Empty-set effects in the comprehension and verification of quantifiers

Experimental results show that the monotonicity of a quantifier (Q) affects processing load. Monotone decreasing Qs are inherently more complex than increasing Qs [2, 6, 4]. It is, however, poorly understood how to account for monotonicity effects. In theoretical work on quantificational complexity in the tradition of [12, 10] Qs such as *less than/more than four* are modeled with the same kind of semantic automata (see also [9] for a generalization to iterated quantifiers with identical (minimal) automata for doubly quantified *every-some* and *no-no* sentences). Our study aimed at testing the recent proposal by [1, 5] who put forward a novel explanation of complexity differences between decreasing and increasing Qs based on the property of having/not having the empty set $\emptyset$ as a witness set. They propose an algorithmic theory in the sense of [8] according to which comprehension of a quantified sentence consists in specifying an algorithm with which it can be verified. Its actual verification is the result of the execution of this algorithm. As in [1], we measured quantificational complexity during comprehension and verification.

[1]'s basic claim is that Qs that do not have the empty set as a witness set such as *most*, *every*, *exactly five* (henceforth non-$\emptyset$-Qs) allow for a substantially simpler verification algorithm than Qs with the empty set as a witness set such as *no* or *less than half* (henceforth $\emptyset$-Qs). Consider, for instance, the proposed verification algorithm for *more than five dots are blue*. In order to verify this sentence it suffices to check whether the instances of the predicate *blue* in the restrictor set are a witness set of the non-$\emptyset$-Q *more than five* and the tuple $\langle$*more than five(dots)* $, +\rangle$ is added to the denotation of *blue*. [1] call this operation simple expansion (**s-exp**). For an $\emptyset$-Q (e.g., *less than five*), however, **s-exp** is not sufficient because it only considers (positive) instances of the predicate *blue*. Here, [1] distinguish two cases which will both lead to a *yes, true* judgment. Expansion with $\langle$*less than five(dots)*$, +\rangle$ is possible iff among the blue entities there is a set of dots small enough to be a witness set of *less than five*, **or**, there are no blue dots, i.e. an empty-set situation. The complex expansion operation (**c-exp**) needed to deal with $\emptyset$-Qs hence involves a distinction of cases, and, as [1] argue, it is the second condition that is difficult to process, because it involves polarity reversal, i.e., the expansion operation adds positive information on the basis of absent positive instances.

Experiment 1: We compared effects of quantificational complexity in monotone increasing, decreasing, and non-monotone Qs in simply quantified sentences (36 German items of the kind *more than/less than/exactly five dots are blue*). 48 participants read and evaluated simply quantified sentences and we measured reading times before they went on to the picture verification stage. Besides *monotonicity* we manipulated how many (n=0 to 11) out of a total of 11 objects of the relevant shape had the target color in a 3×12 within design. The experimental results provide evidence for clear empty set effects during comprehension and during verification. *Less than* sentences were read more slowly than *more than* and *exactly* sentences, an effect consistent across all sentence regions (ANOVAs all $p_{1,2} < .01$). During verification, the evaluation of empty set situations led to significant interactions of the form predicted by [1] both in judgment RTs (ANOVAs $p_{1,2} < .01$) and in error rates (GLME $p < .01$). Evaluation of empty set situations for *less than* led to 25% errors while all other conditions were judged correctly $> 90\%$. Testing the three quantifiers with all ratios of target:non-target objects allows us to rule out a pragmatic alternative explanation of the observed empty set effect based on Gricean reasoning.

Experiment 2: The second experiment demonstrates empty set effects while keeping monotonicity constant. We tested the non-monotone Boolean combinations of Qs *none or three* and *one or three* with pictures showing 0,1,2 or 3 target objects in a 2×4 within design. 48 new participants read 32 sentences such as *none/one or three of the dots are blue* in a self-paced fashion and provided a truth value judgment. The verification data again revealed clear empty set effects. The evaluation of empty set situations for *none or three* was the only condition that led to a significant increase in judgment RT (interaction: ANOVAs $p_{1,2} < .01$) and error rates (interaction: GLME $p < .01$) which reached 30% in this condition. The results of the first two experiments thus confirm [1]'s claim that the empty set property - beyond monotonicity - has a measurable impact on quantificational complexity.

What expansion operation is needed in iterated quantification with multiple Qs? A theorem of [1] is that a multiply quantified statement can be evaluated by means of **s-exp** iff none of its Qs has the empty-set as a witness set. We may thus assume that **c-exp** is used for doubly quantified sentences whenever either Q has the empty-set property. This would predict the same amount of processing load during comprehension whenever one of the Qs has the empty-set property. Difficulty during verification should, however, depend on the rules actually used.

Alternatively, both the subject and the object position could be expanded by choosing the simplest expansion operation for each Q. For example, in the sentence *no boy tickled more than two girls* the object position may be expanded using simple expansion. That is, for every boy x, the tuple $\langle x,$ *more than two(girls)* $, +\rangle$ is added iff the set $\{y|tickle(x, y)\}$ is a witness set of *more than two girls*. Otherwise the tuple $\langle x,$ *more than two(girls)* $, -\rangle$ is added. This would predict that both Qs may – independently of each other – increase difficulty, depending on whether they are $\emptyset$-Qs or not.

