

# Méthode de réordonnancement de réponses par transformation d'arbres : présentation et analyse des résultats

Guillaume Bernard<sup>1</sup>

(1) LIMSI - CNRS, B.P. 133, 91403 ORSAY, France

gbernard@limsi.fr

**Résumé.** Dans cet article nous présentons une évaluation et une analyse des résultats d'une méthode de réordonnancement de réponses pour un système de questions-réponses. Cette méthode propose une sélection des réponses candidates à une question en calculant un coût par transformation d'arbres. Nous présentons une analyse des résultats obtenus sur le corpus Clef 2004-2005 et nos conclusions sur les voies d'amélioration possibles pour notre système.

**Abstract.** This paper describes an evaluation and an analysis of the results of an answers reranking method for a question-answering system. Candidate answers to a question are reordered by computing a tree transform cost. We discuss the results of our evaluation on the Clef 2004-2005 corpus and describe possible improvements to the system.

**Mots-clés :** Recherche d'information, Traitement Automatique des Langues, systèmes de questions-réponses.

**Keywords:** Information Retrieval, Natural Language Processing, Question-Answering systems.

## 1 Introduction

Il est fréquent que l'information recherchée pour répondre à une question soit exprimée en des termes différents de ceux de la question. Les techniques de recherche d'information traditionnelles sont ainsi souvent pénalisées lorsqu'il s'agit de détecter les variations linguistiques telles que la paraphrase (Ligozat, 2006). Notre objectif est de répondre à cette problématique en proposant une méthode de réordonnancement des réponses candidates à une question en utilisant des arbres représentant la question et un texte contenant potentiellement la bonne réponse (Bernard, 2008). Un coût de transformation est calculé entre ces deux arbres. D'autres méthodes se sont déjà appuyées sur des représentations des textes sous forme d'arbres pour identifier la réponse à une question (par exemple (Ligozat, 2005)), qui se basent sur la structure et les dépendances d'arbres syntaxiques.

L'objectif de cet article est de présenter une analyse de cette méthode. Nous présentons le contexte d'application de notre approche, le système Ritel, puis nous décrivons brièvement le fonctionnement de notre méthode. En nous appuyant sur les résultats obtenus lors d'une évaluation précédente, nous proposons une analyse de notre méthode et les conclusions que nous en avons tiré.

## 2 Ritel, un système de questions-réponses oral et interactif

Le projet Ritel (Rosset *et al.*, 2006) du LIMSI sert de contexte pour notre travail. L'un des objectifs de Ritel est de réunir les fonctionnalités d'un système de questions-réponses et d'un système de dialogue oral. Dans cet article nous nous concentrons sur le composant questions-réponses de Ritel. Deux modules du système sont directement liés à notre travail : le module d'analyse des documents et des questions, qui transforme les textes de notre corpus et les requêtes de l'utilisateur en arbres textuels, et le module de recherche d'information.

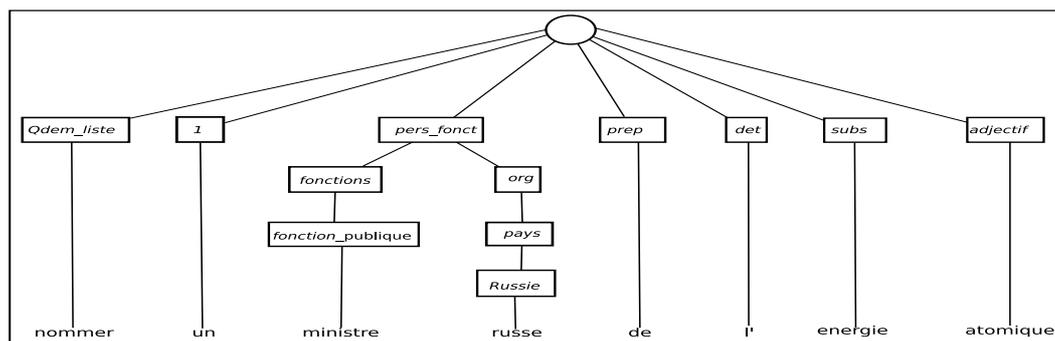


FIG. 1 – Exemple d'arbre construit pour «Nommer un ministre russe de l'énergie atomique.»

