ELLEIPO: Computing elliptical clausal coordinate structures for use in first- and second-language teaching

Karin Harbusch¹ & Denis Krusko¹ & Gerard Kempen²

¹ Computer Science Dep., Univ. of Koblenz-Landau, {harbuschkruskod}@uni-koblenz.de
² MPI Nijmegen & Cognitive Psychology Unit, Leiden University, Gerard.Kempen@mpi.nl

Few course materials for first-and second-language instruction teach adequate rules for the composition of elliptical structures (cf. Xing, 2006: 179). Hence, advanced learners have to search for such rules on their own—in grammar books or on the internet. In this Abstract, we present an e-learning system, called ELLEIPO, that can teach readily understandable, linguistically motivated syntactic rules for one elliptical domain: CLAUSAL COORDINATE ELIPSIS (CCE). We distinguish four types of CCE structures:

- GAPPPING, with three variants called LONG DISTANCE GAPPPING (LDG), SUBGAPPPING, and STRIPPING (subscript g used in the examples below),
- FORWARD CONJUNCTION REDUCTION (FCR; subscript f),
- BACKWARD CONJUNCTION REDUCTION (BCR; b; also known as RIGHT NODE RAISING), and
- SUBJECT GAP WITH FINITE/FRON TED VERB (SGF; s).

The rules we use are adapted from a psycho-linguistically motivated CCE system (Kempen, 2009). ELLEIPO takes parsed non-elliptical (unreduced) sentences as input and can generate all possible elliptical coordinate structures allowed by the rules. The original version took Dutch and German as target languages (Harbusch & Kempen, 2006). The accuracy of the rules has been tested with CCE structures from a written and a spoken treebank for German. Accuracy was 99% for a written corpus of about 50,000 sentences (Harbusch & Kempen, 2007), and 97% for about 18,000 spoken sentences (Harbusch & Kempen, 2009). The same rules, with only minor amendments, turned out to work for Estonian (Harbusch et al., 2009), Hungarian (Harbusch & Bátori, 2013), and Russian (Harbusch & Krusko, 2015).

Here we describe how ELLEIPO can be deployed in e-learning systems for CCE in L1 and L2. The prototype we implemented consists of two successive teaching phases (of which the first one is optional): (1) instruction and training on how to recognize various types of identity of—usually major—constituents shared by unreduced conjoined clauses; and (2) training exercises in which the student applies the rules to sentences presented by the system.

In the first phase, the student learns to recognize syntactic, morphological and referential identity of constituent pairs as necessary precondition to trigger any CCE form:

(I1) **Lemma identity** of major constituents in the conjoined clauses. Only the lemmas need to be identical; morphological properties such as Number/Person of a Verb or Case of a Noun may differ (e.g., the three verbforms go, goes and went are lemma-identical).

(I2) **Grammatical-function identity** of major constituents of the conjuncts (e.g., SGF requires identity of the grammatical Subject).

(I3) **Wordform identity** requires identity of all morphological features (as is the case in FCR).

In the first teaching phase, the individual identity types are taught by visualizing them in syntactic trees (see Figure 1, a case licensing FCR of **dass Hans** in the second conjunct. In order to remain as theory-neutral as possible, we display syntactic structures in a format resembling the TIGER treebank format; Brants et al., 2004). Initially, the learners are presented with trees and attempt to highlight pairs of constituents that are identical according to the above definitions. Elleipo evaluates the responses and provides feedback. When the student has reached a sufficiently high score, the unreduced sentences are presented without trees.

In the second teaching phase, the individual CCE-construction rules are explained with visual support in the form of specially annotated syntactic trees. For reasons of space we only address FCR and BCR rules here. FCR is licensed if both clausal conjuncts start with a wordform- and grammatical function identical sequence of full constituents. BCR is licensed by
Susi heard that Hans had an accident and could die.

Figure 1. German example showing all wordform-identity tags (node subscripts). The CCE application domain is highlighted in blue; red squares show corresponding constituent pairs in the conjoined clauses.

Lemma identity in the right-periphery but, unlike FCR, optionally disregards major constituent boundaries (cf. Mary came before three o’clock, and Mary left after five o’clock vs. *The man came before three o’clock, and the woman left after five o’clock).

The system presents exercises of all CCE types consecutively. Elleipo’s rule system takes care of feedback and error correction. Elleipo can deal with multiple combinations of CCE types in the same pair of coordinated clauses and yields all possible reductions as output. For instance, it generates two CCE structures for the German example in (1): the first as a combination of Gapping and BCR (subscripts g-1 and b-2), the second as Gapping at two clause levels (subscripts g-1 and gg-2).

(1) Monopole sollen_1 geknackt werden_2 und Märke sollen_1 getrennt werden_2
Monopole sollen_1 geknackt werden_2 und Märke sollen_1 getrennt werden_2

‘Monopolies should be shattered and markets split’

In the conference we give a demonstration of our system (which has been implemented in JAVA). In future work, Elleipo will have to be attached to an existing parser that produces parse trees in Elleipo’s input format. This will enable students to test their own sentences.

Literature
Harbusch, K. & Krusko, D. 2015. Distinctive similarity of clausal coordinate ellipsis in Russian compared to Dutch, Estonian, German, and Hungarian. 5th BSNLP, Hissar, Bulgaria.
Xing, J.Z. 2006. Teaching and Learning Chinese as a Foreign Language: A Pedagogical Grammar, Hong Kong Univ. Press, Hong Kong.