

Résolution des références aux documents dans un corpus de dialogues humains

Andrei Popescu-Belis

Université de Genève, ISSCO/TIM/ETI
andrei.popescu-belis@issco.unige.ch

Résumé

Cet article étudie la résolution des références à des entités lorsqu'une représentation informatique de ces entités est disponible. Nous nous intéressons à un corpus de dialogues entre humains, portant sur les grands titres de la presse francophone du jour, et proposons une méthode pour détecter et résoudre les références faites par les locuteurs aux articles des journaux. La détection des expressions nominales qui réfèrent à ces documents est réalisée grâce à une grammaire, alors que le problème de la détection des pronoms qui réfèrent aux documents est abordé par des moyens statistiques. La résolution de ces expressions, à savoir l'attribution des référents, fait quant à elle l'objet d'un algorithme inspiré de la résolution des coréférences. Ces propositions sont évaluées par le biais de mesures quantitatives spécifiques.

Mots-clés : résolution des références, dialogue humain, évaluation quantitative.

Abstract

This article studies the resolution of references to entities in cases when a computational representation of the entities is available. Our data is a corpus of human dialogues about the front pages of one or more daily francophone newspapers. The main goal is to propose a method for the detection and resolution of references made by speakers to newspaper articles. A grammar is used to detect the nominal expressions referring to documents, while statistical methods are used to detect the pronouns referring to documents. An algorithm inspired from coreference resolution is used to solve all these expressions, i.e. to find the documents and articles they refer to. Specific quantitative metrics are applied to evaluate these proposals.

Keywords: reference resolution, human dialogue, quantitative evaluation.

1. Introduction

La compréhension automatique du dialogue entre humains passe par la compréhension des entités auxquelles les locuteurs font référence. Dans cet article, nous étudions les références faites aux documents dans le dialogue oral en situation de réunion. Les réunions de revue de presse auxquelles nous nous intéressons comportent en effet un nombre élevé de références aux journaux et aux articles. Leur compréhension automatique est une étape vers une compréhension plus approfondie de la réunion – par exemple en vue du résumé automatique – et permet également de déterminer, à tout moment de la réunion, le document dont il est question. Par exemple, lorsqu'un locuteur produit l'énoncé suivant : « Le titre de cet article est très étonnant », l'expression *cet article* se réfère à un élément précis d'un document disponible au format électronique. La résolution des références aux documents revient ainsi à identifier chaque expression référentielle (ER) qui fait référence à un élément de document, et à lui associer cet élément.

Nous situerons dans un premier temps ce problème au sein du traitement automatique de langues et résumerons les étapes que nous proposons pour le résoudre (section 2). Après avoir brièvement décrit les données utilisées (section 3), nous énoncerons des solutions pour les étapes de traitement proposées, accompagnées des mesures d'évaluation adaptées : détection des ER nominales (section 4), respectivement pronominales (section 5), puis résolution des ER (section 6). Les résultats obtenus avec différents paramètres de ces méthodes sont ensuite discutés (section 7), suivis d'un aperçu des applications et des perspectives (section 8).

2. Résolution des références à un ensemble connu d'entités

Le rôle cognitif des expressions référentielles dans un discours est de spécifier les entités dont parle le locuteur, ou plus exactement des représentations de ces entités dans l'esprit des interlocuteurs, plus communément appelées entités discursives (Sidner, 1983 ; Reboul *et al.*, 1997). Par conséquent, la résolution des références peut être conçue comme l'opération visant à construire ou à identifier des entités discursives à partir d'une série d'énoncés. La résolution automatique des références nécessite la manipulation par l'ordinateur de représentations des entités discursives, ce qui requiert une quantité notable de connaissances sur le monde et sur la langue. Nous nous intéressons ici aux références produites par les participants à des réunions qui discutent les principaux titres des journaux du jour, disponibles en version numérisée. Après avoir détecté les ER se référant à des documents, en utilisant une transcription des mots prononcés pendant la réunion, leur résolution automatique consiste à associer à chacune des ER un pointeur vers la représentation numérique du document et de l'article correspondant.

