

Review and critique of a GPSG account of gapping in French †

INTRODUCTION

Coordination is an interesting phenomenon of natural languages that must be dealt with in a natural language processing system. In this article we shall focus on the type of coordination which is illustrated in examples (1a-b) below.

- (1) a Pierre offre des fleurs et Luc des chocolats.
Peter offers flowers and Luke chocolates.
- b Pierre offre des fleurs pour Noël et Luc pour Pâques.
Peter offers flowers for Christmas and Luke for Easter.

This type of coordination is called *non-constituent coordination*, since a list of constituents follows the *et* (*and*) instead of a single one. In the coordinated sentence¹ to the right of the *et*, when the verb is omitted, other constituents can or must also be omitted, but some must be kept.

Intuitively coordination establishes a parallelism between propositions. In our GPSG account, isomorphic analysis trees are a reflection of the syntactic parallelism between coordinating and coordinated sentences and this syntactic parallelism induces the semantic parallelism.

LINGUISTIC CONSTRAINTS

We briefly review the constraints that are described in Bouchard, *et al.* [1997]². First of all, gapping is not unique to verbs but applies to all elements marked +v, that is adjectives and adverbs also [Bouchard, *et al.*, 1996], and only those elements. Without biasing the analysis, we can visualize the elision as a gap that can propagate in certain cases. We call the element that triggers the gap the *principal elided element*. The analysis has shown that at least an *emphasis* (subject, controller in the case of infinitives or extracted complement) and a complement or adjunct (in canonical position) of the principal elided element must always remain. Starting from the principal elided element, heads of other +v constituents which directly dominate the principal elided element must be omitted as well as all other constituents on the level except the controller of a gapped constituent.

GPSG ACCOUNT

* Département d'informatique.

** Département de linguistique.

† Research funded by the SSRC (410-97-0419) and the FCAR (98-ER-1198).

¹ In order to simplify the presentation, we only treat binary coordination with *et*. We use the term *coordinating sentence* to refer to the sentence to the left of the *et* (*and*) and *coordinated sentence* to refer to the sentence to the right.

² The paper compares the GPSG and HPSG analyses of the phenomenon.

Our Generalized Phrase Structure Grammar (GPSG) [Gazdar, *et al.* 1985] account of gapping follows the analysis of Russell [1987]. In this analysis, a +G feature identifies a gap. A *syntactic parallelism* is established between coordinating and coordinated sentences: in the analysis tree of the coordinated sentence, omitted verbal heads are marked +NULL and other omitted elements, +ELIDED.

The coordinating sentence is analyzed using the standard grammar rules. A TEMPLATE feature is used to encode the subcategorization information observed in the coordinating sentence; the value of this feature is then used to constrain the analysis of the coordinated sentence as can be seen in the ID rules of Figure 1 which introduce gapping.

$X_1^2[+V, +COORD] \rightarrow H[TEMPLATE \alpha, -G], H[CONJ \text{ et}, TEMPLATE \alpha, +G]$	<i>coordination with gapping</i>
$X_1^2[+V, +COORD] \rightarrow H[TEMPLATE \alpha, -G], H[CONJ \text{ et}, TEMPLATE \beta, -G]$	<i>coordination without gapping</i>

Fig. 1. ID rules for the coordination of +V constituents.

The coordinated sentence is analyzed using ID rules that are generated by the metarules of Figure 2.

MR1	$X_1^2[+V] \rightarrow H, X_2^2, W_1, W_2$	
	↓	
	$X_1^2[+V] \rightarrow H[+NULL], X_2^2, W_1[+ELIDED], W_2$	
MR2	$X_1^2[+V] \rightarrow H, X_2^2, X_3^2[+V], W_1$	<i>if X_2^2 controls X_3^2</i>
		↓
	$X_1^2[+V] \rightarrow H[+NULL], X_2^2, X_3^2[+V, +G], W_1[+ELIDED]$	
MR3	$X_1^2[+V] \rightarrow H, X_2^2[+V], W_1$	<i>if X_2^2 has no local controller</i>
		↓
	$X_1^2[+V] \rightarrow H[+NULL], X_2^2[+V, +G], W_1[+ELIDED]$	

where W_1 and W_2 are (possibly empty) multisets of categories and $W_1[+ELIDED]$ means that all the categories in W_1 are marked +ELIDED.

