiers are processed faster than the proportional quantifiers. This is consistent with computational analysis as only proportional quantifiers demand a recognition mechanism with unbounded internal memory, like a stack in push-down automata (see [3]). Therefore, there is not only a quantitative but also qualitative difference between memory resources which are necessary to compute these two types of quantifiers. This conclusion also follows from the differences in the brain recruitments observed in [1].

1.2 The Present Study

The data obtained so far support the assumption that the difficulty of mental processing of quantifiers depends on the complexity of the corresponding minimal automata. This complexity can be explained by a difference in needed memory resources, e.g., different number of states in the case of various numerical quantifiers. The present paper extends previous results by studying the engagement of working memory during quantifier verification tasks.

We examined three groups of quantifiers: proportional, parity and numerical (high and low rank). We predicted that when subjects are asked to maintain arbitrary information in short-term memory then similar differences between quantifiers should be revealed as those described in [3] as well as in [16]. In particular, the difficulty (indicated by reaction time and accuracy) should decrease as follows: proportional quantifiers, numerical quantifiers of high rank, parity quantifiers, numerical quantifiers of low rank. Additionally, processing of the proportional quantifiers should influence the storage functions. The effect should be stronger in more demanding situation, for instance when the number of elements to be stored in the memory is increasing.

2 Method

2.1 Participants

Sixty native Polish-speaking adults took part in the study. They were volunteers from Warsaw University of Finance and Management undergraduate population. Of these, 18 were male and 42 were female. The mean age was 24 years (SD = 4.75) with a range of 21-40 years. Each subject was tested individually in exchange for partial fulfillment of course credits.

2.2 Materials and Procedure

The general aim of this study was to assess how subjects are judging the truth-value of statements containing natural language quantifiers with an additional memory load. The experiment was a combined task and consisted of two elements. It required participants to verify sentences and to memorize a sequence of single digits for the later recall.

Sentence Verification Task The task consisted of sixty-four grammatically simple propositions in Polish containing a quantifier that probed a color feature of a car on a display, e.g., 'Więcej niż połowa samochodów jest czerwona' (More than half of the cars are red) or 'Parzysta liczba samochodów jest niebieska' (An even number of cars are blue). The same number of color pictures presenting a car park with 15 cars were constructed to accompany the propositions. The colors used for the cars were red, blue, green, yellow, purple and black. Each picture contained objects in two colors (see Figure 1).

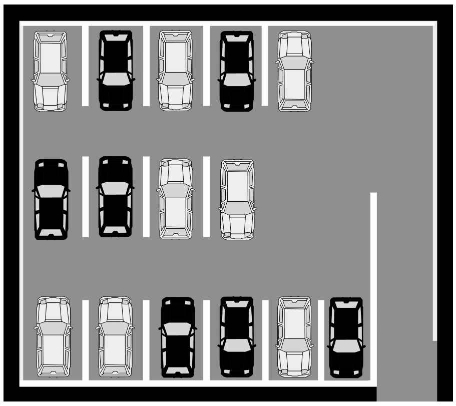

Fig. 1. An example of stimulus used in the first study.

Eight different quantifiers were used in the study. They were divided into four groups:

(1) parity (divisibility) quantifiers (odd, even), DQ;
(2) proportional quantifiers (less than half, more than half), PQ;
(3) numerical quantifiers of relatively low rank (less than 5, more than 4), NQ4/5;
(4) numerical quantifiers of relatively high rank (less than 8, more than 7), NQ7/8.

Each quantifier was presented in 8 trials. Hence, there were in total 64 tasks in the study.

Half of each type of items was true and half false. Propositions were accompanied with a quantity of target items near the criterion for validating or falsifying the proposition. Therefore, these tasks required a precise judgment (e.g. seven targets in 'less than half'). Debriefing following the experiment revealed that none of the participants had been aware that each picture consisted of exactly fifteen objects.

Each quantifier problem involved one 15.5 s event. In the event the proposition and a stimulus array containing 15 randomly distributed cars were presented for 15000 ms followed by a blank screen for 500 ms. Subjects were asked to decide if the proposition accurately described the presented picture. They responded by pressing the button with letter 'p' if true; the button with letter 'f' was pressed if false. The letters refer to first letters of Polish words for 'true' and 'false'.

