

Etaient présents Frédérique Bénard, Maria Candéa, Serge Fleury, Cédric Gendrot, Luiggi Samsonetti.
Bien sûr, toutes vos réactions sont les bienvenues...

DATE DE PROCHAINE REUNION....

Puisque la prochaine réunion aura lieu après les vacances, nous n'avons pas fixé de date précise.
Elle aura lieu dans la 2^{ème} quinzaine de septembre à priori.

!#####

Cette réunion en plein air fut un peu moins technique que les précédentes et nous a permis de faire un petit bilan sur le projet innovant, ainsi qu'une petite réflexion sur toutes les activités à venir... En effet, en plus des aspects techniques, il sera indispensable d'assurer un suivi du projet lorsque la base de données sera prête.
Il FAUT que cette base de données rentre dans les esprits des membres de l'ED... et il nous incombe d'en assurer la publicité. (cf. infra)

Voici les 3 points abordés au cours de cette réunion :

- 1. Débriefing de la journée C-oraux à la BNF, présentation du projet aux VIII RED.
- 2. Catalogage des fiches de Méta-Données
- 3. MKM --> améliorations
- 3. sous à dépenser

1. Débriefing de la journée C-oraux à la BNF, présentation du projet aux VIII RED.

La journée C-Oraux à la BNF a bien mis en évidence l'importance de récolter des corpus oraux et d'en assurer la pérennité.

Je vous rappelle que le guide C-Oraux est téléchargeable à l'adresse suivante (version non définitive) :

http://www.culture.gouv.fr/culture/dglf/corpus_oraux.htm

Un point nouveau de cette journée concernait les aspects juridiques ; personne n'a apporté de réponse définitive à ce problème... Voici malgré tout quelques éléments de réponse :

- l'utilisation d'une petite fiche à faire signer aux locuteurs est suggérée. Ce n'est pas la panacée mais constitue pour l'instant la meilleure solution. (J'ai personnellement fait des enregistrements pour ELDA et j'ai signé une feuille de ce type)
- le guide contient quelques termes juridiques à utiliser permettant ainsi une meilleure compréhension entre juristes et linguistes. Serge a suggéré de mentionner notre projet au service juridique de Paris3 pour voir ...

Restent encore les questions éthiques...

Notons qu'il peut être utile de faire la distinction entre l'archivage de corpus d'étudiants (ou de chercheurs), comme nous le faisons, qui est utile dans une optique universitaire; et l'archivage de langues rares (type LACITO) qui à des visées linguistiques plus ambitieuses et qui est peut-être plus précisément concernée par les questions éthiques.

Les VIII^{èmes} rencontres de l'ED268

Notre présentation au cours de la journée des Rencontres de l'ED268 s'est très bien passé. Nous étions en terrain conquis, certes, mais nous avons pu recueillir l'engouement et les attentes des personnes présentes dans la salle. Il ne faut pas relâcher la pression et continuer ce travail de publicité dès la rentrée (voir point 4.).

lien direct vers l'article :

http://pi-ed268.univ-paris3.fr/publis/RJC-ED_fb_cg_050605.pdf

Nous avons envoyé avec Fred une autre proposition de communication pour les journées Doctorants de Paris 7. Cette communication sera un peu plus axée sur la méthodologie. La prochaine étape serait JEP/TALN...

2. Catalogage des fiches de Méta-Données

Nous avons évoqué dans notre précédente réunion la mise en place imminente d'un catalogue qui sera la première partie visible accessible de notre BDD. Serge en a déjà fait circuler un exemplaire... Cette récolte de Méta-Données permettra surtout de proposer le moteur de recherche principal dont nous parlons depuis quelques réunions maintenant. Un « moteur de recherche » fourni par Michel est en cours d'installation sur le serveur. Pour l'instant, il ne permet que d'effectuer des recherches sur les métadonnées. (Serge vient d'envoyer un mail à ce propos, il va décidément plus vite que moi !)

