

Prise en compte des disfluences dans un système d'analyse syntaxique automatique de l'oral

Rémi Bove¹, Christine Chardenon², Jean Véronis¹

¹ Université de Provence – Équipe DELIC
{remi.bove ; jean.veronis}@up.univ-mrs.fr

² France Télécom Division Recherche et Développement
christine.chardenon@rd.francetelecom.fr

Résumé

Nous présentons dans cette étude un essai de prise en compte des disfluences dans un système d'analyse linguistique initialement prévu pour l'écrit, en vue de la réalisation d'un prototype de traduction parole-parole. À partir d'une étude approfondie sur corpus, nous montrons comment des modifications du lexique et de la grammaire ont permis de traiter les cas les plus simples (pauses remplies, répétitions de mots isolés, etc.). D'autres cas plus complexes comme répétitions et auto-corrrections de syntagmes ont nécessité la mise au point d'un mécanisme de contrôle sémantique permettant de limiter la combinatoire. Cette étude a mis également en évidence la difficulté de traitement de phénomènes tels que les amorces (mots interrompus) et les constructions inachevées, qui pour l'instant restent sans solution satisfaisante.

Mots-clefs : disfluences, analyse syntaxique en dépendances, traitement automatique de l'oral.

Abstract

In this paper we describe an attempt to take speech disfluencies into account in a linguistic analysis system initially designed for written data. Using a detailed corpus analysis, we show how the lexicon and grammar can be modified to solve the simplest cases (such as filled pauses, single-word repeats, and so forth). More difficult cases such as phrasal repeats and self-repairs required the development of a semantic control mechanism in order to avoid combinatorial explosion. This study also reveals the difficulty of processing word fragments and aborted constructs, which receive no satisfactory solution in the current state of the art.

Keywords: disfluencies, parsing, automatic speech processing.

1. Introduction

La reconnaissance de la parole a fait d'immenses progrès au cours des dernières années, mais elle n'atteint pour l'instant ses meilleures performances que dans des conditions très contraintes : mots isolés, domaine à vocabulaire limité ou à large vocabulaire mais avec une qualité sonore élevée et un environnement non bruité. De plus, la qualité de la reconnaissance se dégrade très rapidement sur la parole spontanée, comme dans le cas de la transcription de flux audio en vue de l'indexation, ou d'applications de dialogue homme-machine et de traduction parole-parole visant le grand public. Ces applications ont le plus souvent été abordées à l'aide de méthodes purement statistiques – par exemple dans le projet Audiosurf¹ –, mais il nous semble que le couplage avec des méthodes plus linguistiques pourrait être porteur d'améliorations sensibles.

Traitement de la parole et traitement automatique des langues (TAL) ont cependant des traditions encore relativement séparées. Le traitement de la parole est plutôt concentré sur le

¹ http://www.telecom.gouv.fr/rntl/AAP2001/Fiches_Resume/AUDIOSURF.htm

domaine acoustique-phonétique, tandis que les outils développés en TAL, et en particulier les analyseurs syntaxiques, sont principalement destinés à analyser l'écrit. Les recherches portant sur l'interface entre les deux domaines sont pour l'instant peu nombreuses, et seuls quelques travaux, tels que ceux d'Antoine *et al.* (2003) ont tenté d'enchaîner la reconnaissance et les techniques du TAL comme l'analyse robuste (chunking) ou la construction de dépendances sémantiques guidée par les buts applicatifs.

La chaîne d'analyse linguistique TiLT (Chardenon, 2005) développée par l'équipe Langues Naturelles de France Télécom Division Recherche et Développement a été conçue pour les différentes étapes du traitement de l'écrit (analyse lexicale, syntaxique, sémantique, génération, ...). Toutefois, comme elle est réalisée de façon très modulaire et dissocie de façon très claire les données linguistiques (lexique, grammaire, etc.) du code, nous avons fait le pari qu'elle devait pouvoir être couplée sans trop de difficultés avec un système de reconnaissance de parole également développé chez France Télécom R&D, dans la perspective de la réalisation d'un prototype de traducteur parole-parole. Nous montrons ici comment la chaîne TiLT peut être adaptée au traitement des phénomènes spécifiques à la parole spontanée, et en particulier à celui des disfluences.

