Experimental support for a modal analysis of "at least" and "at most"

Bart Geurts

University of Nijmegen

Introduction

In mathematics, "$n > 2$" and "$n < 3$" are equivalent with "$n \geq 3$" and "$n \leq 2$", respectively, and for a long time it was believed that analogous equivalences hold for (what would seem to be) the linguistic counterparts of "$>$" et al. Arguing against this assumption, Geurts and Nouwen (2007) show that "at least n" is not equivalent with "more than n-1", and that "at most n" is not equivalent with "less than n+1". For example, if these equivalences held, the (a) and (b) sentences should be equally acceptable, which they are not:

(1) a. I will invite at most 2 people: Jack and Jill.
     b. ?I will invite fewer than 3 people: Jack and Jill.

(2) a. I will invite at least 2 people: Jack and Jill.
     b. ?I will invite more than 1 person: Jack and Jill.

According to the alternative analysis proposed by G&N, superlative quantifiers ("at least/most") have a two-part modal meaning. On their account, (3a) means that (a) the speaker is certain that Betty had 3 sherries and (b) considers it possible that she had more. (3b) simply means, as on the standard view, that the number of sherries Betty had exceeds 2:

(3) a. Betty had at least 3 sherries.
     b. Betty had more than 2 sherries.

G&N's analysis of the difference between "at most" and "less than" is analogous.

G&N's theory of "at least/most" makes two kinds of predictions that can be tested by experimental means. First, it predicts that the two types of quantifiers should give rise to different inference patterns. Secondly, it predicts that superlative quantifiers ("at least/most") are more complex and therefore harder to process than comparative ones ("more/less than"). In the following, I will discuss these predictions in some detail, and present two experimental studies designed to test whether the differences predicted by the theory are real.

1. Inference patterns

According to G&N's theory, a valid argument need not remain valid when a comparative quantifier is replaced with its superlative counterpart, and we should expect these differences to be reflected in human reasoning. To illustrate, the following arguments are all valid on the standard account, while on the modal analysis (4b) is invalid and (5b) may be problematic for pragmatic reasons:

(4)   a.   (i) Wilma had 3 beers $\rightsquigarrow$ (ii) Wilma had more than 2 beers.
      b.   (i) Wilma had 3 beers $\rightsquigarrow$ (ii) Wilma had at least 3 beers.

(5)   a.   (i) Wilma had 3 beers $\rightsquigarrow$ (ii) Wilma had fewer than 4 beers.
      b.   (i) Wilma had 3 beers $\rightsquigarrow$ (ii) Wilma had at most 3 beers.

(4b.ii) means, according to G&N, that (a) the speaker is certain that Wilma had 3 beers and (b) considers it possible that she had more. So a speaker who claims (4b.i) as well as (4b.ii) flatly contradicts himself: if Wilma had 3 beers, it is not (epistemically) possible that she had more. (Note that this story presupposes that "3 beers" means "3 beers and no more", an assumption that turns out to be moot, as we will presently see.)

(5b) is rather different. The modal meaning of (5b.ii) is that (a) the speaker considers it possible that Wilma had 3 beers, but (b) rules out the possibility that she had more. Logically speaking, this follows from the premiss that Wilma had 3 beers, but on a pragmatic level there is a tension between that premiss and the (a) component of the conclusion: for it might be argued that, if Wilma had 3 beers, it is not merely possible but (epistemically) necessary that she had.

Suppose someone who believes that φ *must* be the case is asked whether φ *might* be the case. How would he respond? Experimental evidence suggests

that the outcome will be mixed. In the context of an acquisition study, Noveck (2001) asked adult controls whether there might be a parrot in a box that was known to contain a parrot. In one experiment, 35% of participants said there might be, while in another the percentage was as high as 75%. (The main difference between the two experiments was that the latter included a more intensive training session.) Similarly, in an unpublished study, I asked participants to imagine Betty saying: "I'm certain that Fred is drunk", after which they had to decide whether or not this implied that Betty believes that Fred might be drunk. 61% of the participants said that it did.

Although there is a great deal of variation between these studies, they agree that people's responses to this type of task are mixed. Assuming this much is right, the modal analysis of "at most" leads us to expect that people's responses to cases like (5b) will likewise be mixed.

Experiment 1
The first experiment was a paper-and-pencil study, in which 28 native speakers of Dutch were presented with single-premiss arguments in two different categories. Arguments of the first category were designed to test the differences between superlative and comparative quantifiers. Arguments of the second category explored how these quantifiers interact with disjunction.

