Parsing Repairs

MARCEL CORI*, MICHEL DE FORNEL** & JEAN-MARIE MARANDIN*

*Université Paris 7, **EHESS (CELITH) & *CNRS (URA 1028)

Abstract

The paper deals with the parsing of transcriptions of spoken utterances with self-repair. A syntactic analysis of self-repair is given. A single well-formedness principle accounts for the regularities observed in a large corpus of transcribed conversations: a constituent is a well-formed repair iff it can be substituted into the right edge of the tree which represents the syntactic structure of the interrupted utterance. The analysis is expressed in a PS-grammar. An augmentation of the Earley algorithm is presented which yields the correct inputs for conversational processing.

1. Introduction

If natural language understanding systems are ever to cope with transcriptions of spoken utterances, they will have to handle the countless self-repairs (or self-corrections) that abound in them. This is a longstanding problem: "hesitations and false starts are a consistent feature of spoken language and any interpreter that cannot handle them will fail instantly" (Kroch & Hindle 1982:162). See also (Kay et al. 1993). The current assumption is that interruptions and self-repairs should be handled by editing rules which allow the text to be normalized; these rules belong to a kind of adjustment module within the performance device (Fromkin 1973; Kroch & Hindle 1982; Hindle 1983; Labov (pc); Schegloff 1979). We shall lay the foundations of another approach in this paper: interruptions and self-repairs can be directly handled by the syntactic module. Our proposal is based on the observation that "speakers repair in a linguistically principled way" (Levelt 1989:484).

The regular character of self-repair has been emphasized in a number of detailed descriptive studies in different fields: linguistics, conversation analysis, psycholinguistics (Blanche-Benveniste 1987; Fornel 1992a, 1992b; Frederking 1988; Levelt 1983, 1989; Schegloff et al. 1977; Schegloff 1979). Among others, Levelt (1989:487) proposes that "self-repair is a syntactically regular process. In order to repair, the speakers tend to follow the normal rules of syntactic coordination". We have shown elsewhere that self-repair cannot be reduced to a kind of coordination[1]. Nevertheless the forms of self-repair are not only regular but they are submitted to a simple geometric principle of well-formedness. This principle is given a formal representation in a PS-grammar. It opens a fresh perspective on the parsing of non standard inputs with self-repairs: a simple and principled augmentation of a standard parsing algorithm can handle them. We make the point with the Earley algorithm.

2. Characterizing self-repair

2.1 *The overt characteristics of self-repair*

The overt characteristics of self-repair are the following: an utterance is interrupted. The interruption is marked by a number of prosodic or phonetic signals such as cut-offs, pauses, hesitation markers or lengthenings. The interruption is followed by an arbitrary number of constituents which appear to be in a paratactic relation to the interrupted utterance. This is illustrated by the following sample taken from a corpus of transcribed conversations[2]:

(1) a. elle était:: an- mm irlandaise (.) enfin:: de l'Irlande
b. elle ne sort plus de son:: euh studio
c. mais il faudrait que vous passiez par euh:: (.) par le:: par le numéro du commissariat hein
d. je croyais qu'il était euh:: je croyais qu'il était encore là-bas jusqu'à ce soir[3]

We shall use the following shorthand convention in the following: O stands for the interrupted utterance, # for any prosodic or phonetic signal and R for the repair.

[1]The argumentation is summed up in (Cori et al. 1995); it is fully developped in (Fornel & Marandin Forthcoming).

[2]The research is based on an extended corpus of spontaneous self-repairs in French (approximatively 2000 occurences). They are taken from a large body of transcribed audio and video tapes of naturally occurring conversations in various settings (telephone, everyday conversation, institutional interaction, etc.). We refer the reader to (Schegloff et al. 1977; Schegloff 1979) for the transcription conventions of (1).

[3](1.a) She was En- mm Irish (.) from Ireland; (1.b) she doesn't leave her er studio; (1.c) but you should go through (.) through the (.) through the number of the police station; (1.d) I thought that he was er I thought that he was still there till tonight.

In order to limit the word to word glossing of French utterances, we shall use simple forged examples in the following.

