Readers need more than (just) world knowledge to process subset comparatives

Margaret Grant
University of Massachusetts Amherst
Contact: meg@linguist.umass.edu

This research explores the use of conceptual/world knowledge and syntax in processing subset comparatives. Subset comparatives have the form More NP$_1$ than NP$_2$..., where NP$_2$ is a subset of the set described by NP$_1$. In (1a), eagles are a subset of birds.

1. a. More birds than eagles flew over the conservation area.
   b. More birds than airplanes flew over the conservation area.

To distinguish subset comparatives from contrastive comparatives like (1b), comprehenders must use their knowledge of the world or the context (e.g., 2). Syntactic information can also signal a subset comparative. For example, bare plural NP complements to than can be either contrastive or subset comparatives, but full-DP complements to than can only be subsets (compare 3a to 3b).

2. Yankee stadium poses more problems than just empty seats.
   (Brenden Monroe, www.bleacherreport.com, 23 April 2009)

3. a. More birds than an eagle flew over the conservation area.
   b. #More birds than an airplane flew over the conservation area.

Further, subset comparatives often contain just in the than-complement, while contrastive comparatives cannot (4). We investigated these signals of subset comparatives in two studies of eye movements during reading.

4. #More birds than just airplanes/an airplane flew over the conservation area.

In Experiment 1 (N = 36), subjects read items like (5), which were either subset or contrastive comparatives with bare plural or singular indefinite (full-DP) complements to than. We predicted that bare plural subset comparatives (5a) would initially be processed as contrastive, with readers realizing late in processing that a subset interpretation is necessary. Singular indefinite complements to than cannot be contrastive comparatives; therefore in (5c) the subset relationship could be exploited earlier in processing. On Region 4 (complement to than), we found longer Go-Past Times (p$\text{MCMC} < .001$) and more Regressions Out (p < .001) for full-DPs (see (6) for means). While Region 4 effects could be due to length, this penalty persisted on Region 5. Consistent with our prediction, a penalty for bare plural subset comparatives emerged in Regressions Out of Region 6, where this condition had the most regressions (interaction $p = .02$).

5. Experiment 1 Materials (| indicates analysis regions)
   a. It seems that| more flowers$_1$ than| tulips$_4$ grew well$_5$ in the small greenhouse$_6$|
   b. It seems that| more violets$_2$ than| tulips$_4$ grew well$_5$ in the small greenhouse$_6$|
   c. It seems that| more flowers$_2$ than| a tulip$_4$ grew well$_5$ in the small greenhouse$_6$|
   d. It seems that| more violets$_2$ than| a tulip$_4$ grew well$_5$ in the small greenhouse$_6$|

6. Mean Go-Past Times and Regression rates for critical regions, Experiment 1.
In Experiment 2 (N = 24), subjects read items like (7), which manipulated subset vs. contrastive comparatives and the presence/absence of *just*. Go-Past Times on Region 5 showed a penalty for subsets ($p_{MCMC} < .01$) and a trend toward an interaction wherein subset comparatives without *just* were longest (see 8). Regressions Out showed a similar pattern, with effects of subset ($p < .05$), *just* ($p < .05$), and an interaction ($p < .01$). These immediate effects of the subset comparative may be due to the presence of *just* in the experiment leading readers to become aware of subset interpretations. We included trial order in our statistical models, and found interactions of order and *just* such that reading times on Regions 5-6 for conditions with *just* decreased over the course of the experiment. We believe that *just* focused readers’ attention on the following NP, allowing them to interpret the NP as a subset, or dismiss it as anomalous.

7. Experiment 2 Materials ([ ] indicates analysis regions)

8. Mean reading times and regression rates for critical regions, Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>Subset</th>
<th>No Subset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bare Plural</td>
<td>Sg. Indefinite</td>
</tr>
<tr>
<td>Go-Past Time (ms)</td>
<td>Region 4: 344</td>
<td>480</td>
</tr>
<tr>
<td></td>
<td>Region 5: 366</td>
<td>459</td>
</tr>
<tr>
<td></td>
<td>Region 6: 1495</td>
<td>1402</td>
</tr>
<tr>
<td>Regressions Out (%)</td>
<td>Region 4: 18</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Region 6: 38</td>
<td>24</td>
</tr>
</tbody>
</table>

Together, these experiments show that the timecourse of interpretation of subset comparatives depends on the presence and type of syntactic indicator. Although norming data showed that our NP2 s were good members of the NP1 sets (rating 4.81/5 for subsets vs. 2.3 for contrastives), subset information does not appear to guide processing immediately unless there is an overt indication of its importance.