

# Analyse linguistique détaillée pour la Compréhension Automatique de la Parole spontanée

Jérôme Goulian

VALORIA, EA 2593

Université de Bretagne Sud

Site de Tohannic, rue Yves Mainguy, 56000 Vannes, France

jerome.goulian@univ-ubs.fr

---

## Résumé

Un des enjeux de la CHM orale, dès qu'elle aura quitté le champ d'investigation du dialogue fortement finalisé, semble être de pouvoir allier robustesse (face aux spécificités de l'oral), efficacité et couverture de la langue. Cet article tente de montrer qu'une analyse linguistique détaillée peut être menée tout en respectant la contrainte de robustesse imposée. Le système, proposé comme une alternative aux méthodes sélectives, repose tout d'abord sur l'exploitation du pouvoir structurant de la syntaxe au niveau de constituants minimaux non récursifs (*chunks*). Une recherche des relations de *dépendances* entre les têtes lexicales associées à ces unités peut être ensuite envisagée à un niveau sémantico-pragmatique.

---

## 1. Introduction

Au cours de ces dernières années, les systèmes de Communication Homme-Machine orale ont connu des développements et des succès significatifs. On voit ainsi apparaître des systèmes opérationnels qui permettent un dialogue interactif avec l'utilisateur via le téléphone. Ces systèmes concernent des applications aussi diverses que la réservation ferrovière par téléphone (mise en service par les chemins de fer néerlandais suite au projet européen ARISE (Baggia *et al.*, 1999)) ou encore le routage téléphonique grand public (mis en place en 1992 par AT&T (Lokbani & White, 1999)). Les réussites évoquées ci-dessus reposent principalement sur les progrès significatifs qu'a connus la Reconnaissance Automatique de la Parole, où l'on dispose désormais d'approches statistiques à la robustesse très appréciable. Nous nous intéressons ici aux niveaux linguistiques supérieurs de tels systèmes et plus précisément au module de *Compréhension* du langage parlé. Ce module a pour but de construire une représentation sémantique de l'énoncé prononcé, à partir de la séquence de mots (ou plus précisément de la liste des *n*-meilleures hypothèses ou du treillis de mots) issue du module de Reconnaissance. Cette représentation est destinée à être complétée et utilisée par le module de gestion du dialogue.

Le caractère spontané de la langue parlée induit l'apparition d'inattendus structurels (hésitations, répétitions, corrections, inachèvements...) (Blanche-Benveniste *et al.*, 1990) qui cassent la

régularité syntaxique des énoncés. La compréhension de la parole ne peut donc pas s'envisager aisément sous la forme d'une analyse linguistique détaillée. Une alternative utilisée en CHM orale consiste à profiter du caractère finalisé du dialogue pour développer des approches orientées par la tâche. Il s'agit de méthodes *sélectives* qui limitent la compréhension à la recherche de séquences clés (*îlots* ou *segments conceptuels*) dans l'énoncé. Le sens de l'énoncé est alors représenté par une structure de *frame* regroupant un ensemble de rôles prédéfinis instanciés par les segments détectés.

Le principal intérêt de ces méthodes est qu'elles permettent de faire ressortir facilement l'information (dite utile) d'un énoncé tout en ignorant de nombreux problèmes d'agrammaticalité. Le caractère partiel de cette analyse garantit a priori une certaine robustesse d'analyse quelle que soit la modélisation adoptée. Celle-ci peut reposer sur l'utilisation d'une base de règles (utilisation de grammaires hors-contextes augmentées de traits sémantiques –système TINA du MIT (Seneff, 1992)– ou encore de grammaires sémantiques (Bennacef *et al.*, 1994)) ou sur une modélisation stochastique (Levin & Pieraccini, 1995). Ces approches minimalistes (d'un point de vue linguistique) ont donné de très bons résultats pour des applications dédiées à des tâches très spécifiques (renseignement et réservation aérienne et plus généralement consultation de bases de données). Peu de travaux portent en revanche sur des domaines à plus large couverture. Or, comme le souligne (Van Noord *et al.*, 1998), dans la mesure où l'augmentation du domaine d'application va de pair avec des productions orales grammaticalement plus complexes, nous pensons qu'une analyse linguistique détaillée devient essentielle pour le traitement correct des énoncés, notamment pour certains phénomènes, comme les ellipses. De la même manière, notre étude du corpus PARISCORP (Bonneau-Maynard & Devillers, 1998), portant sur le domaine relativement large des renseignements touristiques a montré par exemple que les modaux et les adverbes peuvent être utilement pris en compte pour répondre aux besoins d'une Interaction Homme-Machine naturelle et réellement collaborative.

