Incremental sentence production inhibits clausal coordinate ellipsis: A comparison of spoken and written language

Karin Harbusch¹ & Gerard Kempen²,³

¹Department of Computer Science, University of Koblenz-Landau
²Max Planck Institute for Psycholinguistics, Nijmegen
³Cognitive Psychology Unit, Leiden University

One of the benefits of incremental sentence production is to reduce the working memory capacity needed for advance planning: The planning units can be of considerably smaller size (measured in terms of word length) than in case of non-incremental production. The same advantage has been claimed for the various forms of ellipsis, which preempt the need to plan the detailed shape of one or more constituents and thereby reduce the size of planning units. Because working memory load tends to be higher in spoken than in written language, one expects that speakers, in comparison with writers, will more frequently resort both to incremental production and to the use of elliptical constructions. As far as we know, this prediction is generally borne out for incremental production. However, in two corpus studies into the incidence of clausal coordinate ellipsis in spoken and written English, Meyer (1995) and Greenbaum & Nelson (1999) obtained a data pattern opposite to the prediction. In written clausal coordinations, the proportion of elliptical versions is about twice as high as in spoken coordinations.

Recent treebanks with spoken and written sentences enable us to verify and extend the latter finding. In the CGN treebank (Corpus of Spoken Dutch), the percentage of elliptical versions of clausal coordinations is even ten times lower than in the TIGER treebank with German written texts. Although we realize that the languages of these treebanks are different (albeit closely cognate), and probably the text genres as well, we believe that these new data count as verification of the pattern observed earlier in English corpora.

After an account of our data extraction method, we will present a detailed overview of the incidence of four types of clausal coordinate ellipsis—including Gapping, Forward Conjunction Reduction (FCR), and Right Node Raising (RNR)—in the spoken and written treebanks. Finally, we propose a theoretical explanation of the data pattern based on the assumption that during spontaneous speaking the scope (“window”) of online grammatical planning is basically restricted to one (finite) clause. In producing clausal coordinations, checking the possibility of “forward” ellipsis (Gapping, Forward Conjunction Reduction) requires comparison of form and meaning of two adjacent clauses. As this overtaxes the online planning scope of the sentence production system, speakers prefer to plan the form of second or later conjoined clauses in isolation, that is, without taking the shape of preceding clauses into account and thereby eliminating elliptical options. RNR, the “backward” versions of coordinate ellipsis, is more severely affected in spoken language because it requires the simultaneous presence within the planning window of (nearly) two complete clauses. Indeed, whilst RNR is readily observable in written texts, in spoken language it is a rare phenomenon manifesting itself only in very short clauses.