2.2 *The structural characteristics of self-repair*

The structural features of self-repair are the following:

- A) O is a segment analyzable as a well-formed syntactic unit apart from the fact that one or more sub-constituent(s) may be missing.

- B) R is a segment analyzable as a single syntactic unit. This unit may be lexical, phrasal or sentential. It is usually a maximal projection (X^{max} or S) but need not be. R can be interrupted as O can be; this yields what we call a cascaded repair (§3.2 below).

Note that any analysis which reduces self-repair to coordination presupposes (B). In this connection, note the difference between (2.a) and (2.b):

(2) a. ?? l'homme avec les lunettes a poussé le clown # avec les moustaches a poussé le clown
b. l'homme a donné un coup de poing au # une gifle au clown[4]

The string *avec les moustaches a poussé le clown* does not make up a constituent and thus is not a licit R, whereas *une gifle au clown* is a licit R since it can be treated as a single constituent, a ghost constituent (Dowty 1988), in a coordination and in a question-answer pair; *une gifle au clown* is not a maximal projection.

- C) R depends on O[5]. The dependency between O and R includes two sub-relations:

- C1) R repairs a constituent of O which immediately precedes R. Hence the ill-formedness of *l'homme avec les lunettes a poussé le clown # avec les moustaches.* The PP *avec les moustaches* cannot repair *avec les lunettes* over the VP *a poussé le clown.*

- C2) The choice of the category of R depends on O: R is a licit daughter in O. This is illustrated in (3):

[4](2.a) The man with the spectacles pushed the clown # with the moustache pushed the clown; (2.b) the man gave a punch to the # a slap to the clown.

(2.a) is judged an ill-formed repair by Levelt (1989:489). We have not encountered repairs such as (2.a) in our corpus.

[5]Levelt (1989:486) did observe the fact: "well-formedness of a repair is apparently not a property of its intrinsic syntactic structure. It is rather dependent on its relation to the interrupted original utterance"

(3) a. Les enfants attendent le bateau # le ferry de Marseille
b. Les enfants attendent le # que le bateau vienne
c. Les enfants attendent le # bateau
c′. Les enfants attendent le ferry # de Marseille[6]

Any contemporary theory of coordination puts two constraints on each conjunct: (i) "each conjunct should be able to appear alone in place of the entire coordinate structure" (Sag et al. 1985:165); (ii) each conjunct shares at least one feature with the other (categorial identity being the most frequent case). Self-repair does not have to meet the latter constraint (ii): this is why it cannot be reduced to a coordinate structure. On the other hand, it has to meet the former.

- D) R completes O. The intuition which underlies the notion of repair is the following: when R is interpreted as a repair, R is interpreted as a constituent in O, it may, or may not, replace a constituent partially or completely realized in O. For example the sequences $O\#R$ in (3.a) and in (3.c′) are interpreted as (4) would be; in (3.a) the NP *le ferry de Marseille* replaces *le bateau* whereas in (3.c′) the PP *de Marseille* replaces nothing.

(4) Les enfants attendent le ferry de Marseille

3. Analyzing self-repair

3.1 *Syntactic well-formedness*

Generalization (C) which characterizes the relation holding between O and R can be unified in a single principle, the principle of the right edge (REP)[7]:

(5) A constituent R is a well-formed repair for O iff it can be substituted into the right edge of the interrupted O.

The interrupted part of (3.b) *Les enfants attendent le #* may be repaired with an R of category N *bateau*, NP *le ferry*, VP *espèrent que le bateau viendra* or S *Les enfants espèrent que le bateau viendra* (the constituency requirement involves categorial identity) or with S'[que] *que le bateau vi-*

[6](3.a) The children wait for the boat # the ferry to Marseille; (3.b) that the boat arrives.

[7]Principle (5) is reminiscent of the Major Constituent Constraint on gapping (Hankamer 1973; Gardent 1991).

On the status of the right edge for discourse processes, see (Gardent 1991; Prüst 1993).

enne (in accordance with the sub-categorization requirement of the verb *attendre*). This is illustrated in Figure 1[8].