2. Méthodologie

Le terme de *disfluences* regroupe un certain nombre de phénomènes spécifiques à l'oral, qui reflètent le *travail de formulation* du locuteur (Morel et Danon-Boileau, 1998) : hésitations, répétitions, inachèvements, etc. Contrairement à ce que peu laisser penser la connotation quelque peu négative du terme, les disfluences sont des phénomènes tout à fait normaux et habituels de la parole spontanée, qui correspondent à la mise en œuvre en temps réel des structures de la langue. Pour Blanche-Benveniste (2003), elles sont souvent le reflet d'une séparation par le locuteur entre la syntaxe et le lexique : par exemple, les structures syntaxiques peuvent être déjà en place, sans que l'accès au lexique ait été totalement effectif, comme dans l'exemple (1) ci-dessous :

*je voudrais connaître euh le détail **des des** titres* (1)

Les disfluences ne doivent pas être considérées comme des accidents aléatoires, des faiblesses imputables au locuteur, n'ayant aucune valeur fonctionnelle ni grammaticale. En effet, nous avons observé dans le cadre de ce travail et dans diverses études antérieures (par exemple Henry, Campione et Véronis, 2004 ; Campione et Véronis, 2005), que les disfluences n'interviennent pas au hasard mais apparaissent en fonction de contraintes syntaxiques très précises. Blanche-Benveniste (2003) soutient même que les répétitions et les hésitations ont une réelle valeur fonctionnelle, servant d'indice de la mise en place des syntagmes par le locuteur. Il nous paraît donc préférable d'adopter une stratégie d'analyse qui ne se contente pas de « gommer » les disfluences, mais qui les intègre véritablement au traitement syntaxique global, selon l'approche que nous avons proposée dans Benzitoun *et al.* (2004) et Benzitoun et Véronis (2005). Il n'y a toutefois à notre connaissance que très peu de travaux théoriques sur la question de l'intégration des phénomènes de disfluences dans les grammaires (voir cependant McKelvie, 1998 ; Guénot, 2005).

Il nous a paru préférable, à ce stade de l'étude, de travailler sur des corpus manuellement transcrits, plutôt que sur des sorties du système de reconnaissance vocale, afin de cerner les difficultés qui étaient véritablement imputables à l'intégration des disfluences dans la chaîne de traitement TiLT. En effet, les inévitables erreurs que produit un système de reconnaissance sur de la parole spontanée brouillent l'analyse et rendent difficile la mise au point de

solutions. Le travail sur corpus transcrits nous place donc dans une situation idéale pendant la phase de prototypage et de mise au point des grammaires.

Notre étude se base sur l'analyse de deux corpus issus de projets de France Télécom R&D : Vocalia Dialogue Bourse et Transat. Le premier projet concerne le développement d'un système de dialogue homme-machine destinés à simuler des transactions boursières ; le second vise la mise au point d'un prototype de traduction parole-parole pour des interactions client/vendeur dans le domaine pharmaceutique. Le corpus Vocalia comprend environ 20000 mots ; le corpus Transat, encore en cours d'élaboration, comprend pour l'instant 4500 mots.

Parmi les phénomènes traités, nous n'avons pas intégré les pauses silencieuses, que certains auteurs rangent dans les disfluences. Toutefois, une grande partie des pauses silencieuses a simplement un rôle de « ponctuation » de l'énoncé (Henry, 2002), et participe donc au contraire à la *fluence*, en facilitant le repérage des structures syntaxiques et des « chunks » : ces pauses ne constituent pas un problème particulier pour l'analyseur. Nous avons montré par ailleurs (Campionne et Véronis, 2005) que lorsque les pauses silencieuses interviennent dans un contexte de disfluence, elles ne sont quasiment jamais isolées, mais interviennent en conjonction avec une autre marque du travail de formulation (en particulier le *eah* d'hésitation).

Par contre, nous avons effectué un traitement des « marqueurs discursifs » (*bon, ben, alors, quoi, etc.*) : nous ne considérons *pas* ces marqueurs comme des disfluences, mais comme des signaux particuliers propres à la communication orale (dont le statut est d'ailleurs fort peu clair, cf. Teston et Véronis, 2004). Ils sont toutefois extrêmement fréquents dans la parole spontanée, et très souvent associés aux disfluences, en particulier aux hésitations (*eah bon, alors eah, etc.*). Leur traitement était donc incontournable dans notre étude.

