

# Résolution automatique d'anaphores grammaticales

**Encadrement :** Juan-Manuel Torres <[juan-manuel.torres@univ-avignon.fr](mailto:juan-manuel.torres@univ-avignon.fr)>  
Vincent Labatut <[vincent.labatut@univ-avignon.fr](mailto:vincent.labatut@univ-avignon.fr)>  
Luis Adrian Cabrera Diego <[luis-adrian.cabrera-diego@alumni.univ-avignon.fr](mailto:luis-adrian.cabrera-diego@alumni.univ-avignon.fr)>

D'après Wikipedia : « En grammaire, une *anaphore* est un mot ou un syntagme qui, dans un énoncé, assure une reprise sémantique d'un précédent segment appelé *antécédent*. L'anaphore est un procédé fondamental qui participe à la cohérence d'un texte. »

En d'autres termes, une anaphore est un groupe de mots qui fait référence à un passage du texte situé avant lui. Le sens de ce groupe de mots correspond à celui de ce passage. Par exemple :

Jean n'avait pas de stylo : je lui ai prêté le mien.

Le pronom possessif le mien est une anaphore : on parle d'anaphore *pronominale*. Son antécédent est le nom stylo. Ici, l'anaphore permet d'éviter une répétition lexicale. Sans elle, la phrase serait moins élégante : Jean n'avait pas de stylo : je lui ai prêté mon stylo.

Une anaphore peut donner lieu à certaines ambiguïtés :

Mon voisin a adopté un gros chien. Il n'est pas très sympathique.

Le pronom personnel il est une anaphore. Mais quel est son antécédent ? Est-ce le voisin, ou bien est-ce le gros chien, qui n'est pas très sympathique ?

Mon voisin a adopté un gros chien. Cet animal n'est pas très sympathique.

Le syntagme nominal cet animal est également une anaphore, mais, dans ce cas, un doute moindre plane sur l'identité de son antécédent : cette construction rend plus vraisemblable que ce soit le gros chien qui soit peu sympathique, sans toutefois écarter un éventuel effet humoristique du locuteur, qui qualifierait le voisin d'animal.

Mon voisin a adopté un gros chien. Ce dernier n'est pas très sympathique.

Le syntagme nominal ce dernier est également une anaphore, qui ne laisse cette fois plus d'ambiguïté. Cette anaphore est dite *nominale*. Pour être complet, il faut aussi mentionner les anaphores *adverbiales*, par exemple :

Mon voisin a enfermé son chien dans son garage. Là, il ne me dérangera plus.

La bonne résolution des ambiguïtés générées par les anaphores est relativement naturelle pour un être humain. Cependant, la définition d'une méthode de traitement automatique, qui serait exécutable efficacement par un ordinateur, est un problème ouvert.

Le but de ce projet est de définir une telle méthode, capable de traiter des textes en français, et de l'implémenter sous la forme d'un logiciel libre, qui sera mis à disposition de la communauté. Il y a là un réel besoin, car aucun logiciel de ce type n'existe actuellement. Si le français est la priorité, il est aussi possible de s'intéresser à l'adaptation de l'outil à d'autres langues, dans un second temps.

L'étudiant devra explorer les deux principales approches pour la résolution d'anaphores : les méthodes à base d'apprentissage artificiel [1, 2, 3] et celles à base de règles [4, 5]. Il faudra également considérer une solution hybride, mélangeant les deux approches [6]. L'évaluation de l'outil, ainsi que la phase d'apprentissage pour la première méthode, nécessitent la constitution d'un corpus. Nous avons actuellement un petit corpus, qu'il faudra étendre. L'étudiant devra participer au processus d'annotation manuelle.

En fonction du temps restant, il sera alors possible d'envisager différentes applications de l'outil, liées au résumé automatique, à l'analyse textuelle ou aux bases de connaissances (Wikipédia, WordNet, BabelNet, etc.).

## Références

- [1] C. Aone, S. W. Bennett. 1995. Evaluating automated and manual acquisition of anaphora resolution strategies. In Proceedings of 33th Annual Meeting of the Association for Computational Linguistics, 122–129.
- [2] J. McCarthy, W. Lehnert. 1995. Using decision trees for coreference resolution. In Proceedings of the 14th International Joint Conference on Artificial Intelligence, 1050–1055.
- [3] W. M. Soon, H. T. Ng, D. C. Y. Lim. 2001. A Machine Learning Approach to Coreference Resolution of Noun Phrases. Computational Linguistics, Vol. 27(4), 521–544.
- [4] J. R. Hobbs. 1978. Resolving Pronoun References. Lingua, Vol. 44, 311–338.
- [5] S. E. Brennan, M. W. Friedman, C. J. Pollard. 1987. A Centering Approach to Pronouns. In Proceedings of the 25th Annual Meeting on Association for Computational Linguistics (ACL '87), 155–162.
- [6] S. Lappin, H. Leass. 1994. An algorithm for pronominal anaphora resolution. Computational Linguistics, Vol. 20(4), 535–562.