Dans la section suivante, nous décrivons et proposons, comme alternative aux méthodes sélectives, un système de compréhension composé de deux modules principaux s'appuyant sur des techniques d'analyse structurelle robustes et des formalismes développés pour le TALN.

## 2. Compréhension de la parole et analyse linguistique détaillée

### 2.1. Architecture du système

Le système proposé s'articule autour de deux analyses successives :

- Une première analyse majoritairement **syntactique**. Cette analyse **partielle** ne cherche pas à porter de jugement syntaxique global sur la totalité de l'énoncé mais doit simplement permettre un découpage en groupes syntaxiques minimaux non récursifs (GN, GP, ...) appelés *chunks* (Abney, 1991).
- Une deuxième analyse permettant de rattacher les différents constituants minimaux issus de la première passe. Cette analyse dépend avant tout d'une information lexicale plutôt que de la catégorie syntaxique des constituants ; elle recherche les **dépendances** entre les têtes lexicales associées aux constituants à un niveau **sémantico-pragmatique**. Une étape intermédiaire d'étiquetage sémantique à un niveau lexical des groupes syntaxiques issus de la première analyse est donc requise.

Un exemple complet d'analyse d'un énoncé est donné section 3 pour illustrer plus précisément les actions réalisées par chacune de ces étapes, détaillées ci-après.

## 2.2. Analyse syntaxique partielle en chunks

### 2.2.1. Objectifs

L'analyse syntaxique partielle est de notre point de vue un préalable essentiel à l'analyse sémantique détaillée de l'énoncé. Les méthodes sélectives sont basées sur le constat qu'il est inutile de s'attacher à une grammaticalité jugée illusoire en situation de dialogue oral. Or les études linguistiques de corpus oraux (Blanche-Benveniste *et al.*, 1990) attestent cependant que les inattendus structuraux de l'oral présentent des régularités sur lesquelles une analyse automatique peut utilement se reposer (Goulian, 1998). On peut noter par exemple que si les énoncés, les mots ou les groupes de mots peuvent rester inachevés à n'importe quel instant de la production orale, la répétition ou la reprise reprend toujours au début du syntagme avorté ou enrichi<sup>1</sup>. Utiliser le pouvoir structurant de la syntaxe au niveau local des syntagmes, sans chercher à effectuer de rattachements syntaxiques entre eux, nous semble ainsi une manière de réaliser une analyse linguistique riche de l'énoncé tout en préservant une robustesse appréciable. L'avantage offert par ailleurs par cette analyse syntaxique partielle est d'être suffisamment générale pour ne pas être trop dépendante d'une application particulière. En effet, un nombre réduit d'étiquettes syntaxiques, indépendantes de l'application, et un jeu de règles d'une grammaire hors-contexte sont suffisants pour caractériser les relations de sous-catégorisation nécessaires à la modélisation de la structure syntaxique du syntagme.

### 2.2.2. Définition des chunks

Nous proposons d'adapter la définition des *chunks* donnée par Abney (Abney, 1991) au cadre du dialogue oral. Ces "constituants minimaux non récursifs" définis par Abney, correspondent grossièrement à la notion linguistique de constituant si ce n'est qu'aucun attachement n'est effectué. Par exemple les chunks verbaux n'incluent jamais leurs arguments mais se contentent d'identifier la complexité de la structure verbale (chunk verbal infinitif, interrogatif, etc...). Nous proposons ainsi sept groupes de chunks dont le dernier est spécifique au dialogue oral :

- GV (Groupes Verbaux) : GV<sub>i</sub> (interrogatif), GV<sub>inf</sub> (infinitif), GV<sub>m</sub> (modaux).
- GAdj (Groupes Adjectivaux) : tout adjectif ou groupe gouverné par un adjectif ("les plus chers" par exemple) ;
- GN (Groupes Nominaux) : tous les groupes nominaux classiques, intégrant les groupes adjectivaux immédiats (i.e. qui les suivent dans l'ordre syntagmatique) ainsi que les déterminants et les adjectifs numériques ("deux places", "les repas complets" seront étiquetés GN), les pronoms personnels font aussi partie de cette catégorie ;
- GP (Groupes Prépositionnels) : tout groupe nominal gouverné par une préposition, y compris les compléments du nom ;
- Gadv (Groupes Adverbiaux) : tous les adverbes (jamais intégrés dans d'autres chunks à ce niveau de l'analyse) ;
- COO (Coordinations) : les conjonctions de coordination "et", "ou", etc. ;
- COR, HES : une partie des inattendus structuraux de l'oral (répétitions, reprises, corrections, hésitations) ; au cours de l'analyse, ces inattendus peuvent être inclus<sup>2</sup> dans des chunks spécifiques qui seront traités lors de la seconde étape. On aura par exemple des

