

Résolution anaphorique intégrée à une analyse automatique de discours d'un corpus oral retranscrit

Alain RÉGNIER

LPL CNRS, Université de Provence,
29 Avenue R. Schuman 13621 Aix-en-Provence
alain.regnier@lpl.univ-aix.fr

Résumé. Nous présentons une résolution anaphorique intégrée à une analyse automatique de discours. Cette étude traite des anaphores pronominales et des anaphores zéro. Notre analyse est basée sur trois approches : une analyse basée sur les contraintes, une analyse fonctionnelle et une analyse dynamique. Pour évaluer la faisabilité et la fiabilité de notre approche, nous l'avons expérimentée sur un corpus de 97 histoires produites à l'oral par des enfants. Nous présentons le résultat de cette évaluation.

Abstract. We present an anaphora resolution integrated in a discourse analysis. This study deals with pronoun anaphora and zero anaphora. Our analysis is based on three approaches. A constraint based rule analysis, a functional approach and a dynamic analysis. In order to evaluate the feasibility of our approach and its reliability we have experimented with a corpus of 97 speech stories produced by children. We present here the results of our evaluation experiment.

Mots-clés : analyse de discours, résolution anaphorique, anaphore pronominales, anaphores zéro, grammaires de propriétés, grammaire fonctionnelle, analyse dynamique, discours oral.

Keywords: discourse analysis, anaphora resolution, pronoun anaphora, zero anaphora, property grammars, functional grammar, dynamic analysis, speech discourse.

1 Introduction

Nous traitons d'une résolution anaphorique intégrée au sein des autres traitements linguistiques. Plus généralement nous essayons d'intégrer au mieux les différentes analyses linguistiques au sein de l'analyse de discours. Cette résolution anaphorique s'inscrit dans un projet de réalisation d'un analyseur de discours pour classification automatique de textes (Nouali *et al.*, 2005). Pour réaliser cet analyseur, nous sommes confronté à la nécessité de résoudre certaines anaphores. Il s'agit des anaphores pronominales sujet et objet mais aussi des anaphores zéro. Notre présentation du projet met en exergue l'application des mêmes principes pour chaque niveau de l'analyse linguistique. Ces principes sont l'application de contraintes, une approche dynamique et la hiérarchisation de fonctions sémantiques, syntaxiques et pragmatiques. Dans un premier temps nous présentons l'application de ces principes pour le découpage en chunks et leurs rattachements, ainsi que pour l'assignation des fonctions, ensuite nous aborderons l'analyse anaphorique et les résultats obtenus.

Les sections vont s'articuler de la façon suivante. La section 2 dégage les propositions sur lesquelles nous avons axé notre analyse du discours. La section 3 décrit le fonctionnement de l'analyseur en présentant les trois types de mise à jour actuellement implémentée. La section 4 présente l'analyse utilisée pour la résolution des anaphores pronominales et des anaphores zéro. La section 5 présente les résultats.

2 Proposition

L'analyse que nous proposons est à la croisée de trois types d'approche. C'est une approche par contraintes et nous utilisons le paradigme des Grammaires de Propriétés tel qu'il est décrit dans (Blache, 2001). Nous utilisons aussi l'apport des constructions au sein des Grammaires de Propriétés (Blache, 2005). C'est aussi une approche fonctionnelle pour la place importante que nous accordons aux fonctions. Nous nous référons pour cela à la Grammaire Fonctionnelle de S. C. Dik (Dik, 1997). C'est une analyse de discours dynamique pour, à terme, être à même d'intégrer facilement des inférences sémantiques comme la SDRT le propose (Asher, 1993).

Nous appliquons les mêmes principes de contraintes, les principes fonctionnels et les principes de l'analyse dynamique pour tous les niveaux des traitements linguistiques. De ce fait, à chaque niveau, on applique des contraintes comme les contraintes de linéarité par exemple. On observe aussi un ordre prioritaire de fonctions. Ces fonctions sont les fonctions sémantiques, syntaxiques et pragmatiques définies en Grammaire Fonctionnelle. L'analyse est incrémentale, le contexte sélectionne les contraintes que l'on applique et chaque entité nouvelle tente de s'intégrer à la structure précédemment créée.