L'analyse produite par Ritel est non contextuelle (Villaneau *et al.*, 2007). Concrètement, nous n'utilisons aucune information sur le dialogue pour typer nos textes, et chaque phrase des documents est traitée séparément. Le but est de typer les mots ou groupes de mots en procédant à des regroupements selon le sens de ces mots. Les types des nœuds peuvent appartenir à différentes catégories : entités nommées (une personne, un lieu, etc.), entités grammaticales (des verbes, des prépositions), ou entités spécifiques (par exemple des couleurs, comme “rouge”). Plus généralement, les informations contenues dans les nœuds de nos arbres sont plus d'ordre sémantique que syntaxique. L'analyse est faite de façon hiérarchique : un groupe de mots sera décomposé en plusieurs sous-nœuds typés, et l'ensemble de ces sous-nœuds sera typé par un nœud global. La structure des arbres textuels résultant de l'analyse est plutôt étendue (plus large que profonde). Les arbres ont en moyenne 3 à 4 niveaux de profondeur. Nous donnons dans la figure 1 un exemple de l'arbre textuel. On peut voir que les informations contenues dans les nœuds sont sémantiques (*pers\_fonct*) et syntaxiques (*prep*). Les nœuds sont cependant plus souvent typés sémantiquement que syntaxiquement. Par ailleurs, le découpage d'un nœud en sous-nœud est basé sur des structures sémantiques.

Le système RI/QA (Rosset *et al.*, 2007) de Ritel est composé de trois modules : la sélection des documents intéressants, la sélection des passages de texte contenant potentiellement la réponse à la question traitée, et la sélection des réponses candidates. Au début de la recherche, la question est analysée pour déterminer le type de réponse attendu et le système génère un Descripteur De Recherche (DDR) qui permet de guider la recherche de la réponse. Les DDRs représentent les éléments critiques et secondaires de la question ainsi que le type attendu de réponse. Les éléments critiques correspondent aux nœuds importants de la question, et les éléments secondaires aux nœuds dont la présence est plus optionnelle. Dans la figure 1, les éléments critiques seraient “ministre russe” et “atomique”, et “énergie” serait un élément secondaire. Chacune de ces trois étapes utilise les informations fournies par le DDR. Le premier module calcule un score pour chaque document selon les différentes entités du DDR et retourne une liste de documents. Le système sélectionne alors les passages des documents pouvant contenir la réponse et retourne

une liste de passages ordonnée. Enfin, à partir de cette liste de passages, le système retourne une liste de réponses courtes extraites de ces passages. Les réponses sont ordonnées selon un score.

Pour qu'un extrait de texte soit identifié comme pouvant contenir la réponse à la question, il doit contenir les éléments critiques du DDR. S'il contient les éléments secondaires, le score global du passage sera plus élevé. Cette méthode est comparable à une approche par sacs de mots typés et transformables. Les scores calculés par le système sont proches des scores de compacités (Gillard *et al.*, 2007). Une des limitations des DDRs est qu'ils ne conservent pas toutes les informations. Ainsi, les éléments critiques et secondaires sont conservés, mais pas les prépositions par exemple. Cependant, le problème principal des DDRs reste la perte de la structure de la question : les mots sont classés par importance (critique ou secondaire) et il n'y a pas de gestion du contexte de chaque mot ou de relation avec les autres éléments. Dans l'exemple suivant, les réponses évaluées sont "2005" et "1992".

*Question : En quelle année l'ouragan Andrew s'est-il manifesté ?*

*Texte : Avec Katrina en 2005 l'ouragan Andrew, qui a frappé la Floride en 1992, a été l'un des cyclones les plus destructeurs.*

Le système actuel pourrait donner "2005" comme bonne réponse car elle est proche de l'élément critique "ouragan Andrew" et que "2005" est fréquent dans les différents documents du corpus. Or, avec une gestion des relations entre les différents éléments du texte, on peut déduire que le pronom interrogatif "qui" est lié à "ouragan Andrew" et que la bonne réponse est 1992. Notre objectif est de proposer une méthode pouvant répondre à ces problèmes. Notre méthode intervient en sortie du système RI/QR. Nous récupérons en entrée la question posée par l'utilisateur sous la forme d'un arbre typé, la liste des réponses candidates ordonnée selon un score et la liste des passages du texte d'où les réponses ont été extraites.