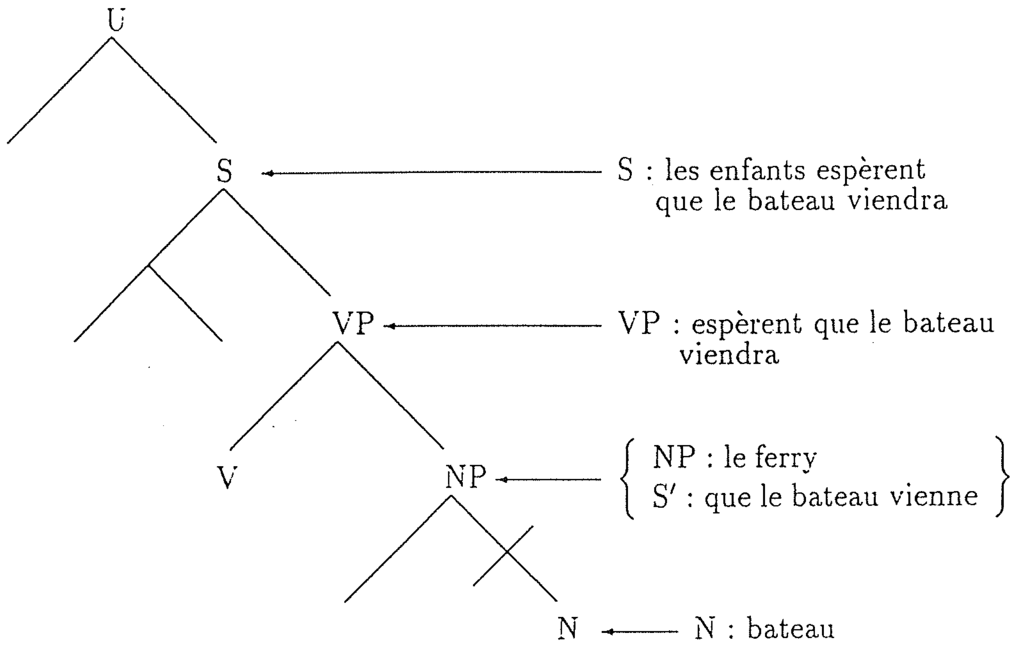

Fig. 1: *Illustration of the licensing principle.*

Principle (5) prevents ill-formed repairs such as (2.a) above. It accounts for all types of self-repair (reformulation, lemma substitution and restart)[9].

3.2 *Cascaded repair*

A repair R itself can be interrupted and it can be repaired. Examples are given in (1.a) and (1.c) where the phonetic signal is followed by a "string of Rs"[10]. The sequence can be schematized as $O\#R_1\#R_2 \ldots R_m$. The REP needs not be augmented or modified to handle this case once we have made precise the structures acting as O in the cascade. For example:

(6) Les enfants attendent le # le bateau de # qui va à Marseille

R_1 (*le bateau de*) can be substituted into O. R_2 (*qui va à Marseille*)

[8] The category U stands for Utterance.

[9] On the contrary, the reduction of self-repair to coordination leads to distinguish three different processes (De Smedt & Kempen 1987).

[10] Blanche-Benveniste (1987) proposed that the Rs form a coordinate structure. See Fornel & Marandin (Forthcoming) for counter argumentation.

cannot be (?? *les enfants attendent # qui va à Marseille*). On the other hand, it can be substituted into the "new" configuration N which is obtained by substituting R_1 into O (*les enfants attendent le bateau de*).

Cascaded repairs result from the iteration of repair. Repair always involves only one O and one R at a time. The tree obtained by substituting R_1 into O gives N which becomes the O for repair R_2 and so forth.

3.3 *Interpreting O#R*

The interpretation of $O\#R$ is built on the tree N obtained by substituting R into O. Thus R is treated as a repair. For example, the interpretation of an utterance such as (3.b) *Les enfants attendent le # que le bateau vienne* discards the interrupted NP *le #* and is derived from the tree N which is the repaired tree: *Les enfants attendent que le bateau vienne.*

The main implication for the interpretation of $O\#R$ is the following: the recovery of the interpretation is parallel to the licensing of the category of R. Once R is recognized as a constituent of O, no specific rule of interpretation has to be called for; the configuration N is interpreted exactly in the same way that a canonical configuration would be[11].