3 L'analyseur

L'analyseur de discours a été implémenté en O'CAML par Gilles Régnier selon les spécifications que nous avons établies. Il construit une représentation du discours au moyen de trois types de mise à jour. La première mise à jour construit les segments de texte qui correspondent aux chunks d'un chunk parser. La deuxième concerne les rattachements des chunks entre eux. Cette mise à jour leur attribue de fait un rôle. La troisième est une mise à jour fonctionnelle. Elle met à jour les fonctions sémantiques, une fonction syntaxique et des fonctions pragmatiques. Après ces mises à jour une phase de filtrage appelée inhibition est activée pour limiter les effets de l'explosion combinatoire des interprétations proposées.

Le lexique utilisé est le lexique du Laboratoire Parole et Langage d'Aix-en-Provence que l'on a enrichi d'informations sémantiques comme la structure des prédicats, des restrictions de sélection, et des indications sur la classe sémantique à laquelle appartient l'item lexical.

3.1 La mise à jour des chunks

La frontière des constituants est déterminée par l'évaluation de chaque frontière de mot. Chaque mot entrant est évalué et attribue un score sur sa frontière gauche et/ou sa frontière droite. Il peut aussi apparaître dans un contexte spécifique et être intégré à une construction qui modifie ces scores. La règle 1 est un exemple extrêmement simplifié de hiérarchie pour la frontière

gauche d'un syntagme nominal donné à titre d'illustration. Celui-ci indique que les éléments les plus susceptibles d'ouvrir un syntagme nominal sur leur gauche, sont par ordre décroissant, un déterminant puis un adjectif cardinal puis un adjectif ordinal puis un nom.

(1) Det > Card > Ord > N

L'analyse est dynamique. La détermination des frontières des chunks constitue des contextes. Les contraintes sont appliquées en fonction des contextes ainsi créés. Par exemple les contraintes de linéarité sont différentes en fonction du type de chunk sur lesquelles on les applique. Si l'on est au sein d'un syntagme nominal, les contraintes de linéarité spécifiques au syntagme nominal sont appliquées.

3.2 La mise à jour des rattachements

La mise à jour des rattachements relie les différents chunks entre eux. Lorsqu'un chunk est constitué, l'analyseur essaie de le rattacher à la structure. L'attachement est établi selon un ordre de priorité basé sur les fonctions linguistiques qui accorde une préférence sur certaines structures. Pour la linéarité des fonctions des arguments du verbe par exemple, l'ordre de préférence indique que le sujet précède (sauf cas particuliers) les compléments du verbe.

L'approche est dynamique. Chaque mise à jour réévalue l'ensemble des rattachements de la représentation du ou des prédicats en cours. Cependant seules certaines positions de la structure restent ouvertes au rattachement. Par exemple un pronom sujet qui précède un verbe verrouille le point de rattachement de la position sujet de ce verbe. La réévaluation des rattachements s'opère donc sur une structure qui est partiellement constituée.

On observe des contraintes sur la structure comme par exemple la contrainte de projectivité (les arcs des rattachements ne peuvent se croiser). On applique aussi des contraintes syntaxiques sur la ou les prépositions qui peuvent introduire la sous-catégorisation.

3.3 La mise à jour des fonctions

La mise à jour des fonctions assigne à certains termes les fonctions de la Grammaire Fonctionnelle. La fonction syntaxique *sujet* et les fonctions pragmatiques *Thème*, *Topique* et *Focus* sont centrales pour notre résolution anaphorique comme nous le verrons en section 4.

Les fonctions sémantiques assignent par exemple les rôles d'*agent*, de *patient* aux arguments des prédicats.

La mise à jour des fonctions syntaxiques ne concerne pour le français, conformément à la Grammaire Fonctionnelle de S. C. Dik, que la fonction *sujet*. Cette assignation est utilisée lors des rattachements. La fonction *Sujet* est aussi mise à contribution pour la résolution des anaphores zéro et des anaphores pronominales.

La mise à jour des fonctions pragmatiques assigne les fonctions de *Topique*, de *Focus* et de *Thème* de la Grammaire Fonctionnelle. La fonction pragmatique *Thème* est assignée au syntagme nominal des extractions à gauche. La fonction pragmatique *Focus* est assignée au syntagme nominal ou prépositionnel d'une construction clivée.