3. Traitement

3.1. Enrichissement du lexique

La première étape de notre travail a consisté à enrichir le lexique général de TiLT. Nous avons intégré les marqueurs discursifs en créant une catégorie grammaticale particulière MARQDISC. En effet, le fonctionnement des marqueurs discursifs est très particulier, et leur attribuer une des catégories existantes telle qu'adverbe n'était pas satisfaisant. La plupart des mots existaient déjà dans le lexique : *bon* était adjectif et nom, *quoi* pronom, etc. Ils font donc désormais l'objet d'une nouvelle entrée du type :

bon ((MARQDISC)

Nous avons également ajouté une entrée pour la pause remplie (ou « *eah* » d'hésitation) qui n'existait pas dans le lexique général. La pause remplie est une unité généralement autonome, qui prend place dans l'énoncé indépendamment des autres unités.

je souhaite acheter eah cinquante actions Michelin (2)

On remarquera que ces adjonctions ont été faites dans le cadre d'une extension du système vers l'oral, mais elles seront utiles même dans certaines applications écrites de la chaîne de traitement, puisqu'on les retrouve dans des situations d'écrit informel (forum, chats, etc.).

3.2. Adaptation des règles grammaticales

L'exploitation de ces nouvelles entrées du lexique s'est ensuite faite par l'introduction de nouvelles règles de grammaire. Notre travail a essentiellement consisté à enrichir la grammaire de dépendances construite pour le traitement de l'écrit.

Par exemple, pour les marqueurs discursifs, la règle (a) :

```
(a) RègleAttachement MD02_ORAL MD_ORAL
Schéma GV-CT|GV-PT[...] >> MARQ_DISC
ConditionsPrincipal (SY_OTY_MARQ_DISC/!)
AutresConditions ((P += SY_OTY_MARQ_DISC))
```

permet de rattacher un marqueur discursif (MARQ_DISC) en début d'énoncé à l'unité ou au syntagme qui le suit (GV-CT|GV-PT|...), en précisant certaines conditions : l'élément suivant ne doit pas être également un marqueur discursif (ConditionsPrincipal (SY_OTY_MARQ_DISC/!)) (dans ce cas une autre règle s'applique) et l'on indique la présence du marqueur au sein du groupe ainsi formé (AutresConditions ((P += SY_OTY_MARQ_DISC))).

Une stratégie similaire a été appliquée pour les répétitions et autocorrections. Par exemple, la règle (b) :

```
(b) RègleAttachement DETDEF_ORAL REPETITION_ORAL
Schéma GN-D >/> GN-D
```

permet de préciser que la règle concerne le phénomène de répétition (REPETITION_ORAL), et qu'elle porte sur la répétition de déterminants définis (DETDEF_ORAL). Ainsi lorsque deux déterminants définis (GN-D) se succèdent, le second va « consommer » (>/>) le premier pour ne former qu'un arbre.

Les figures 1 et 2 donnent respectivement un exemple de représentation syntaxique arborescente avant et après mise en place des règles pour la phrase « *ben je je viens vous voir* » (mêlant marqueurs et répétition « simple »). Notons que la figure 1 comporte trois « arbres » partiels ; ceux-ci se retrouvent correctement regroupés une fois les règles mises en place (figure 2).

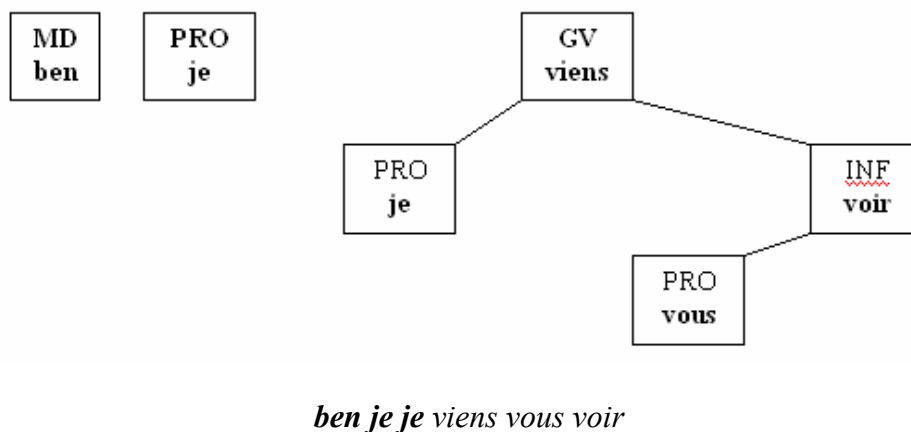
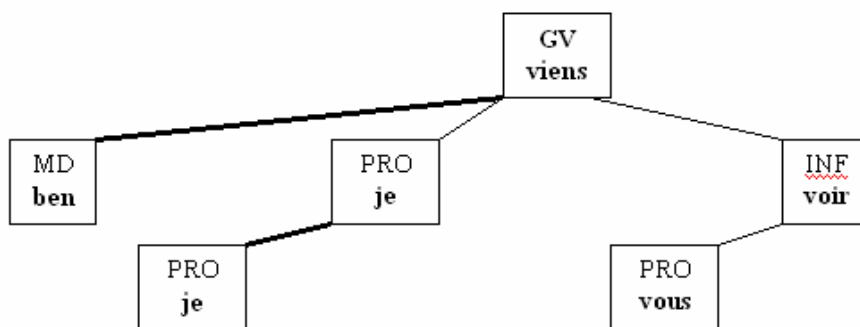