Memory Task At the beginning of each trial the subjects were presented a sequence of digits consisting of four or six elements from the range between 0 and 9. After completing the sentence verification task they were asked to recall the string. Each quantifier type was accompanied by the same number of four and six digits.

3 Results

3.1 Sentence Verification Task

ANOVA with type of quantifier (4 levels) and number of digits to memorize (2 levels) as two within-subject factors was used to examine differences in means in reaction time and accuracy of sentence verification task. Greenhouse-Geiser adjustment was applied where needed.

The analysis of reaction time indicated that two main effects – of quantifier type ($F(2.282, 134.62) = 41.405$; $p < 0.001$; η^2=0.412) and of number of digits ($F(1, 59) = 4.714$; $p < 0.05$; η^2=0.075) as well as quantifier × digits interaction ($F(2.544, 150.096) = 2.931$; $p < 0.05$; η^2=0.05) – were significant (see Figure 2).

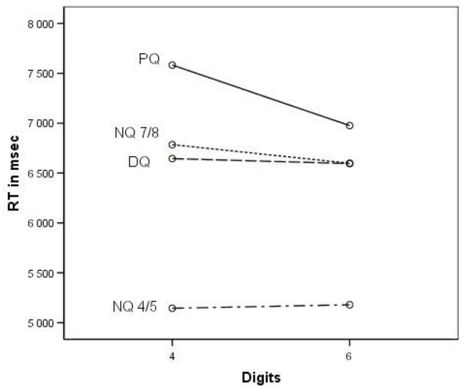

Fig. 2. Mean RT in 4- and 6-digit memory load conditions.

For simple effects we analyzed differences between quantifiers separately for two memory conditions. We found that mean reaction time was determined by quantifier type while subjects were maintaining 4 digits in memory. Pairwise comparisons among means revealed that PQ were solved longer than other types of quantifiers while NQ 4/5 were processed shorter than the rest of quantifiers; finally, there was no difference between DQ and NQ 7/8. In 6-digit condition we also found a significant effect – NQ 4/5 had shorter average RT than other quantifiers. One-way ANOVA revealed that only PQ differed between memory load conditions (see Table 2).

Quantifier	M (4-digit)	M (6-digit)
PQ	7582	6976
DQ	6644	6595
NQ 7/8	6784	6598
NQ 4/5	5144	5179

Table 2. Mean RT in milliseconds for each quantifier type.

The main effects of quantifier type ($F(2.574, 151.867) = 22.238$; $p < 0.001$; η^2=0.275) and of digits ($F(1, 59) = 4.953$; $p < 0.05$; η^2=0.078) were found in accuracy. All four types of quantifiers differed significantly from one another besides DQ and NQ 7/8 (see Table 3 for mean score). In 4-digit condition all quantifiers were performed worse (M = 6.22) than in 6-digit condition (M = 6.43).

Quantifier	M
PQ	5.57
DQ	6.36
NQ 7/8	6.45
NQ 4/5	6.93

Table 3. Mean (M) of the accuracy for each type of quantifier.

Summing up, we observed that in 4-digit memory load condition proportional quantifiers were solved longer and poorer than other types of quantifiers. On the other hand, numerical quantifiers with low rank were performed shorter and better than others. There was no difference between parity quantifiers and numerical quantifiers of high rank.

In 6-digit condition we observed lower average reaction time of numerical quantifiers of low rank in comparison with proportional, parity and numerical quantifiers of high rank, which had equal means. Analysis of accuracy showed the following increase of difficulty: numerical quantifiers of low rank, then parity quantifiers and numerical quantifiers of high rank (the same level), and finally proportional quantifiers.

Finally, the accuracy on all types of quantifiers was better in 6-digit condition. However, as we will see in the next section there was a significant drop in recalling task.

3.2 Memory Task

ANOVA with two within-subject factors was used to examine how strings of digits (2 levels: four and six elements) were recalled with respect to quantifier type (4 levels) they were accompanied by. Greenhouse-Geiser adjustment was applied where needed.