---

1. Les enrichissements lexicaux, très fréquents à l'oral, se caractérisent par l'instanciation de plusieurs syntagmes à la même position syntaxique (place dans l'ordre syntagmatique). L'exemple suivant illustre ce phénomène : "je cherche *un restaurant un restaurant indien un restaurant assez bon marché*".

2. s'ils ne sont pas ignorés : il s'agit d'une analyse partielle

chunks marquant les reprises (COR) pour des structures comme “non”, “pardon”, etc. et des chunks marquant les hésitations (HES). Ce choix de ne pas passer par une phase de prétraitement de ces inattendus, mais de les intégrer dans des chunks particuliers, nous semble judicieux pour pouvoir marquer, dans la mesure du possible, ces phénomènes et aider ainsi à leur traitement ultérieur, notamment pour les enrichissements lexicaux.

### 2.2.3. *Mise en oeuvre*

De nombreuses méthodes récentes d'analyse structurelle de surface (shallow parsing) ont démontré leur capacité à analyser de façon robuste des textes portant sur des domaines relativement larges (Abney, 1996) (Aït-Mokhtar & Chanod, 1997). En particulier, des systèmes comme FASTUS (Hobbs *et al.*, 1997), ont montré que les expressions régulières compilées en automates à états finis peuvent permettre une analyse linguistique significative (ou du moins suffisante au niveau des constituants) efficace et rapide de ce type de textes. Notre analyse résulte de l'application d'une séquence finie et ordonnée de transducteurs. Chaque transducteur est utilisé pour introduire (dans la chaîne d'entrée étiquetée) des marqueurs de délimitation autour des instances d'un chunk particulier. Les chunks sont décrits par un ensemble de règles exprimées au moyen d'expressions régulières. Ces expressions intègrent un opérateur spécifique exprimant une relation de remplacement contrainte par la direction de l'analyse (ici gauche-droite) et favorisant la détection des instances maximales des expressions (Karttunen *et al.*, 1996) (Karttunen, 1996). Nous utilisons, pour la définition de cet opérateur et la compilation des expressions régulières, l'utilitaire Fsa (Van Noord, 1997).

A titre d'exemple, l'énoncé suivant “*Quels sont les restaurants indiens euh plutôt pas chers près de la gare non pardon près de l'hôtel Caumartin*” conduira au découpage (présenté de manière simplifiée ici) : “GVi [quels sont] GN [les restaurants GAdj (indiens)] HES [euh] Gadv [plutôt] GAdj [pas chers] GP [près de GN (la gare)] COR [non pardon] GP [près de GN (l'hôtel Caumartin)]”.

### 2.3. *Étiquetage sémantique des têtes lexicales associées aux chunks*

L'étiquetage sémantique autour d'une tête lexicale des chunks est une étape indispensable pour leur rattachement sur des critères sémantico-pragmatiques. Il faut noter que chacun de ces groupes syntaxiques, selon son contexte, correspond à une **unité conceptuelle unique** de la tâche considérée. Ainsi un GN correspondra à un *objet* du domaine (noté Obj), un GP à une *propriété* portant sur un objet (notée Pte), etc. A titre d'exemple, le GP “*près de l'hôtel Caumartin*” sera étiqueté autour de la tête “*près de*” comme étant une propriété de type *Loc* (localisation) et ayant comme argument l'objet référent de type *hôtel* et de nom *Caumartin*. Ces étiquettes sont issues de l'étude du corpus PARISCORP (Bonneau-Maynard & Devillers, 1998) et ont été validées sur 1000 énoncés.

### 2.4. *Rattachement sémantico-pragmatique*

Cette étape réalise une analyse lexicalisée qui a pour but de :

- caractériser les relations de *dépendances* entre chunks, qui vont correspondre aux relations prédicats-arguments de la représentation sémantique finale. La résolution des ambiguïtés laissées par l'analyse en chunks et cette recherche de dépendances s'effectuent sur des critères sémantico-pragmatiques.
- analyser et traiter de manière robuste les phénomènes de l'oral. Une analyse partielle doit, là encore, rester possible.