Figure 1. Représentation syntaxique *avant* mise en place des règles



ben je je viens vous voir

Figure 2. Représentation syntaxique *après* mise en place des règles

Les répétitions se font parfois avec auto-correction, souvent pour corriger le genre, le nombre, ou l'éllision :

je n'arrive pas à le à l'aval (3)

L'exemple ci-dessus est un cas de répétition d'**unité simple** : l'empan de la répétition est un mot unique, presque exclusivement un mot-outil (article, préposition, etc.) monosyllabique introducteur de « chunk » (Henry, Campione et Véronis, 2004). Ce type de répétition produit un « chunk » inachevé (dans l'exemple ci-dessus, le premier *je*). Il témoigne généralement de la difficulté de mise en place du lexique (ici le choix du verbe). Un autre type de répétition concerne des **unités complexes**, souvent des « chunks » entiers

euh je voudrais connaître l'analyse technique l'analyse technique sur le CAC-4 (4)

De tels cas sont plus complexes à traiter, particulièrement lorsqu'ils font intervenir une autocorrection :

je voudrais savoir la la valeur la cotation de l'action euh Aventis (5)

Il faut être capable de repérer que certains syntagmes doivent être regroupés et que la recherche des fonctions syntaxiques qu'ils remplissent doit tenir compte de ce regroupement. Dans l'exemple ci-dessus, l'arbre représentant le syntagme *la valeur* est attaché à l'arbre représentant *la cotation*, qui remplit la fonction d'objet direct du verbe *savoir*. Dans l'exemple (6), si l'arbre représentant *j'aimerais* est attaché, comme élément repris, à l'arbre *je voudrais*, alors les autres éléments de l'énoncé pourront se rattacher naturellement à ce dernier arbre, sans modification autre de la grammaire :

donc j'aimerais euh je voudrais de l'aspirin (6)

On peut ajouter dans la grammaire des règles permettant d'attacher un groupe (nominal, verbal, ...) à un autre groupe du même type, comme il a été fait pour les introducteurs de « chunks ». Toutefois, des suites consécutives de déterminants ou prépositions ne permettent guère d'autres interprétations que celle d'une répétition, alors que pour des syntagmes complexes, ce type de règle est beaucoup trop tolérant et serait susceptible de générer une explosion combinatoire. On ne peut en effet pas imposer comme contrainte que la source et la cible de la répétition soient de structure absolument identique. L'exemple ci-dessous (variante du 5) est tout à fait légitime :

je voudrais savoir la [la valeur] [le volume d'échanges] de l'action euh Aventis (7)

Nous avons commencé à intégrer des critères de taille et de structure destinés à limiter la combinatoire. En effet, il semble, en première approximation, que des arbres très profonds ou très larges aient peu de chances d'être la source d'auto-corrections. Cette hypothèse est cohérente avec un modèle psychologique qui attribue à un mécanisme de self-monitoring la production des auto-corrections (Levelt, 1989) : interruption et mise en place d'une auto-correction demandent un délai relativement bref au locuteur, incompatible avec la durée de structures complexes. Grâce à un outil de gestion des grammaires développé dans l'équipe Langues Naturelles, nous avons par exemple pu modifier automatiquement les règles de grammaire exploitées par l'analyseur de manière à ce que les arbres représentant des groupes verbaux avec compléments ne puissent plus être utilisés par des règles de reprise de groupes verbaux. Une étude plus approfondie en corpus s'impose toutefois pour déterminer précisément les contraintes.

