How a linguistic theory ensues
from systematic grammaticality ratings by native speakers

We report our experience with six rounds of determining CCE rules applying field studies. CCE/Clausal Coordinate Ellipsis covers the following types of systematic elisions in two and-coordinated clauses:

1. **Gapping:** *John eats apples and his sons eat cherries*

1a. **LDG/Long-Distance Gapping:**

   *John wants to clean his car and Ann wants to clean her bike*

1b. **Subgapping:** *John was killed and his sons were wounded*

1c. **Stripping:** *John sleeps and Susan sleeps too*

2. **FCR/Forward-Conjunction Reduction:** *Sue opened the door and Sue went out*

3. **BCR/Backward-Conjunction Reduction:**

   *Ben came before one o’clock and Anne left after one o’clock*

4. **SGF/Subject Gap with Finite/Fronted Verb:**

   *Into the forest went the hunter and the hunter shot a hare.*

In grammar books or second-language teaching-material, rules for the composition of elliptical structures play a secondary role at best. Therefore, we propose an easy to handle field-study design to attain CCE rules in a new target language.

As starting point, the psycholinguistically motivated CCE theory by Kempen (2009) was implemented in a language-independent manner to generate all elisions for an unreduced syntactic input structure (cf. crossed-out words in examples (1)-(4)). A parallel treebank (~100 examples per language) yields translations of sentences that systematically vary conditions to apply/block CCE. For instance, example (4) without subject-verb inversion licenses FCR instead of SGF. Fronting the object in the second conjunct blocks CCE (reductions like pronominalizations (*a hare, *s/he shot*) are not at issue here). Currently, the parallel treebank comprises Dutch/German/Estonian/Hungarian/Polish/Russian. The translations solely preserve the structure—cf. Duczmal (1985) mentioning a similar strategy to make translations sound natural in the target language allowing for high-quality grammaticality ratings.

In preparation of the experiments, the treebank becomes translated into the new target language. The main idea behind the field-study design is based on the fact that the treebank examples account for fine-grained aspects of CCE-rule application. Thus, informants itemize unconscious rules of a linguistic theory by means of judging/producing individual CCE examples in their mother tongue. Two experiments systematically examine the hypothetical rules:

**Experiment 1/Overgeneration**

Native speakers rate all automatically produced elisions for the sentences of the treebank on a three-valued scale: grammatical+meaning identical with the unreduced sentence/dubious/ungrammatical, or not meaning preserving. The votes demarcate valid CCE rules from those to be modified—e.g. Dutch/German SGF does not license sentence-initial objects.

**Experiment 2/Undergeneration**

For each unreduced sentence in the treebank, another group of informants produces all paraphrases they can think of. No linguistic restrictions are stipulated—not to overtax the test subjects. Before evaluating these results, all non-elliptical paraphrases (e.g. pronominalization) are excluded from the list to identify cases not covered by the hypothetical rules—e.g. Polish BCR phenomenon absent from grammars.

According to our experience, just five informants per experiment suffice to adjust the CCE rules. Additionally, we report promising results with online experiments.

**Literature**