Présentatif <i>Voici</i> ...	
Information	$\left[\begin{array}{l} \text{OPERATEUR DECL} \\ \text{PRAG} \left[\begin{array}{l} \text{FONCTIONS [TOPIQUE []]} \\ \text{ACTE-P } \textit{Presenter} \end{array} \right] \\ \text{SEM} \left[\begin{array}{l} \text{PRED}_{E_1} \left[\begin{array}{l} \text{VOICI} \\ \text{[]} \end{array} \right] \end{array} \right] \\ \text{SYNT} \left[\begin{array}{l} \text{P} \left[\begin{array}{l} \text{V } \textit{Voici} \\ \text{SN []} \end{array} \right] \end{array} \right] \end{array} \right]$
Propriétés	$\begin{array}{l} \text{Const} = \{ \text{V[voici]}, \text{SN}_1 \} \\ \text{Lin} : \text{V} \prec \text{SN}_1 \\ \text{Unic} = \{ \text{SN}_1 \} \\ \text{V} \Rightarrow \text{SN}_1 \end{array}$

FIG. 1 – Construction du présentatif *Voici* ...

La fonction pragmatique *Topique* est assignée de façon très diverse. Elle peut être déterminée au moyen de constructions particulières. La figure 1 représente la construction *Voici* qui s'applique sur des phrases comme l'exemple 2.

(2) Voici un ver.

Lorsque les propriétés définies dans l'ensemble des propriétés sont satisfaites, les informations associées de la construction sont propagées dans la structure. On peut observer que les informations pragmatiques de cette construction assignent la fonction de *Topique* à un argument du verbe *Voici*. Celui-ci est un syntagme nominal précédé par *Voici*. La résolution anaphorique que nous proposons utilise ce type d'information pragmatique.

La fonction *Topique* peut aussi être assignée au syntagme nominal tête d'une phrase nominale. Dans la majeure partie des cas, en l'absence de construction particulière, elle est assignée au syntagme auquel est assignée la fonction syntaxique *Sujet*.

3.4 La phase d'inhibition et la sortie de l'analyseur

Au fur et à mesure de la construction des éléments et de l'application des contraintes une évaluation de la structure est faite. Deux types d'évaluation sont utilisés. Le premier concerne la linéarité et le deuxième concerne le nombre de rattachements. Ces évaluations sont indépendantes. Une interprétation est éliminée lorsque la différence entre son score et celui de l'interprétation la mieux évaluée dépasse un seuil fixé. Cette évaluation permet d'éliminer en cours d'analyse les interprétations que l'on estime les moins cohérentes.

Pour observer les résultats de l'analyseur un ensemble d'informations est extrait des objets construits et exprimé dans un graphe orienté. La figure 2 est une visualisation d'un graphe issu d'un des segments de texte de notre corpus. Les numéros qui précèdent chaque extrait de texte indiquent l'ordre d'apparition dans le segment de texte. Les chunks verbaux sont représentés par des rectangles. Les conjonctions de coordination sont représentées par des triangles. Les ellipses représentent les syntagmes nominaux et les syntagmes prépositionnels. Le cercle représente certains adverbes (comme ceux de lieu et temps) identifiés comme satellite (selon la terminologie de la Grammaire Fonctionnelle). L'élément tête de chaque chunk est représenté entre crochets.

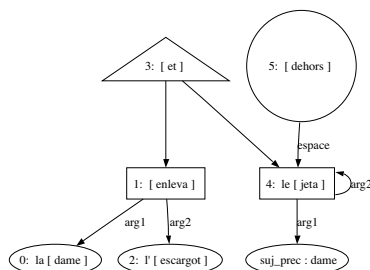


FIG. 2 – Sortie d'une interprétation de la phrase *La dame enleva l'escargot et le jeta dehors*

4 La résolution anaphorique

La résolution anaphorique concerne deux types d'anaphore : les anaphores pronominales et les anaphores zéro. Des études (dont (Ide & Cristea, 2000)) ont montré des relations entre la structure du discours (notamment RST, théorie des veines) et la résolution anaphorique. La structure de discours contraint l'accessibilité des référents. Or l'analyseur ne fournit pas beaucoup d'informations sur la structure rhétorique du discours. (Tetreault & Allen, 2004) ont mené une expérience sur un corpus de dialogues dont la structure de discours a été annotée. Ils en concluent que la prise en compte de la structure du discours n'apporte pas statistiquement d'améliorations significatives. La structure de discours n'étant que partiellement réalisée par notre analyseur, cela ne devrait pas générer un impact important sur nos résultats.