The most important critical items and results are given in (6)-(8), where the numbers in square brackets are the percentages of participants that endorsed the preceding argument:

(6)   a.   Wilma had 3 beers $\overset{?}{\leadsto}$ Wilma had more than 2 beers.   [1]

     b.   Wilma had 3 beers $\overset{?}{\leadsto}$ Wilma had at least 3 beers.   [.48]

(7)   a.   Wilma had 3 beers $\overset{?}{\leadsto}$ Wilma had fewer than 4 beers.   [.92]

     b.   Wilma had 3 beers $\overset{?}{\leadsto}$ Wilma had at most 3 beers.   [.58]

(8)   a.   Wilma had 3 beers $\overset{?}{\leadsto}$ Wilma had 2 or 3 beers.   [.22]

     b.   Wilma had 3 or 4 beers $\overset{?}{\leadsto}$ Wilma had at least 3 beers.   [.96]

The main findings are the contrasts between (6a) and (7a), on the one hand, and (6b) and (7b), on the other. While the standard theory doesn't predict these contrasts, G&N's account does. However, there is one wrinkle in the data: although the response rates for (6a) and (6b) are significantly different, as predicted, it was unexpected that the argument in (6b) should still be endorsed at a fairly high rate. I believe this is because, in the context

of arguments involving the quantifier "at least 3 beers", some participants adopted an "at least" interpretation of the expression "3 beers"—an explanation that is confirmed by the fact that 48% of the participants endorsed the reversed variant of (6b), as well:

(9)   Wilma had at least 3 beers $\overset{?}{\leadsto}$ Wilma had 3 beers.

(See Geurts 2006 for an analysis of number words which predicts that, while the dominant interpretation of "3 beers" is exact, there is a recessive "at least" construal, as well.) Additional support for this explanation was obtained in a follow-up study, in which we compared plain "3 beers" with "exactly 3 beers", expecting that the latter would reduce the rate of positive responses. This turned out to be the case:

(10)   a.   Wilma had 3 beers $\overset{?}{\leadsto}$ Wilma had at least 3 beers.   [.58]
       b.   Wilma had exactly 3 beers $\overset{?}{\leadsto}$ Wilma had at least 3 beers.   [.21]

(11)   a.   Wilma had at least 3 beers $\overset{?}{\leadsto}$ Wilma had 3 beers.   [.58]
       b.   Wilma had at least 3 beers $\overset{?}{\leadsto}$ Wilma had exactly 3 beers.   [0]

Note, finally, the contrast between the response rates for the arguments in (6b) and (8b). On the face of it, the only difference between these arguments is that the premiss of first argument is stronger (i.e. more informative). Nevertheless, whereas the second argument is accepted by practically everybody, the first argument is rejected half of the time. This phenomenon makes little sense from a logical point of view, but is readily explained by G&N's theory. The sentence "Wilma had 3 or 4 beers" conveys that the speaker is certain that Wilma had 3 beers and considers it possible that she more than 3 beers, and according to the modal analysis this is precisely what the conclusion says.

2. Complexity

The most straightforward prediction made by G&N's theory is that the meanings of superlative quantifiers are more complex than those of their comparative counterparts. If this is true, it is natural to expect that, at some stage in the process of language learning, children will have problems with superlative but not with comparative quantifiers. Experimental data reported by Musolino (2004) suggest that this is indeed the case. Musolino

presented 5-year-old children with collections of cards showing various numbers of smiley faces or stars. In each trial, children had to select the cards meeting one of the following descriptions (as before, the rates of correct responses are given in square brackets):

(12)   a. cards with exactly 2 {smiley faces/stars}   [1]
           b. cards with more than 2 {smiley faces/stars}   [.88]
           c. cards with at least 2 {smiley faces/stars}   [.54]
           d. cards with at most 2 {smiley faces/stars}   [.50]

(Unfortunately, Musolino's materials didn't include items with "fewer than".) So, in this study, children had no problems with "exactly 2" and "more than 2", while with superlative quantifiers they performed at chance level, thus confirming the predictions made by G&N's analysis.

The aim of our second study was to see whether adults, too, find superlative quantifiers harder to process than comparative ones. Since offline tasks are ill-suited for this purpose, an online design was used.

Experiment 2
In this experiment, participants had to decide whether a given scene was correctly described by a sentence containing either a superlative or a comparative quantifier. The scenes simply consisted of letters; the sentences used were:

(13)   a. There are more than 2 A's.
           b. There are at least 3 A's.
           c. There are fewer than 3 A's.
           d. There are at most 2 A's.

The experiment was presented on a notebook computer. On each trial, the target sentence was displayed first. Participants were instructed to press the space bar once they had read and understood the sentence. The sentence would then be replaced with a scene, and participants had to decide whether the sentence was true or false of the scene.

Reading times as well as decision times were recorded. The former turned out not to differentiate between conditions, which is to say that we found no evidence that sentences with "at least/most" take longer to read than their counterparts with "more/fewer than". However, we also found that, in the decision stage, the "at least/most" conditions were slower than the

"more/fewer than" conditions. This is in line with the prediction that the meanings associated with the former are more complex.

References

Geurts, B. (2006). Take "five": the meaning and use of a number word. In S. Vogeleer and L. Tasmowski (Eds.), *Non-definiteness and plurality*, pp. 311–329. Amsterdam/Philadelphia: John Benjamins.

Geurts, B. and R. Nouwen (2007). "At least" et al.: the semantics of scalar modifiers. Forthcoming in *Language*.

Musolino, J. (2004). The semantics and acquisition of number words: integrating linguistic and developmental perspectives. *Cognition* 93: 1–41.

Noveck, I. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition* 78: 165–188.