3.3. Ajout du contrôle sémantique

Pour limiter l'attachement d'arbres correspondant à des syntagmes nominaux ou verbaux, il nous a semblé intéressant d'introduire une notion de contrôle basé sur la sémantique. Par exemple, dans l'exemple (6), c'est la proximité des sens d'« aimer » et de « vouloir » (dans un contexte applicatif donné) qui va permettre de valider l'hypothèse que « je voudrais » est une reprise de « j'aimerais ». En effet, un ou plusieurs sens peuvent être associés aux verbes « aimer » et « vouloir » : *amour* / *souhait* pour « aimer », et *souhait* pour « vouloir ». La proximité est établie pour cet exemple à partir du sens *souhait*. Nous allons maintenant détailler ce mécanisme de contrôle

L'analyseur TiLT est basé sur le formalisme des grammaires de dépendance, tout nœud d'un arbre syntaxique correspond donc à un mot (simple ou composé) de l'énoncé initial. L'analyse syntaxique s'effectue après une phase d'analyse lexicale, qui attribue à tout mot de l'énoncé un ensemble d'interprétations lexicales, en fonction du lexique utilisé. La notion d'interprétation lexicale correspond à un ensemble d'informations telles que lemme, description morpho-syntaxique, mais aussi « unité sémantique ».

L'unité sémantique permet de représenter le sens de l'interprétation. Les unités sémantiques sont organisées dans un réseau sémantique dans lequel elles sont reliées par des relations d'hyponymie (is-a), d'appartenance, etc. Nous admettons que l'héritage multiple est possible, mais pas les cycles. Il est alors possible de définir la notion de proximité sémantique entre deux unités sémantiques us_1 et us_2 . Cette proximité vaut 1 si les unités sémantiques sont identiques, et elle est calculée en fonction de paramètres tels que nombre, sens et type des relations à parcourir pour passer de l'une à l'autre dans le réseau. Ce type de calcul peut poser problème si le réseau sémantique n'est pas d'une densité uniforme, mais il existe des solutions pour pallier cela (Resnik, 1999).

La construction de l'arbre syntaxique consiste à créer des relations entre des mots. La construction d'une relation entre deux mots implique un filtrage des interprétations lexicales de chacun de ces mots, pour ne conserver que celles qui valident la relation. À tout nœud d'un arbre syntaxique se trouve donc associé un ensemble d'unités sémantiques représentant les sens possibles de ce nœud. Ceci étant posé, l'algorithme de calcul de proximité d'arbres sera le suivant :

- Calculer les ensembles d'unités sémantiques associés au premier arbre, en filtrant les unités considérées comme non pertinentes. Fusionner ces ensembles en un ensemble E1. Les unités considérées comme non pertinentes sont en particulier celles associées aux mots grammaticaux.

- Calculer les ensembles d'unités sémantiques associés à l'arbre candidat pour constituer l'auto-correction ou la répétition, en filtrant les unités considérées comme non pertinentes. Fusionner ces ensembles en un ensemble E2.
- Calculer l'intersection entre les ensembles.
- Si cette intersection est un ensemble non vide de cardinal N, calculer la proximité comme :

$$P = N / \text{card}E1$$

- Si cette intersection est vide, pour toute unité u_{1i} de E1, calculer sa proximité sémantique $P(u_{1i}, u_{2j})$ avec toute unité de E2.

$$P(u_{1i}, E2) = \min_j (P(u_{1i}, u_{2j}))$$

$$P(E1, E2) = \frac{1}{\text{card}E_1} \sum_i P(u_{1i}, E2)$$

Ce calcul est évidemment approximatif, dans la mesure où il met au même niveau tous les sens des différents nœuds d'un arbre, mais il permet de garder une complexité de calcul raisonnable pour un résultat suffisant du point de vue de la représentativité de la distance sémantique entre deux arbres. La prise en compte de ce calcul par l'analyseur en dépendances a nécessité une légère évolution du code de ce dernier, pour accepter des règles d'attachement requérant le contrôle sémantique, pour calculer les ensembles de concepts décrits ci-dessus, pour appeler les méthodes de calcul de similarité entre concepts développées dans un module « réseau sémantique » de la chaîne TiLT et pour prendre en compte un seuil en deçà duquel on considère que deux arbres ne peuvent être reprise l'un de l'autre

Il est à noter qu'un arbre identifié comme « repris » ne disparaît pas de l'analyse syntaxique globale de l'énoncé et que ce choix d'attachement pourrait être remis en cause dans une couche de traitement ultérieur, sémantique et pragmatique. Par conséquent, il n'est pas problématique d'être un peu trop tolérant sur ces attachements.