Fig. 2. The gapping metarules.

As an example, we illustrate in Figures 3a and 3b the result of the interaction of the ID rules which license the analysis tree of sentence (2) below.

(2) Pierre permet à ses enfants de manger la glace et Luc _ à ses parents _ la tarte.

Peter allows his children to eat the ice cream and Luke _ his parents _ the pie.

The V^2 constituents in the trees 3a and 3b are distinguished by a double digit index. The first digit of the index i is 0 for the coordination as a whole, 1 for the coordinating sentence and 2 for the coordinated sentence. For any given i , the second digit ranks along a given branch the V^2 constituents starting from 1 at the leaves of the tree. We observe in figures 3a and 3b that there is a one to one correspondence between the V^2 nodes in the coordinating and coordinated sentences.

$$V_{1i}^2 \leftrightarrow V_{2i}^2 \quad 1 \leq i \leq 4$$

Constituents which have been omitted in the coordinated sentence are marked as +NULL if verbal heads, and +ELIDED otherwise. It is in this sense that the analysis trees of the GPSG account are said to be *isomorphic*.

The constituent V^2_{23} is licensed by an ID rule produced by MR2 since *à ses enfants* (*his children*) controls *manger la glace* (*to eat the ice cream*) and the constituent V^2_{21} is licensed by an ID rule generated by MR1.

Since it can be easily shown that the ID rules generated by the metarules over generate greatly, a crucial part of our analysis is the TEMPLATE feature which is used to constrain the allowable gapped coordinated sentences. In the analysis presented in Bouchard, *et al.* [1997], TEMPLATE is a FOOT feature the propagation of which is governed by a new TEMPLATE principle patterned on the FFP.

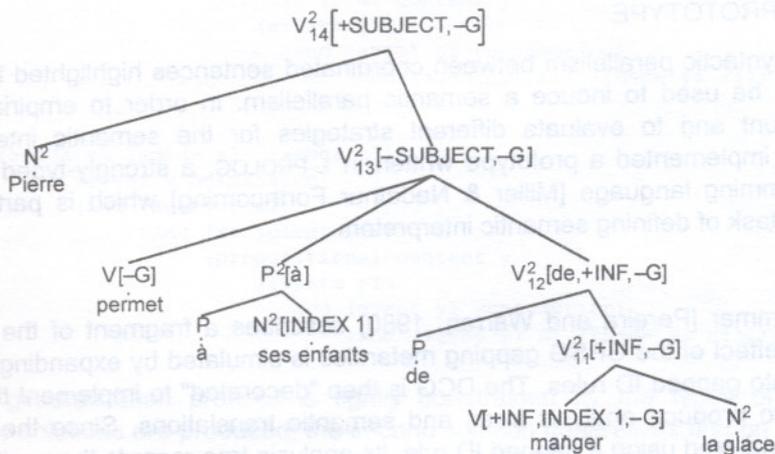


Fig. 3a. Pierre permet à ses enfants de manger la glace (*Peter allows his children to eat the ice cream*).

