The process of evaluating comparative quantifiers

Introduction: Quantifiers are nowadays often treated as decomposable instead of atomic expressions (overview: Szabolcsi 2010). Hackl (2002) analyzed comparative quantifiers using lexemes originally proposed for comparatives with gradable adjectives. For sentence (1-a), he proposed the LF in (1-b), where ***er*** compares the quantities denoted by ***n*** and ***d-many***(***dots***, ***blue***).

(1) a. *More than n dots are blue.*
 b. [***er n***][[***d-many dots***][***blue***]]

Seuren-Rullmann ambiguities (Rullmann 1995, Heim 2006) and *cross polar nomalies* (Büring 2007) motivated decomposition of '*less than*' comparatives using ***er*** and the *antonymizer* ***little***. In these approaches, the meaning of '*less*' cannot be expressed as a semantic primitive but has to be derived from ***er*** and ***little*** (but see Beck, 2013; Doetjes *et al.*, 2011, for accounts without this kind of decomposition). Carrying this over to quantifiers, the sentence in (2-a) receives the LF in (2-b).

(2) a. *Less than n dots are blue.*
 b. [[***little er***]***n***][[***d-many dots***][***blue***]]

I present experimental evidence for these semantic representations from a picture verification experiment. The dependent variables are response times and proportions of errors. To make predictions about these, a processing theory for verification is derived from the symbolic representations (1-b) and (2-b) by introducing a minimal set of processing assumptions that are well-established in cognitive psychology.

Processing theory: Based on studies on the mental representations of numbers and numerosities (Dehaene 2007), it is assumed that the quantities ***n*** and ***d-many***(***dots***,***blue***) are represented as bell-shaped functions f_X and f_Y, respectively. We think of these as *probability density functions* of random variables X and Y (cf. Pietroski *et al.* 2009). X is distributed normally with mean $\mu_X = n$ and very small standard deviation σ_X (almost exact representation). In contrast, Y has a (right-skewed) log-normal distribution with parameters μ_Y and σ_Y. The maximum of f_Y (see (3-a)) lies at m which is the number of target objects (e.g., blue dots); and σ_Y is considerably larger than 0.

Applied to its arguments, ***er*** computes a representation of the difference between the compared quantities (cf. Link's (1990) *theory of relative judgment*), i.e. the density function f_{X-Y} of $X - Y$. This computation uses cross-correlation, $(f \star g)(x) = \int_{-\infty}^{\infty} f(\tau)g(x+\tau)d\tau$, as in (3-b) and can be approximated using biologically plausible algorithms (Deneve *et al.* 2001). Because σ_X is very small, ***er*** approximates a shift of f_Y as in (3-c). For *'less than n'*, there is an additional computational step. The antonymizer ***little*** reflects f_{X-Y} at the origin as in (3-d).

(3) a. $f_Y(x) = (x\sigma_Y\sqrt{2\pi})^{-1}e^{-\frac{(lnx-\mu_Y)^2}{2\sigma_Y^2}}$
 b. $\boldsymbol{er}(f_X, f_Y)(x) = (f_X \star f_Y)(x) = f_{X-Y}(x)$
 c. $\lim_{\sigma_X \to 0} f_{X-Y}(x) = f_Y(x+n)$
 d. $(\boldsymbol{little}(f_{X-Y}))(x) = f_{X-Y}(-x)$

The function f_{X-Y} (or ***little***(f_{X-Y}) for *'less than'*) feeds into the decision process. The intuition is that the mean activities $I_{1/2}$, specified in (4-a/b) (or $I'_{1/2}$ for *'less than'*), of two neural populations (favoring the yes- and no-response, respectively, and subject to random fluctuations) are accumulated over time until some confidence threshold is reached (*leaky competing accumulator model* of Usher & McClelland 2001). Bogacz *et al.* (2006) show that this closely approximates the classical *drift diffusion model* (*DDM*, Ratcliff 1978; stochastic differential equation in (5)) with drift rate $\mu = (I_1 - I_2)/\sqrt{2}$. Intuitively, the drift rate is the mean step size towards one of the response alternatives per time unit. Mean decision times and proportions of errors decrease when the absolute value of μ is increased. The Wiener process W is scaled by c and represents white noise.