Experiment 3: To test these predictions we had 72 new participants read 72 German sentences of the form Q_1 *boys tickled* Q_2 *girls* (see [1] for a more detailled description of this experiment). Q_1 was either Aristotelian *each* (non-$\emptyset$) or *no* ($\emptyset$) or a superlative Q *at least one* (non-$\emptyset$) or *at most one* ($\emptyset$). Q_2 was either *more than two/three* (non-$\emptyset$) or *less than two/three* ($\emptyset$). After each sentence, a set diagram was displayed for which participants had to provide a truth-value judgment. We tested three types of diagrams: 0-, 1- and 3-set diagrams which showed no, one or three boys tickling Q_2 girls. Accordingly, the experiment employed a factorial 2 (*empty-set* Q_1) $\times 2$ (*type of* Q_1) $\times 2$ (*empty-set* Q_2) $\times 3$ (*diagram*) within design. On the object region, conditions with two non-$\emptyset$-Qs were read faster than conditions involving $\emptyset$-Qs (ANOVAs: *empty-set* Q_1 $\times$ *empty-set* Q_2, $p_1 < .05$; $p_2 = .09$). This is in line with the hypothesis that participants chose **c-exp** whenever one of the two Qs had the empty-set property. In both subject and object position, $\emptyset$-Qs led to a slowdown of judgments times as reflected by significant main effects of *empty-set* Q_1 ($p_{1/2} < .01$) and *empty-set* Q_2 ($p_{1/2} < .01$) but no reliable interaction. The empty set effects of the individual Qs thus added up. In particular, this pattern of results can be explained if we assume that both Qs are expanded using the simplest possible algorithm, i.e. the second theoretical alternative sketched above. Furthermore, consistent with recent proposals from the semantic literature (e.g., [3]) superlative Qs were more difficult to process than Aristotelian Qs consistently across the reading and verification stages (main effects of *type of* Q_1, all $p_{1/2} < .01$) suggesting that they are inherently more complex. As in Exps. 1 & 2, errors were limited to those conditions which required the evaluation of a $\emptyset$-Q with an empty predicate. This time, polarity reversal led to performance below chance level in sentences with the two $\emptyset$-Qs *at most* and *less than* ($\chi^2(1) = 5.06$; $p < .05$).

Conclusions: In line with [1] our experiments show that $\emptyset$-Qs are more difficult to comprehend and to verify than non-$\emptyset$-Qs and that this difficulty is limited to the the evaluation of empty set situations. Exps. 1 & 2. investigated the online comprehension and verification of simply quantified sentences. The cumulative increase of complexity caused by the empty set property of the individual Qs in the verification data of Exp. 3 calls for an expansion operation that incorporates the simplest algorithm for each quantifier. In general, our findings speak in favour of multiple factors independently affecting quantificational complexity, among them the empty set property but also superlativity as well as others known from experimental research on quantificational complexity [7, 11].

[1] O. Bott, U. Klein, and F. Schlotterbeck. Witness sets, polarity reversal and the processing of quantified sentences. In M. Aloni, M. Franke, and F. Roelofsen, editors, *Proceedings of the 19th Amsterdam Colloquium 2013*, pages 59–66, ILLC, Amsterdam, 2013.

[2] B. Geurts, N. Katsos, J. Moons, and L. Noordman. Scalar quantifiers: logic, acquisition, and processing. *Language and Cognitive Processes*, 2010.

[3] B. Geurts and R. Nouwen. "At least" et al.: the semantics of scalar modifiers. *Language*, 83(3):533–559, 2007.

[4] B. Geurts and F. van der Slik. Monotonicity and processing load. *Journal of Semantics*, 22:97–117, 2005.

[5] U. Klein. Quantification, witness sets and conservativity. Proceedings of the 29th West Coast Conference on Formal Linguistics, pages 125–133. Cascadilla Press, 2012.

[6] J. Koster-Moeller, J. Varvoutis, and M. Hackl. Verification procedures for modified numeral quantifiers. In N. Abner and J. Bishop, editors, *Proceedings of the 27th West Coast Conference on Formal Linguistics*, Somerville, MA, 2008. Cascadilla Proceedings Project.

[7] C. T. Mcmillan, R. Clark, P. Moore, C. Devita, and M. Grossman. Neural basis for generalized quantifiers comprehension. *Neuropsychologia*, 43:1729–1737, 2005.

[8] Y. Moschovakis et al. Sense and denotation as algorithm and value. *Lecture notes in logic*, 2:210–249, 1994.

[9] S. Steinert-Threlkeld and T. F. I. III. Iterating semantic automata. *Linguistics and Philosophy*, 36:151–173, 2013.

[10] J. Szymanik. *Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language*. PhD thesis, University of Amsterdam, Amsterdam, 2009.

[11] J. Szymanik and M. Zajenkowski. Comprehension of simple quantifiers: Empirical evaluation of a computational model. *Cognitive Science*, 34(3):521–532, 2010.

[12] J. van Benthem. *Essays in Logical Semantics*. Reidel, 1986.