

Titre de la thèse / *Title*

Modélisation informatique de structures dynamiques de segments textuels pour l'analyse de corpus. / *Data-processing modeling of dynamic structures of textual segments for the analysis of corpus.*

Résumé / *Abstract*

L'objectif de la thèse est de proposer un modèle informatique pour représenter, construire et exploiter des structures textuelles. Le modèle proposé s'appuie sur une représentation du texte sous la forme d'un plan lexique/occurrences augmenté de systèmes d'annotations lexicales et contextuelles, modèle dont une implantation a été réalisée dans le logiciel SATO dont on présente les fonctionnalités et l'organisation interne. La présentation d'un certain nombre de travaux rend compte du développement et de l'utilisation du logiciel dans divers contextes.

La prise en charge formelle des structures textuelles et discursives trouve un allié dans le langage de balisage XML et dans les propositions de la Text Encoding Initiative (TEI). Formellement, les structures construites sur les segments textuels correspondent à des graphes. Dans le contexte d'une analyse textuelle en élaboration, ces graphes sont multiples et partiellement déployés. La résolution de ces graphes, au sens du rattachement des nœuds à des segments textuels ou à des nœuds d'autres graphes, est un processus dynamique qui peut être soutenu par divers mécanismes informatiques. Des exemples tirés de la linguistique textuelle servent à illustrer les principes de l'annotation structurelle. Des considérations prospectives sur une implantation informatique d'un système de gestion de l'annotation structurelle sont aussi exposées.

The objective of the thesis is to propose a data-processing model to represent, build and exploit textual structures. The suggested model relies on a «type/token» form of text representation extended by systems of lexical and contextual annotations. This model's establishment was carried out in the SATO software -- of which the functionalities and the internal organization are presented. Reference to a number of works give an account of the development and use of the software in various contexts.

The formal assumption of the textual and discursive structures find an ally in the beaconing XML language and the proposals of the Text Encoding Initiative (TEI). Formally, the structures built on the textual segments correspond to graphs. In a development driven textual analysis context, these graphs are multiple and partially deployed. Their resolution, within the fastening of the nodes to textual segments or that of other graphs, is a dynamic process which can be sustained by various data-processing mechanisms. Examples drawn from textual linguistics are used to illustrate the principles of structural annotation. Prospective considerations for the data-processing establishment of a management system of the structural annotation are also exposed.

Mots clés / *Key Words*

Analyse de texte assistée par ordinateur ; analyse de discours ; modèle SATO ; annotation structurelle ; TEI ; textométrie. / *Computer Aided Text Analysis ; discourse analysis ; SATO model ; structural annotation ; TEI ; textometry.*