Compréhension Automatique de la Parole et TAL : une approche syntaxico-sémantique pour le traitement des inattendus structuraux du français parlé

Jérôme Goulian, Jean-Yves Antoine, Franck Poirier
VALORIA , EA2593 – Université de Bretagne Sud
Site de Tohannic, rue Yves Mainguy, 56000 Vannes
jerome.goulian@univ-ubs.fr

Résumé - Abstract

Dans cet article, nous présentons un système de Compréhension Automatique de la Parole dont l'un des objectifs est de permettre un traitement fiable et robuste des inattendus structuraux du français parlé (hésitations, répétitions et corrections). L'analyse d'un énoncé s'effectue en deux étapes : une première étape générique d'analyse syntaxique de surface suivie d'une seconde étape d'analyse sémantico-pragmatique, dépendante du domaine d'application et reposant sur un formalisme lexicalisé : les grammaires de liens. Les résultats de l'évaluation de ce système lors de la campagne d'évaluation du Groupe de Travail *Compréhension Robuste* du GDR I3 du CNRS nous permettent de discuter de l'intérêt et des limitations de l'approche adoptée.

This paper discusses the issue of how a speech understanding system can be made robust against spontaneous speech phenomena (hesitations and repairs). We present a spoken French understanding system. It implements speech understanding in a two-stage process. The first stage achieves a finite-state shallow parsing that consists in segmenting the recognized sentence into basic units (spoken-adapted *chunks*). The second one, a Link Grammar parser, looks for inter-chunks dependencies in order to build a rich representation of the semantic structure of the utterance. These dependencies are mainly investigated at a pragmatic level through the consideration of a task concept hierarchy. Discussion about the approach adopted is based on the results of the system's assessment in an evaluation campaign held by the CNRS.

Mots-clefs – Keywords

communication orale homme-machine, compréhension automatique de la parole, répétitions, corrections, analyse syntaxique partielle, grammaires de liens.
spoken man-machine dialog, speech understanding, repairs, shallow parsing, link grammars.

1 Introduction

Parmi les problèmes auxquels sont confrontés les systèmes de traitement automatique de la langue parlée se trouvent incontestablement les inattendus structuraux dus au caractère spontané des productions orales. D'un point de vue purement descriptif, citons parmi ces phénomènes :

- Les hésitations. Ce sont les manifestations les plus courantes des recherches de dénomination. Les plus simples vont de la courte pause dans la prononciation à des pauses plus longues comblées par des interjections (*euh*) ou des appuis du discours (*donc*, *alors*).

- Les répétitions et les corrections. Qu'il s'agisse pour le locuteur de préciser sa pensée ou de se corriger, ces phénomènes se caractérisent par un phénomène d'entasssement paradigmatique (Blanche-Benveniste, 1990), illustré par l'exemple suivant :

  je cherche *un*
  un restaurant *euh*
  un restaurant français *près de*
  près de la gare

  On en distingue deux types suivant la présence ou non d'éléments (interjections ou appuis du discours) entre le ou les groupes de mots répétés ou repris. Notons par ailleurs que l'enrichissement successif d'un lexème préalablement prononcé, est un phénomène de recherche de dénomination fréquemment utilisé.

Le traitement de ces phénomènes oraux s'est envisagé, à la suite des travaux de Hindle (Hindle, 1983), par des étapes de pré-analyse, ayant pour objectifs la détection puis le filtrage ou la correction automatique des reprises et répétitions orales. Ont ainsi été développés des techniques robustes de détection automatique de motifs fondés sur les mots et leurs catégories syntaxiques (Bear *et al.*, 1992) ou des modèles de langage stochastiques spécifiques prenant en compte marqueurs lexicaux et informations acoustiques (Heeman & Allen, 1999). Ces approches ont donné d'assez bons résultats pour la détection d'une grande majorité de ces phénomènes ; cependant, leurs concepteurs reconnaissent la nécessité d'une intervention plus importante d'informations syntaxiques et sémantiques pour d'une part éviter la surgénérativité et l'élimination abusive de structures informatives sciemment utilisées par le locuteur[1], et d'autre part assurer un traitement correct de l'ensemble de ces phénomènes. Cet article présente un système de compréhension de la parole, *ROMUS*, dans le cadre d'une communication homme-machine finalisée portant sur le renseignement touristique. Ce système n'intègre pas d'étape de pré-traitement de ces phénomènes. Leur traitement est envisagé au cours du processus d'analyse du sens de l'énoncé ; processus qui prend en compte des connaissances syntaxiques et sémantiques.