2.1. Décomposition du problème

La résolution des références aux documents nécessite deux étapes préparatoires, la transcription des dialogues et la construction de la structure logique des documents, suivies de deux étapes clés qui font l'objet du présent article : la détection des ER se référant aux documents, et la reconnaissance du référent de chaque ER. Afin de ne pas introduire une source d'erreurs significative dans l'évaluation des méthodes que nous proposons, nous utiliserons dans la suite les transcriptions manuelles des dialogues. Nous utiliserons également les structures logiques de référence (correctes) pour ne pas introduire de biais supplémentaire. Ici, les documents discutés lors des réunions sont les premières pages de journaux francophones, principalement *Le Monde* (France), mais aussi *Le Soir* (Belgique), *Le Devoir* (Québec), *La Presse* (Tunisie). Une structure logique comprenant les principaux éléments du contenu peut être dérivée automatiquement du format PDF avec de bonnes performances (Hadjjar *et al.*, 2004 ; Popescu-Belis & Lalanne, 2004, section 3.2). Chaque première page sera représentée comme un ensemble d'*articles*, chaque article comprenant un champ *titre*, un champ *auteur* (lorsque celui-ci est explicite), un champ *contenu* et éventuellement une ou plusieurs *images*.

L'attribution des référents consiste, d'un point de vue informatique, à attribuer à chaque ER de la transcription un pointeur vers le document correspondant, puis vers l'article correspondant parmi ceux de la première page. Une attribution plus précise, que nous n'aborderons pas ici, permettrait d'associer à l'ER la composante de l'article qu'elle désigne (titre, auteur, etc.) lorsque cela est le cas. Nous proposons ici des méthodes pour les deux tâches principales de la résolution des références aux documents, qui sont : (1) l'identification des ER faisant référence aux documents, et (2) l'attribution d'un référent à chacune de ces ER. La séparation de ces deux tâches est courante pour le traitement des coréférences (Hirschman, 1997 ; van Deemter

& Kibble, 2000). Deux cas doivent être distingués pour la première tâche. Nous proposerons d'abord (section 4) une méthode d'identification des ER nominales qui se réfèrent aux documents, fondée sur des règles d'appariement syntaxiques. Cette méthode n'est toutefois pas applicable pour identifier les pronoms qui réfèrent aux documents, puisque ceux-ci ont la même forme que ceux qui réfèrent à d'autres entités, et c'est pourquoi nous proposerons une méthode statistique fondée sur des indices de surface.

2.2. Domaines d'application possibles

La tâche de résolution des références aux entités a déjà été abordée dans plusieurs applications du traitement automatique des langues, par exemple la compréhension de textes ou le dialogue homme-machine. La tâche est à la portée des systèmes informatiques lorsque le contexte est suffisamment simple pour se prêter à une modélisation informatique (Gaizauskas & Humphreys, 1997 ; Landragin *et al.*, 2002) ou bien lorsque les représentations utilisées sont volontairement simplifiées (Luperfoy, 1992 ; Popescu-Belis *et al.*, 1998).

Il faut distinguer les liens référentiels, qui relient une ER à une entité discursive, des liens dits de coréférence et des liens anaphoriques, qui sont des liens entre ER. La coréférence est une relation entre deux ER qui spécifient la même entité (Hirschman, 1997 ; van Deemter & Kibble, 2000), et l'anaphore est une relation asymétrique entre deux ER, dites respectivement ER antécédente et ER anaphorique, par exemple un groupe nominal défini et un pronom qui le reprend (Mitkov, 2002). L'anaphore peut être accompagnée de coréférence, mais elle peut aussi être plus complexe, comme dans le cas de l'anaphore associative ou indirecte. La résolution des références diffère donc de la résolution des coréférences ou des anaphores.

Le cas traité ici présente des similarités avec certains problèmes rencontrés dans la génération de textes, ou dans les dialogues de commande homme-machine, deux cas où les ensembles de référents sont connus du système. Le problème de la génération est l'inverse du nôtre, puisque les ER doivent être composées à partir d'une représentation du référent, y compris pour les références aux documents (Paraboni & van Deemter, 2002). Dans les dialogues de commande, par exemple pour manipuler un environnement virtuel, ce sont seulement les ER du locuteur humain qui doivent être reliées à l'entité discursive qu'elles évoquent, et toute erreur influence immédiatement la suite du dialogue (Landragin *et al.*, 2002).

