

11.10.02

Lieu : Annexe de l'Université Paris 8 - 15, rue Catulienne, 93500 Saint-Denis

Présents : Christian, Fanch, Annick, Annelies, Ivani, Marie-Anne, Patrice, Boris, Nathalie, Sophie Dalle.

1. Bilan financier

Une fois les dernières dépenses terminées, nous avons prévu de nous retrouver tous pour une soirée festive de fin de projet, qui pourra avoir lieu le 6, 7 ou 8 décembre, au même moment que la conférence finale des projets COGNITIQUE du Ministère.

2. Corpus

Annick propose la numérisation au format DVD. Marie-Anne doit se renseigner pour commander 20 DVD vierges à Paris 8.

Numérisation différente pour les linguistes (3 caméras) et pour les informaticiens de l'IRIT : sans montage, 1 seule caméra à la fois (l'une après l'autre), en CD, format MPEG 1.

Boris peut repartir du LIMSI avec (il passe 2 jours au LIMSI pour travailler à la mise en place de la vidéo dans l'éditeur).

Boris a besoin de tester des extraits courts de plusieurs locuteurs.

Sur 1 CD : 5 mn de film environ

Sur 1 DVD : tout le corpus normalement

Information :

Le logiciel ANVIL (au format JAVA) présenté à la conférence « Handicap et Technologies » le 04.10. par Jean-Claude Martin du LIMSI semble très intéressant. Il s'agit de communication multimodale et intègre l'éditeur de partition, la vidéo et éventuellement le traitement statistique. Les membres « colins » du LIMSI commencent un projet avec lui dans cette direction.

3. Le point sur l'avancement des travaux

a) analyse linguistique du corpus

A inclure dans le rapport final :

- Ivani : corpus en LSP fait dans cadre du projet, comparaison de 3 LS différentes.
- transcription manuelle du Cheval par Khadra : l'extraction sur images par Ivani, la transcription par Gwenaëlle, la transcription partielle par MA en anglais (pour Deaf Way)
- 2 tableaux récapitulatifs par MA et Ivani
- explication de l'élaboration du corpus

Question de Patrice : quelle analyse faire du corpus ? MA propose un traitement statistique couplé à une microanalyse sur les points pertinents.

b) Mise en forme et diffusion du corpus

Annick propose pour la « valorisation des résultats » du rapport final d'inclure nos objectifs pour la diffusion du corpus. Le LIMSI a contacté l'association ELRA : diffusion de corpus linguistiques, qui propose des contrats très précis de durée et modalités financières variables.

Question : savoir qui a les droits du corpus : P8 (existe-t-il un service de reproduction des corpus) ou le Ministère ? Se renseigner à l'UMR 7023 de P8 (Clive Perdue)

Annick nous rappelle le contenu de la fiche d'engagement du locuteur. Pour plus d'info, voir www.legifrance.fr ou la cellule juridique du CNRS.

c) L'éditeur

Les linguistes de Paris 8 demandent de tester l'éditeur avec Christophe, la présentation à la conférence « Handicap et Technologies » le 03.10. ayant été trop brève pour se faire une idée précise. Nous attendons également que la vidéo soit incluse à l'éditeur.

d) Info : Corpus ARC-LSF

Corpus ARC-LSF : 4 locuteurs sourds enregistrés presque dans les mêmes conditions que le corpus LS COLIN, enregistrement à St Jacques, avec 2 caméras.

Consignes très ciblées : verbes directionnels (dans des énoncés simples) 15 verbes directionnels à conjuguer (des procès à une main, deux mains...). Contraintes liées plus au traitement de l'image que sur les aspects linguistiques. Les linguistes de Paris 8 demandent s'ils pourraient consulter et avoir un exemplaire de ce corpus.

LS-COLIN – Rapport Final PLAN

Rédaction du rapport final en équipe : Format RTF pour TOUS

Date butoir des rapports par chaque responsable des points : le 06 novembre

Rapport à envoyer le 15 novembre 2002.

Relecture de l'ensemble du rapport : [Christian CUXAC](#)

Mise en page final : [Boris](#)

1. PRESENTATION DU PROJET - [Marie-Anne](#)

2. ETAT DE L'ART ET DEMARCHE RETENUE ([Nathalie](#) et Ivani)

2.1 Les différentes transcriptions manuelles (histoire Blanche-Neige)

2.2 Système d'écriture et transcription (à compléter par [Annelies](#) avec le rapport pdf de la conférence LREC)

2.3 Démarche – Grille de Christian : indices de segmentation

3. CONCEPTION ET ELABORATION DU CORPUS Responsable : [Annick](#) (aidée de Marie-Anne)

4. REALISATION DE L'EDITEUR DE PARTITION - [Annelies](#) (Assistée par Patrice Dalle)

5. EXPLOITATION DU CORPUS - [Ivani](#) et Nathalie

Justification de pourquoi on s'est éloigné des objectifs du départ

6. RESULTATS OBTENUS – LES ACQUIS EN MATIERE DE COGNITION

[Christian](#), Annie et Fanch (et Ivani, Marie-Anne, Nathalie)

Apport croisé : encore à venir (intervention d'Annelies)

7. CONCLUSION ET PERSPECTIVES [Christian, Patrice et Annelies](#)

- Vers LS-COLIN 2

Citer les éventuels autres projets : Projet avec J. Claude MARTIN : Action spécifique : la problématique de la transcription de corpus multimodaux (vidéo, parole audio, ...)
Action STIC.

ARC LSF (L'INRIA) – contact de quelqu'un de Bordeaux : Christian Retoré

- Les besoins de continuer

- PUBLICATIONS : en annexe

- 1) Important : Citer la création du site WEB du projet – créé par Patrice Dalle, donner ou pas le mot de passe ?? (ce qui nous obligerait à tout relire et éplucher ce qu'on a écrit !!)
- 2) A corriger : dans l'article collectif de Toulouse, ajouter la mention explicite du soutien financier du Ministère pour le projet LS-COLIN.