

Conjunction reduction and gapping in clause-level coordination: an inheritance-based approach

GERARD KEMPEN

Nijmegen Institute for Cognition and Information, University of Nijmegen, Nijmegen, The Netherlands

Received August 12, 1991

Revision accepted September 24, 1991

Comput. Intell. 7, 357-360 (1991)

Three important contraction phenomena in clause-level coordination are backward conjunction reduction (BCR), forward conjunction reduction (FCR), and gapping. They are exemplified by sentences (1)–(3) respectively.

- (1) John stole ... and Peter bought *a bike*
- (2) *John* stole a bike and ... sold it immediately
- (3) John *stole* a bike and Peter ... a car

By substituting the italicized words for the ellipses, one obtains the non-contracted counterparts. It is generally recognized that these phenomena obey entirely different laws (e.g., see the literature surveys by Van Oirsouw (1987), Goodall (1987), and Pijls and Kempen (1986)), but the empirical evidence around them is so complex that no satisfactory linguistic analyses, let alone syntactic generation algorithms, have materialized.

1. Coordination and self-correction

In order to fill this gap I propose a novel approach which was inspired by a psycholinguistic observation originally due to Levelt (1983, p. 78). In an extensive study of speech errors, he registered retracing targets in so-called retracing repairs. The three elements of such corrections are reparandum (i.e., the original utterance containing an error), editing term (“... uh ...”, “... I mean ...”, “... or rather ...”), and repair text (cf. (4)). The assumption is that the speaker interrupts the original utterance, signals this to the listener by means of a pause and (or) an editing term, backtracks to an earlier point in the utterance, and reformulates it from there. Levelt observed that only certain positions in the reparandum qualify as potential retracing targets. He proposed a rule for demarcating these positions, in which the three elements of retracing repairs are likened to the left-hand conjunct, the conjunction, and the right-hand conjunct of a coordination, as shown in (4) and (5). The retracing target in (4) is the position just before the verb. The fragment preceding this point (i.e., the subject NP) is incorporated into the final utterance and, spliced together with the repair (VP), forms a correct sentence. (The incorporated fragment is not uttered a second time.)

- | | | |
|----------------------------|---------------------|----------------------------|
| (4) <u>John bought ...</u> | uh | <u>... stole a bike</u> |
| <i>reparandum</i> | <i>editing term</i> | <i>repair</i> |
| and | | |
| (5) <u>John bought ...</u> | and | <u>... stole a bike</u> |
| <i>left-hand conjunct</i> | <i>conjunction</i> | <i>right-hand conjunct</i> |

Levelt’s rule — I need not go into details here — places the position marked by > as potential retracing target in sentence (6) because the corresponding coordination (7) is ungrammatical. Examples (8) and (9) show that the retracing target may be located at sentence onset.

- (6) This > man bought ... uh ... boy stole a bike
- (7) *This man bought ... and ... boy stole a bike
- (8) > This man bought ... uh ... this boy stole a bike
- (9) > This man bought ... and this boy stole a bike
- (10) John stole two ... and ... bought three *books in that shop*

This similarity between retracing repairs on one hand and conjunction reduction on the other suggests analogous treatments of both. Backward conjunction reduction (the first ellipsis in (5) and (10)) corresponds to a retracing repair where part of the reparandum is withheld and postponed until it surfaces as the final fragment of the repair text. And forward conjunction reduction (the second ellipsis in (5) and (10)) can be construed as incorporation of material which precedes the retracing target, into the reformulated clause.

A second type of self-corrections (the so-called non-retracing repairs; cf. Van Wijk and Kempen (1987)) involves the reformulation of one or a few major constituents only. Examples (11) and (12) show that they have their coordination counterparts as well, namely in gapping constructions.

- (11a) I saw his son yesterday, you know, the one with the blond hair
- (11b) I *saw* his eldest son *yesterday*, and ... their youngest one ... (too)
- (12a) One car crashed into the other, I mean, the Jaguar into the Porsche
- (12b) A BMW *crashed* into a Volvo yesterday, and a Jaguar ... into a Porsche ...

The tripartite correspondence emerging between self-correction and coordination is summarized in Table 1. In the following sections, I hope to demonstrate that grafting the syntax of coordination onto the process of self-repair opens up an interesting perspective on the study of contraction phenomena.