La distinction entre auto-correction et énumération est extrêmement délicate. L'exemple (8) peut-être une auto-correction ou une énumération, comme c'est clairement le cas dans l'exemple (9) :

je voudrais de l'aspirine du paracétamol et puis je voudrais aussi des couches (8)

je voudrais de l'aspirine des pastilles pour la toux et puis je voudrais aussi des couches (9)

Il est probable que dans de nombreux cas, des indications prosodiques (intonation, allongement syllabique) peuvent servir d'indicateurs aidant à la désambiguïsation. Nous ne sommes pas certains pour l'instant de la possibilité d'obtenir de telles informations du système de reconnaissance, ni de la fiabilité qu'elles pourraient avoir. Dans un souci de robustesse, il nous semble intéressant d'exploiter au maximum les traitements qui peuvent être faits en aval de la reconnaissance vocale. Dans le cas de la phrase (7), le calcul de proximité sémantique permettra d'affecter une meilleure probabilité à l'interprétation comme auto-correction : les produits ayant la même fonction, il est probable que le locuteur ne réclame pas les deux (mais on ne peut en fait complètement exclure cette possibilité). Ce ne sera pas le cas pour la phrase (8). Mais là encore, ces choix pourront être remis en cause lors d'une phase de validation pragmatique, et nécessiter une demande de confirmation de la part du système.

4. Phénomènes non traités

Le contrôle sémantique n'est pas suffisant non plus pour analyser d'autres phénomènes tels que les amorces (appelés aussi fragments de mots) ou encore les inachèvements. Les amorces correspondent à l'interruption du mot avant sa prononciation complète et constituent des événements langagiers d'une fréquence non négligeable (environ une amorce tous les 250 mots selon l'estimation de Pallaud (2002)) :

oui le d- je veux le détail (10)

Le traitement des amorces implique des modifications au niveau du système de reconnaissance vocale. En effet, pour l'instant, celui-ci fournit le mot (évidemment erroné) le plus proche dans son lexique en fonction du modèle de langage sous-jacent, et le système TiLT n'a pas de moyen de savoir qu'il s'agissait d'une amorce. Ce problème constitue un défi pour tous les systèmes de reconnaissance à l'heure actuelle, car le nombre d'amorces potentielles est très grand (pratiquement chaque suite de phonèmes commençant chaque mot plein), et la fréquence de chaque amorce trop faible pour pouvoir entraîner efficacement un modèle de langage statistique. Il est probable que d'autres critères (prosodiques notamment) doivent être pris en compte

Les inachèvements apparaissent comme des échecs de la part du locuteur qui l'obligent à changer de construction (Benveniste, 2000) ou des changements d'intention en milieu d'énoncé. Par exemple, dans une phrase, un locuteur peut commencer une construction qu'il abandonne pour enchaîner un énoncé sans lien direct avec le précédent :

qu'est-ce que vous avez d'efficace contre + bon je ne sais pas si ce que j'ai c'est vraiment la grippe mais (11)

Le système n'est pas en mesure pour l'instant de traiter les inachèvements, car une modification des règles qui permettrait à la plupart des syntagmes de rester incomplets aboutit à une explosion combinatoire extrêmement rapide. Face à ces phénomènes, l'analyseur ne peut que proposer des arbres syntaxiques partiels recouvrant l'énoncé, mais ne permettant pas d'aboutir à une solution globale pour celui-ci.

5. Conclusion

Nous avons présenté dans cette étude un essai de prise en compte des disfluences dans un système d'analyse linguistique initialement prévu pour l'écrit, en vue de la réalisation d'un prototype de traduction parole-parole. Une étude approfondie sur corpus nous a permis de dégager le fonctionnement des différents types de disfluences, qui ne constituent pas un « bruit » de nature aléatoire, mais obéissent à des contraintes précises. Des modifications du lexique et de la grammaire ont permis de traiter les cas les plus simples (pauses remplies, répétitions de mots isolés, etc.). D'autres cas plus complexes comme la répétition de syntagmes ont nécessité la mise au point d'un mécanisme de contrôle sémantique permettant de limiter la combinatoire. Cette étude montre la faisabilité de l'intégration des phénomènes de l'oral dans l'analyse syntaxique, mais de nombreux problèmes restent évidemment à résoudre. En particulier, la tâche s'annonce très plus délicate pour d'autres phénomènes tels que les amorces (mots interrompus) et les énoncés inachevés.

