

Proxem développe depuis 2006 la plateforme de traitement linguistique, **Antelope** (*Advanced Natural Language Object-oriented Processing Environment*). Cette plateforme, utilisable sans contrainte pour la recherche et l'enseignement, a été téléchargée (sur www.proxem.com) par plus de 600 internautes à ce jour. Antelope permet l'analyse syntaxique et sémantique de corpus de volume important, en intégrant plusieurs composants préexistants ainsi que des données linguistiques de large couverture provenant de différentes sources. Un effort d'intégration permet néanmoins d'offrir une plateforme homogène. Antelope inclut les modules suivants :

- Etiquetage morphosyntaxique (*Part-of-speech tagging*),
- Analyse syntaxique avec deux analyseurs robustes pour l'anglais : le *Link Grammar Parser* (basé sur des règles) ou le *Stanford Parser* (probabiliste) qui produisent une forêt d'arbres ou des graphes de dépendances.
- Détection de compléments de temps et de lieu,
- Analyse sémantique permettant de détecter des prédicats,
- Résolution d'anaphores et de coréférences,
- Lexique sémantique riche, basé sur les données de *WordNet*, enrichi par d'autres ressources de large couverture,
- Détection d'expressions multi-mots ("*Abraham Lincoln*", "*come back*"),
- Désambiguïsation lexicale utilisant plusieurs heuristiques,
- Calcul de mesures de similarités entre mots et concepts...

Antelope vise à être simple à mettre en œuvre. Les principaux modules disposent de paramètres macroscopiques par défaut, privilégiant un traitement rapide ou précis. Un utilisateur expert a la possibilité de jouer sur des paramètres plus fins, ou d'inclure ses propres algorithmes dédiés à une tâche donnée. Le niveau de traitement le plus complet permet d'extraire d'un texte des connaissances à un niveau sémantique.

Antelope préserve autant que possible les ambiguïtés au niveau lexical, syntaxique et sémantique. Un élément porteur d'ambiguïté est toujours associé à un score. Cette notion est mise en œuvre au niveau de l'analyse syntaxique (pondération des arbres de la forêt), de la désambiguïsation lexicale (sens possibles d'un mot), du calcul des anaphores (antécédents possibles) et de l'interprétation sémantique d'une construction syntaxique.

L'architecture technique de la plateforme repose sur des « bonnes pratiques » de Génie logiciel, facilitant la modularité du logiciel et la réutilisation de composants. Antelope dispose de mécanismes permettant la montée en charge (traitements multitâches).

Développée initialement pour la langue anglaise, Antelope est en cours d'adaptation au français, et intègre maintenant l'analyseur syntaxique *TagParser* et les données de WOLF (WordNet Libre du Français).