2. Inheritance relationships between clauses

How can the relationship between reparandum and repair, and the analogous one between left-hand and right-hand members of a coordination, be characterized? In particular, what exactly is meant by *incorporation* of fragments of one clause into the next? Linguistic students of coordination have sought an answer in two directions: “deletion under identity” of constituents shared by the clauses, and direct generation of pairs of full and elliptic clauses via special rules (phrase-structure, categorial, or otherwise) without deleting any elements. Neither route has proven very successful (see the reviews by Goodall (1987) and Van Oirsouw (1987)).

TABLE 1. Contraction phenomena in self-correction and coordination: correspondences

Coordination	Self-correction
Backward conjunction reduction	Post-interrupt fragment of reparandum is delayed and surfaces as final fragment of repair
Forward conjunction reduction	Reparandum fragment preceding retracing target is incorporated into revised clause
Gapping	Small number of major constituents of the reparandum is revised; unchanged parts of the original clause are incorporated into revised clause

The affinity between self-repairs and coordination suggests a third route. A speaker who is engaged in correcting an already uttered clause may be compared to the editor of a manuscript who, instead of copying the old manuscript in full, meanwhile replacing inadequate passages by new text, simply sets up pointers to the correct parts of the original manuscript and only writes out the new text fragments. The revised manuscript thus becomes a mixture of new passages and pointers to original passages. In computer science parlance, the revised manuscript is a *new instantiation* of the original manuscript from which it *inherits* all properties that are not specified locally, that is, within the new instantiation. This concept of *default inheritance* is what I meant by "incorporation" in the previous section. The original and the revised clause are seen as objects, the major constituents as properties, and in the inheritance relationship between these clausal objects the revised clause is inheritor ("client") which "delegates" certain properties to the original clause object ("proxy").

Transferred to the coordination domain, the inheritance model designates the left-hand and right-hand conjuncts as proxy and client respectively. For instance, the right-hand conjunct of (12*b*), a case of gapping, is a clausal object with two constituents inherited from the left-hand clause (finite verb and modifier) and two locally specified constituents (subject NP and modifier PP) which overwrite their left-hand counterparts.

A simple informal notation can make the inheritance relationship explicit. In (13) and (14) I have written the successive members of a coordination beneath one another on different lines. A pair of brackets surrounding an ellipsis denotes inheritance of the corresponding constituent from the previous line. Symbols following opening brackets designate grammatical categories or functions of the inherited constituents.

(13) A BMW crashed into a Volvo yesterday
and a Jaguar [VFin ...] into a Porsche [Mod ...]

(14) John > stole two ϕ
and [Subj ...] bought three *books in that shop*

A somewhat extended form of this notation also encompasses FCR and BCR; e.g., (10) can be rendered as (14). The symbols > and ϕ denote, respectively, retracing target and post-interrupt fragment. The analysis shows that the right-hand conjunct inherits the subject NP. As for terminology, I will refer to the two inheritance patterns underlying gapping and FCR by the terms *whole-clause inheritance* (WCI) and *clause-initial inheritance* (CII) respectively.¹ Instead of "retracing target" (>), I will often use the term

¹In the remainder of this paper I will not discuss BCR anymore. Needless to say that it is not an inheritance phenomenon: parents don't inherit from their children! See also Pijls and Kempen (1986).

juncture because it marks the splice where the beginning of the left-hand conjunct is joined to the right-hand conjunct.

Before entering the empirical arena I wish to point out that the remainder of this paper will be based on evidence from Dutch rather than English. Dutch, my native language, features considerable word order variation, in particular between main clauses (SVO, VSO) and subordinate clauses (SOV). This property helps to rule out incorrect hypotheses. Although the analyses put forward below apply to Dutch only, they offer a vantage point for the study of coordination and for the generation of coordinate structures in cognate languages such as German and English.

3. Inheritance in clauses with one verb

It is a well-known observation that, at least in single-verb clauses, the choice between gapping and FCR hinges on the presence or absence of a verb on the right-hand side of the coordination. Right-hand clauses which contain an explicit (overt) verb follow the FCR pattern (CII); without an overt verb only gapping is permitted (WCI). This generalization holds for finite main and subordinate clauses (see (15*a-b*) and (16*a-c*); non-finite clauses will be discussed in Sect. 4).