### 3 Réordonnement des réponses candidates

Plusieurs systèmes de questions-réponses utilisent des approches à base d'arbres syntaxiques pour répondre aux questions. Les documents et les questions posées par l'utilisateur sont ainsi transformés en arbres structurés. Les nœuds des arbres peuvent être enrichis avec des informations sémantiques, voire des relations sémantiques avec d'autres nœuds de l'arbre. Ces méthodes ont l'avantage de faciliter la comparaison entre un texte et une question et de repérer les mots ou groupes de mots en relation dans les deux arbres. On parle généralement de *points d'ancrage*. Par ailleurs, ce type d'approche a l'avantage de faciliter la manipulation des textes, par le biais de la manipulation des nœuds des arbres (Magnini & Kouylekov, 2006). Ces méthodes sont de ce fait adaptées pour générer des paraphrases (Braz *et al.*, 2005). Notre méthode est basée sur le calcul du coût d'une transformation entre l'arbre de la question et l'arbre du passage contenant la réponse. Nous comparons les deux arbres pour trouver la suite d'opérations permettant de transformer l'arbre du texte en l'arbre de la question avec le plus faible coût. Les extraits des documents étant souvent très longs, nous avons décidé d'extraire le segment contenant la réponse candidate. Pour le moment, nous nous servons de la ponctuation pour délimiter ce segment. Nous procédons également à des modifications sur l'arbre de la question. Pour chaque réponse avec le score le plus fort, le passage du texte d'où elle a été extraite est connu. On supprime alors de l'arbre de la question les nœuds correspondant aux pronoms interrogatifs («qui», «comment», etc.) et on ajoute un nœud avec la réponse à évaluer.

Des opérateurs de manipulation d'arbres sont utilisés. Cette approche est similaire à celle présentée par (Magnini & Kouylekov, 2006), avec deux différences fondamentales, concernant les opérateurs de transformation et les fonctions de coût associées aux opérateurs. Nous utilisons

quatre opérateurs de transformation : suppression d'un nœud, insertion d'un nœud et substitution d'un nœud par un autre, déjà utilisés par Magnini et Kouylekov, auxquels nous ajoutons l'opérateur de déplacement, notre objectif étant de couvrir davantage de phénomènes linguistiques. Le passage d'une phrase en forme active à la forme passive peut être mieux couvert par l'opérateur de déplacement. Nous déterminons les éléments à supprimer en identifiant les relations entre les nœuds du texte et de la question : équivalence pour des nœuds identiques, équivalence de lemmes si la forme lemmatisée des deux nœuds est identique et de synonymie<sup>1</sup>. A chacun des opérateurs de transformation est associée une fonction de coût. Magnini et Kouylekov utilisent des arbres syntaxiques pour représenter leurs textes, qui ont la caractéristique d'être très structurés. Leurs fonctions de coût sont basées sur cette structure. Les arbres typés retournés par l'analyseur de Ritel sont moins profonds et fournissent des informations plus sémantiques que syntaxiques.

Nous avons choisi d'avoir une approche basée sur le contexte global de nos arbres. Nous voulons tenir compte de l'ensemble du sens de la phrase lors de l'application d'une opération. Notre objectif est d'avoir un coût dépendant du contexte et du changement linguistique impliqué par l'opération. Dans l'exemple suivant, il serait coûteux de supprimer "vice" car cela change complètement le sens de l'extrait du texte :

*Question : Nommer un ministre russe de l'énergie atomique.*

*Texte : et aux règles de non-prolifération des armes nucléaires, a déclaré le vice-ministre de l'énergie atomique Viktor Mikhailov.*

Les fonctions de coût sont pour le moment assez simples. Le contexte d'une opération est représenté par le type des nœuds voisins gauche et droite. La structure générale de nos fonctions est identique : une suite de conditions portant sur le type du nœud sur lequel est appliqué l'opérateur, le coût de chaque branche étant déterminé par le contexte du nœud.

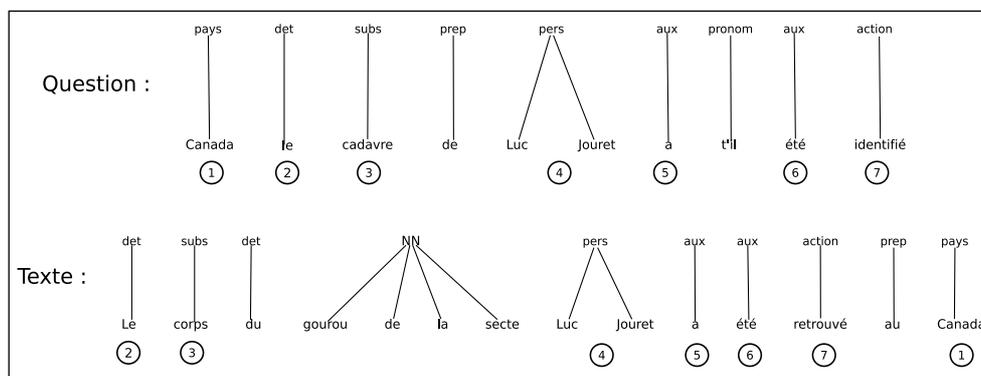


FIG. 2 – Points d'ancrage entre les arbres de la question et le texte - Points d'ancrage 1, 2, 4, 5 et 6, nœuds identiques - Points d'ancrage 3 et 7, relation de synonymie