Je vous rappelle que notre objectif était de pouvoir retrouver grâce à ce moteur de recherche des annotations (par exemple les occurrences du mot « Amérique »). Cet aspect faisait partie des requêtes formulées pendant les VIII^{èmes} rencontres de l'ED268. Actuellement, si l'intégralité de l'annotation est insérée dans la partie « description » des métadonnées, il est alors possible d'y accéder de manière un peu détournée.

La prochaine étape sera de mettre en place des liens entre le catalogue et les annotations de fichiers afin de pouvoir opérer une recherche sur une sorte de dictionnaire de formes annotées.

Luigi a mentionné l'existence de moteurs de recherche existants sur Internet, je reprends un extrait de son mail :

- freebank : <http://freebank.loria.fr/corpus.php>

- clapi : <http://corpusgric.univ-lyon2.fr:8000/corpus/>

- et pourquoi pas talkbank <http://talkbank.org> (mais il faut installer avant :-)

Selon Serge, cet aspect ne devrait pas poser de problème... Tant mieux, car c'est un point très accrocheur pour les éventuels utilisateurs !

3. MKM --> améliorations

Serge est en train d'améliorer MKM, une nouvelle version est déjà disponible :

<http://pi-ed268.univ-paris3.fr/files/makeMETADATA-W32-1.08.zip>

Serge est en train d'insérer un manuel convivial pour l'utilisation de MKM.

Je pense que Fred est également en train de s'en occuper puisque cela fera probablement partie de son mémoire.

En effet, un problème à l'heure actuelle reste que le remplissage de MKM n'est pas très facile pour un néophyte, ceci n'étant pas dû à MKM mais plutôt à la fiche DC/OLAC qui n'est pas toujours très intuitive !

Il faudrait que chacun d'entre nous soit familiarisé avec le remplissage de MKM. Les personnes souhaitant contribuer à la base de données entreraient en contact avec l'un d'entre nous, notamment en fonction de la discipline concernée... nous devrions donc tous être en mesure de répondre aux questions diverses.

Rappelons que Serge continue à mettre en place 2 versions de MKM : une version .exe (offline) et une seconde identique mais sur internet (online)

4. les sous à dépenser ...

Il nous reste à vérifier si les 3000 euros qui ont été versés l'année dernière ont été complétés par un 2^{ème} virement de 3000 euros

Pour l'instant, les propositions d'achat sont les suivantes :

- 2 Marantz (à réserver pour les doctorants et +) (devis à fournir par Cédric)

- 4-5 Mini Discs (comme celui de Thierry) avec micro de rechange (devis à fournir par Thierry)

- Logiciels Cordial pour 13 machines (devis à fournir par Luigi)

5 – Programme pour la rentrée – suite du projet !!

Suite de la publicité indispensable... en effet la base de données ne sert à rien si elle n'est pas utilisée. Comment faire ?

- La réunion de rentrée de l'ED268 ... Nous ferons à nouveau un peu de publicité !!
- Fred intégrera dans son mémoire une synthèse : la marche à suivre en quelques points pour contribuer à la base de données.
- Un nouveau cours sera intégré aux prochains savoir-faire (date provisoire 27 janvier 2006) : un point sur la base de données, comment en profiter, comment y contribuer ? + un cours sur Transcriber que nous conseillons !
- Un nouveau cours dans le cursus TAL « Linguistique de corpus » permettra de familiariser les étudiants avec l'archivage de corpus de manière générale ainsi que la base de données de l'ED bien sûr. Les prochains cours de Phonétique/Phonologie devraient intégrer des TDs à rendre qui consistent en un petit corpus étiqueté ... à insérer dans la base de données !!

Nous comptons sur vous tous qui demandez aux étudiants de fournir des corpus pour vos cours :

- o essayez d'insérer systématiquement ces corpus dans la base de données
- o indiquez les synthèses/consignes existantes,
- o indiquez les prochains savoir faire
- o proposez matériel disponible