Nous proposons d'adapter pour ce module les *Link Grammars* (Sleator & Temperley, 1991). En envisageant la structure de l'énoncé uniquement selon les *relations* diverses que ses éléments entretiennent entre eux, ce formalisme lexicalisé, inspiré des *grammaires de dépendances*, garantit une certaine souplesse d'analyse (Goulian, 1998). Les travaux de (Grinberg *et al.*, 1995) ont par ailleurs montré la capacité des Link Grammars à analyser de façon robuste des corpus de dialogues oraux. Dans ce formalisme, les liens correspondent à des connecteurs étiquetés (ici par des informations pragmatiques issues de l'analyse de corpus) qui peuvent s'accrocher à droite ou à gauche de la tête lexicale du mot considéré. La contrainte forte sur laquelle repose une analyse avec ce formalisme est celle de *planarité* i.e. deux arcs ne peuvent pas se croiser. Cette contrainte ne semble pourtant pas gênante pour une analyse du français parlé, langage à ordre peu variable (Antoine & Goulian, 1999).

### 3. Exemple d'analyse complète d'un énoncé

Nous illustrons schématiquement dans cette section les étapes d'analyse successives de l'énoncé (1) conduisant à la représentation sémantique finale donnée par (Fig. 1)

(1) *Quels sont les horaires d'ouverture non de fermeture du Louvre et du musée comment déjà du musée Grévin*

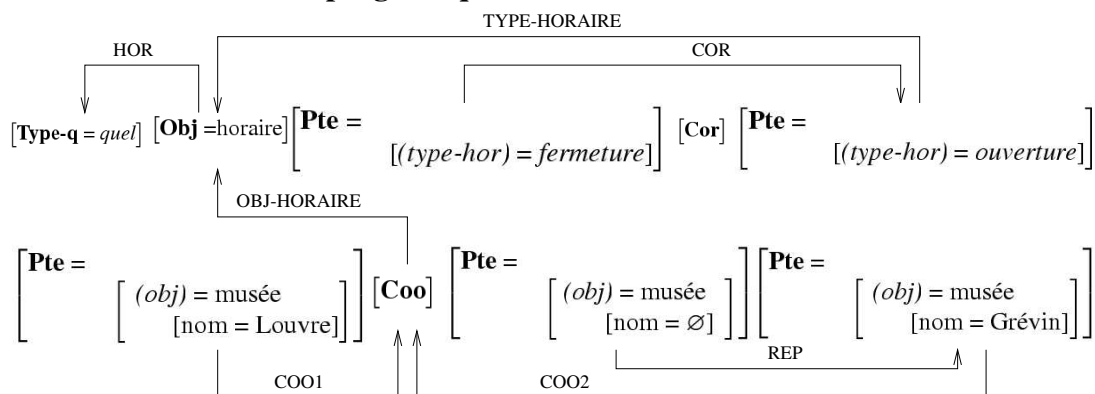
#### 1. analyse en chunks :

$$[Quels\ sont]_{GV_i} [les\ horaires]_{GN} [d'ouverture]_{GPN_{om}} [non]_{COR} [de\ fermeture]_{GPN_{om}} [du\ Louvre]_{GPN_{om}} [et]_{COO} [du\ musée]_{GPN_{om}} comment\ déjà [du\ musée\ Grévin]_{GPN_{om}}$$

#### 2. étiquetage sémantique des têtes lexicales des chunks :

$$\begin{array}{l} [\text{Type-q} = \text{quel}] [\text{Obj} = \text{horaire}] \left[ \begin{array}{l} \text{Pte} = \\ [(type-hor) = \text{fermeture}] \end{array} \right] [\text{Cor}] \left[ \begin{array}{l} \text{Pte} = \\ [(type-hor) = \text{ouverture}] \end{array} \right] \\ \left[ \begin{array}{l} \text{Pte} = \\ [(obj) = \text{musée} \\ [\text{nom} = \text{Louvre}]] \end{array} \right] [\text{Coo}] \left[ \begin{array}{l} \text{Pte} = \\ [(obj) = \text{musée} \\ [\text{nom} = \emptyset]] \end{array} \right] \left[ \begin{array}{l} \text{Pte} = \\ [(obj) = \text{musée} \\ [\text{nom} = \text{Grévin}]] \end{array} \right] \end{array}$$

#### 3. rattachement sémantico-pragmatique des chunks :



### 4. Perspectives

Cet article a tenté de montrer qu'une analyse linguistique détaillée peut être menée pour la CAP, en adaptant des méthodes et des formalismes du TALN. L'analyse syntaxique partielle proposée est à l'heure actuelle en cours de développement <sup>3</sup>.