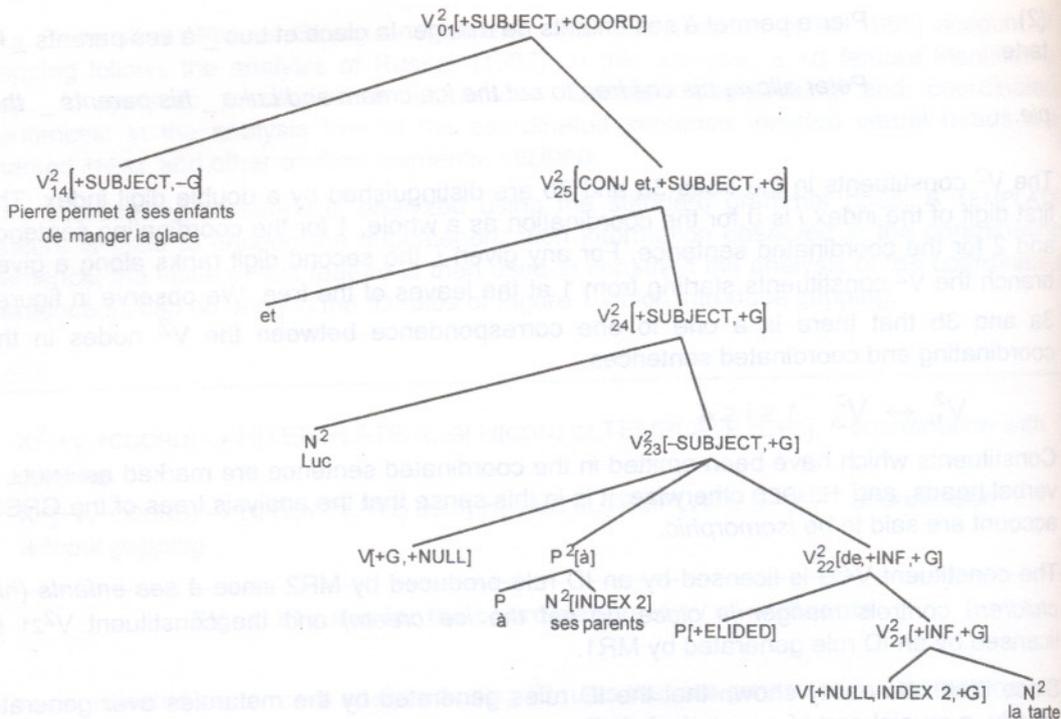


Fig. 3b. Pierre permet à ses enfants de manger la glace et Luc, à ses parents la tarte
(Peter allows his children to eat the ice cream and Luke, his parents the pie).

λ -PROLOG PROTOTYPE

The strong syntactic parallelism between coordinated sentences highlighted by the GPSG analysis can be used to induce a semantic parallelism. In order to empirically test our GPSG account and to evaluate different strategies for the semantic interpretation of gapping, we implemented a prototype written in λ -PROLOG, a strongly-typed higher-order logic programming language [Miller & Nadathur Forthcoming] which is particularly well-suited to the task of defining semantic interpreters.

A DCG grammar [Pereira and Warren, 1980] describes a fragment of the grammar of French. The effect of the GPSG gapping metarules is simulated by expanding by hand the +V ID rules into gapped ID rules. The DCG is then "decorated" to implement the TEMPLATE feature and to produce analysis trees and semantic translations. Since the coordinated sentence is analyzed using a gapped ID rule, its analysis tree records the position of all the constituents (marked +NULL or +ELIDED) of the gap. The analysis of a gapped coordinated sentence is constrained by the TEMPLATE feature: the value of this feature percolates first up (in the coordinating sentence) and then down (in the coordinated sentence).

Semantic interpretation in the prototype is based on a simple truth functional model in which the semantic interpretation of a sentence is represented by a logical formula. The formula corresponding to a sentence is a conjunction of atomic formulae which are existentially quantified explicitly [Davidson, 1989].

The semantic interpretation of the gapped constituents in the coordinated sentence is obtained by transferring the semantic interpretations of the corresponding elements in the coordinating sentence. One way of achieving this is by applying the anti-unifiers (generalizations) of the semantic interpretation of the coordinating sentence to the semantic interpretation of the coordinated sentence. This solution effectively implements a proposal presented by Sag, *et al.* [1985]. The coordinated sentence of example (2) is analyzed as an iteration of X^2 constituents, the semantic interpretations of which are represented as a reversed list

```
(is-object tarte :: is-individual parents :: is-individual luc :: nil)
```