2 Le système ROMUS

Notre système de compréhension adopte une démarche identique à celle des outils d'analyse syntaxique robuste développés pour le TAL. Ces systèmes s'articulent en général en deux étapes : une première étape repérant des structures minimales et une seconde étape calculant des structures ou relations plus complexes (Ejerhed, 1993). Nous adaptons ainsi le principe de *chunking* (i.e. segmentation en constitutants minimaux non-récursifs (Abney, 1991)) à l'analyse des énoncés oraux. Cette première étape repose uniquement sur des informations grammaticales et syntaxiques encodées au moyen d'automates à états finis. Elle est donc entièrement portable d'une application à une autre ; notre système répondant ainsi en partie à la question de la généricité des systèmes actuels de compréhension de la parole (Hirschman, 1998). Les segments extraits sont ensuite mis en relation les uns avec les autres par une analyse reposant sur une grammaire de liens. Les dépendances entre segments sont cette fois principalement envisagées à un niveau sémantico-pragmatique[2].

2.1 Segmentation syntaxique partielle

Etiquetage Chaque mot de l'énoncé fourni par le module de reconnaissance de la parole reçoit préalablement une étiquette correspondant à sa catégorie grammaticale. Le lexique est représenté

[1] C'est le cas par exemple des enrichissements lexicaux. Un exemple en est donné (Bear *et al.*, 1992) par la structure suivante *flights daily flihts* dans l'énoncé *show me flights daily flights*, qui peut être faussement détectée comme une répétition suivant le motif (M1 X M1).

[2] Nous faisons ici référence au domaine de l'application et non à la prise en compte du dialogue.

sous forme d'un automate à états finis déterministe permettant un temps d'accès réduit et un faible volume de stockage. Il comporte actuellement environ 45000 mots. Un ensemble réduit de 25 étiquettes grammaticales est utilisé pour la segmentation. Certaines informations morphologiques encodées dans le lexique, comme le temps des verbes, sont néanmoins conservées. L'ambiguïté lexicale est gérée par l'application de règles de désambiguïsation contextuelles encodées au moyen de transducteurs. Une attention particulière a été portée à la désambiguisation d'expressions fréquentes à l'oral (le mot *quoi* par exemple, employé comme interjection). Toutefois, une trop forte désambiguïsation à cette étape risque d'entraîner des erreurs très pénalisantes pour la suite de l'analyse. La version actuelle de notre système présente un compromis acceptable (taux de décision de 86.4%[3] contre 95% pour l'étiqueteur utilisé au moment de l'évaluation). En cas d'ambiguïté, les différentes séquences sont conservées.

Segmentation Outre les chunks classiques (chunks nominaux, chunks verbaux, etc.), nous avons choisi de caractériser deux types de segments minimaux : les expressions langagières spécifiques et indépendantes de l'application (date, heure, prix), et les segments constitués des *marques* (interjections, appuis du discours) des inattendus structuraux de l'oral. Chacun de ces segments peut être décrit par des expressions régulières mettant en jeu les étiquettes grammaticales précédemment évoquées. Ces expressions sont compilées en transducteurs que l'on rend déterministes (Roche & Schabes, 1997). Chacun de ces transducteurs est utilisé en cascade pour introduire dans l'énoncé des marqueurs de délimitation autour de ces groupes. La tête lexicale des segments grammaticaux classiques est également marquée lors de cette étape. L'ambiguïté de segmentation est gérée par l'heuristique de maximisation des segments détectés. Chacune des séquences de l'étape de pré-étiquetage est ainsi segmentée et conservée pour être traitée en parallèle par les niveaux suivants de l'analyse. À l'intérieur d'une séquence, les mots non intégrés dans des groupements sont éliminés.

[je*\$_{pr}$] [cherche*\$_{vb-conj,présent}$] un\$_{art-indéfini}$ [un\$_{art-ind}$ restaurant*\$_n$] [euh\$_{hes}$] [un\$_{art-ind}$ restaurant*\$_n$] [français*\$_{adj}$] près-de\$_{prep}$ [près-de*\$_{prep}$ la\$_{art-def}$ gare\$_n$]

Figure 1: Segmentation de l'énoncé *"je cherche un un restaurant euh un restaurant français près de près de la gare*". Les segments sont représentés entre crochets. Les étiquettes utilisées pour la segmentation sont indiquées après l'anti-slash. La tête lexicale est marquée par*.