La figure 2 illustre le fonctionnement de notre approche. Dans un souci de lisibilité, les arbres sont représentés dans une forme simplifiée (pas de sous-nœuds et de nœud racine). Dans le contexte de l'exemple, on compare le texte *Le corps du gourou de la secte Luc Jouret a été retrouvé au Canada* avec la question *Où le cadavre de Luc Jouret a-t-il été identifié ?*. La réponse à évaluer est "Canada", que l'on a insérée à la place de "Où", qui est un pronom locatif. Les

<sup>1</sup>On dispose d'un dictionnaire de synonymes de 15 000 entrées qui est une ressource interne au LIMSI. A chaque entrée sont associés ses synonymes triés selon leur sens général. La synonymie est détectée en comparant les synonymes d'un nœud avec les synonymes de l'autre nœud.

points numérotés sur la figure 2 représentent les points d'ancrage entre les deux arbres. Ainsi, les points 1, 2, 4, 5 et 6 sont des relations de nœuds identiques, tandis que les points 3 et 7 sont des relations de synonymie. À partir de ces points d'ancrage, notre système peut calculer les suites d'opérations pour transformer l'arbre du texte en l'arbre de la question. Les fonctions de coût associées à chaque opérateur permettent de déterminer la suite d'opérations la moins coûteuse. Dans le cas présent, une suite d'opération possible est :

- 1) Suppression des nœuds *du*, *gourou de la secte* et *au*.
- 2) Déplacement de *Canada* avant le nœud *le*.
- 3) Substitution synonymique du nœud *corps* par le nœud *cadavre*.
- 4) Substitution synonymique du nœud *retrouvé* par *identifié*.
- 5) Insertion du nœud *de* avant le nœud *Luc Jouret* et du nœud *t'il* avant le nœud *été*.

## 4 Évaluation du système

### 4.1 Cadre expérimental

Type de la réponse	#questions	Méthode proposée		Système Ritel		Rappel
		Acc	MRR	Acc	MRR	
raison	15	21.4%	0.24	21.4%	0.27	50.0%
couleur	1	0.0%	0.33	0.0%	0.33	100.0%
lieu	57	29.1%	0.4	54.5%	0.60	72.7%
nombre	42	24.8%	0.35	50.0%	0.57	72.5%
organisation	31	18.4%	0.27	47.4%	0.51	60.5%
personne	62	25.8%	0.35	36.4%	0.42	63.6%
définition d'une organisation	30	30.0%	0.44	46.7%	0.6	86.7%
définition d'une personne	41	17.6%	0.32	47.1%	0.55	73.5%
inconnu	49	17.1%	0.22	26.8%	0.31	53.7%
date	50	37.1%	0.40	47.9%	0.57	72.9%
divers	16	0.0%	0.15	18.7%	0.25	50.0%
total	395	24.0%	0.34	41.5%	0.48	65.8%

TAB. 1 – Résultats de l'évaluation selon la catégorie de la question

Nous avons évalué notre méthode sur les corpus de questions et de documents des campagnes Clef 2004 et Clef 2005 (395 questions au total). Trois mesures ont été utilisées pour évaluer notre approche : l'Accuracy, le MRR et le Rappel. L'Accuracy correspond au pourcentage de questions où la réponse retournée en première position est la bonne réponse. Le MRR (Mean Reciprocal Rank) permet d'évaluer la position moyenne des bonnes réponses. Il est calculé en faisant la moyenne des rangs réciproques de chaque question, ce dernier étant obtenu en calculant l'inverse du rang de la bonne réponse à la question. Par exemple, si la bonne réponse à une question est au rang 4, son rang réciproque est de 1/4. Le Rappel correspond au pourcentage de questions dont la liste des réponses retournées contient la bonne réponse. Pour cette évaluation nous avons fixé le nombre maximum de réponses retournées à 10. Nous séparons nos résultats selon le type de la réponse attendue à une question. Cette classification est déterminée automatiquement lors de la génération du DDR. Nous présentons ces résultats dans le tableau 1.