The semantic interpretation of the coordinating sentence is then generalized upon, given the semantic interpretation of the X^2 fragments in the coordinated sentence.

```
?- genL (exists y\
  and (event y permission)
    (and (agent y pierre)
      (and (recipient y enfants)
        (propositional-content y
          (exists y1\
            and (event y1 consumption)
              (and (agent y1 enfants) (theme y1
glace))))))
  (is-object tarte :: is-individual parents :: is-individual luc ::
nil) R.
```

```
R = lambda x\ lambda x1\ lambda x11\ exists y\
  and (event y permission)
    (and (agent y x)
      (and (recipient y x1)
        (propositional-content y
          (exists y1\
            and (event y1 consumption)
              (and (agent y1 x1) (theme y1 x11))))))
;
```

```
R = lambda x\ lambda x1\ lambda x11\ exists y\
  and (event y permission)
    (and (agent y x1)
      (and (recipient y x)
        (propositional-content y
          (exists y1\
            and (event y1 consumption)
              (and (agent y1 x) (theme y1 x11))))))
;
```

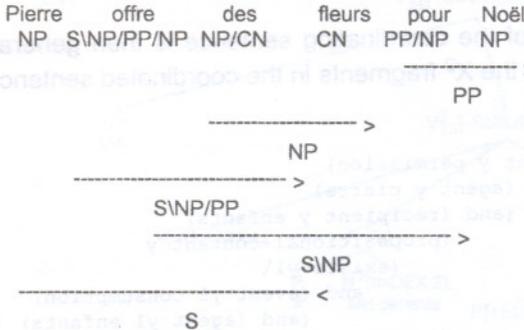
Even if the generalization process is tightly constrained by the types of the domain elements, two solutions are produced: the second solution generalizes first on the *recipient*, instead of the *agent*, and is unacceptable. The generalization of the semantic interpretation of the coordinating sentence is applied to the list of semantic interpretations of the X^2 fragments in the coordinated sentence to produce the interpretation of the coordinated sentence.

CCG ACCOUNT OF GAPPING

Steedman [1990] has presented an account of gapping in the framework of Combinatorial Categorical Grammar (CCG)³, an enrichment of pure Categorical Grammar (CG) with linguistically motivated rules borrowed from Combinatory Logic.

As an illustration, we show the CCG analysis of sentence (1b) above.

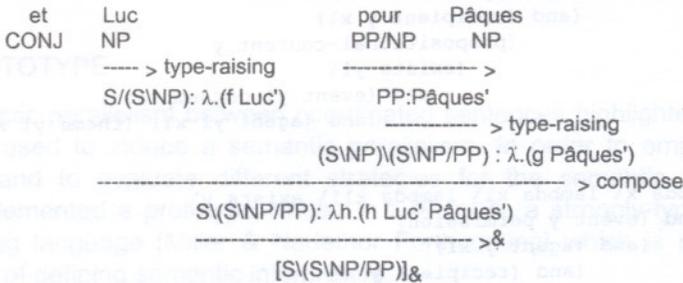
The analysis of coordinating sentence is straightforward.



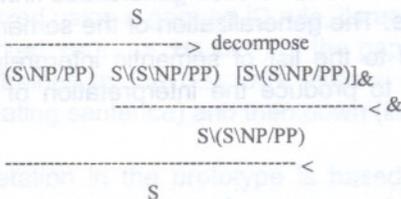
This analysis yields the following semantic interpretation.

(offre' Pierre' fleur' Noël')

The coordinated sentence is analyzed as follows.



The syntactic type of the coordinated sentence is that of an incomplete sentence S containing a subject NP and a PP (i.e., the verb and the object NP are missing). The crucial step in the derivation is the type-raising of the PP constituent. The next step in the analysis is a *decompose* step required in order to conjoin the two coordinated sentences.