3. Description des données et des annotations

Le corpus de 22 réunions utilisé ici a été enregistré dans une salle de l'Université de Fribourg (Suisse) équipée d'un ensemble de caméras et de microphones (Lalanne *et al.*, 2004). Chaque réunion mentionne entre un et quatre journaux, la moyenne étant proche de deux ; la durée moyenne est de 15 minutes environ, avec presque toujours quatre participants.

Les dialogues ont été transcrits manuellement et sauvegardés au format XML (Barras *et al.*, 2001), et l'annotation a été enrichie avec les informations référentielles également au format XML (Salmon-Alt & Romary, 2004). Afin de préparer des données pour l'évaluation, nous avons annoté manuellement toutes les ER référant aux documents, et avons codé les documents et articles correspondant à chacune d'entre elles (Popescu-Belis & Lalanne, 2004, section 4). Chaque ER est ainsi encadrée de balises `<er>`, ce qui permet de lui associer un numéro d'index unique au sein du dialogue correspondant, par exemple : `<er id="7">cet article</er>`. Ensuite, un bloc séparé d'annotations à la fin de chaque transcription code le document et

l'article correspondant à chaque ER en utilisant son numéro d'index, par exemple : `<ref er-id="7" doc-file="LeMonde030404.xml" doc-id="//Article[@ID='3']"/>`.

Les dialogues contiennent 437 ER se référant aux articles ; ces expressions sont donc relativement fréquentes dans ce type de réunion. Une analyse fréquentielle des types d'ER observés (Popescu-Belis & Lalanne, 2004, section 5.1) indique une préférence pour les mots du domaine journalistique et les noms propres des journaux : « la une du Monde », « le journal », « l'article suivant », « le grand titre », « un autre point », « le dessin ». On trouve également un nombre important de pronoms personnels et démonstratifs : « ils », « il », « celui-là ». Nous avons considéré que les usages quasi impersonnels des pronoms *il* ou *ils*, comme dans « ils disent que... », faisaient référence à l'auteur de l'article et donc, par métonymie, à l'article respectif.

L'accord entre annotateurs pour les liens ER/articles est élevé. Après concertation, nous avons pu atteindre un accord complet pour l'annotation des documents, et un accord d'environ 97 % pour les liens ER/articles (Popescu-Belis & Lalanne, 2004, section 4.2.3).

4. Détection des ER nominales

4.1. Construction et application de la grammaire

Les connaissances linguistiques sur la forme des ER nominales, combinées avec les observations faites sur le corpus, nous ont permis de construire et de tester une grammaire d'identification des ER nominales référant aux documents.

Nous avons utilisé l'environnement XML CLaRK (Simov *et al.*, 2004) pour définir une grammaire fondée sur des règles d'appariement de patrons en vue de reconnaître les ER référant aux documents, en fonction de leurs constituants et des mots du voisinage. CLaRK permet aussi d'effectuer préalablement la segmentation en mots, tout en conservant les annotations XML préexistantes, et en marquant les ER reconnues avec des balises XML. La grammaire que nous avons écrite consiste en 25 règles environ, exprimées à l'aide d'expressions régulières. Par exemple, une série de mots qui correspond à l'expression « (un)? (premier|dernier|petit|autre)? point » sera automatiquement entourée des balises `<er>` et `</er>`. Nombre de ces règles contiennent des disjonctions logiques et des items optionnels, et c'est pourquoi les règles couvrent plusieurs centaines de formes d'ER possibles. Nous avons testé plusieurs variantes de notre grammaire, et avons choisi une version qui privilégie le taux de rappel – l'une des mesures de performance utilisées.

4.2. Mesure d'évaluation pour la détection

Évaluer la détection des ER est un tâche aisée seulement si on la définit comme la comparaison exacte (tout ou rien) des limites des ER trouvées par la grammaire par rapport aux limites correctes. Or, il est parfois trop restrictif de fixer strictement les limites « correctes » d'une ER. Par exemple, lorsque l'ER comporte une proposition relative, détecter la relative ou bien seulement la tête nominale de l'ER peut convenir également du point de vue de la résolution des références.