2.2 Analyse sémantico-pragmatique par grammaire de liens

Extraction des dépendances locales Afin de pouvoir envisager le rattachement de ces différents segments selon des critères sémantico-pragmatiques, ceux-ci sont transformés en structures arborescentes qui rendent compte de manière compacte de toutes les informations extraites par l'analyse précédente. La racine de ces arbres est composée d'un triplet $< RS, T, M >$. RS est le Rôle Sémantique du groupement vis à vis de l'application ; ce rôle est dérivé, en première approximation, de la catégorie grammaticale du segment[4] : un groupe nominal aura le rôle sémantique d'*Objet*, un groupe adjectival celui de *Propriété*, etc. T correspond à la Tête lexicale du segment. Les groupes "*des hôtels*" et "*une pizzeria*" auront respectivement pour T les mots *hôtel* (la tête lexicale) et *restaurant* (une pizzeria n'est pour l'application qu'un établissement de restauration particulier). M est l'ensemble des propriétés Morphologiques qui pourront intervenir dans le rattachement des constituants. Les autres informations (la propriété *Spécialité : pizza* par exemple) sont exprimées sous forme de dépendances (feuilles ou noeuds de l'arbre). Elles n'interviendront pas dans le rattachement.

[3] La précision, évaluée sur un échantillon similaire de 1200 énoncés, est de 96.7%

[4] Cette catégorisation sémantique dépend en partie de l'application. Ainsi le groupe prépositionnel "*proche-de la gare*" se verra attribuer le rôle de "*Propriété-Localisation*" car le seul autre rôle de *proche-de* dans l'application, à savoir un ordre de grandeur sur un prix (proche-de 5 euros) aurait été préalablement étiqueté comme tel.

Chunk	RS	T	M	Dépendances locales
Verbal *voudrais savoir*	Acte-Modal	savoir	temps:conditionnel	Modalité vouloir
Nominal *une pizzeria*	Objet	restaurant	indéfini	Quantité — Spécialité 1 — pizza
Prépositionnel *près-de la gare*	Propriété-Localisation	près-de	défini	Obj-Loc gare
Prep. Heure *avant 15 heures*	Propriété-Heure	avant	$\emptyset$	Val-heure — Val-minute 15 — 0

Table 1: Triplets $< RS, T, M >$ et dépendances locales associées à différents segments.

Rattachement sémantico-pragmatique Il est le résultat d'une analyse lexicalisée caractérisant les relations de dépendances entre segments. Celles-ci vont correspondre aux relations prédicat-argument de la représentation sémantique finale. Notre analyse utilise une grammaire de liens (Sleator & Temperley, 1991). Les liens entre segments, représentés par les triplets $< RS, T, M >$, s'expriment au moyen de connecteurs étiquetés et orientés qui peuvent s'associer deux à deux pour former une relation valide. À chaque entrée du dictionnaire est associée une formule décrivant les relations autorisées entre l'élément concerné et les éléments situés à sa droite et à sa gauche dans l'énoncé. Les deux entrées suivantes signifient que l'élément E_2 doit entretenir une relation R avec un autre élément situé soit à sa gauche dans l'énoncé soit à sa droite (opérateurs $^-$ et $^+$ respectivement). La succession $-E_2\ E_1-$ est en ce sens valide car ces deux éléments (dans cet ordre) peuvent être reliés par R.

$$(E_1) < SR_1, T_1, P_1 >: R^-; \qquad (E_2) < SR_2, T_2, P_2 >: R^- \ or \ R^+;$$

On distingue deux types de relations principales :

- les relations spécifiques à l'application comme par exemple la relation `catégorie` qui peut lier les deux segments minimaux $< Objet, hotel, ind >$ et $< Objet, etoile, def >$;
- les relations exprimant des constructions syntaxiques génériques comme le traitement des coordinations, des relatives, des procédés d'extraction. Par exemple, l'entrée du dictionnaire correspondant à la coordination logique *et* est de la forme :

  $$< Coo, et, \emptyset >: (COO^-_{X_1} \ and \ COO^+_{X_2}) \ and \ X^-;$$

  où $X1, X2 \subset X$ dénotent des relations *Et-compatibles* au sens de X. Dans l'énoncé *je voudrais une chambre double et avec douche*, le marqueur *et* peut relier *double* et *avec douche* car les deux relations *Catégorie-chambre* (entre chambre et double) et *TypeBain-chambre* (entre chambre et douche) sont compatibles pour la coordination logique *et* selon la relation X englobante *Propriété-chambre*. Les entrées du dictionnaire de ces marqueurs sont générées automatiquement à partir des différents ensembles *Et*-compatibles.