4.1 Les relations anaphoriques pronominales

Les anaphores pronominales traitées sont restreintes aux pronoms clitiques de la troisième personne ; *il, elle, ils, elles* en tant que pronom sujet et *le, la, lui, les, leur* en tant que pronom préverbal complément. La résolution ne concerne pas les anaphores abstraites (Amsili *et al.*, 2006).

Pour résoudre ces anaphores, on a utilisé les principes de la théorie du centrage (Grosz *et al.*, 1995). Cette théorie décrit le fonctionnement de la structure attentionnelle du discours qui contribue à la cohérence discursive. Il s'agit de la mise en correspondance de deux ordres de centres attentionnels : les centres anticipateurs et les centres rétrospectifs. Lorsque deux énoncés se succèdent le premier énoncé contient des centres attentionnels ordonnés du plus saillant au moins saillant. Les centres rétrospectifs vont se connecter sur les centres anticipateurs et ce plus favorablement vers les plus saillants. Les centres rétrospectifs les plus à même de se connecter sur les centres anticipateurs sont les pronoms anaphoriques. La théorie du centrage énonce des règles pour ordonner les centres attentionnels de sorte à pouvoir les appairer.

Dans notre cas, la hiérarchie des antécédents est basée sur les fonctions pragmatiques de la grammaire fonctionnelle (Dik, 1997) Part 1, Chap.13, pp309–338. L'ordre utilisé est donné dans l'exemple 3. Cette hiérarchie indique que l'élément auquel on a assigné la fonction *Thème* est prioritaire sur un élément auquel on a assigné la fonction *Focus*. Ce dernier est prioritaire sur un élément auquel on a assigné la fonction *Topique*. Les autres arguments sont ordonnés selon

leur ordre d'apparition dans la phrase.

(3) Fonction Thème > Fonction Focus > Fonction Topique > autres arguments

Les différentes fonctions sont identifiées par les constructions spécifiques comme indiqué dans la partie 3.3. L'identification d'une construction spécifique comme l'extraction à gauche permet d'assigner la fonction *Thème*. La fonction *Focus* est identifiée grâce aux constructions clivées. La fonction Topique est identifiée par des constructions diverses comme *Il était une fois ..., il y a ..., c'est un ...*.

Les centres rétrospectifs qui nous intéressent sont les pronoms. Nous traitons deux types de pronoms. Les pronoms sujet et les pronoms objet. Ces deux types de pronoms sont ordonnés dans le cas où les deux types sont en présence simultanément dans le même énoncé. Comme la règle 4 l'indique, nous favorisons les pronoms sujet que nous affectons en priorité à l'antécédent le plus saillant.

(4) pronoms sujet > pronoms complément

Dans la théorie du centrage, la cohérence est locale et les règles ne s'appliquent que sur deux phrases consécutives. (Alshawi, 1987) propose un ordre de saillance au moyen de pondérations. Ceci permet d'étendre les rapports entre pronoms et antécédents sur plusieurs phrases. Chaque apparition d'un référent de discours ajoute un nombre de points qui varie selon la position de ce référent dans la phrase. Ce capital de points associé au référent diminue au fur et à mesure des phrases s'il n'est pas cité. Une illustration de ce type de pondération est présentée dans (Lappin & Leass, 1994) et pour le français dans (Victorri, 2005).

Nous appliquons une méthode mixte qui d'une part utilise une hiérarchie attentionnelle entre les énoncés successifs et qui d'autre part pondère les référents du discours sur des empan plus larges que deux phrases successives. Nous mémorisons dans une pile les référents du discours selon un ordre de saillance dépendant de leur occurrence dans le discours. Cependant notre pondération ne dépend pas de l'étiquette syntaxique des éléments comme pour (Lappin & Leass, 1994). Elle dépend de l'ordre de saillance dans chaque énoncé. L'élément le plus saillant d'un énoncé reçoit toujours la même pondération quelle que soit sa position syntaxique. De plus notre pondération comprend des exceptions. Comme nous allons le voir plus loin, la fonction *Thème* manipule directement la hiérarchie des référents. De même une analogie structurelle entre deux énoncés successifs favorise la coréférence. Enfin nous accordons une pondération plus forte pour les référents qui apparaissent sous forme de pronom car c'est l'indice d'un ancrage plus fort dans le discours.