Il paraît donc utile d'autoriser une certaine variabilité des limites des ER, comme dans les conventions des évaluations MUC-7 (Hirschman, 1997), avec la différence que l'annotation de référence de MUC-7 contenait les limites maximales *et* minimales autorisées pour considérer qu'une ER a été détectée par le système, alors que nous fixons seulement les frontières maxi-

males. Ainsi, si la grammaire détecte une ER à l'intérieur des frontières d'une ER correcte, le programme d'évaluation considère que l'ER a bien été trouvée. Si aucun fragment n'est détecté à l'intérieur d'une ER correcte, le programme compte une erreur de *rappel*. Si une frontière d'ER est détectée à l'extérieur de toute ER correcte (même si elle en intersecte une), elle sera comptée comme une erreur de *précision*. Les taux de rappel et de précision sont ensuite normalisés respectivement par le nombre total d'ER correctes, et par le nombre total d'ER trouvées par le système. Enfin, la méthode doit être ajustée pour prendre en compte le cas des ER imbriquées. Un sous-produit de cette méthode est ainsi la construction d'une correspondance entre les ER correctes et celles détectées, qui sera utilisée pour l'évaluation (section 6.2). Les résultats obtenus pour la détection des ER, utilisant ces mesures, sont discutés dans la section 7.1.

5. Détection des pronoms qui réfèrent aux documents

On rencontre dans le corpus environ 90 occurrences de *il* et *ils* qui réfèrent aux documents/articles ou à leurs auteurs, et environ 830 occurrences de ces mêmes pronoms qui ne réfèrent pas aux documents. Or, la méthode de détection des ER nominales décrite ci-dessus (section 4.1) s'applique difficilement à la détection des pronoms qui réfèrent aux documents, car ceux-ci ne changent pas de forme selon leur référent, et, en outre, les constructions syntaxiques dans lesquelles ceux-ci interviennent présentent une plus grande variabilité que dans le cas des ER nominales. Il faudrait donc construire manuellement un très grand ensemble de règles pour modéliser les collocations caractérisant les pronoms se référant aux documents.

Une possibilité de désambiguïsation des pronoms pourrait passer par la séparation de tous les pronoms impersonnels (Danlos, 2005), puis par la résolution de toutes les anaphores pronominales dans les dialogues (Strube & Müller, 2003) afin de distinguer *via* les antécédents les pronoms qui réfèrent aux documents. La complexité de cette méthode et son taux d'erreur probable ne justifient pas son emploi dans notre cas.

Nous construisons alors de façon automatique des règles de classification des pronoms *il* et *ils* fondées sur la présence ou non de certains mots indices dans leur voisinage – à savoir dans une fenêtre de N mots centrée sur le pronom. Afin de laisser au système statistique le choix de ces mots indices, nous commençons l'entraînement avec, comme traits, tous les mots observés au moins deux fois dans le corpus au voisinage des pronoms *il* et *ils*, et avec $N = 9$. Plusieurs façons de représenter les co-occurrences des mots-indices avec les pronoms sont étudiées, notamment le codage précis de la position – numérique (entre -4 et 4) ou symbolique (e.g. 'pos+2') – ou bien le codage de la simple présence ou absence du mot-indice, sans spécifier sa position. La largeur N de la fenêtre peut varier aussi, entre 3 et 9 mots, *i.e.* de 1 à 4 mots avant et après le pronom.

Pour la tâche de classification des pronoms selon qu'ils réfèrent à un document ou non, nous étudions trois familles de méthodes : arbres de décision, k plus proches voisins, et machines à vecteur support, grâce au système Weka (Witten & Frank, 2000). Toutes ces méthodes sélectionnent implicitement, après apprentissage, les traits utiles à la classification, comme nous le verrons avec les résultats dans la section 7.2.

