

Clausal Coordinative Ellipsis in Hungarian in comparison to Dutch, Estonian and German

Karin Harbusch & István Bátori

Universität Koblenz-Landau, Computer Science Dept.

{harbusch|batori}@uni-koblenz.de

In the following, we prove that the psycholinguistically motivated theoretical framework of clausal coordinate ellipsis generation by Kempen (2009) and its implementation (called ELLEIPO which was originally programmed in JAVA for Dutch and German (Harbusch & Kempen, 2006) and which has been extended for Estonian (Harbusch et al., 2009)) can be easily tailored to Hungarian.¹ Given the fact that the investigated languages belong to two rather different families of languages (Estonian and Hungarian are Finno-Ugric languages whereas Dutch and German belong to the Indo-European languages), not much overlap for the individual elision rules for the two language families would be expected. However, the similarities for Dutch, Estonian and German CCE generation are remarkable. As linguists might object that Estonian is heavily influenced by German, verifying the rules for Hungarian as well underpins the claim that the rules can be generalized for Finno-Ugric languages.

Table 1 delineates the four CLAUSAL COORDINATE ELLIPSIS (CCE) types (GAPPING (including LONG-DISTANCE GAPPING (LDG), SUBGAPPING and STRIPPING), FORWARD and BACKWARD CONJUNCTION REDUCTION (FCR and BCR) and SUBJECT GAP WITH FINITE/FROTTED VERB (SGF); cf. first column) and illustrates them by Hungarian examples. Table 2 outlines the rule set licensing the individual constructions in Dutch/Estonian/German.

All forms of GAPPING (cf. examples (1) to (4)) are characterized by elision of the posterior member of a pair of LEMMA-IDENTICAL² Verbs (N.B. that word order is not essential as in sentence (1); the position of the elided Verb need not be peripheral but is often medial, as in (2) through (4)). Every non-elided constituent (“REMNANT”) in the posterior conjunct should pair up with a constituent in the anterior conjunct that has the same grammatical function but is not coreferential. Stated differently, the members of such a pair are CONTRASTIVE. In LDG, the remnants originate from different clauses (more precisely: from different clauses that belong to the same SUPERCLAUSE; a superclause is a hierarchy of finite or nonfinite clauses that—with the possible exception of the topmost clause—do not include a Subordinating Conjunction). In FCR, elision affects the posterior token of a pair of LEFT-PERIPHERAL strings consisting of one or more wordform-identical major constituents³. BCR is almost the mirror image of FCR as it deletes the anterior member of a pair of RIGHT-PERIPHERAL lemma-identical word strings (‘to-steal’ in (7)); however, BCR may elide PART OF a major constituent. SGF elides the Subject of the posterior conjunct in a main clause, when in the anterior conjunct the wordform-identical Subject follows the Finite Verb (Subject-Verb inversion).

In our presentation, we show that Hungarian obeys these rules with few deviations. For instance, example (8) is unacceptable in Dutch; in Hungarian Gapping, superclauses are not relevant (cf. sentence 9⁴—which is ruled out in Dutch, Estonian and German).

(9) *A rendőrség reméli hogy a tuntutők haza mennek és*
The police hopes that the demonstrators home go and

a tuntutők remélik_g hogy a rendőrség [haza-mennek]_{gg}
the demonstrators that the police

‘The police hopes that the demonstrators go home and *vice versa*’

Table 1. CCE examples in Hungarian. Struck-out text represents elisions.

¹ In the talk, we give a live demo of the ELLEIPO system (see www.uni-koblenz.de/~harbusch/Welcome-to-Elleipo.html).

² For LEMMA IDENTITY, only the lexical entries (citation form) of the constituents have to be identical. In contrast, WORDFORM IDENTITY requires, in addition, identity of their morphological features. COREFERENTIAL CONSTITUENTS refer to the same discourse entity or entities, irrespective of whether or not they include the same lemma(s).

