

Exploitation d'un dictionnaire de langue dans un environnement informatique dédié à l'apprentissage : le cas du TLFi au lycée

Chrysta Pélissier, Claire Becker

Laboratoire ATILF : Analyse et Traitement Informatique de la Langue Française
UMR 7118, Université de Nancy 2 - 44 av. de la Libération, 54 063 Nancy, <http://www.atilf.fr/>
Chrysta.Pelissier@atilf.fr, Claire.Becker@atilf.fr

Résumé

L'objectif de ce document est de présenter notre projet qui vise à exploiter les grandes ressources linguistiques pour l'enseignement et l'apprentissage. Plus spécifiquement, nous nous attachons à exploiter un dictionnaire de langue, le Trésor de la Langue Française informatisé (TLFi) et la base textuelle Frantext, développés par l'ATILF, dans le cadre de la préparation des épreuves anticipées de Français du baccalauréat.

A partir d'un texte extrait d'une base textuelle qui regrouperait l'ensemble des textes étudiés en classe, l'élève a la possibilité, pour certains mots du texte, d'obtenir la ou les définitions correspondantes, extraites du TLFi.

Mots clés : dictionnaire électronique ; définition ; désambiguïsation contextuelle ; baccalauréat ; français ; lycée.

1. Introduction

Notre travail s'intègre dans le cadre de la valorisation des ressources linguistiques développées par le laboratoire. La notion de ressource linguistique regroupe l'ensemble des documents sur support papier et sur support informatique susceptibles d'apporter des informations participant à l'enrichissement des connaissances que peut avoir l'humain sur la langue française. Ces ressources peuvent prendre la forme par exemple de dictionnaires de langue ou de bases textuelles. Nos recherches portent plus particulièrement sur l'utilisation de ces ressources pour l'enseignement et l'apprentissage de la langue française. Les instructions officielles du primaire, du collège et du lycée préconisent l'utilisation de ces ressources. Par exemple, les instructions du ministère de 2002 [4] incitent les enseignants du lycée à faire rechercher aux élèves différentes informations dans des bases de données, sur Internet ou encore à consulter des dictionnaires.

2. Cadre de notre recherche

Notre démarche de recherche s'effectue à deux niveaux. D'une part, nous sommes en train d'établir une méthodologie qui vise à favoriser l'utilisation des ressources linguistiques telles qu'elles sont actuellement disponibles par les enseignants et les apprenants de langue [5]. D'autre part, nos travaux portent sur le développement de ressources dérivées des ressources linguistiques existantes, plus adaptées aux besoins du public visé [2]. Le travail que nous présentons ici fait référence au second cas et s'intègre au projet e-Lorraine.

Conduite sur trois années, l'opération e-Lorraine s'est achevée en mars 2002 par la connexion au reste du monde via Internet, d'un ensemble de machines mises à la disposition des élèves dans les établissements secondaires du territoire lorrain. Aujourd'hui, l'enjeu est de mettre à la disposition des enseignants et des apprenants, sur la plate-forme e-Lorraine (<http://www.e-lorraine.net>), un ensemble de contenus pédagogiques destinés à aider les élèves à préparer les épreuves du baccalauréat. Pour l'ATILF, il s'agit de mettre à la disposition des lycéens des ressources linguistiques adaptées à leurs besoins.

3. Les ressources linguistiques de l'ATILF

Parmi les ressources linguistiques développées par l'ATILF, nous nous intéressons plus particulièrement au TLFi (Trésor de la Langue Française Informatisé <http://atilf.atilf.fr/tlf.htm>) et à Frantext (<http://atilf.atilf.fr/frantext.htm>). Le TLFi [1] est la version informatisée du Trésor de la Langue Française. Il se caractérise par la richesse (100 000 entrées, 270 000 définitions, 430 000 exemples) et la diversité (informations diachroniques et synchroniques pour chaque mot) des informations qu'il propose. Frantext est une base de textes littéraires du XIV^{ème} au XX^{ème} siècle. L'utilisateur peut obtenir à l'aide de cette base une liste de textes écrits par un auteur particulier,

appartenant à une période donnée ou encore contenant un mot particulier. La mise à disposition de ces deux ressources prises séparément pour la préparation du bac de français ne semble pas pertinente. En effet, les textes étudiés en classe pour l'épreuve orale ne sont pas tous présents dans la base Frantext. De plus, pour chacun des textes, différentes informations peuvent aider l'élève à comprendre et interpréter chaque texte (figures de style, sens de certains mots, etc.). Ces informations ne sont pas actuellement présentes pour chacun des textes de la base. Ainsi, une nouvelle ressource destinée à favoriser la préparation du bac de français est actuellement en cours de développement à l'ATILF. Cette ressource correspond à une combinaison de Frantext et du TLFi. Il s'agit d'une base textuelle composée des textes étudiés par les lycéens pour le bac. Pour chaque texte, l'élève peut en particulier obtenir pour certains mots du texte la définition du TLFi qui semble la plus appropriée.