Nous avons fait une deuxième évaluation de notre approche, que nous ne présentons pas ici, où le nombre maximum de réponses retournées a été fixé à 50 : l'Accuracy moyenne régresse. On peut en déduire que le trop grand nombre de réponses candidates perturbe notre approche. Une évaluation a été réalisée sur la méthode présentée précédemment de Magnini et Kouylekov à Clef 2005 (Tanez *et al.*, 2005). Le système présenté a été testé sur des tâches monolingues pour l'italien et le bulgare. Dans le cas de l'italien, deux évaluations ont été faites pour tester différents paramètres. Avec cette approche, ils obtiennent en tâche monolingue sur le bulgare une Accuracy de 27.5%, le meilleur système ayant obtenu 32.03%<sup>2</sup>. Nous comptons faire prochainement une adaptation de notre système pour le faire fonctionner sur la tâche RTE à laquelle Magnini et Kouylekov ont participé. Leurs résultats se situaient dans la moyenne générale. Nous pourrions ainsi comparer notre approche qui est basée sur une analyse sémantique.

## 4.2 Analyse de l'évaluation

Le tableau 1 montre que notre méthode dégrade les résultats obtenus par Ritel quelle que soit la catégorie de questions, avec une baisse moyenne de 17%. L'origine de ces résultats peut être expliquée par l'état élémentaire des fonctions de coût associées aux opérateurs. Les principaux défauts sont le manque de précision du calcul du coût linguistique de chaque opération et la mauvaise gestion du contexte d'application d'une opération qui est limitée aux nœuds voisins et non à l'ensemble de l'arbre. L'analyse que nous avons conduit porte sur l'origine des résultats obtenus par notre approche. Certaines catégories semblent mieux adaptées à notre approche, où l'on peut noter une baisse des résultats moindres : "personne" et "date" (baisse d'environ 10%)<sup>3</sup>.

## 5 Étude du corpus de questions et des résultats obtenus

Après avoir analysé les résultats obtenus par notre système lors de l'évaluation, et émis un certain nombre d'hypothèses sur les origines des scores obtenus (Bernard, 2008), nous avons décidé de procéder à une étude approfondie des résultats obtenus. Nous nous sommes concentrés sur les catégories de question qui nous semblaient les plus pertinentes, et pour chacune des questions, nous avons analysé les causes du réordonnement produit par notre méthode. Nous en avons déduit des voies d'amélioration pour notre approche.

### 5.1 Corpus étudié

Nous nous sommes concentrés sur les catégories les plus importantes en nombre de questions : "date", "nombre", "personne" et "lieu". De plus, ces catégories semblent appropriées à notre approche, car elles semblent être plus sujettes à des problèmes de proximité syntaxique. La catégorie des organisations n'a pas été étudiée car la version de l'analyseur disponible pour cette étude ne détectait pas leurs frontières de manière fiable. On arrive donc à un ensemble de questions étudiées de 211 pour "date", "nombre", "personne" et "lieu". Sur ces 211 questions, 59 ont été mises de côté car la bonne réponse n'est pas contenue dans la liste des réponses candidates, ce qui porte le nombre de questions étudiées à 152.

### 5.2 Méthodologie

Nous avons travaillé par catégorie de question pour essayer de déterminer les phénomènes propres à certaines classes de questions et les phénomènes plus généraux. Pour chaque question, nous avons la liste des réponses ordonnées par Ritel et la liste des réponses réordonnées par

<sup>2</sup>Nous ne pouvons pas comparer directement avec Ritel, qui obtient 38% en tâche monolingue sur le français uniquement sur Clef 2005. Notre approche obtient une Accuracy de 17%.

<sup>3</sup>Les résultats sont différents de ceux présentés dans (Bernard, 2008) car entre-temps le système Ritel a évolué. Nous n'avons pas encore pris en compte cette évolution.

notre approche. À partir des passages de texte d'où étaient extraites les réponses, nous pouvions alors déterminer les causes de nos résultats, aussi bien pour un réordonnement positif que négatif. Nous avons aussi comme informations les suites d'opérations utilisées lors des transformations des arbres pour réordonner les réponses.

Plus généralement, nous avons analysé les points suivants pour chacune des questions :

- **Identification des correspondances entre la question et la phrase contenant la réponse :**
  - recherche des équivalences entre un groupe de mots et un autre ;
  - identification de la pertinence d'une synonymie (selon le contexte d'une phrase, deux mots ne sont pas forcément synonymes) ;
  - identification des regroupements entre mots pouvant apporter de l'information ;
- **Implémentation de mécanismes spécifiques selon le type de la question ;**
- **Informations perdues lors de l'extraction de la phrase contenant la réponse :**
  - pertinence de l'information perdue ;
  - proximité de l'information par rapport à la phrase contenant la réponse que l'on a extraite ;
  - présence d'anaphores ;
  - apport du contexte global du document d'où a été extraite la phrase pour les transformations d'arbre.