6. Résolution des ER : l'attribution des référents

6.1. Description de l'algorithme

Nous avons implémenté un algorithme inspiré de nos travaux sur la coréférence (Popescu-Belis, 2003) afin de calculer, pour chaque ER détectée, le document et l'article qui est le plus probablement son référent. L'algorithme gère deux variables globales, qui sont le journal courant et l'article courant – une simplification des « éléments focaux » (Grosz *et al.*, 1995). L'algorithme résout les ER par ordre d'apparition dans la transcription. Pour chaque ER, on cherche d'abord le document auquel cette ER fait référence, si plusieurs documents ont été discutés lors de la réunion. On considère qu'un changement du journal courant peut s'opérer seulement par l'usage d'une ER citant le nom du journal, auquel cas l'ER reçoit ce document comme référent. Sinon, le journal courant reste le même, et sera associé à l'ER courante. Pour la première ER d'une réunion, si celle-ci ne contient pas le nom d'un journal, le journal le plus fréquent (*Le Monde*) lui sera associé par défaut.

Pour ce qui est de l'attribution de l'article à chaque ER, notre algorithme possède une liste d'ER typiquement anaphoriques, par exemple *il*, *l'article* (sans spécifieur), *cet article*, *l'auteur* (sans spécifieur), etc. Si l'ER à traiter figure dans cette liste, son référent sera l'article courant. Si elle n'y figure pas, elle est considérée comme non-anaphorique, et l'algorithme cherche dans le contexte voisin de l'ER des indices lexicaux, puis teste les appariements possibles entre ce contexte et les différents articles du journal courant. L'article qui présente le maximum de similarités lexicales est choisi comme référent de l'ER, et devient aussi l'article courant. Les importances relatives des différentes similarités lexicales – par exemple entre les mots de l'ER et les mots du titre, ou ceux de l'article – peuvent être analysées de façon expérimentale, de même que la taille du contexte droit/gauche à examiner pour l'appariement (section 7.3). Naturellement, ces heuristiques simples ne prétendent pas couvrir toute la variabilité du phénomène de la référence, mais – comme nous le verrons à la section 7.3 – fournissent de bons résultats pour l'attribution des référents.

6.2. Évaluation de la résolution des ER

Contrairement à la résolution des coréférences (Popescu-Belis, 2003), la résolution des références aux documents peut être évaluée simplement en termes de *correction*, en examinant si chaque ER est ou non correctement liée au journal et à l'article auquel elle fait référence. On peut calculer un pourcentage de correction pour le rattachement ER/document et un autre taux, nécessairement inférieur, pour le rattachement ER/article.

Toutefois, lorsque la résolution des ER est combinée avec la détection automatique des ER, l'évaluation de la résolution doit prendre en compte les « bruits » et les « silences » du processus de détection. Les ER surnuméraires (bruit) ne peuvent être évaluées en termes de références aux documents, puisqu'en réalité elles n'y réfèrent pas, donc nous les ignorons dans le calcul de la correction ajustée. En revanche, les ER correctes ignorées (silence) sont comptées dans la correction ajustée comme des liens manquants, donc elles diminuent le taux d'ER correctement rattachées, bien qu'elles n'aient pas été vues par le résolveur. L'implémentation de l'évaluateur se fonde nécessairement sur l'algorithme d'appariement des ER détectées avec celles correctes mentionné dans la section 4.2, et calcule le taux de liens corrects pour la tâche de rattachement ER/document et ER/article.

7. Analyse des résultats obtenus

7.1. Performances pour la détection des ER nominales

La performance initiale de la grammaire de reconnaissance des ER est, en termes de rappel (R), précision (P) et F-mesure : $R = 0.65$, $P = 0.85$ et $F = 0.74$. L'analyse expérimentale permet d'évaluer le mérite de certaines règles d'appariement. Par exemple, en ajoutant une règle qui marque tous les pronoms de troisième personne comme se référant aux documents, la précision et la F-mesure diminuent nettement : $R = 0.71$, $P = 0.52$ et $F = 0.60$. Ce fait justifie une méthode de détection distincte pour les pronoms qui se réfèrent à des documents. De même, ajouter une règle qui marque tous les démonstratifs comme se référant à des documents décroît encore les performances : $R = 0.70$, $P = 0.46$ et $F = 0.56$. Il apparaît cependant que pour le présent corpus de réunions, les démonstratifs *celui-ci* et *celui-là* sont presque toujours employés pour faire référence à des articles. Par conséquent, les meilleurs scores sont obtenus après avoir ajouté cette règle de reconnaissance des démonstratifs : $R = 0.68$, $P = 0.88$ et $F = 0.76$. Le rappel et la précision de la grammaire pourront être mieux évalués et améliorés lorsque davantage de données de test seront disponibles, pour éviter une « sur-adaptation » à un corpus donné.