Notre dictionnaire comporte environ 1000 entrées, rendant compte de 36 requêtes (demande de *tarif*, d'*horaire*, etc.) et 158 concepts (objets et propriétés de l'application).
L'algorithme d'analyse est fondé sur une exploration dynamique de tous les liens possibles calculés à partir de l'ensemble des disjonctions associées à chaque élément de l'énoncé. Il est gouverné par une contrainte de planarité. Cette analyse est partielle : elle autorise la formation d'îlots (au moins deux éléments reliés entre eux mais pas avec le reste de l'énoncé). Un système de coût permet de classer les différents graphes valides. Nous privilégions, par ordre

de coût décroissant, les analyses complètes, les analyses partielles avec îlots puis les analyses partielles avec des éléments isolés. En cas d'égalité, un heuristique privilégie les analyses ayant caractérisé le moins de dépendances longue-distance.

Construction de la représentation sémantique finale Elle découle du parcours du graphe non-orienté obtenu. Cette étape participe également au traitement des inattendus (cf. 3).

3 ROMUS et le traitement des inattendus structuraux

Répétitions et corrections de mots La segmentation syntaxique partielle de l'énoncé permet un premier traitement de quelques répétitions : les mots qui ne sont pas intégrés dans un segment à l'issue de cette étape sont éliminés. Ce "filtrage" n'est pas dû au hasard : chaque reprise d'un syntagme s'effectue systématiquement au début de celui-ci (déterminants et prépositions toujours repris). Il est donc justifié de ne pas tenir compte de ces amorces de syntagmes. Le fait de conserver les différentes séquences à l'issue de l'étiquetage grammatical permet de déléguer aux niveaux supérieurs de l'analyse le traitement correct de certaines structures particulières comme *un des hôtels*, qu'il faudrait comprendre comme un hôtel parmi plusieurs [5].

Répétitions et corrections non marquées portant sur des segments Leur traitement s'effectue en deux étapes. L'analyseur sémantique se contente d'identifier les relations sémantiques entre les segments en présence. La grammaire de liens utilisée permet d'exprimer le phénomène d'entassement paradigmatique au moyen d'un opérateur spécifique. Par exemple, le segment *une chambre* peut être relié avec plusieurs éléments selon la relation *TypeBain-chambre*[6]. Lorsqu'un élément est relié à plusieurs autres selon la même relation, il s'agit d'une répétition, d'une correction ou d'une énumération. Cette ambiguïté peut être levée lors du parcours du graphe :

- Si les triplets $< RS, T, M >$ des deux éléments sont équivalents, une répétition est identifiée. Dans le cas d'un enrichissement lexical, le parcours du graphe permet la fusion des propriétés des deux éléments si elles sont Et-compatibles.
- Si les relations en jeu ne sont pas Et-compatibles, une correction est identifiée. L'ordre des segments intervient alors : seul le second est conservé.

Ce traitement se fonde ainsi à la fois sur les informations extraites par l'étape syntaxique et sur des connaissances liées à l'application. Lorsqu'il n'est pas possible, l'ambiguïté est conservée dans la repésentation sémantique pour être levée le cas échéant par l'interprétation contextuelle.

Répétitions et corrections marquées portant sur des segments Le même mécanisme est utilisé. Toutefois, leur traitement est facilité par l'exploitation, dès la construction des liens, des marques des hésitations et des corrections. Celles-ci sont en effet traitées de la même manière que les coordinations. Pour que des liens soient bâtis, il faut que les relations en jeu soient compatibles selon ces différentes marques : *non*, *euh*, etc. Elles permettent par ailleurs une restriction des analyses possibles et l'instanciation plus aisée du phénomène concerné[7].

4 Évaluation

Une première version du système *ROMUS* a été évaluée lors de la campagne d'évaluation *"par défi"*, du groupe de travail 5.1 "compréhension robuste" du GDR I3 du CNRS[8]. Le but essentiel

[5] Dans ce cas, le mot *un* est étiqueté soit comme article soit comme adjectif numéral. Les qualificatifs associés à ce genre de structure (*parmi ..., dont vous ..., etc.*) permettent entre autre le classement prioritaire des analyses construites à partir de la seconde hypothèse.