## 5.3 Résultats de l'analyse

### 5.3.1 Problèmes liés aux sous-nœuds

Lorsque nous avons commencé à concevoir notre méthode, l'analyseur de Ritel produisait des arbres très peu profonds. Désormais, il y a une meilleure définition des types et de leur hiérarchie, et nous avons en moyenne une profondeur de 4 niveaux. Notre méthode n'est pour le moment pas adaptée pour traiter ce type d'arbres. L'un de nos problèmes est la non-prise en compte des relations d'équivalences et de synonymies entre les sous-nœuds des arbres : un nœud du texte peut très bien contenir un sous-nœud en rapport avec un nœud de la question mais ils ne seront pas mis en correspondance. Dans l'exemple ci-dessous, "prix Nobel" contenu dans l'arbre de la question n'est pas mis en relation avec celui de la phrase car ce dernier est un sous-nœud.

*Question* : En quelle année Thomas Mann a-t-il obtenu le <\_Prix> prix Nobel </\_Prix> ?

*Texte* : En 1929, six jours après son premier contrat, il assiste à Stockholm à la remise du <\_prix> <\_Prix> prix Nobel </\_Prix> <\_type\_prix> de littérature </\_type\_prix> </\_prix> à Thomas Mann avant de tirer le portrait de George Bernard Shaw à Londres.

De même, on peut étendre ce problème aux relations synonymiques dans les sous-nœuds. Dans l'exemple suivant, la relation de synonymie entre les sous-nœuds "cyclone Andrew" et "ouragan Andrew" n'est par exemple pas détectée. Les types sont identiques, mais la synonymie ne fonctionne pas car elle est faite à un niveau global. Il faut donc descendre dans les nœuds pour valider la relation.

*Question* : En quelle année l' <\_phen\_meteo> <\_decl> ouragan </\_decl> <\_meteo> <\_nom> Andrew </\_nom> </\_meteo> </\_phen\_meteo> s'est-il manifesté.

*Texte* : Elles ont en mémoire les importantes pertes qu'elles ont enregistrées après le passage du <\_phen\_meteo> <\_decl> cyclone </\_decl> <\_meteo> <\_nom> Andrew </\_nom> </\_meteo> </\_phen\_meteo> en 1992 ou après le tremblement de terre en Californie en 1994.

Une première approche pour traiter les sous-nœuds serait de prendre en compte le sous-arbre sur lequel est appliquée l'opération. Ainsi, on ne substituerait plus un nœud mais un sous-arbre par un autre sous-arbre. On identifierait d'abord les opérations élémentaires au niveau des

feuilles des sous-arbres du texte, puis on remonterait dans l'arbre en générant des opérations plus globales. Si les feuilles d'un sous-arbre du texte sont en relation avec le même sous-arbre de la question, on procéderait alors à la substitution. Le coût serait donc calculé en fonction de l'équivalence entre les structures des deux arbres : type sémantique des différents nœuds proche ou équivalent, peu de nœud pour lesquels on ne trouve pas de relations ...

### 5.3.2 Perte d'information liée à l'extraction de la phrase

Un deuxième problème souvent observé est l'extraction de la phrase contenant la réponse et la perte d'informations essentielles qui en résulte. Ce problème est directement lié à notre fonction de coût de suppression, qui pour le moment est élémentaire et ne pénalise pas suffisamment les textes contenant la mauvaise réponse, et la gestion du contexte.

Pour le moment, l'impact de la suppression sur l'ensemble du calcul du coût est délibérément minoré car cela faussait trop les résultats : les passages du texte comportaient souvent un grand nombre de nœuds et la suppression prenait plus d'importance que les autres opérations. La fonction de coût associée se doit donc d'être bien plus fine. En fixant certaines limites dans l'ensemble de l'extrait du texte (selon la ponctuation mais aussi certaines conjonctions comme "et", "que", etc.), il serait possible d'avoir des bornes pour la suppression. Dans l'exemple ci dessous nous avons procédé à une segmentation manuelle. Le texte en entier est illustré, et non pas la phrase extraite par notre méthode, qui correspondrait alors au tout dernier segment. Si l'on identifie l'ensemble des relations entre le texte et la question, on détecte l'existence d'une relation d'équivalence sur les nœuds "World Trade Center" et "attentat" entre la question et le segment 5 qui disparaît si on ne traite que la phrase avec la réponse. Les segments traités dans ce cas là sont le cinquième et le huitième.

