
RECONNAISSANCE, TYPAGE ET TRAITEMENT DES COREFERENCES DES TOPONYMES FRANÇAIS ET DE LEURS GENTILES PAR DICTIONNAIRE ELECTRONIQUE RELATIONNEL

Claude Belleil

Institut de Recherche en Informatique de Nantes

Université de Nantes

Faculté des Sciences, rue de la Houssinière, 44000 Nantes - belleil@irin.univ-nantes.fr

Ces travaux proposent le traitement automatique des toponymes français et des noms d'habitants qui leur sont associés : les gentiles.

Après un premier chapitre consacré à un état de l'art, nous introduisons dans le chapitre 2, le problème des noms propres en français. Le chapitre 3 est consacré à l'exposé de notre méthodologie. Celle-ci se décompose en deux phases principales : la reconnaissance et le typage, la résolution des coréférences.

Les chapitres suivants sont consacrés à la présentation des outils informatiques que nous avons construits pour résoudre les problèmes de reconnaissance, de détermination du type et de calcul des coréférences.

Dans le chapitre 4, nous présentons le travail de collecte en détaillant les résultats de l'enquête nationale que nous avons conduite. Ensuite, nous justifions notre choix d'une base de données relationnelle pour réaliser le dépouillement, le contrôle et la validation de ces informations. En conclusion de ce chapitre, nous notons que, pour des raisons de performance et de technique d'interrogation, la base de données relationnelle ne peut être utilisée directement dans le cadre d'un processus d'analyse automatique. Nous expliquons de quelle façon elle a été utilisée pour générer de façon automatique un dictionnaire électronique relationnel des noms de lieux et d'habitants.

La construction des outils de reconnaissance, de typage et de résolution des coréférences repose sur la technique des automates et des transducteurs. Nous présentons ensuite les propriétés d'un transducteur particulier que nous avons mis au point : le transducteur d'identifiants. Dans le chapitre 6, nous présentons le processus de reconnaissance et de typage. Enfin, dans le chapitre 7, nous exposons notre méthode de détection et de calcul des coréférences par fermeture transitive de la matrice booléenne associée. Ceci nous amène à détailler l'ensemble des informations sur les noms propres reconnus que notre système est susceptible de transmettre à un analyseur classique.

MODELISATION DE LA PERTINENCE EN RECHERCHE D'INFORMATION MODELE CONCEPTUEL, FORMALISATION ET APPLICATION

Nathalie Denos

Laboratoire CLIPS-IMAG

Domaines Universitaire, BP 53, 38041 Grenoble Cedex 9, France

Email : Nathalie.Denos@imag.fr

Les systèmes de recherche d'information ont pour fonction de permettre à l'utilisateur d'accéder à des documents qui contribuent à résoudre un problème d'information. Le système peut être vu comme un instrument de prédiction de la pertinence des documents du corpus pour l'utilisateur. Les indices traditionnellement utilisés par le système pour estimer cette pertinence sont de nature thématique, et sont fournis par l'utilisateur sous la forme d'un ensemble de mots-clés : la requête. Ainsi la conception classique des systèmes de recherche d'information limite la modélisation de la pertinence à ses aspects "objectifs" et thématiques.

Cependant la pertinence est une notion très complexe, qui dépend de la façon dont l'interaction avec les documents se déroule. Mon travail consiste à intégrer les connaissances acquises sur la pertinence, notamment sur la nature et la structure du processus cognitif d'évaluation de la pertinence chez les individus, dans la conception des systèmes de recherche d'information. Le principe fondateur en est que l'utilisateur détient le monopole du sens qu'il cherche à exprimer, et que le système doit lui fournir les outils lui permettant de formuler ce sens en le positionnant dans son champ de compétences.

Ma thèse s'appuie sur les travaux réalisés sur les facteurs de la pertinence pour les individus, pour prendre en compte les facteurs qui relèvent 1) de l'utilisation interactive du système et 2) de la nécessité d'adaptation de la fonction de

correspondance à la situation de recherche particulière dans laquelle l'utilisateur se trouve. Trois nouvelles fonctions du système de recherche d'information sont ainsi définies en termes d'utilisation du système : permettre la détection de la pertinence des documents retrouvés, permettre la compréhension des raisons de leur pertinence système, et permettre de procéder à une reformulation du problème d'information dans le cadre d'un processus itératif de recherche.

ETUDE DES MARQUEURS DISCURSIFS DANS LE DIALOGUE FINALISE

Nathalie Colineau

*Laboratoire de Communication Langagière et Interaction Personne-Système
Université Joseph Fourier, BP. 53, 38041 Grenoble Cedex 09
tel : 04 76 51 45 10 fax : 04 76 44 66 75
Email : Nathalie.Colineau@imag.fr*

L'étude que nous avons réalisée se place dans le cadre des analyses de discours, avec pour objectif de donner une interprétation des énoncés produits, en s'appuyant sur des marques linguistiques, prosodiques, actionnelles et contextuelles.

L'idée sur laquelle repose l'étude de ces marques est qu'un locuteur donne dans son énonciation suffisamment d'"indices" pour être compris, indices qui permettront à l'interlocuteur d'interpréter ce que le locuteur a voulu dire. En effet, il apparaît que ces indices linguistiques reflètent l'organisation voulue par le locuteur et que l'allocutaire les exploite comme guides lors de l'interprétation des énoncés.

Dans un premier temps, une analyse linguistique a été faite sur un corpus de dialogues oraux finalisés. Ce travail a permis de déterminer un ensemble de marques pertinentes, et de répertorier leurs différents emplois dans chaque contexte d'énonciation. Trois types de marques ont été retenus : les marques intervenant dans la détermination des actes de dialogue, les marques de connexion dont le rôle est de marquer la structure du discours et les marques de cohésion rendant compte des chaînes thématiques. Cependant les marques linguistiques à elles seules ne permettent pas dans tous les cas de déterminer l'acte de dialogue accompli, d'autres marques de nature extra-linguistique doivent être intégrés dans l'analyse. En effet, la situation dans laquelle sont produits les énoncés et le rôle de chacun des partenaires du dialogue contribuent aussi à la détermination des actes de dialogue accomplis. Afin d'intégrer ces informations dans l'analyse, un modèle des enchaînements d'énoncés a été élaboré, tenant compte à la fois du type d'actes de dialogue accomplis à chaque tour ainsi que du rôle occupé par le locuteur dans l'interaction.

Dans un second temps, un modèle informatique a été élaboré comme outil pour tester la pertinence de ces marques et ainsi la validité de notre approche. La tâche du système consiste à partir des marques discursives relevées, à déterminer si l'énoncé produit est une requête, une question, une information donnée, etc.

Enfin, pour valider notre approche et montrer la pertinence des marques discursives collectées, nous avons analysés trois corpus de dialogues oraux. Nous montrons ainsi le comportement réglé de ces marques et l'importance de leur prise en compte dans l'interprétation des énoncés.