Nous avons utilisé la pondération suivante. Pour chaque énoncé, le premier élément de l'ordre des antécédents reçoit 100 points et 120 s'il s'agit d'un pronom. Le deuxième élément reçoit 50 points et 60 s'il s'agit d'un pronom. Le passage d'un énoncé au suivant a pour effet de multiplier tous les scores par 0,6. Si les exemples 5, 6 et 7 se suivent alors en 5, *Jean* a le score de 100. En 6, *Herbert* a le score de 100 et *Jean* qui est lié au pronom *le* a le score de 120 (soit $0,6 \times 100 + 60$). De ce fait, selon notre pondération, en 7, *Il* est attribué à *Jean*.

(5) Jean marche dans la rue.

(6) Herbert le salue.

(7) Il hoche la tête.

(8) Il l'interroge.

Cette affectation au moyen de pondération comprend des exceptions. C'est le cas lorsqu'un énoncé dont le verbe a deux arguments instanciés est suivi d'un énoncé de même structure dont les deux arguments sont pronominalisés. On attribue alors par analogie structurelle les arguments dans leur ordre respectif quelle que soit leur pondération. Par exemple si la phrase 8 succédait à 6, on attribue *Herbert* à *Il* et *Jean* à *l'* même si *Jean* a une plus forte pondération que *Herbert*.

Une autre exception concerne la fonction pragmatique de *Thème*. Cette fonction est prioritaire sur la saillance. Nous l'interprétons comme une modification de cette saillance. L'élément auquel est assignée la fonction *Thème* obtient le score de l'élément le plus élevé de la pile et se voit attribuer prioritairement le centre rétrospectif le plus élevé. Quel que soit le score des antécédents dans l'énoncé qui précède l'exemple 9, nous apparions *Le monsieur* à *il*.

(9) Le monsieur, il le prend.

D'autres règles spécifiques sont appliquées comme l'assignation de la fonction topique au complément du verbe, si le pronom "on" est sujet. On utilise des contraintes comme la contrainte d'accord en genre et en nombre. Cette contrainte est très importante. (Beaver, 2004) qui utilise la théorie de l'optimalité pour une résolution d'anaphore classe l'accord comme la contrainte de rang le plus fort.

Lorsque l'analyseur propose plusieurs analyses pour une même séquence, il faut faire un choix pour transmettre à la séquence de texte suivante les antécédents en cours. Afin de résoudre ce problème et permettre la transition, nous avons choisi de sélectionner la solution majoritaire proposée dans l'ensemble des interprétations. Nous avons utilisé le même procédé pour présenter les résultats. Au sein d'une séquence textuelle, si plusieurs interprétations sont proposées seul le résultat majoritaire est retenu.

4.2 Les relations anaphoriques zéro

Les anaphores zéro ne concernent que les sujets omis substituables par des pronoms clitiques sujet à la troisième personne. La résolution est simple et uniquement basée sur la fonction syntaxique *Sujet*. Lorsque le sujet est manquant dans un énoncé, on lui attribue le sujet de l'énoncé précédent en vérifiant les contraintes d'accord. La figure 2 illustre la résolution d'une anaphore zéro. L'analyseur attribue pour sujet du verbe *jeta* l'élément tête du sujet de la phrase qui précède *dame*.

5 Résultats

Le corpus utilisé est une partie d'un corpus constitué par Monique Vion et Annie Colas (Vion & Colas, 2000). Pour produire ce corpus, des bandes dessinées ont été présentées à des enfants. On a demandé aux enfants de raconter l'histoire que présentent ces images. Ce corpus est actuellement intégré au corpus CHILDES. Nous avons retenu pour l'évaluation une série de ces histoires appelée "Ver et escargot". D'autres séries nous ont servi de corpus d'observation et d'entraînement. Nous avons utilisé les transcriptions des enfants de 7, 9 et 11 ans. Cela représente un ensemble de 97 textes.