(15*a*) Ik > koop vanmiddag een CD en [Subj ...]
beluister hem vanavond
I buy this-afternoon a CD and
listen-to it tonight

(15*b*) (als) ik > vanmiddag een CD koop en
[Subj ...] hem vanavond beluister
(if) I this-afternoon a CD buy and
it tonight listen-to

(15*c*) *(als) > ik vanmiddag een CD koop en jij
[Obj ...] vanavond beluistert
(if) I this-afternoon a CD buy and you
[Obj ...] tonight listen-to

(15*d*) *(als) ik > vanmiddag een CD koop en [Subj ...]
[Obj ...] vanavond beluister

(15*e*) *Ik > koop vanmiddag een CD en [Subj ...]
beluister [Obj ...] vanavond

(16*a*) Ik koop vanmiddag een CD en jij [VFin ...]
[Mod ...] een LP

(16*b*) (als) ik vanmiddag een CD koop en jij [Mod ...]
een LP [VFin ...]

(16*c*) (als) ik vanmiddag een CD koop en [Subj ...]
vanavond een LP [Vfin ...]

(16*d*) *Ik koop vanmiddag een CD en jij leent
[Mod ...] [Obj ...]
I buy this-afternoon a CD and you borrow

(17*a*) Ik werk in Nijmegen en hij [VFin ...] in Leiden
(I work in Nijmegen and he in Leyden)

- (17b) Hij zat vóór mij en zij achter mij
(He was sitting in front of me and she behind me)
- (17c) Hij werkt in Nijmegen en (in) Leiden
(He works in Nijmegen and (in) Leyden.)
- (18a) *(dat) hij naar > de muziek luisterde en
[Subj ...] [Obj ...] leuk vond
(that) he to the music listened and
liked
- (18b) (dat) hij de muziek > beluisterde en [Subj ...]
[Obj ...] leuk vond

The basic modes of operation of clause-initial and whole-clause inheritance in FCR and gapping can be deduced from the examples in (15)–(18). Starting with CII, its functioning can be summarized in the following statements:

- (i) The juncture (>) is located before the verb or a major constituent preceding the verb.
- (ii) All pre-juncture (but no post-juncture) major constituents are inherited by the right-hand conjunct.
- (iii) The subcategorization frame of the right-hand verb is complete, that is, all obligatory complements are either locally mentioned or inherited.

Sentences (15c–e) are ungrammatical because *beluister* cannot inherit the direct object, an obligatory complement, from left: *een CD* is a *post-juncture* constituent.

The operation of WCI (gapping) is simple: any non-verbal major constituent of the left-hand clause is a potential proxy. This is demonstrated by examples (16a–c). Example (16d) verifies that as soon as a new verb comes into play, the contraction pattern switches from gapping to FCR. Major clause constituents in the form of PPs cannot be contracted *unless they are adjacent*. Repetition of the underlined words is obligatory in (17a, b), but not in (17c).

Example (18) illustrates a further aspect of CII. The juncture is always located in front of a full major constituent (subject, direct or indirect object, adverbial or prepositional modifier, etc.). The right-hand conjunct cannot inherit *part* of a major phrase. This is attempted in (18a), but fails due to the verbs having different subcategorization frames. This problem does not arise in paraphrase (18b) because *beluisterde* has an NP, not a PP, as direct object.

CII and WCI as defined here are only moderately sensitive to word order, as is desirable in the context of a language like Dutch. Gapping does not even require the right-hand constituents to have the same order as their left-hand counterparts. In the right-hand member of (19), subject and temporal modifier have swapped places.

- (19) De eerstejaars doen deze week tentamen en volgende
week [VFin ...] de tweedejaars [Obj ...]
The freshmen take this week exam and next
week the sophomores

Another FCR phenomenon, also related to Dutch word order variability, cannot be covered by CII. The front position of Dutch main clauses is not only open to the subject but can be filled by other constituents quite easily. Consider (20a), which seems to imply inheritance of the subject NP from a *post-juncture* position, prohibited by CII. This diagnosis is premature, though. In (20b), *meteen* (*at once*) has been replaced by *behendig* (*nimbly*), which cannot be semantically interpreted as a modifier to *riep* (*shouted*).

Nevertheless, (20b) is an acceptable sentence. This implies that the juncture is located at sentence onset, thus preventing rightward inheritance of *behendig* (20c). By the same token, (20d) must be the correct analysis of (20a).