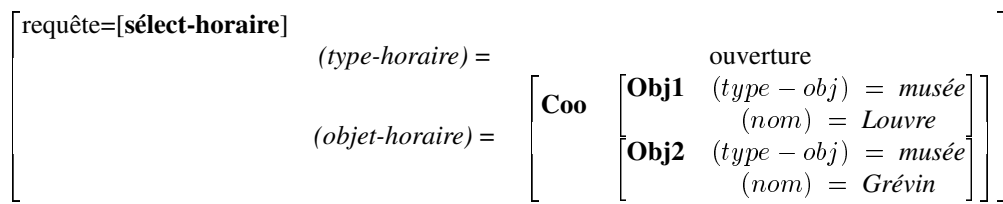


FIG. 1 – Représentation sémantique finale de l'énoncé (1)

## Références

- ABNEY S. (1991). Parsing by chunks. In *Principle Based Parsing*. In R. Berwick, S. Abney and C. Tenny, Eds., Kluwer Academic Publishers.
- ABNEY S. (1996). Partial parsing via finite-state cascades. In J. CARROLL, Ed., *Workshop on Robust Parsing ESSLLI'96*, p. 8–15.
- ANTOINE J. & GOULIAN J. (1999). Le français parlé spontané est-il un langage à ordre variable? In *Actes des Journées Internationales de Linguistique Appliquée, JILA'99, Nice, France*.
- AÏT-MOKHTAR S. & CHANOD J.-P. (1997). Incremental finite-state parsing. In *Procs. of ANLP'97, Washington*, p. 72–79.
- BAGGIA P., KELLNER A., PERRENOU G., POPOVICI C., STURM J. & WESSEL F. (1999). Language modelling and spoken dialogue systems - the arise experience. In *Actes Eurospeech'99, Budapest*.
- BENNACEF S., BONNEAU-MAYNARD H., GAUVAIN J., LAMEL L. & MINKER W. (1994). A spoken language system for information retrieval. In *Procs. of ICSLP'94, Yokohama, Japon*.
- BLANCHE-BENVENISTE C., BILGER M., ROUGET C. & VAN DEN EYNDE K. (1990). *Le français parlé; études grammaticales*. CNRS Editions, Paris.
- BONNEAU-MAYNARD H. & DEVILLERS L. (1998). *Acquisition, Transcription et Annotation du Corpus Pariscorp*. Rapport interne, Action de Recherche Concertée "Dialogue Oral" (Arc B2) de l'AUF.
- GOULIAN J. (1998). Analyse robuste du français parlé. Mémoire de D.E.A., I.N.P.G, Université Joseph Fourier Grenoble I.
- GRINBERG D., LAFFERTY J. & SLEATOR D. (1995). *A robust parsing algorithm for Link Grammars*. Rapport interne, CMU-CS-TR-95-125, CMU, USA.
- HOBBS J., APPLLET D., BEAR J., ISRAEL D., KAMEYAMA M., STICKEL M. & TYSON M. (1997). Fastus: A cascaded finite-state transducer for extracting information from natural-language text. In ROCHE & SCHABES, Eds., *Finite State Devices for Natural Language Processing*. MIT Press, MA.
- KARTTUNEN L. (1996). Directed replacement. In *34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz.
- KARTTUNEN L., CHANOD J.-P., GREFFENSTETTE G. & SCHILLER A. (1996). Regular expressions for language engineering. *Natural Language Engineering*, 2(4), 305–328.
- LEVIN E. & PIERACCINI R. (1995). Concept-based spontaneous speech understanding system. In *Procs. of Eurospeech'95, Madrid, Spain*, p. 555–558.
- LOKBANI M. & WHITE S. (1999). La reconnaissance de la parole. *La recherche*, 319, 82.
- SENEFF S. (1992). Tina: a natural language system for spoken language applications. *Computational Linguistics*, 18(1), 61–86.
- SLEATOR D. & TEMPERLEY D. (1991). *Parsing English with a Link Grammar*. Rapport interne, CMU-CS-91-196, CMU, USA.
- VAN NOORD G. (1997). Fsa utilities: A toolbox to manipulate finite-state automata. In D. RAYMOND, D. WOOD & S. YU, Eds., *Automata Implementaton*, p. 87–108. Springer Verlag.
- VAN NOORD G., BOUMA G., KOELING R. & NEDERHOF M.-J. (1998). Robust grammatical analysis for spoken dialogue systems. *Natural Language Engineering*, 1, 1–48.