Une mise en forme des textes a été nécessaire. Les balises spécifiques de CHILDES ont été enlevées et certaines conventions de transcription ont été corrigées. Par exemple lorsque le pronom *il* était prononcé *i* le transcripteur l'a écrit *i*. Certaines disfluences ont été corrigées. Il s'agit de répétitions, nous n'avons conservé que la partie droite de ces répétitions. Avant d'effectuer l'analyse nous nous sommes assuré que tous les mots de notre corpus renvoient à une ou plusieurs entrées de notre lexique. Les ajouts étaient composés d'interjections et de noms propres.

Au sein de ces histoires nous avons identifié 671 anaphores pronominales et 31 anaphores zéro. Les textes ont été soumis à une évaluation de trois sujets. Pour chaque pronom et anaphore zéro, ils ont indiqué le nom de l'antécédent. Ils avaient le choix entre trois solutions. S'ils identifiaient l'antécédent, ils indiquaient le nom de l'antécédent. Dans le cas de pronoms pluriels à antécédents multiples, ils pouvaient les lier au moyen d'un signe plus. Si aucun antécédent n'était identifié ou si une ambiguïté des antécédents était manifeste, ils notaient un point d'interrogation. 10% des antécédents n'ont pas obtenu d'accord total entre les sujets quelle que soit la réponse pour les anaphores pronominales. Et 3% n'ont pas obtenu d'accord unanime pour les anaphores zéro.

Anaphores pronominales	Précision
Anaphores pronominales consensuelles	86 %
Anaphores pronominales non consensuelles	95%

TAB. 1 – Évaluation des anaphores pronominales

Sur les 671 anaphores pronominales, 18 n'ont pas été prises en compte puisqu'elles ne présentaient pas d'intérêt pour la suite de nos travaux. Ce sont par exemple des anaphores incluses dans une subordonnée relative. 35 anaphores supplémentaires n'ont pas été prises en compte puisqu'elles ont été jugées ambiguës ou sans antécédent par les sujets. Les résultats du tableau 1 indiquent la précision obtenue sur 616 anaphores. Les anaphores consensuelles sont celles dont les sujets ont désigné un antécédent à l'unanimité. Les non consensuelles sont celles où seule la majorité des sujets a désigné l'antécédent attendu. L'essentiel des erreurs a lieu dans des configurations où les relations entre les pronoms et leurs antécédents ne sont pas ambiguës.

Les résultats du tableau 2 portent sur l'ensemble des anaphores pronominales (consensuelles et non consensuelles confondues) et les anaphores zéro. Les résultats semblent satisfaisants lorsqu'on les compare à d'autres travaux sur la résolution anaphorique comme (Trouilleux, 2002) et (Tetreault, 2001). Peut-être, la nature du corpus facilite-t-elle notre tâche. Il s'agit pour nous d'un corpus oral alors que les exemples cités sont appliqués sur des corpus écrits. L'observation de ce corpus montre les textes oraux peuvent être difficiles pour la résolution anaphorique puisque 10% des textes présentent des difficultés de résolution pour les sujets interrogés. Il n'y a par exemple pas d'antécédent pour une partie des pronoms. Cependant lorsque la résolution est possible, il semble que l'information est présentée de façon plus conforme aux prédictions de la théorie du centrage que les textes écrits. A ce titre nous avons favorisé les anaphores interprastiques et la continuité thématique comme la théorie du centrage le préconise. Cela nous semble plus approprié aux corpus oraux. On peut remarquer aussi que la couverture des analyses citées sur les textes écrits est aussi plus large puisqu'elles intègrent une résolution des adjectifs possessifs.

Type d’anaphore	Précision
Anaphores pronominales	87 %
Anaphores zéro	93%

TAB. 2 – Évaluation générale

6 Conclusion

L’intégration de différents niveaux linguistiques au sein de l’analyse de discours permet de profiter des informations des différents traitements pour résoudre des problèmes discursifs. L’analyse de discours met à disposition des informations fonctionnelles issues de constructions syntaxiques spécifiques. Nous avons pu observer que ces informations, comme la fonction sujet, peuvent remonter dans l’analyse et améliorer la résolution de problèmes discursifs comme la résolution anaphorique.

