

TRAITEMENT DE DONNÉES LINGUISTIQUES SUR ORDINATEUR

Guy BUCHHOLTZER

Afin de faciliter les recherches linguistiques, nous avons mis au point pour l'ERA 431 d'Ethnolinguistique Amérindienne quelques programmes informatiques de traitement de données linguistiques.

Ces programmes sont principalement constitués dans le but d'étudier - au niveau de masses très importantes de données telles que les récits mythiques, traditions orales, etc. - et à l'aide d'un ensemble assez restreint de faits, de règles, de notions et de données diverses, la masse considérable de propriétés, lois et constructions du discours.

Ces programmes ont été écrits pour l'étude de la langue **kwakiutl** (Colombie britannique) mais peuvent être adaptés à tous les types de langues.

Il s'agit principalement :

- d'un programme de concordance qui trie toutes les occurrences d'une lexie, selon un ordre alphabétique, alphanumérique ou autre, suivant son ordre d'occurrence dans le texte. Chaque lexie est donnée avec son contexte droit et gauche, et se trouve indexée par rapport à l'ensemble du texte
- d'un programme de statistique adaptée à des problèmes lexico-statiques
- d'un programme d'analyse morphologique
- d'un programme d'extraction et de traitement de différentes unités d'analyse linguistiques
- de quelques programmes d'édition et de correction de textes.

Ces différents programmes permettent d'étudier les constructions du discours, ses caractéristiques sémantiques, syntaxiques, et logiques.

Dès à présent, ils permettent de constituer les bases nécessaires à l'étude lexicologique de thesaurus importants et au contrôle des données.

Nous donnerons une description détaillée ainsi que la méthode et les conditions d'utilisation de ces programmes dans AMERINDIA.