

Séminaire de l'UMR STL : *Constitution et exploration de corpus*

PLANNING DES PROCHAINES RÉUNIONS (AU 22/11/2006)

L'objectif de ce séminaire est de permettre à tous d'utiliser des corpus existants, de constituer un corpus pour documenter une recherche en cours, d'explorer des données accumulées (éventuellement sans avoir préalablement réfléchi à leur structuration) ou encore d'aligner des données bilingues pour pouvoir analyser en parallèle un segment exprimé dans plusieurs langues, voire différentes traductions dans une langue donnée d'un même segment, voire d'aligner un signal sonore/vidéo à sa transcription orthographique ou phonétique et à des annotations.

Le travail fait durant ces séances ne nécessite aucune compétence informatique préalable. Mais il sera d'autant plus intéressant pour les plus néophytes qu'il sera interactif. Merci donc, au moins dix jours avant, de pointer les questions techniques ou méthodologiques auxquelles vous souhaitez que nous abordions durant le séminaire, les objets linguistiques que vous aimeriez voir traiter ou les travaux que vous avez déjà engagés et qui pourraient nourrir la discussion sur chaque thème.

Si durant les séances nous considérons que nous voulons nous donner plus de temps pour travailler sur un thème ou qu'il est utile d'en aborder d'autres, nous pourrions prévoir de nouvelles dates de réunion. En fonction des travaux qui seront engagés, une journée d'études pourra être programmée au printemps.

Date	Thème de la séance
mardi 9 janvier de 10h à 13h salle A5 206 ¹	<i>Frantext (1)</i>
jeudi 11 janvier de 9h30 à 13h30 salle A5 208	<i>Initiation à Praat (1)²</i>
jeudi 11 janvier de 14h30 à 17h30 salle A5 206	<i>Constitution de corpus multilingues à partir de traductions saisies dans des traitements de textes et des mémoires de traduction</i>
vendredi 12 janvier de 9h30 à 13h30 salle A5 208	<i>Initiation à Praat (2)</i>
mardi 23 janvier de 10h à 13h salle A5 206	<i>Frantext (2)</i>
vendredi 26 janvier de 10h à 13h salle A5 206	<i>Présentation de corpus bilingues alignés (dont le Dutch Parallel Corpus) et de leur exploration avec un concordancier bilingue (ParaSearch)</i>
vendredi 9 février de 16h à 19h salle A5 206	<i>Constitution de corpus multilingues à partir de documents repérés sur le Web</i>
vendredi 9 mars de 16h à 19h salle A5 206	<i>Présentation de différentes techniques d'alignement de textes traduits réunis dans un corpus multilingue</i>
mercredi 28 mars de 10h à 13h et de 14h à 17h salle A5 206	<i>Structuration de corpus monolingues en XML</i>
vendredi 13 avril de 10h à 13h salle A5 206	<i>Constitution de corpus à partir de CD-R de presse</i>
jeudi 3 mai de 10h à 13h salle A5 206	<i>Étiquetages morphosyntaxiques de corpus avec le TreeTagger et avec Cordial Analyseur puis exploration de ces corpus étiquetés (1)</i>
jeudi 3 mai de 14h à 17h salle A5 206	<i>Structuration en XML de corpus multilingues de textes alignés</i>
vendredi 4 mai de 10h à 13h salle A5 206	<i>Étiquetages morphosyntaxiques de corpus avec le TreeTagger et avec Cordial Analyseur puis exploration de ces corpus étiquetés (2)</i>
vendredi 4 mai de 16h à 19h salle A5 206	<i>Réflexion méthodologique sur les transcriptions pour constituer des corpus</i>

¹ Pour rejoindre la salle : <http://stl.recherche.univ-lille3.fr/siteherges/LTTAC/Lille3-Campus-universitaire.htm>

² Si vous souhaitez y participer, merci de nous l'indiquer à Cyril Auran et à moi et de prévoir de vous munir d'un enregistrement de 5 mn au format Wave.