- (20a) Meteen > klom ze achterop en [Mod ...]
riep [Subj ...] “Rijden maar!”
At-once climbed she on-the-back and
shouted “Move it!”
- (20b) *Behendig > klom ze achterop en
[Mod ...] riep [Subj ...] “Rijden maar!”
Nimbly climbed she on-the-back and
shouted “Move it!”
- (20c) > Behendig klom ze achterop en riep “Rijden
maar!”
- (20d) > Meteen klom ze achterop en riep “Rijden maar!”
- (20e) Vanmorgen kwam hij ziek thuis en bleef
de hele dag in bed
This-morning came he sick home and stayed
the whole day in bed
- (20f) *Vanmorgen > kwam hij ziek thuis en [Mod ...]
bleef hij de hele dag in bed

In order to allow the right-hand verb access to its subject, I will assume that the *front position of a right-hand finite clause inherits the subject from the left-hand conjunct unless this position is occupied by another major constituent*. (The latter is the case in (21) and makes repetition of *ik* on the right mandatory.)

- (21) *Dit brood > lust ik niet en [Obj ...] gooi weg
(I don’t like this bread and throw it away)
This bread like I not and throw away

The proposal also implies that the clause-initial adverb in (20e) is not semantically interpreted as a modifier of the right-hand conjunct. Indeed, (20e) does not sound semantically anomalous. But insertion of an explicit subject into the right-hand clause (*hij* in (20f)) is predicted to cause a semantic anomaly, as indeed it does.

4. Inheritance in multiple-verb clauses

What happens if the left-hand member of a clausal coordination contains several verbs, for instance, a main verb and an auxiliary, or a finite verb plus the head verb of an embedded non-finite clause? Which inheritance pattern will prevail when one of the verbs is changed but the other remains intact? And does it matter which of the verbs is changed? Coordination (22a), whose right-hand side inherits both the finite matrix verb and the non-finite complement verb, is comparable to single-verb gapping cases.² However, consider (22b) where the non-finite verb *schrijven* is overwritten by *uitgeven* (*to publish*). This sentence is unacceptable, presumably because the object of *uitgeven* cannot be inherited from left. Inserting a direct object (*boeken*) indeed restores grammaticality. Apparently, the right-hand

²Notice that *boeken*, the direct object of the complement verb, is taken to be a major constituent in the matrix clause. This assumption is indeed made in certain grammatical formalisms. According to Segment Grammar (De Smedt and Kempen 1991), for instance, *boeken* belongs to the complement clause in functional structure, but is “raised” into the matrix clause in constituent structure. Here the NP appears as a sister of the subject, finite verb, etc.

verb wants its subcategorization frame to be complete and the applicability of CII statement (iii) (see Sect. 3) must be extended to the WCI domain.

- (22a) Ik wil artikelen schrijven en jij [VFin ...]
boeken [VInfin ...]
I want articles to-write and you
books
(I want to write articles and you ... books)
- (22b) *Ik wil artikelen schrijven en jij [VFin ...] [Obj ...]
uitgeven
- (22c) Ik wil artikelen schrijven en jij boeken uitgeven
- (22d) (dat) ik artikelen wil schrijven en jij boeken uitgeven

The proposed treatment has the advantage of being applicable to the subordinate equivalents of (22a-c) as well, despite Dutch word order rules which cause the constituents of certain embedded non-finite clauses to be interspersed between those of the matrix (clause union; compare (22c-d)).

- (23a) Ik wil deze CD vanmiddag kopen en
vanavond beluisteren
I want this CD this-afternoon to-buy and
tonight to-listen
- (23b) Ik wil [CompS deze CD > vanmiddag kopen]
en [CompS [Obj ...] vanavond beluisteren]
- (23c) *Ik wil [CompS deze CD > vanmiddag kopen]
en jij ... [CompS [Obj ...] > vanavond beluisteren]
- (23d) ?Ik wil [CompS deze CD vanmiddag kopen]
en jij ... [CompS hem vanavond beluisteren]
- (23e) Ik wil [CompS vanmiddag deze CD kopen]
en jij ... [CompS de opnamen vanavond beluisteren]

In (23a), WCI has left two complement clauses in *adjacent positions*. This opens up new ellipting possibilities (as was the case in PP example (17c)). Analysis (23b) shows that the complement clauses display the FCR pattern with rightward inheritance of *deze CD*. If adjacency is disturbed (23c), ungrammaticality ensues. Example (23c) can be made less unacceptable by adding the explicit direct object *hem* (*it*) in the right-hand complement clause, thus removing FCR from the complement clause (23d). But this introduces an unaccented constituent, which is undesirable in gapping cases. The non-pronominal direct object *de opnamen* (*the recordings*) solves this problem and produces a fully grammatical sentence (23e).