The analysis indicated main effect of digits ($F(1, 59) = 90.646$; $p < 0.001$; η^2=0.606) and digits × quantifier interaction ($F(3, 177) = 4.015$; $p < 0.05$; η^2=0.065) (see Figure 3).

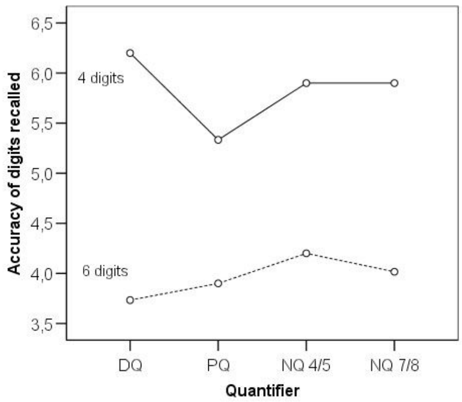

Fig. 3. Accuracy of 4- and 6-digit recall with respect to quantifier type.

To examine the interaction effect we compared recall accuracy for 4 and 6 digits. Significant differences between two situations for each level of second variable were obtained. Performance on digit recall with respect to quantifier type was also analyzed separately for 4- and 6-digit strings. In the former condition digits accompanying PQ were memorized worse in comparison with other determiners, while in the latter condition we did not observe any differences (see Table 4).

Number of digits	M (PQ)	M (DQ)	M (NQ 7/8)	M (NQ 4/5)
4	5.33	6.20	5.90	5.90
6	3.90	3.73	4.01	4.20

Table 4. Means of recalling accuracy with respect to quantifier type.

4 Discussion

Our study assessed quantifier verification task with additional memory load conditions. Obtained data revealed that in the 4-digit load condition the most difficult were proportional quantifiers (the longest RT and the poorest accuracy). Subjects performed better on numerical quantifiers with low ranks than on the other determiners, and finally there were no differences between parity quantifiers and numerical quantifiers of high rank. The results support our predictions and are consistent with the previous findings in [3] and [17].

We expected similar effects in 6-digit memory load condition. This hypothesis was only confirmed with respect to sentence verification accuracy. The score increased in all types of quantifiers but differences between them remained at the same level as in 4-digit condition. Moreover, we observed that numerical quantifiers of low rank had the lowest average reaction time. Proportional, parity and numerical quantifiers of high rank had equal means.

The discrepancy between performances under two memory load conditions needs explanation. We believe that the analysis of digits retrieval sheds some light on the obtained data. The real differences between quantifiers occurred only in 4-digit condition. Holding six elements in memory was probably too difficult in face of processing secondary task. The decrease of accuracy in digits recall with simultaneous increase in performance on quantifier verification task could be described as a trade-off between processing and storage (see [15, 14]).

Another interesting observation concerns proportional quantifiers. In 4-digit condition the strings of numbers accompanying this class of quantifiers were recalled worst. However, in the case of 6-digit memory load there were no differences among quantifier types. It is worth to put those results together with the data on the reaction time for proportional quantifier verification. The mean RT decreased because subjects focused only on the sentence verification task ignoring the recalling task. This may be interpreted as supporting the hypothesis, following from the computational model, that working memory engagement in the case of proportional quantifier processing is qualitatively different than in the processing of quantifiers corresponding to finite-automata.

An interesting result is tied up with numerical quantifiers and the number of states in the corresponding minimal automata. In [5] it has been hypothesized that the number of states is a good predictor of cognitive load. Indeed, our current results show the difference between numerical quantifiers of low and high ranks. This fact strongly supports that claim.