*Question avec la réponse insérée* : Février 1993 l' **attentat** à la bombe contre le **World Trade Center**

*Texte découpé* :

- (1) Un terroriste présumé extradé (2) d'un pays asiatique a été inculpé mardi (3) à New York  
(4) pour sa participation présumée (5) à l'**attentat** contre le **World Trade Center**,  
(6) ont rapporté les télévisions locales. (7) Cet acte terroriste avait fait 6 morts (8) en février 1993.

Dans le cas présent, les bornes de suppression seraient insérées avant le segment 5 et après le segment 7. Tous les nœuds non compris entre ces bornes seront supprimés. Cette méthode permet donc d'identifier les segments contenant l'information utile. Il est cependant clair que plus un nœud est éloigné des bornes, plus la suppression doit être pénalisante : un nœud contenant un élément critique de la question ne valide pas forcément la réponse évaluée s'il est éloigné du nœud contenant la réponse. Une première idée serait d'avoir une gestion plus poussée du contexte en incluant des relations entre segments pour déterminer le coût des suppressions. On pourrait ainsi modéliser plus précisément l'impact des opérations. Par exemple, le segment 5 décrit la cible d'un attentat, et il est en relation avec le segment 7 qui décrit les conséquences de cet attentat. Le segment 7 est en relation avec les segments 5 et 6 car il contient la date de l'attentat. L'utilisation de relations sémantiques telles qu'elles sont décrites dans (Rosset *et al.*, 2007) serait une possibilité pour traiter ce type de problème.

### 5.3.3 Ajout de relations entre les nœuds

Au cours de notre analyse, il est souvent apparu nécessaire d'ajouter des relations anaphoriques ou sémantiques entre nos différents nœuds. Les problèmes concernant les anaphores semblent plus fréquents dans les questions de type "personne" que dans les autres catégories. La nature de ces problèmes est assez diverse : pronoms personnels ("il") faisant référence à une personne,

ou des références plus complexes. Dans l'exemple suivant, La phrase contenant la réponse à la question ne contient pas d'informations sur "Jacques Delors". Par contre la phrase suivante parle de "Jacques Delors" comme étant le "prédécesseur" de Jacques Santer ("son prédécesseur"). Il faudrait pouvoir identifier l'adjectif passif "son" comme faisant référence à "Jacques Santer". De plus, il faut mettre en relation "Jacques Delors" avec "prédécesseur". Il faut identifier la relation d'antonymie entre "succéda" et "prédécesseur" qui permet enfin de confirmer que "Jacques Santer" est le successeur de "Jacques Delors".

*Question* : Qui **succéda** à Jacques Delors à la présidence de la Commission européenne ?

*Texte contenant la réponse* : Les ministres des affaires étrangères des Quinze, réunis en conférence intergouvernementale, ont officiellement nommé lundi les 20 membres de la commission européenne présidée par le Luxembourgeois **Jacques Santer**.

*Phrase suivante* : **Son prédécesseur**, le Français Jacques Delors, a fait ses adieux après dix ans de présidence.

Il faudrait donc à la fois disposer d'un module de traitement des anaphores<sup>4</sup>, avec notamment des règles d'inférences pour déterminer les relations entre les différents nœuds, mais aussi enrichir notre dictionnaire de synonymes en prenant en compte la nature des nœuds.

L'ajout de relations sémantiques se révèle nécessaire, particulièrement pour des relations de type "géographique". L'obtention de bons résultats dans les questions de type "lieu" semblent particulièrement dépendante de ces relations. Dans l'exemple suivant, la mauvaise réponse "Russie" sera invalidée étant donné que "participe" n'implique pas que Halifax se trouve en Russie. Plus loin, "en nouvelle Ecosse" permet de déterminer où se trouve vraiment "Halifax". Cette approche requiert à la fois de créer des relations sur les verbes (relation ternaire dans ce cas entre "Russie", "participera" et "au sommet de Juin") ainsi qu'une gestion des prépositions de lieu comme "à" ou "en".

*Question* : Où se trouve Halifax ?

*Question avec la réponse insérée* : Russie se trouve Halifax.

*Texte* : La Russie participera au sommet de Juin, à Halifax, en nouvelle Ecosse, sur les mêmes bases que l'an dernier.