Les approches utilisées, les contraintes, les fonctions et l’approche dynamique se combinent au sein de notre analyse. Elles sont applicables à tous les niveaux d’analyse. Elles ne se confinent pas à un seul de ces niveaux.

L’observation des erreurs nous amène à quelques pistes d’amélioration de l’existant. Par exemple, la résolution des pronoms à antécédents multiples et la gestion de la cataphore n’étaient pas implémentées. Elles ont produit des erreurs. On peut envisager des améliorations en tenant compte de la diversité des fonctions pragmatiques. On pourrait affiner la pondération de la saillance des référents en exploitant mieux les différentes fonctions. On peut continuer à améliorer l’analyse de discours pour contraindre les énoncés afin de diminuer le nombre d’interprétation et simplifier la détermination des résultats.

D’autres erreurs demanderaient un investissement considérable notamment pour les résolutions qui font appel à notre connaissance du monde. Dans la suite des énoncés 10 et 11 l’antécédent de *Il* en 11 nous semble être plus favorablement relié au patient plutôt qu’à l’agent du verbe *jeter*. Et cela quel que soit le poids de *Il* en 10. Cette intuition fait appel à notre connaissance du monde.

(10) *Il le jette.*

(11) *Il tombe.*

D’autres problèmes liés à la résolution anaphorique pourraient se poser si nous changions de format ou de genre de texte. Par exemple les textes analysés ont en moyenne une douzaine de séquences de texte. Si les textes étaient plus longs, la pondération d’un référent du discours pourrait augmenter considérablement et ne plus permettre de transition. Nous devrions prévoir une valeur plafond que la pondération ne pourrait pas dépasser.

Références

ALSHAWI H. (1987). *Memory and context for language interpretation*. New York : Cambridge University Press.

- AMSILI P., DENIS P. & ROUSSARIE L. (2006). Anaphores abstraites en français : représentation formelle. *Traitement Automatique des Langues*, **46**(1), 15–39.
- ASHER N. (1993). *Reference to abstract objects in discourse*. Dordrecht : Kluwer Academic Publishers.
- BEAVER D. (2004). The optimization of discourse anaphora. *Linguistics and philosophy*, **27**(1), 3–56.
- BLACHE P. (2001). *Les Grammaires de Propriétés : des contraintes pour le traitement automatique des langues naturelles*. Paris : Hermès Sciences.
- BLACHE P. (2005). Property grammars : A fully constraint-based theory. In H. CHRISTIANSEN & AL., Eds., *Constraint Solving and Language Processing*. Springer.
- DIK S. C. (1997). *The theory of functional grammar. 2 Volumes. Part 1 : The Structure of the Clause. Part 2 : Complex and Derived Constructions*. Berlin : Mouton de Gruyter.
- GROSZ B., JOSHI A. & WEINSTEIN S. (1995). Centering : a framework for modelling the local coherence of discourse. *Computational Linguistics*, **21**(2), 203–226.
- IDE N. & CRISTEA D. (2000). A hierarchical account of referential accessibility. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, ACL 2000*, p. 416–424, Hong-Kong.
- LAPPIN S. & LEASS H. J. (1994). An algorithm for pronominal anaphora resolution. *Computational Linguistics*, **20**(4), 535–561.
- NOUALI O., RÉGNIER A. & BLACHE P. (2005). Classification de courriers électroniques : une approche par apprentissage basée sur des modèles linguistiques. *Revue d'intelligence artificielle*, **19**(6), 885–912.
- TETREAULT J. R. (2001). A corpus-based evaluation of centering and pronoun resolution. *Computational Linguistics*, **27**(4), 27–4.
- TETREAULT J. R. & ALLEN J. F. (2004). Dialogue structure and pronoun resolution. In COLIBRI, Ed., *Proceedings of DAARC 2004*, p. 7–12, Lisbonne.
- TROUILLEUX F. (2002). A rule based pronoun resolution system for french. In COLIBRI, Ed., *Proceedings of DAARC 2002*, Lisbonne.
- VICTORRI B. (2005). Le calcul de la référence. In HERMÈS, Ed., *Sémantique et traitement automatique du langage naturel*, p. 133–172, Paris.
- VION M. & COLAS A. (2000). Mode de recueil et outil d'analyse d'un corpus de parole spontanée étudié d'un point de vue psycholinguistique. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence*, **19**, 155–167.