What will happen in multiple-verb cases when the right-hand conjunct introduces a new matrix verb? Extrapolating from the single-verb cases treated in Sect. 3, we predict FCR and CII. Indeed, the ungrammaticality of (24a-b) shows that no post-juncture constituents of the left-hand clause are inherited. The grammatical alternatives to (24a) is (24c), with an explicit direct object on the right.

- (24a) *Ik > verwacht [CompS deze week
een artikel te schrijven]
en ... hoop [CompS volgende week
[Obj ...] te versturen]
(*I expect to write an article this week and hope to
mail ... next week)
- (24b) *Ik > verwacht [CompS deze week
een artikel te schrijven]

en ... hoop [CompS volgende week
een recensie [VInfin ...]]
(*I expect to write an article this week and hope ...
a review next week)

- (24c) Ik > verwacht deze week een artikel te schrijven en ...
hoop het volgende week te versturen

Finally, without presenting the supporting empirical evidence I assume that auxiliaries (future, perfective, modal) can be given a treatment similar to complement taking verbs with clause union such as *willen* (*to want*). The examples in (25) may suffice (FCR in (25a-c), gapping in (25d-f)).

- (25a) (dat) ik > vanmiddag een CD heb gekocht
en [Subj ...] hem vanavond zal beluisteren
(that) I this-afternoon a CD have bought
and it tonight will listen-to
- (25b) (dat) ik deze CD > vanmiddag heb gekocht en
[Subj ...] [Obj ...] vanavond zal beluisteren
- (25c) *Ik > heb vanmiddag een CD gekocht en zal
[obj ...] vanavond beluisteren
- (25d) Ik heb een CD gekocht en jij [VFin ...] een LP
[VInfin ...]
- (25e) *Ik heb een CD gekocht en jij [VFin ...] [Obj ...]
beluisterd
- (25f) Ik heb een CD gekocht en jij [VFin ...] een LP
beluisterd

5. Conclusion

The inheritance-based approach modeled after the self-correction process appears to be a viable one. Not only the examples given here support this conclusion but also the fact that, although as yet incomplete, it accounts for a large part of the rich collection of observations brought together by Geerts *et al.* (1984; Chap. 27.5) in their Standard Grammar of Dutch.

Finally, although I have refrained from using the conceptual machinery or the terminology of an established grammatical formalism, I do not believe that all existing linguistic schools of thought are equally compatible with the ideas put forward here. But, due to space limitations, I have to leave this for the reader to assess.

- DE SMEDT, K., and KEMPEN, G. 1991. Segment grammar: a formalism for incremental sentence generation. *In* Natural language generation in artificial intelligence and computational linguistics. Edited by C.L. Paris, W.R. Swartout, and W.C. Mann. Kluwer Academic Publishers, Boston, MA.
- GEERTS, G., HAESERYN, W., DE ROOIJ, J., and VAN DEN TOORN, M.C. 1984. *Algemene Nederlandse Spraakkunst*. Wolters-Noordhoff, Groningen, The Netherlands.
- GOODALL, G. 1987. *Parallel structures in syntax: coordination, causatives, and restructuring*. Cambridge University Press, Cambridge, MA.
- LEVELT, W. 1983. Monitoring and self-repair in speech. *Cognition*, 14: 41-104.
- PIJLS, F., and KEMPEN, G. 1986. Een psycholinguïstisch model voor syntactische samentrekking. *De Nieuwe Taalgids*, 79: 217-234.
- VAN OIRSOUW, R. 1987. *The syntax of coordination*. Croom Helm, London, United Kingdom.
- VAN WIJK, C., and KEMPEN, G. 1987. A dual system for producing self-repairs in spontaneous speech: evidence from experimentally elicited corrections. *Cognitive Psychology*, 19: 403-440.