3.4 *Parsing self-repairs*

The analysis allows a simple solution to the problem of parsing an input $O\#R$. The relevant features are the following: (i) R is a licit daughter of O and (ii) R is a daughter on the right edge of O according to the REP. (The REP restricts the choice of categories for R). Thus the input $O\#R$ can be parsed with a classical algorithm such as Earley and as easily as any other input.

Moreover, the same kind of ambiguity encountered in the parsing of canonical inputs arises: attachment ambiguity. For example, *Marie* in (7.a) can be substituted to *Paul* or to *la femme de Paul*: likewise in (7.b) *le professeur Tournesol (...)* can be substituted under S' or U.

(7) a. Jean aime la femme de Paul # Marie
b. Tournesol m'a dit que l'élève # le professeur Tournesol m'a dit que l'élève n'était pas au point[12]

[11]Here lies the other drawback of the reduction of self-repair to coordination: in a coordinate structure, the well-formedness constraints are distinguished from the interpretative rules which depend on the choice of the conjunctions. If self-repair were a kind of coordination, its semantics should be given a separate and specific formulation. This does not seem plausible.

We propose in (Fornel & Marandin Forthcoming) a heuristic rule that minimizes the attachment ambiguity.

4. Representing self-repair

self-repair receives a straightforward formal representation in a PS-grammar. We first define the notions of interrupted tree and right subtree.

Let $G = \langle V_T, V_N, \mathcal{R}, U \rangle$ be a CF-grammar where V_T is a terminal vocabulary, V_N a non-terminal vocabulary, $U \in V_N$ the axiom, and where the rules are numbered from 1 to n. Each rule i is $left(i) \rightarrow right(i)$; $\lambda(i)$ is the length of $right(i)$; $right_j(i)$ is th j-th symbol in $right(i)$. We assume that there are no rules with $right(i)$ being the empty string.

An elementary tree is associated with each rule. Complex non punctual trees are represented by leftmost derivations: $A = (i_1 \ldots i_p)$. $root(A)$ is the label of the root of A.

Definition 1 An *interrupted tree*, written $\mathcal{A} = (i_1 \ldots i_{k-1} i_k[l] i_{k+1} \ldots i_p)$, is such that the l-th leaf of i_k is a terminal leaf of the tree $A = (i_1 \ldots i_k \ldots i_p)$ (i.e., a leaf labelled with a symbol taken in V_T), all nodes preceding this leaf (according to the precedence order) dominate terminal leaves of A, and all nodes following this leaf are leaves of A.

Definition 2 An *elementary right subtree (ERS)* of an interrupted tree $\mathcal{A} = (i_1 \ldots i_{k-1} i_k[l] i_{k+1} \ldots i_p)$ is defined as follows:

(i) i_1 is an ERS of $\mathcal{A}$

(ii) if i_j is an ERS of $\mathcal{A}$ and if $right(i_j) = \alpha Y$, Y being a non-terminal symbol, if all non-terminal leaves of i_j are roots of elementary trees in $\mathcal{A}$, then the last one of these elementary trees, i_{j+s}, is also an ERS of $\mathcal{A}$. If $j + s = k$, we must have $l \geq \lambda(i_{j+s}) - 1$.

Definition 3 If i_r is an ERS of $\mathcal{A}$, then $(i_r \ldots i_p)$ is a *right subtree* of $\mathcal{A}$.

Right edge principle. We consider an interrupted tree $O = (i_1 \ldots i_k[l] \ldots i_p)$ such that $root(O) = U$ and a tree $R = (j_1 \ldots j_q)$.

(8) R is a well formed repair for O iff either $root(R) = U$ or there is an ERS i_r of O and a rule ξ in the grammar such that $left(i_r) = left(\xi)$ and $right(i_r) = \rho X$ and $right(\xi) = \rho\ root(R)$ with $X \in V_N \cup V_T$.