(4) a. $I_1 = P(X - Y \geq 0.5) = \int_{0.5}^{\infty} f_{X-Y}(x)dx,$ $(I'_1 = \int_{0.5}^{\infty}(\boldsymbol{little}(f_{X-Y}))(x)dx)$
 b. $I_2 = P(X - Y < 0.5) = 1 - I_1$ $(I'_2 = 1 - I'_1)$

(5) $dY = \mu dt + cdW,$ $Y_0 = 0$

Experimental design: 48 participants read German versions of sentences like (1-a) and (2-a) with numerals *'four'*, *'six'*, *'eight'* and *'ten'* self-paced. Then, pictures of randomly distributed shapes of two colors were presented. Participant had to provide a truth-value judgment by pressing one of two buttons. The number of target objects was equal to the number mentioned in the sentence. Thus, the correct response was always "no". Response key assignments were counterbalanced between participants. Every picture was combined once with *'more than n'* and once with *'less than n'*. Sentence-picture pairs were distributed over eight experimental lists making sure that participants saw each picture only once. Each participant provided ten judgments per condition. 343 fillers were added to each list leading to a balanced proportion of expected yes/no-judgments. Response times and proportions of errors and were recorded.
Predictions: Because of the skewness of f_{X-Y}, drift rates are larger for the *'less than'* as compared to the *'more than'* conditions. This predicts less errors for the former than for the latter. Response times are the sum of *non-decision times* (including, e.g., information encoding and motor response) and *decision times* (time needed for the decision process). Due to larger drift rates, *'less than'* conditions have shorter *decision times*, but the decision process starts later due to the additional processing step for ***little*** (part of *non-decision time*). This additional step prolongs response times for the *'less than'* conditions by a constant amount of time.
Results: The proportions of errors and response times for correct judgments are provided in the table below. Mixed effects model analyses revealed significant main effects of *quantifier type* (*'more than'* vs. *'less than'*) in the proportions of errors ($p < .01$) and in the response times ($p < .01$). The *'less than'* conditions led to less errors but longer response times (even for correct judgments) than the *'more than'* conditions. Apart from these effects, the larger numerals led to more errors ($p < .001$) and longer response times ($p < .001$) than smaller ones. Finally, there was a marginal interaction ($p = .052$) between the *quantifier type* and the *size of the numeral* in the proportions of errors.

	proportions of errors				response times in ms			
	four	*six*	*eight*	*ten*	*four*	*six*	*eight*	*ten*
more than	0.04	0.08	0.13	0.18	1345	2088	2890	3481
less than	0.04	0.10	0.07	0.13	1413	2132	2966	3571

Discussion: The experimental results are in line with the presented processing theory. As predicted by the *DDM*, *'less than'* led to fewer errors than the *'more than'* conditions. In addition, the prolonged response times of the former conditions are taken as evidence for the additional processing step introduced by the antonymizer ***little***. Since the processing theory was derived from (1-b) and (2-b) via a minimal set of processing assumptions, the results also provide evidence for these symbolic representations. The strength of the present account is that it combines insights into the compositional semantic representations of comparatives with insights into decision processes involved in comparing quantities. Other theories have difficulty explaining the results. The *automata model* of quantifier verification predicts no differences between *'more than n'* and *'less than n'*. Theories that assume negation in *'less than'* quantifiers can explain their prolonged response times, but their low proportions of errors remain unexplained.

As a further empirical test, I am currently conducting a *speed accuracy trade-off* experiment, where participants provide judgments within certain time windows. The dependent variable are proportions of errors. It is predicted that, early after the onset of the (200ms) picture presentation, judgments are at chance because information encoding is not finished and the decision process has not started, yet. For *'less than'*, this phase is predicted to last longer than for *'more than'* because an additional encoding step (***little***) is involved. In later time windows, error rates for *'less than'* should drop faster as compared to *'more than'* due to a larger drift rate (faster decision process).

References: Beck. S. (2013) *Journal of Semantics.* Bogacz, R. *et al.*. (2006), *Psychological Review 113(4).* Büring, D. (2007). *Proceeding of SALT XVII.* Dehaene, S. (2007). *Sensorimotor foundations of higher cognition.* Deneve, S. *et al.* (2001). *Nature Neuroscience 4(8).* Doetjes, J. *et al.* (2011). *Proceedings of SALT XIX.* Hackl, M. (2002). *Proceedings of the 20th WCCFL.* Heim, I. (2006). *Proceeding of SALT XVI.* Link, S. (1990). *Journal of Mathematical Psychology 34.* Pietroski, P. *et al.* (2009). *Mind and Language 24.* Ratcliff, R. (1978). *Psychological Review 85(2).* Rullmann, H. (1995). *Maximality in Semantics of Wh-Constructions.* Szabolcsi, A. (2010). *Quantification.* Usher, M. & McClelland, J. L. (2001). *Psychological Review 108(3).*