Références

BENZITOUN C., CAMPIONE E., DEULOFEU J., HENRY S., SABIO F., TESTON S., VALLI A., VÉRONIS J. (2004). « L'analyse syntaxique de l'oral : problèmes et méthode ». In *Journée d'étude de l'ATALA*

- « EVANS: Méthodes et outils pour l'évaluation des analyseurs syntaxiques » (15 mai 2004) [<http://www.up.univ-mrs.fr/veronis/pdf/2004-Atala-analyse-oral.pdf>].
- BENZITOUN C., VÉRONIS J. (2005). « Problèmes d'annotation d'un corpus oral dans le cadre de la campagne EASY ». In *Actes de TALN 2005*. Dourdan : 13-16.
- BLANCHE-BENVENISTE Cl. (2000). *Approches de la langue parlée en français*. Ophrys, Paris.
- BLANCHE-BENVENISTE Cl. (2003). « La naissance du syntagme dans les hésitations et répétitions du parler ». In J.-L. Araoui (éd.), *Le sens et la mesure. Hommages à Benoît de Cornulier*. Honoré Champion, Paris : 40-55.
- CAMPIONE E., VÉRONIS J. (2005). « Pauses and hesitations in French spontaneous speech. Disfluency in spontaneous speech ». In *DiSS'2005*. Aix-en-Provence : 43-46.
- CANDÉA M. (2000). *Contribution à l'étude des pauses silencieuses et des phénomènes dits « d'hésitation » en français oral spontané*. Thèse de doctorat, Université de Paris III.
- CHARDENON C. (2005). « Analyse syntaxique en dépendance et évaluation ». In *Actes de TALN 2005*. Dourdan : 25-28.
- GALA PAVIA N. (2003). *Un modèle d'analyseur syntaxique robuste fondé sur la modularité et la lexicalisation de ses grammaires*. Thèse de Doctorat, Université de Paris XI Orsay.
- HENRY S. (2002). « Quelles répétitions à l'oral ? Esquisse d'une typologie ». In *Actes des 2es Journées de Linguistique de Corpus* (sous presse). Lorient. [<http://www.up.univ-mrs.fr/delic/papiers/Henry-2002lorient.pdf>]
- HENRY S., CAMPIONE E., VÉRONIS J. (2004). « Répétitions et pauses (silencieuses et remplies) en français spontané ». In *Actes des XXV^e Journées d'Étude sur la Parole (JEP'04)*. Fès (Maroc) : 261-264.
- HENRY S., PALLAUD B. (2004). « Amorces de mots et répétitions : des hésitations plus que des erreurs en français parlé ». In *Actes des JADT 2004*. Presses universitaires de Louvain, Louvain-la-Neuve. [www.cavi.univ-paris3.fr/lexicometrical/jadt/jadt2004/pdf/JADT_082.pdf]
- LEVELT W.J.M. (1989). *Speaking. From intention to articulation*. The MIT Press, Cambridge.
- MOREL M.-A., DANON-BOILEAU L. (1998). *Grammaire de l'intonation. L'exemple du français*. Ophrys, Paris.
- MCKELVIE D. (1998). « The syntax of disfluency in spontaneous spoken language ». In G. Sampson et D. McCarthy (éds), *Corpus Linguistics, Reading in a widening discipline*. London.
- PALLAUD B. (2002). « Les amorces de mots comme faits autonymiques en langage oral ». In *Recherches sur le Français Parlé* 17 : 79-102.
- RESNIK P. (1999). « Semantic Similarity in a Taxonomy : An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language ». In *Journal of Artificial Intelligence Research* 11 : 95-130.
- TESTON S., VÉRONIS J. (2004). « Recherche de critères formels pour l'identification automatique des particules discursives ». In *Journée d'étude de l'ATALA « Modéliser et décrire l'organisation discursive à l'heure du document numérique »*. La Rochelle [<http://www.up.univ-mrs.fr/veronis/pdf/2004-LaRochelle-Part-Disc.pdf>].
- VALLI A., VÉRONIS J. (1999). « Étiquetage grammatical de corpus oraux : problèmes et perspectives ». In *Revue Française de Linguistique Appliquée* IV (2) : 113-133.