7.2. Performances pour la détection des pronoms référant aux documents

Les premières expériences avec l'apprentissage automatique pour la détection des pronoms ont permis de départager certaines méthodes et certains traits utilisés, en dépit de la faible quantité de données (environ 90 exemples positifs et 830 négatifs). Nous utilisons une méthode de validation croisée, fondée sur dix tirages aléatoires séparés de 80 % des données comme jeu d'entraînement et 20 % comme jeu de test, en respectant la proportion d'exemples positifs et négatifs, ce qui nous permet de calculer les scores moyens et leurs variances.

Les meilleurs scores sont obtenus par les machines à vecteurs support (SVM), une méthode connue pour ses bonnes performances même avec une quantité réduite de données d'entraînement. Parmi les différentes façons d'utiliser les mots voisins des pronoms comme indices, on observe des scores semblables pour toutes celles qui préservent l'information de position des mots, et qui considèrent une fenêtre d'au moins deux mots avant et après le pronom. Ce sont donc surtout les mots adjacents qui permettent la désambiguïsation des pronoms ; les indicateurs les plus fiables sont les constructions « ils parlent » et « ils disent », utilisées le plus souvent à valeur impersonnelle pour évoquer le contenu d'un article.

Les valeurs des scores pour la détection des pronoms *il* ou *ils* référant aux documents peuvent être exprimés en termes de rappel (R) et de précision (P), donc de F-mesure, ou bien en termes de κ (*kappa*), métrique qui tient compte de la possibilité de détection par hasard ; cette possibilité est ici assez élevée, étant donné la grande asymétrie entre les deux classes. Les meilleurs scores obtenus, avec une fenêtre de cinq mots centrée sur les pronoms, sont : $\kappa = 0.60 \pm 0.07$, $R = 0.70 \pm 0.10$, $P = 0.60 \pm 0.06$ et $F = 0.64 \pm 0.06$. Ces performances sont encourageantes, notamment si l'on observe que κ est proche de la limite inférieure des scores « acceptables » d'accord entre juges humains.

La machine à vecteurs support obtenant ces scores utilise un noyau polynomial de degré 1. La variation du degré, celle des coefficients maximaux, ou le changement du noyau en une fonction RBF, ne permettent pas d'améliorer les performances de cette méthode de classification. L'utilisation des arbres de décision (construits par l'algorithme C4.5) ou la méthode des k plus

proches voisins ne permettent pas d'améliorer les scores ; les meilleurs arbres de décision ont des scores comparables aux SVM.

7.3. Performances pour la résolution des ER

L'algorithme de résolution des ER, appliqué d'abord aux ER *correctes*, atteint 97 % de correction pour l'identification des documents auxquels réfèrent les ER, à savoir 428 ER sur 437. Si l'on examine seulement les réunions où sont discutés plusieurs documents, la correction est de 93 %. Pour ce qui est de l'identification correcte des articles, le score est de 67 %, à savoir 303 ER sur 437. Si nous comptons seulement le score des ER pour lesquelles le document a été correctement identifié au préalable, la correction est de 68 % (301 ER sur 428) ; le score change peu puisque la plupart des associations ER/document sont de toute façon correctement résolues. Ces scores élevés prouvent la pertinence de notre algorithme, surtout si on les compare à ceux de méthodes triviales (*baseline*) : par exemple, si on choisit constamment l'article principal de la première page comme référent de l'ER, la correction n'est que de 18 % environ.