[6] C'est ce même mécanisme qui permet également d'exprimer les objets de requêtes multiples. La compréhension de ce type de requêtes étant rarement intégrée dans les systèmes de compréhension actuels.

[7] La marque *euh* ne suffit toutefois pas seule à lever l'ambiguïté entre éumération et correction.

[8] GDR-PRC-I3, Pôle Parole, G.T. 5.1., http://www.univ-ubs.fr/valoria/antoine/Gt51/Eval_defi.html

de cette campagne d'évaluation est de pouvoir porter un diagnostic sur les systèmes des différents participants en regard des approches adoptées ; l'évaluation se veut donc essentiellement qualitative mais reste néanmoins objective (Antoine *et al.*, 2002). Chaque système a été évalué sur un ensemble de 1200 énoncés, intégrant des phénomènes linguistiques difficiles à traiter. La comparaison directe des résultats entre les différents systèmes participants n'est toutefois pas aisée ; chacun des systèmes adressant un domaine d'application différent[9]. La version de *ROMUS* testée a obtenu un taux global d'erreur de 15%. L'analyse des sources d'erreurs permet de dresser un premier bilan des limitations et avantages de l'approche adoptée.

- Avec 20% des erreurs constatées, le système a été très sensible aux erreurs de reconnaissance simulées, notamment dans les cas de substitution ou de suppression d'une préposition. Dans 75% des cas ces perturbations sont intervenues à l'intérieur d'un chunk, cassant ainsi sa régularité[10] sans que celui-ci soit entièrement repris dans le reste de l'énoncé comme c'est le cas généralement lors de perturbations dues au locuteur. En dehors de ces cas, la segmentation syntaxique est rarement à l'origine des erreurs observées.
- Le système a été particulièrement robuste face aux phénomènes de mouvement des constituants et aux structures complexes (requêtes multiples, extractions, etc.) avec moins de 5% d'erreurs sur chacun de ces phénomènes.
- Les résultats sont enfin prometteurs en ce qui concerne la tolérance face aux inattendus structuraux. 8% des énoncés qui en contenaient (sur 525) ont été mal compris. Ces erreurs, proviennent majoritairement d'erreurs dans l'étiquetage grammatical des mots de l'énoncé. La version du système testée reposait sur un étiqueteur priviligiant la décision, les répétitions et surtout les faux-départs ont entraîné des erreurs d'étiquetage très pénalisantes. Les règles de désambiguïsation actuellement utilisées sont plus souples.

Références

ABNEY S. (1991). Parsing by chunks. In *Principle Based Parsing*. In R.Berwick, S.Abney and C.Tenny, Eds., Kluwer Academic Publishers.

ANTOINE J.-Y. *et al.* (2002). Predictive and objective evaluation of speech understanding: the challenge evaluation campaign of the i3 speech workgroup of the french cnrs. In *LREC'02*. à paraître.

BEAR J., DOWDING J. & SHRIBERG E. (1992). Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialogue. In *ACL'92*, p. 56–63.

BLANCHE-BENVENISTE C. (1990). *Le français parlé ; études grammaticales*. CNRS Editions, Paris.

EJERHED E. (1993). Nouveaux courants en analyse syntaxique. *T.A.L.*, **34.1**, 61–82.

HEEMAN P. & ALLEN J. (1999). Speech repairs, intentional phrases and discourse markers : modeling speakers' utterances in spoken dialogue. In *Computational Linguistics, 25(4)*, p. 527–573.

HINDLE D. (1983). Deterministic parsing of syntactic nonfluencies. In *21st Annual Meeting of the Association for Computational Linguistics, ACL'83*, p. 123–128.

HIRSCHMAN L. (1998). Language understanding evaluation: lessons learned from muc and atis. In *Actes de LREC'98, Grenade, Espagne*, p. 117–122.

ROCHE E. & SCHABES Y. (1997). *Finite state Language Processing*, chapter Deterministic Part-of-Speech Tagging with Finite State Transducers, p. 205–239. MIT Press.

SLEATOR D. & TEMPERLEY D. (1991). *Parsing English with a Link Grammar*. Rapport interne, CMU-CS-91-196, CMU, USA.

[9] Pour une comparaison plus directe, nous envisageons d'utiliser la méthodologie DCR.

[10] Les méthodes stochastiques –repérage de segments conceptuels– connaissent les mêmes problèmes.