Repaired tree. A repaired tree N is obtained by substituting R for a right subtree of O: $N = (i_1 \ldots i_{r-1} \xi j_1 \ldots j_q)$. R is then a right subtree of N.

[12](7.a) Jean loves the wife of Paul # Marie; (7.b) Tournesol told me that the student # Professor Tournesol told me that the student was not ready.

Note that lexical repair is not a special case: it corresponds to the case of a punctual R tree.

Cascaded repairs. We have two sequences:

(i) $N_0, N_1, \ldots, N_m$ where $N_0, N_1, \ldots, N_{m-1}$ are interrupted trees and N_m a complete tree such that $root(N_0) = U$, and

(ii) $R_1, \ldots, R_m$ where $R_1, \ldots, R_{m-1}$ are interrupted trees (interrupted repairs), and R_m is a complete tree.

Condition (8) is verified for each pair $\langle N_{i-1}, R_i \rangle$. N_i is a new tree obtained from N_{i-1} and R_i. N_m is the repaired tree of the cascade.

5. An augmented Earley algorithm for repair

We show how to augment the Earley algorithm (Earley 1970) to parse interrupted inputs with repairs.

5.1 *String representations*

Let LF be a set of lexical forms. A *categorization function* associates a set of terminals with each lexical form u: $cat(u) \subset V_T$.

A *representation* of a string $u_1 u_2 \ldots u_n \in LF^*$ is given by a tree A such that $root(A) = U$ and such that if the ordered sequence of the leaves of A is $z_1 z_2 \ldots z_n$, then for each i, $z_i \in cat(u_i)$.

A string may be represented by an interrupted tree $\mathcal{A} = (i_1 \ldots i_k[l] \ldots i_p)$ by taking $z_1 z_2 \ldots z_q$ where z_q is the lth leaf of i_k.

5.2 *Augmentations to the standard algorithm*

A type is added in the definition of the states:

- *right* (vs *left*) indicates whether an elementary tree is an ERS;
- *cut* distinguishes the states involved in the building of interrupted trees.

We add the following to the definition of the operations:

- **predict**: [1.1.2] and [1.2] below are added to send into the set S_{m+1}, which contains the initial states for R, all elementary trees which may dominate R.
- **scan**: [2.1.2], [2.2] and [2.3] are added to handle the replacement of punctual subtrees of O.

- **complete**: [3.2] is added in order to obtain a representation of the interrupted trees O in addition to the straightforward output of the algorithm: the repaired trees N.

5.3 *The augmented algorithm*

The input data of the algorithm is a grammar $G = \langle V_T, V_N, \mathcal{R}, U\rangle$ and a string $u_1u_2\ldots u_m\ \#u_{m+2}u_{m+3}\ldots u_{m+p+1}$ where each $u_i \in LF$. We add to the grammar a rule numbered 0 such that $right(0) = U$ and $left(0) \notin V_N$.

The algorithm builds a sequence of sets, $S_0, S_1, \ldots, S_{m+p+1}$, made of states. A state is a 5-uple $\langle q, j, k, t, \alpha\rangle$ where q is a rule, j is a position in $right(q)$ $(0 \leq j \leq \lambda(q))$, k is a set number $(0 \leq k \leq m+p+1)$, t is a type ($right$, $left$ or cut), α is a string (the current result). The initial state $\langle 0,0,0, right, \varepsilon\rangle$ is entered into S_0.

Consider the state $\langle q, j, k, t, \alpha\rangle \in S_i$. The three Earley operations are now the following:

[1] **Predict:**
  [1.1] If $t \neq cut$ and $j < \lambda(q)$ then
    for each rule q' such that $left(q') = right_{j+1}(q)$
      [1.1.1] add $\langle q', 0, i, t', \varepsilon\rangle$ to S_i
        if $t = left$ or $j < \lambda(q) - 1$ then $t' = left$ else $t' = right$.
      [1.1.2] If $i = m$ and $j > 0$ and $j = \lambda(q) - 1$ and $t = right$ then
        add also $\langle q', 0, i, right, \varepsilon\rangle$ to S_{m+1}.
  [1.2] If $i = m$ then
    [1.2.1] if $j > 0$ then for each $\langle q', j', k', right, \beta\rangle \in S_k$
    such that $right_{j'+1}(q') = left(q)$ and $j' = \lambda(q') - 1$.
      for each rule ξ such that $left(q') = left(\xi)$ and $right(q') = \rho X$
      and $right(\xi) = \rho Y$,
        for each rule r such that $left(r) = Y$,
          add $\langle r, 0, k, right, \varepsilon\rangle$ to S_{m+1}.
    [1.2.2] if $j = \lambda(q)$ and $right_j(q) \in V_T$ then for each rule ξ
    such that $left(q) = left(\xi)$ and $right(q) = \rho X$ and $right(\xi) = \rho Y$,
      for each rule r such that $left(r) = Y$,
        add $\langle r, 0, m, right, \varepsilon\rangle$ to S_{m+1}.

[2] **Scan:**
if $t \neq cut$ then:
  [2.1] If $i \neq m$ and $i < m + p + 1$ and $j < \lambda(q)$ and $right_{j+1}(q) \in cat(u_{i+1})$

then
[2.1.1] add $\langle q, j+1, k, t, \alpha\rangle$ to S_{i+1}.
[2.1.2] If $i = m-1$ and $t = right$ and $j = \lambda(q) - 1$ and $j \neq 0$ then
for each rule ξ such that $left(q) = left(\xi)$ and $right(q) = \rho X$
and $right(\xi) = \rho Y$,
if $Y \in cat(u_{m+2})$ then add $\langle \xi, j+1, k, t, \alpha\rangle$ to S_{m+2}.
[2.2] If $i = m$ and $t = right$ and $j = \lambda(q) - 1$ and $j \neq 0$
and $right_{j+1}(q) \in cat(u_{m+2})$ then
add $\langle q, j+1, k, t, \alpha\rangle$ to S_{m+2}.
[2.3] If $i = m$ and $j > 0$ then
for each $\langle q', j', k', right, \beta\rangle \in S_k$ such that $right_{j'+1}(q') = left(q)$
and $j' = \lambda(q') - 1$,
for each rule ξ such that $left(q') = left(\xi)$ and $left(q') = \rho X$
and $right(\xi) = \rho Y$
if $Y \in cat(u_{m+2})$ then add $\langle \xi, j'+1, k', right, \beta\rangle$ to S_{m+2}.

[3] **Complete:**
[3.1] If $j = \lambda(q)$ and $t \neq cut$ then
for each $\langle q', j', k', t', \beta\rangle \in S_k$ such that $right_{j'+1}(q') = left(q)$,
add $\langle q', j'+1, k', t', \beta q\alpha\rangle$ to S_i.
[3.2] If $i = m$ and $\lambda(q) \geq j > 0$ then
for each $\langle q', j', k', t', \beta\rangle \in S_k$ such that $right_{j'+1}(q') = left(q)$.
[3.2.1] If $t \neq cut$ and $right_j(q) \in V_T$ then
add $\langle q', j'+1, k', cut, \beta q[j]\alpha\rangle$ to S_m.
[3.2.2] If $t = cut$ then
add $\langle q', j'+1, k', cut, \beta q\alpha\rangle$ to S_m.

If $\langle 0, 1, 0, right, \alpha\rangle$ belongs to S_{m+1+p} then a new tree N is given as α. If $\langle 0, j, 0, cut, \beta\rangle$ belongs to S_m then an interrupted tree O is given as β. The tree β represents the substring $u_1 u_2 \ldots u_m$. Note that if the grammar is left-recursive, it may happen that there are an infinite number of interrupted trees.

5.4 *Remarks*

The algorithm can be easily extended to handle cascaded repairs. A simpler version[13] should be used to provide the inputs for the understanding com-

[13] Without *cut* as a value of type in the definition of a state and without [3.2] in the definition of Complete.

ponent: it only yields the repaired trees. The repaired trees are the relevant inputs for building discourse units, i.e., sentential turn constructional units in conversation[14]. For example, they allow the possible completion of the current turn and make transition to the next turn possible.