## 6 Conclusions et perspectives

Nous avons présenté une méthode ayant pour objectif d'améliorer la sélection des réponses candidates. Cette approche est fondée sur un calcul du coût de transformation entre arbres issus de l'analyse et représentant la question et un texte pouvant contenir la réponse. Nous avons évalué cette méthode sur les corpus Clef 2004 et 2005 et analysé les résultats. Notre approche donne pour le moment de mauvais résultats, principalement à cause du fonctionnement élémentaire de nos fonctions de coûts. Nous avons donc mené une analyse en étudiant les résultats obtenus dans les catégories de questions les plus pertinentes. Nous avons décrit des améliorations à apporter à notre système. L'analyse que nous avons conduite suggère différentes voies d'amélioration. La prise en compte de la profondeur de nos arbres et de leur structure est essentielle pour gérer nos relations d'équivalence et de synonymie entre arbres. Par ailleurs, une meilleure gestion de la synonymie semble nécessaire pour détecter les bonnes réponses. Le dictionnaire que nous utilisons actuellement ne contient pas assez d'informations, et nous étudions des méthodes pour l'enrichir automatiquement. Il s'agirait dans un premier temps d'ajouter des synonymes proches à chaque entrée du dictionnaire, comme "électeurs" et "élisent". Actuellement, le dictionnaire

---

<sup>4</sup>Nous faisons actuellement des évaluations préliminaires pour intégrer un module d'anaphores dans notre approche. Il ne traite pour l'instant que les pronoms personnels "il", "elle" et "lui".

ne modélise les synonymes qu’entre les mots du même type syntaxique. Une première approche pour enrichir ce dictionnaire serait d’utiliser wiktionary.org et d’en extraire automatiquement les synonymes. L’extraction des phrases contenant la bonne réponse ne donne pas de bons résultats pour le moment, et nous perdons assez fréquemment des informations critiques. Ainsi, il semble important d’avoir une meilleure gestion du contexte de nos documents par le biais de la détection des relations entre les nœuds d’un arbre. Nous nous intéressons entre autre aux relations de type anaphorique (Vicedo & Ferrandez, 2000). Nos perspectives incluent également l’étude des techniques dites d’implication textuelle (*textual entailment*) pour améliorer nos résultats (Dagan *et al.*, 2006). Enfin, nous comptons utiliser cette méthode pour améliorer l’interaction entre un utilisateur et un système interactif exploitant les opérations appliquées pour trouver une réponse.

## Références

- BERNARD G. (2008). Réordonnancement de réponses par transformation d’arbre pour un système de question-réponse oral interactif. In *RJCRI’08 (Rencontre des Jeunes Chercheurs en Recherche d’Informations)*, p. 361–368.
- BRAZ R., GIRJU R., PUNYAKANOK V., ROTH D., & SAMMONS M. (2005). Knowledge representation for semantic entailment and question-answering. *IJCAI-05 Workshop on Knowledge and Reasoning for Question Answering*.
- DAGAN I., GLICKMAN O., & MAGNINI B. (2006). The PASCAL Recognising Textual Entailment Challenge. *Lecture Notes in Computer Science*, **3944**, 177–190.
- GILLARD L., BELLOT P. & EL-BÈZE M. (2007). D’une compacité positionnelle à une compacité probabiliste pour un système de questions/réponses. In *Conférence en Recherche d’Information et Applications (CORIA)*, p. 271–286.
- LIGOZAT A.-L. (2005). Apport de l’analyse syntaxique des phrases dans un système de questions-réponses. *Traitement automatique des langues*, **46(3)**, 103–125.
- LIGOZAT A.-L. (2006). *Exploitation et fusion de connaissances locales pour la recherche d’informations précises*. PhD thesis, LIMSI - Université Paris-Sud 11.
- MAGNINI B. & KOUYLEKOV M. (2006). Tree edit distance for recognizing textual entailment : Estimating the cost of insertion. In *the Second PASCAL Challenges Workshop on Recognising Textual Entailment*.
- ROSSET S., GALIBERT O., ADDA G. & BILINSKI E. (2007). The Limsi QAST Systems : Comparison between human and automatic rules generation for question-answering on speech transcriptions. In *Automatic Speech Recognition and Understanding (ASRU)*, p. 647–652.
- ROSSET S., GALIBERT O., ILLOUZ G. & MAX A. (2006). Interaction et recherche d’information : le projet Ritel. *Traitement Automatique des Langues*, **46(3)**, 155–179.
- TANEZ H., KOUYLEKOV M., MAGNINI B., NEGRI M. & SIMOV K. (2005). Exploiting Linguistic Indices and Syntactic Structures for Multilingual Question Answering : ITC-irst at CLEF 2005. In *CLEF 2005*, p. 390–399.
- VICEDO J. L. & FERRANDEZ A. (2000). Importance of pronominal anaphora resolution in question answering systems. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL ’00)*, p. 555–562 : ACL.
- VILLANEAU J., ROSSET S. & GALIBERT O. (2007). Semantic relations for an oral and interactive question-answering system. In *Semantic Representation of Spoken Language (SRS7)*, p. 12–19.