Les meilleurs scores sont obtenus lorsque seul le contexte droit des ER (les mots suivant l'ER) est considéré pour l'appariement avec les articles. Empiriquement, on trouve que le nombre optimal de mots à rechercher dans le contexte droit est d'environ dix. De plus, un appariement entre une ER et le titre d'un article a une valeur prédictive beaucoup plus importante qu'un appariement entre des mots du contexte de l'ER et des mots du corps de l'article. Enfin, si l'on supprime le mécanisme de suivi du « document courant » (la mémoire de l'élément focal), la correction de l'association ER/article tombe à 60 %, ce qui illustre l'utilité de ce mécanisme.

7.4. Performances combinées

Lorsque les modules de détection et de résolution s'enchaînent, leurs erreurs se combinent d'une manière *a priori* imprévisible. Les résultats du module de résolution qui opère sur les résultats du module de détection sont bien inférieurs à une combinaison multiplicative des taux d'erreurs : la correction des liens ER/document atteint 60 % (265 ER sur 437) et la correction des liens ER/article atteint seulement 32 % (141 ER sur 437). Si l'on calcule cette dernière correction seulement sur le sous-ensemble des ER correctement attachées à leur document, le score atteint 46 % (123 ER sur 265). Si la combinaison des taux d'erreur avait été multiplicative, les scores auraient dû atteindre, respectivement, environ 73 % et environ 50 %. L'enchaînement des modules conduit ainsi à une dégradation accrue des performances, et cela même sans avoir introduit dans la chaîne de traitement les erreurs (probables) de reconnaissance vocale.

La baisse des scores est probablement due à la nature de l'algorithme de résolution des ER, qui est fondé sur la mémorisation de l'article courant. Ainsi, la résolution correcte de chaque ER dépend souvent de la résolution correcte de la précédente. Or, nous n'avons pu introduire à ce stade les résultats du détecteur de pronoms dans l'enchaînement des modules, puisque ce détecteur a été développé par entraînement sur ces mêmes données¹. Par conséquent, une quantité importante d'ER correctes ne sont pas détectées, notamment les pronoms, et donc l'algorithme de résolution devient incapable de suivre correctement l'article courant.

¹ Il n'aurait donc pas été correct, d'un point de vue méthodologique, d'introduire ce module dans la chaîne et de l'évaluer sur les données d'entraînement. Or, la quantité de données ne permet pas une séparation entre les données d'entraînement et les données de test.

8. Applications et perspectives

La résolution des références aux documents permet de construire des liens entre deux modalités importantes pour l'enregistrement des réunions : la parole et les documents écrits. Les liens référentiels des dialogues vers les documents permettent d'aligner ces deux modalités, en d'autres termes, de préciser à chaque instant du dialogue quels sont les documents et les articles discutés. Naturellement, étant donné la relative rareté des ER se référant aux documents, d'autres méthodes sont nécessaires pour réaliser un alignement exhaustif, tels l'alignement thématique et l'alignement fondé sur les citations (Lalanne *et al.*, 2004 ; Lalanne & Ingold, 2004). Les références aux documents constituent alors l'une des catégories d'indices possibles.

Du point de vue de la résolution de la référence, les méthodes proposées ici peuvent plus généralement s'appliquer à tout problème de résolution des ER sur un domaine de référence limité, connu à l'avance. Une extension possible consiste à déterminer automatiquement tous les référents mentionnés dans les documents (notamment les personnages humains), puis à résoudre toutes les ER s'y référant. Ces recherches peuvent également trouver une application au problème de la structuration des dialogues, par exemple à la segmentation thématique, en utilisant les références aux documents comme des indices de cohérence thématique. Les applications dans le domaine du traitement des dialogues par l'ordinateur, pourront permettre ainsi, à terme, l'intégration d'une base de données multimodale de réunions au sein d'un système d'information d'entreprise.