6. Conclusion

The main result of the study is the following: even though the configuration "interrupted utterance + repairs(s)" does not belong to the syntactic repertoire of French, it is submitted to a syntactic well-formedness condition. The REP is a simple and unified account of the regularity of self-repair. It comes into line with Schegloff's observation (1979:277): "the effect [of successful repair] is the resumption of the turn-unit before the repair initiation or, if the repair operation involves reconstruction of the whole turn-unit, production of the turn-unit to completion". It gives a more adequate content to Levelt's claim: "speakers repair in a linguistically principled way". Thanks to the augmentation of the Earley algorithm, we claim that parsers can parse repairs in a syntactically principled way.

REFERENCES

Blanche-Benveniste, Claire. 1987. "Syntaxe, choix de lexique, et lieux de bafouillage". *DRLAV* 36-37.123-157. Paris: Université de Paris-VIII.

Cori, Marcel., M. de Fornel & J.-M. Marandin. 1995. "Analyse syntaxique de l'auto-réparation". *Colloque Le Traitement Automatique du Langage Naturel*, 209-219. Marseille.

De Smedt, Koenraad. & G. Kempen. 1987. "Incremental sentence production, self-correction and coordination". *Natural Language Generation* ed. by G. Kempen, 365-376. Dordrecht: Kluwer Academic.

Dowty, David. 1988. "Type raising, functional composition and non-constituent conjunction". *Categorial Grammars and Natural Language Structures* ed. by R. Oehrle et al., 153-197. Dordrecht: D. Reidel.

Earley, Jay. 1970. "An Efficient Context-Free Parsing Algorithm", *Communications of the Association for Computing Machinery* 13:2.94-102.

Fornel, Michel de. 1992a. "The Return Gesture: Some Remarks on Context, Inference, and Iconic Gesture". *The Contextualization of Language* ed. by P. Auer & A. Di Luzio, 159-176. Amsterdam: John Benjamins.

[14]See Schegloff (1979).

——. 1992b. "De la pertinence du geste dans les séquences de réparation". *Les formes de la conversation* ed. by Conein et al., 119-154. Paris: CNET/CNRS.

—— & J.M. Marandin. Forthcoming. "L'analyse grammaticale des auto-réparations".

Frederking, Robert. 1988. *Integrated Natural Language Dialog*. Dordrecht: Kluwer.

Fromkin, Victoria. 1973. *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Gardent, Claire. 1991. *Gapping and VP Ellipsis in a Unification-Based Grammar*. Ph.D. dissertation, University of Edinburgh, Edinburgh. Scotland.

Hankamer, Jorge. 1973. "Unacceptable Ambiguity". *Linguistic Inquiry* 4.17-68.

Hindle, Donald. 1983. "Deterministic Parsing of Syntactic Non-fluencies". *Proceedings of the 21st Meeting of the Association of Computational Linguistics (ACL'83)*, 123-128.

Kay, Martin, M. Gawron & P. Norvig. 1993. "Verbmobil: A Translation System for Face-to-face Dialog", (= *CLSI Lectures Notes*, 33). Chicago: Chicago University Press.

Kroch, Anthony & D. Hindle. 1982. "On the Linguistic Character of Non-Standard Input". *Proceedings of the 20th Meeting of the Association of Computational Linguistics (ACL'82)*, 161-163.

Levelt, William. 1983. "Monitoring and Self-Repair in Speech". *Cognition* 14.41-104.

——. 1989. *Speaking: from Intention to Articulation*. Cambridge, Mass.: MIT Press.

Prüst, Hub. 1993. *On Discourse Structuring, VP Anaphora and Gapping*. Ph.D. dissertation, University of Amsterdam.

Sag, Ivan, Th. Wasow, G. Gazdar & S. Weisler. 1985. "Coordination and how to distinguish categories". *Natural Language and Linguistic Theory* 3:2.117-171.

Schegloff, Emanuel. 1979. "The relevance of repair to syntax-for-conversation" *Syntax and Semantics 12*. ed. by T. Givon, 261-286. New-York: Academic Press.

——, G. Jefferson & H. Sacks. 1977. "The preference for self-correction in the organization of repair in conversation". *Language* 2.361-382.