³We refer the interested reader to [Steedman 1993, 1996]

After application of the coordination rule we have

$S \backslash (S \backslash NP / PP) : \lambda h.(and (h Pierre' Noël') (h Luc' Pâques')) :: (e \rightarrow e \rightarrow t) \rightarrow t$

$S \backslash NP / PP : \lambda x. \lambda y.(offre' x fleurs y) :: (e \rightarrow e \rightarrow t)$

which yields by application and β -reduction the semantic interpretation of the complete sentence

$S : (and (offre' Pierre' fleurs' Noël) (offre' Luc' fleurs' Pâques')) :: t$

We observe that the semantic value of $S \backslash NP / PP$, $\lambda x. \lambda y.(offre' x fleurs y)$, is a generalization of (offre' Pierre' fleurs' Noël') and thus the CCG account of gapping is equivalent to the third strategy of semantic transfer which was explored in our λ -PROLOG prototype.

Unfortunately the elegant CCG account can also be used to derive an ungrammatical sentence such as the following

(3) *Pierre offre des fleurs à Noël et Luc.

**Peter offers flowers at Christmas and Luke.*

with the following semantic interpretation.

$S : (and (offre' Pierre' fleurs' Noël') (offre' Luc' fleurs' Noël')) :: t$

However, this semantic result would be an appropriate interpretation of the sentence

(3') Pierre offre des fleurs à Noël et Luc aussi.

Peter offers flowers at Christmas and Luke also.

CONCLUSION

In our GPSG account of gapping, the parallelism between coordinating and coordinated sentences is reflected in the isomorphic syntactic analysis trees. This GPSG account does not over generate, at the cost of some heavy machinery (3 metarules, the TEMPLATE feature and associated principle). The syntactic parallelism is used to induce a semantic parallelism between coordinated sentences. The CCG account clearly over generates and there seems to be no simple way to prevent it from doing so. As we have shown, if we analyze the right conjunct as an iteration of X^2 constituent, the λ -PROLOG prototype also over generates. Thus the λ -PROLOG prototype allowed us to show that a satisfactory analysis of gapping relies both on syntax and semantics. Semantic transfer by default unification [Emirikian, *et al.* Forthcoming] produces the same effect as the *decompose* rule of CCG but since the syntax of the right conjunct is filtered by the gapped ID rules constrained by the TEMPLATE feature, there is no over generation of gapped conjuncts. Hence the elegance of the CCG account is obtained at a price.

REFERENCES

Bouchard, L. H., L. Emirikian & C. Fouqueré. (1997). Les coordinations avec gapping : formalisation en GPSG et HPSG. *Traitement Automatique des Langues*, 38(2), pp. 145-168.

Davidson, D. (1989). *Essays on Actions and Events*. Oxford: Clarendon Press, 304 p.

Emirikian, L. & L. H. Bouchard. (Forthcoming). Semantic interpretation of gapped coordination in French. In *CIL'97*, Paris, Elsevier, 18 p.

Gazdar, G., E. Klein, G. Pullum & I. Sag. (1985). *Generalized Phrase Structure Grammar*. Cambridge MA: Harvard University Press, 276 p.

Miller, D. & G. Nadathur. (Forthcoming). Higher-Order Logic Programming. In D. Gabbay, C. Hogger & A. Robinson (Ed.), *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vol. 5. Oxford UK: Oxford University Press.

Pereira, F. & D. H. D. Warren. (1980). Definite Clause Grammars for Language Analysis: A Survey of the Formalism and a Comparison With Augmented Transition Networks. *Artificial Intelligence*, **13**, pp. 231-278.

Russell, G. (1987). *Verbal Ellipsis in English Coordinate Constructions*. Ph.D. Thesis, Cambridge University.

Sag, I. A., G. Gazdar, T. Wasow & S. Weisler. (1985). Coordination and How To Distinguish Categories. *Natural Language and Linguistic Theory*, (3), pp. 117-171.

Steedman, M. (1990). Gapping as Constituent Coordination. *Linguistics and Philosophy*, **13**(2), pp. 207-263.

Steedman, M. (1993). Categorical Grammar. *Lingua*, **90**, pp. 221-258.