Finally, let us briefly discuss a problematic case. The relation between parity and numerical quantifiers of high rank is somewhat unclear. In our previous study [3] we observed a significant difference in reaction time between those two types of quantifiers. However, the size effect of the difference was smaller than in other pairwise comparisons among quantifiers. Can the computational model account for the discrepancy? It draws an analogy between states and stack, on the one hand, and working memory resources, on the other hand. The difference between parity and numerical quantifiers can not be explained in that way. Minimal automata corresponding to parity quantifiers have two states while in the case of numerical quantifiers one needs in principle more. However, the critical factor might be that numerical quantifiers unlike parity quantifiers correspond to automata without loops (see Table 1). Clearly, 2-state automata with loops are more complex than 2-state acyclic machines (corresponding to Aristotelian quantifiers) and indeed our previous research has shown a difference between the two quantifier groups [3]. However, drawing only from the computational model it is by no means obvious which factor adds more to cognitive difficulty: additional states or loops[3]. This constitutes one of the most interesting problems for our approach (see [2] for a more detailed discussion). A future research focusing on neurocognitive modeling of quantifier comprehension could help in clarifying the interrelations among computational aspects and their cognitive correlates. The aim would be to pin down the specific cognitive mechanisms responsible for quantifier comprehension, taking into account factors like the role of central executive, attentional costs, storage functions as well as aspects of representing and approximating quantities, like distant effect (see e.g. [18]). After all, quantifiers might be viewed as a way of embedding number system [19] in natural language. The perspective needs to be carefully investigated in the future.

References

1. McMillan, C., Clark, R., Moore, P., Devita, C., Grossman, M.: Neural basis for generalized quantifiers comprehension. Neuropsychologia **43** (2005) 1729–1737
2. Troiani, V., Peelle, J., Clark, R., Grossman, M.: Is it logical to count on quantifiers? Dissociable neural networks underlying numerical and logical quantifiers. Neuropsychologia **47** (2009) 104–111
3. Szymanik, J., Zajenkowski, M.: Comprehension of simple quantifiers. Empirical evaluation of a computational model. Cognitive Science **in press** (2009)
4. van Benthem, J.: Essays in logical semantics. Reidel (1986)
5. Szymanik, J.: A comment on a neuroimaging study of natural language quantifier comprehension. Neuropsychologia **45** (2007) 2158–2160
6. Baddeley, A.: Working memory and language: an overview. Journal of Communication Disorders **36** (2003) 189–208
7. Baddeley, A., Hitch, G.: Working memory. In Bower, G., ed.: The psychology of learning and motivation. New York: Academic Press (1974) 47–90
8. Baddeley, A.: Working memory. Oxford: Oxford University Press (1986)
9. Conway, A., Engle, R.: Individual differences in working memory capacity - more evidence for a general capacity theory. Memory **6** (1996) 122–125
10. Daneman, M., Green, I.: Individual differences in comprehending and producing words in context. Journal of Memory and Language **25** (1986) 1–18
11. Just, M., Carpenter, P.: A capacity theory of comprehension: Individual differences in working memory. Psychological Review **99** (1992) 122–149

[3] Notice, that necessity of using loops suggests more complex verification strategies, e.g., people can try to pair object or just count all of them and then divide by 2. There might be no obvious minimal startegy corresponding to the one coded by the automata. The hypothesis would be that if train in using the minimal startegy after a while subjects will improve on parity quantifiers, even performing better than on numerical quantifiers.

12. King, J., Just, M.: Individual differences in syntactic processing: The role of working memory. Journal of Memory and Language **30** (1991) 580–602
13. Daneman, M., Merikle, P.: Working memory and language comprehension: A meta-analysis. Psychonomic Bulletin and Review **3** (1996) 422–433
14. Duff, S., Logie, R.: Processing and storage in working memory span. The Quarterly Journal of Experimental Psychology **54** (2001) 31–48
15. Daneman, M., Carpenter, P.: Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior **19** (1980) 450–466
16. Troiani, V., Peelle, J., Clark, R., Grossman, M.: Magnitude and parity as complementary attributes of quantifier statements. Neuropsychologia **47** (2009) 2684–2685
17. Szymanik, J., Zajenkowski, M.: Improving methodology of quantifier comprehension experiments. Neuropsychologia **47** (2009) 2682–2683
18. Pietroski, P., Lidz, J., Hunter, T., Halberda, J.: The meaning of 'most': semantics, numerosity, and psychology. Mind and Language **24**(5) (2009) 554–585
19. Dehaene, S.: The Number Sense: How the Mind Creates Mathematics. Oxford University Press, USA (1999)