Références

- BARRAS C., GEOFFROIS E., WU Z. & LIBERMAN M. (2001). Transcriber : development and use of a tool for assisting speech corpora production. *Speech Comm.*, **33**(1-2), 5–22.
- DANLOS L. (2005). ILIMP : Outil pour repérer les occurrences du pronom impersonnel 'il'. In M. JARDINO, Ed., *Actes de TALN 2005 (Traitement automatique des langues naturelles)*, p. 123–132, Dourdan : ATALA / LIMSI.
- GAIZAUSKAS R. & HUMPHREYS K. (1997). Using a semantic network for information extraction. *Natural Language Engineering*, **3**(2-3), 147–169.
- GROSZ B. J., JOSHI A. K. & WEINSTEIN S. (1995). Centering : A framework for modeling the local coherence of discourse. *Computational Linguistics*, **21**(2), 203–225.
- HADJAR K., RIGAMONTI M., LALANNE D. & INGOLD R. (2004). Xed : a new tool for extracting hidden structures from electronic documents. In *Workshop on Document Image Analysis for Libraries*, Palo Alto, CA.
- HIRSCHMAN L. (1997). *MUC-7 Coreference Task Definition*. Rapport MITRE Corp.
- LALANNE D. & INGOLD R. (2004). Documents statiques et multimodalité : l'alignement temporel pour structurer des archives multimédias de réunions. *Document numérique*, **8**(4), 65–89.
- LALANNE D., MEKHALDI D. & INGOLD R. (2004). Talking about documents : revealing a missing link to multimedia meeting archives. In *Document Recognition & Retrieval XI, IS&T/SPIE's Annual Symposium on Electronic Imaging*, San Jose, CA.
- LANDRAGIN F., SALMON-ALT S. & ROMARY L. (2002). Ancrage référentiel en situation de dialogue. *Traitement Automatique des Langues*, **43**(2), 99–130.
- LUPERFOY S. (1992). The representation of multimodal user interface dialogues using discourse pegs. In *ACL 1992*, p. 22–31, Newark, DE.
- MITKOV R. (2002). *Anaphora Resolution*. Londres : Longman.

- PARABONI I. & VAN DEEMTER K. (2002). Towards the generation of document deictic references. In K. VAN DEEMTER & R. KIBBLE, Eds., *Information Sharing : Reference and Pre-supposition in Language Generation and Interpretation*, p. 329–352. Stanford, CA : CSLI.
- POPESCU-BELIS A. (2003). Evaluation-driven design of a robust reference resolution system. *Natural Language Engineering*, **9**(3), 281–306.
- POPESCU-BELIS A. & LALANNE D. (2004). Reference resolution over a restricted domain : References to documents. In *ACL 2004 Workshop on Reference Resolution and its Applications*, p. 71–78, Barcelone.
- POPESCU-BELIS A., ROBBA I. & SABAH G. (1998). Reference resolution beyond coreference : a conceptual frame and its application. In *Coling-ACL 1998*, p. 1046–1052, Montréal.
- REBOUL A., BRIFFAULT X., BALKANSKI C., GAIFFE B., POPESCU-BELIS A., ROBBA I., ROMARY L. & SABAH G. (1997). *Le projet CERVICAL : Representations mentales, référence aux objets et aux événements*. Rapport interne, LORIA, Nancy, et LIMSI, Orsay.
- SALMON-ALT S. & ROMARY L. (2004). RAF : Towards a reference annotation framework. In *LREC 2004*, p. 119–122, Lisbonne.
- SIDNER C. (1983). Focusing in the comprehension of definite anaphora. In M. BRADY & R. BERWICK, Eds., *Computational Models of Discourse*, p. 267–330. Cambridge : MIT Press.
- SIMOV K., SIMOV A., GANEV H., IVANOVA K. & GRIGOROV I. (2004). The CLaRK system : XML-based corpora development system for rapid prototyping. In *LREC 2004*, p. 235–238, Lisbonne.
- STRUBE M. & MÜLLER C. (2003). A machine learning approach to pronoun resolution in spoken dialogue. In *ACL 2003*, p. 168–175, Sapporo.
- VAN DEEMTER K. & KIBBLE R. (2000). On coreferring : Coreference in MUC and related annotation schemes. *Computational Linguistics*, **26**(4), 629–637.
- WITTEN I. & FRANK E. (2000). *Data Mining : Practical Machine Learning Tools with Java Implementations*. San Francisco, CA : Morgan Kaufmann.