4. Exploitation du TLFi

Pour permettre à l'apprenant (pour chaque texte) d'obtenir pour certains mots la définition du TLFi qui lui correspond, nous sommes en train de mener en parallèle trois types de travaux :

1. Nous établissons à partir du TLFi une base de connaissances lexicales structurée,
2. Nous déterminons les informations lexicales présentes dans les articles du TLFi qui semblent intervenir dans le cadre de la détermination pour un mot donné d'une ou plusieurs définitions,
3. Nous définissons un ensemble de règles qui doit permettre de sélectionner la (ou les) définition(s) du TLFi la (les) plus pertinente(s).

Premièrement, il semble que les articles TLFi ne soient pas structurés de manière homogène. En effet, pour certaines entrées, des définitions peuvent apparaître au niveau le plus haut de l'article (I, II, III ; A, B, C), précédées généralement d'un indicateur (registre de langue, domaine technique,...) mais pour d'autres, les premières définitions n'apparaissent qu'au second voire au troisième niveau d'organisation (1, 2, 3 ; a, b, c). Aucune information n'est alors présente au niveau supérieur (I, II, III ; A, B, C) permettant à l'utilisateur de déterminer si la définition qu'il cherche se situe dans cette section (I) ou une autre (II ou III).

Si l'on prend l'exemple de la définition du mot « enseignement », aucune information particulière n'est présente au niveau de A. L'utilisateur ne peut donc pas définir rapidement si la définition qu'il cherche par exemple se situe dans A- ou B-. Il devra parcourir l'ensemble des informations associées à A-1 puis 2.

Deuxièmement, il semblerait, d'après nos premières analyses, que parmi l'ensemble des connaissances lexicales présentes pour chaque entrée dans le TLFi, seules certaines d'entre elles (les domaines techniques ou généraux, les connaissances mises entre crochets, les indicateurs et les termes utilisés dans la définition elle-même) participent à la détermination de la définition la plus appropriée pour un mot donné [3].

Troisièmement, il semble que les règles qui permettent d'associer à un mot une définition particulière fassent intervenir deux types de connaissances :

1. Des connaissances présentes dans le TLFi (domaines, connaissances mises entre crochets, indicateurs, définition),
2. Des connaissances associées au texte. Parmi elles, nous pouvons citer par exemple le fait que les textes présents dans cette base soient tous des textes littéraires, que certains soient des poèmes, d'autres des extraits de pièces de théâtre ou de romans. Ces textes se caractérisent tous également par une liste de champs lexicaux (corps, mort, vocabulaire péjoratif...), de thèmes (voyage, amour, idéal...), de tonalités (lyrique, satirique,...), les figures de styles notées dans le texte (métaphores, métonymie,...).

5. Conclusion

Ce travail sur l'exploitation du TLFi dans le cadre d'un traitement automatique dans un environnement d'aide à l'apprentissage de la littérature française n'en est qu'à ses débuts. Nous avons pour l'instant centré nos recherches sur l'analyse des connaissances linguistiques présentes dans la ressource. Ce travail nous permet d'une part de caractériser le TLFi et d'autre part de définir les connaissances sur lesquelles le raisonnement de détermination de la (ou des) définition(s) associés aux mots d'un texte va pouvoir s'appuyer.

Références

- [1] Dendien, J. and Pierrel, J-M., 2003 : « Trésor de la Langue Française informatisé », Traitement Automatique de la Langue, vol. n°44-n°2/2003, Hermès Sciences Edition, 28 p
- [2] Dubois, C. 1998: Initier à la lecture méthodique en seconde à l'aide de l'informatique, revue de l'EPI n° 92, décembre 1998.
- [3] Ide, N. et Véronis, J. 1998 : Modeling Lexical Databases, colloque international Euralex, 1998
- [4] Ministère 2002 : Programmes de Français, classe de première, séries générales et technologiques et programme de Littérature, classe de terminale, série littéraire, Collection Lycée, Ministère de l'éducation nationale, Direction de l'enseignement scolaire, CNDP
- [5] Pélissier et al. 2004 : Pélissier C., Jadelot C., Pierrel J-M 2004 : « Méthodologie liée à l'utilisation de grandes ressources linguistiques dans le cadre de l'apprentissage : le cas du TLFi en Français au cycle 3 », colloque international Euralex, Lorient, France, juillet 2004.