

Parallelism and ellipsis in comparatives

Katy Carlson, Morehead State University

Comparative constructions as in (1) have many possible syntactic continuations, including bare NP ellipsis (a), VP Ellipsis (b), and full clauses (c).

- (1) Julianna brought her corgi to class more often than...
- a. her dachshund/Maria.
 - b. Maria did.
 - c. Maria brought her poodle.

This project explores their use and properties by examining the frequency of different comparative structures within the Corpus of Contemporary American English (COCA, Davies 2008-present). The corpus data shows that ellipsis structures are much more frequent in comparative contexts than full clauses, with bare NP ellipsis (or stripping) being the most frequent structure by far. Interestingly, bare NPs contrast with the subject of the previous clause in 80% of corpus examples, but these NPs preferentially contrast with the object in processing of ambiguous sentences. In addition, although experimental studies might concentrate on ambiguous bare NP ellipsis, a large percentage of the ellipsis examples were disambiguated by syntactic structure, plausibility, and/or a variety of types of parallelism between contrasted NPs. We suggest a strong role for both focus structure and parallelism in the observed frequencies.

To form the comparative corpus of interest, instances of “more Adverb than” comparatives were extracted from COCA for the top 23 adverbs in 2015; errors and idioms were excluded. The final dataset included 4400 examples from 23 adverbs (e.g., *often*, *quickly*). We hand-coded the syntactic category after *than*, the sentence role of NPs (i.e., the position of contrast in the complete first clause), and similarities between NPs in form and semantic domain.

First, we analyzed the syntax of the continuation type. Bare NP ellipsis made up 50% of the data, and was most common in academic and magazine writing. The second highest category was VP Ellipsis (18%, including the interesting sub-category of inverted VP ellipsis, as in *...than did Maria*). Overall, less than 10% of the continuations were full clauses. On the basis of these patterns, we suggest that the repetition involved in complete clauses, and the deaccenting that would be needed in the prosodic representation, makes clauses dispreferred in comparatives compared to structures that retain mostly contrastive information.

Then we analyzed the grammatical role of the NP continuations. Of the 2260 examples with bare NP ellipsis, over 80% of the NP remnants contrasted with the subject of the previous clause and less than 15% with the object. A small percentage of the examples were ambiguous between subject and object contrasts. Overall, 3490 examples included an NP immediately following *than*, and of these, almost 90% of those NPs were subjects (this was unambiguous for VP ellipsis and full clauses but potentially ambiguous for bare NPs).

If frequency in the input were important to the processing preferences of comprehenders, these distributions suggest that processors encountering a comparative construction should expect a bare NP, and upon encountering an NP at the start of the continuation, they should strongly expect it to be a subject, if frequency in the input is important to their preferences. However, prior research with ambiguous bare NP comparative ellipsis (Carlson 2002) has found a bias toward object contrast: comprehenders are more likely to interpret bare NPs to be objects, contrastive with the object in the antecedent. Carlson (2002) further found that this object

contrast bias could be strengthened and weakened, through the manipulation of parallelism and accent placement. Taken together, we conclude that the bias in the processing of comparative ellipsis sentences is driven by the expected focus position on the object in the antecedent, rather than generalizations based on input frequency.

In order to substantiate the author's categorization of bare NPs as subject or object contrasts, two norming studies were carried out. In one, three undergraduate researchers were trained to indicate which earlier NP contrasted with the bare NP and then given a set of 556 ambiguous examples (i.e., there were between two and four first-clause NPs to contrast with in each). They included 30% of examples categorized as object contrast and the rest subject contrasts. Agreement with the original categorization by the author was 97%. Secondly, a naïve set of participants (N=41) were given a questionnaire study of over 400 bare NP examples from the corpus, with 27% object examples. Again, they were asked to choose which of two or three first-clause NPs contrasted with the bare NP remnant. The choice of contrasts by participants was 94% the same as the author's. Accuracy was slightly lower for object contrasts, which may be due to the majority of subject contrasts. These two norming studies support the claim that the majority of sentences were correctly categorized using the cues available.

These cues to NP role/contrast included animacy, plausibility, and a range of properties that could be called parallelism: the form of the NPs (pronouns, names, definite or indefinite descriptions) and various formal or meaning-related similarities between the NPs. At least one type of parallelism appeared in 80% of bare NP ellipsis examples from the corpus with subject contrast, and 90% of those with object contrast. For example, a notable subset of the items (10%) had antonyms for the bare NP and its contrast within the first clause; around 30% used the same head noun with different modifiers; and around 30% used what might be called contrast markers or paired determiners (*these/those, this/that, some/others, any, else*) or pronouns. Many examples used multiple types of parallelism to clarify the intended contrast with an earlier NP.

In common with other corpus studies of ellipsis sentences (e.g., Bos & Spenser 2011, Miller 2014), this work has found that structures which we might expect to be common and standard are not always the most frequent. For comparatives of this particular type, unelided sentences were in a small minority, though clearly unelided clauses are generally more common than elided clauses. The observed frequency of bare NP ellipsis expressing subject contrast does not match with observed biases in processing. Further, although bare NP ellipsis sentences are in principle ambiguous, and thus might be dispreferred to VP ellipsis options whose auxiliary verb clarifies the subject position of an NP, in practice they were usually disambiguated through parallelism if not already disambiguated by plausibility.

Bos, J., & Spenser, J. (2011). An annotated corpus for the analysis of VP ellipsis. *Language Resources & Evaluation* 45, 463-494.

Carlson, K. (2002). *Parallelism and prosody in the processing of ellipsis sentences* (Outstanding dissertations in linguistics series). New York: Routledge.

Davies, M. (2008-). The Corpus of Contemporary American English: 520 million words, 1990-present. Retrieved from <http://corpus.byu.edu/coca/> in 2015.

Miller, P. (2014). A corpus study of pseudogapping and its theoretical consequences. In C. Pinon (ed.), *Empirical Issues in Syntax and Semantics* 10, 73-90.