³ We use the term MAJOR CONSTITUENT of a clause in a broad sense that includes Head Verb (Main, Copula or Auxiliary), Arguments (e.g. Subject, Direct and Indirect Object, and Non-finite Complement Clause), Adjuncts (Adverbial Modifier, including Adverbial Clause), and Subordinating Conjunctions (i.e. the Complementizer in Complement Clauses—*that*, *whether*—or the Subordinator in Adverbial Clauses—*while*, *although*, *when*, etc).

⁴ N.B. the example works the same when the second Subject is replaced by ‘the organizers’. Thus, the elision does not result from strong semantic expectations overruling detailed syntactic checking.

GAP-PING (g)	(1) <i>Feri Budán lakik és Pesten laknak_g a fiai</i> Feri in-Buda lives and in-Pest live his-sons 'Feri lives in Buda and his sons live _g in Pest'
LDG ((g) ⁺ g)	(2) <i>Ma akarja Jansci az autóját mosni és [ma akarja]_g Zsuszi a kerékpárját [mosni]_{gg}</i> Today should Jansci the car clean and today should Zsuszi the bike 'Today, Jansci should clean the car and today _g Zsuszi should _g [clean] _{gg} the bike'
SUBGAP-PING (sg)	(3) <i>A menekülő tábornokot egy szabómester felismerte és [...] a tömeg felkoncolta</i> The escaping general _{ACC} has master-tailor recognized and the crowd lynched 'A master-tailor recognized and the crowd lynched the escaping general'
STRIP-PING (str)	(4) <i>Péter aludni tud hajón és Jansci [aludni tud hajón]_{str}</i> Péter sleep can in-the-bus and Jansci 'Péter can sleep in the bus and Jansci [can sleep in the bus] _{str} too'
FCR (f)	(5) <i>Az almát szeretem és [az almát]_f gyakran eszem</i> The apples I-like and the apples often I-eat 'The apples I like and [the apples] _f I often eat' (6) <i>Zsuszi hallotta hogy Béla gyakran biciklit lop és</i> Zsuszi hears that Béla often bikes steals and <i>[hogy Béla gyakran biciklit]_f ad el</i> (N.B. unreduced: <i>elad</i>) sells 'Zsuszi hears that Béla often bikes steals and [that Béla often bikes] _f sells'
BCR (b)	(7) <i>Jansci megpróbált autót [lopni]_b és Péter megpróbált_g biciklit lopni</i> Jansci tries cars and Peter bikes to-steal 'Jansci tries to steal cars and Peter tries _g [to steal] _{gg} bikes' (N.B. LDG is also possible here)
SGF (s)	(8) <i>A levest megeszi Jansci és Jansci_s lefekszik</i> The soup eats Jansci and Jansci lies-down 'Jansci eats the soup and lies down'

Table 2. Basic elision conditions for the four clausal coordinate ellipsis (CCE) types in Dutch, Estonian, German and Hungarian.

CCE type	Elision conditions
GAPPING	Lemma identity of Verb & contrastiveness of remnants
LONG-DISTANCE GAPPING (LDG)	Gapping conditions in <i>superclause</i>
SUBGAPPING	Gapping conditions & VP remnant in second conjunct
STRIPPING	Gapping conditions & Only one non-Verb remnant
FORWARD CONJUNCTION REDUCTION (FCR)	Form identity & left-peripherality (within clause boundaries) of major clausal constituents
BACKWARD CONJUNCTION REDUCTION (BCR)	Lemma identity & right-peripherality, possibly disregarding major constituent boundaries
SUBJECT GAP IN CLAUSES WITH FINITE/FRONTED VERBS (SGF)	Form-identical Subject & first conjunct starting with Verb/Modifier/Adjunct & FCR applied if licensed

References

- Harbusch, K. & Kempen, G. 2006. Elleipo: A module that computes coordinative ellipsis for language generators that don't. *Procs. of 11th EACL*, Trento.
- Harbusch, K., Koit, M. & Õim, H. 2009. A comparison of clausal coordinate ellipsis in Estonian and German. *Procs. of 12th EACL*, Athens.
- Kempen, G. 2009. Clausal coordination and coordinate ellipsis in a model of the speaker. *Linguistics*, 4.