

Chère et cher collègue,

Le comité de pilotage de l'IRCOM, consortium "Corpus Oraux et Multimodaux", vous propose, pour la prochaine réunion plénière du comité scientifique, un nouveau format qui permettra d'articuler le bilan des activités du consortium avec trois ateliers autour de thèmes fondamentaux pour le consortium.

Cette réunion sera ouverte à **toute la communauté des chercheurs et enseignants-chercheurs** intéressés aux corpus oraux et multimodaux.

Le programme de la journée comprend :

- un atelier "**Métadonnées**" pour discuter d'un jeu commun de descripteurs, à partir des propositions initiées lors de la table ronde sur les formats du 23 juin et d'exemples pris dans les différents corpus oraux du projet Orféo.
- un atelier "**Annotations**" à partir des travaux du groupe 1 "Corpus linguistiques et finalités scientifiques" qui ré-interrogera les pratiques d'étiquetage dans l'annotation des corpus oraux et plus précisément les interactions entre pratiques d'étiquetage, théorie et objectifs descriptifs, à la lumière des langues peu décrites, non écrites et/ou d'une autre modalité (langues des signes) mais aussi des corpus multimodaux.
- un atelier "**Archivage des corpus**" d'après les travaux du groupe 6 "Stockage sécurisé et mutualisation des corpus oraux et multimodaux" illustrera les différentes étapes de l'archivage. Ces étapes vont du dépôt jusqu'à la consultation des ressources déposées, en passant par le nommage des fichiers, les formats de fichiers, le référencement et la valorisation. Les spécificités des principaux acteurs, BnF, Cocoon et Ortolang/SLDR, seront présentées à travers des exemples concrets de corpus.

Ces ateliers seront organisés en deux sessions de même format pour vous permettre de participer à deux ateliers sur les trois proposés.

La journée se conclura par un bilan des ateliers et une session de discussion.

Nous vous invitons à diffuser largement cette information et à convier d'autres personnes de vos unités qui seraient intéressées à développer ces thématiques.

La réunion se déroulera au centre **FIAP – Jean Monnet , 30 rue Cabanis – 75014 – PARIS**

<http://www.fiap.asso.fr/>

Métro ligne 6, stations Saint-Jacques ou Glacière.

RER B : Station Denfert Rochereau

Bus : N°21

Pour vous inscrire envoyer un courriel à l'adresse suivante : <secretariat-general@ling.cnrs.fr>
AVANT LE 15 SEPTEMBRE 2014

Vous pouvez consulter le site de l'IRCOM : <http://ircom.huma-num.fr/site/accueil.php>

A titre exceptionnel et à destination des jeunes chercheurs intéressés pour participer à cette réunion, une demande de prise en charge des frais de transport peut être faite à l'adresse mail ci-dessus. Pour être prise en compte, cette demande doit parvenir avant le 15 septembre 2014.

CONSEIL SCIENTIFIQUE IRCOM

Lundi 22 septembre 2014 de 9h15 à 17h00

Centre FIAP Jean Monnet - 30 rue Cabanis - 75014 Paris

Ordre du jour

08h 45 – 09h 15 - ACCUEIL

Matinée : Présentation et réalisation d'IRCOM et ateliers

09h 15 – 09h 30 :

- Présentation générale : structure, fonctionnement, réalisations - Dominique Boutet

09h 30 – 11h 00 :

- Présentations des 5 groupes de travail du consortium
 - *Corpus linguistiques et finalités scientifiques*
 - *Interopérabilité*
 - *Multimodalité et modalité visuo-gestuelle*
 - *Questions juridiques, droits des personnes et des producteurs de corpus*
 - *Stockage sécurisé et mutualisation des corpus oraux et multimodaux*

11h 00 – 11h 30 - PAUSE

11h 30 – 13h 00

- Atelier Métadonnées

Christophe Benzitoun et Carole Etienne. Cet atelier se propose de discuter d'un jeu commun de métadonnées qui donnerait les informations nécessaires sur les corpus oraux pour faciliter leur réutilisation par des chercheurs ou des enseignants. En évitant de reprendre les débats sur les catégorisations ou de remettre en cause les descriptions existantes, cette étude cherche à définir les bonnes pratiques en matière de métadonnées. Cette réflexion s'inscrit dans la continuité des téléconférences en février et de la table ronde consacrée aux formats communs en juin et bénéficiera de nombreux exemples parmi les corpus du projet Orfeo.

- Atelier Archivages

Olivier Baude (pour Cocoon), Audrey Viault (département Archives et Manuscrits, BnF), Jean-Marie Pierrel et Jérôme Blanchard (ORTOLANG/SLDR). Cet atelier présentera les étapes de l'archivage et de la valorisation de corpus oraux et multimodaux à travers quelques réalisations qui illustrent les spécificités de chaque institution. Les participants pourront proposer des corpus pour une simulation du processus d'archivage adapté à leur cas.

- Atelier Annotations

Amina Mettouchi, Anne Lacheret Brigitte Garcia et Paola Pietrandrea. Cet atelier propose d'aborder les questions suivantes, aux fondements mêmes des pratiques d'annotation. Dans quelle mesure les hypothèses et les objectifs qui sous-tendent les travaux en linguistique de corpus peuvent-ils/doivent-ils être explicités pour justifier le jeu d'étiquettes utilisées pour les annotations ? Les questions auxquelles se heurte l'annotation ne doivent-elles pas servir de levier pour impulser des propositions théoriques engageant de nouveaux modes de représentation linguistique ? Jusqu'à quel point peut-on tenir le pari de l'approche a-théorique ou « non-aprioristique » dans l'annotation des corpus ? A partir de quel niveau d'analyse et pourquoi celle-ci ne peut-elle plus être uniquement fixée sur des critères de formes ? N'y a-t-il pas un noyau fonctionnel a minima incontournable ? Si oui, de quelle nature (phonémique, morphologique, syntaxique...) ? Y a-t-il à cet égard des spécificités liées aux différentes langues, au type de modalité utilisée, etc. ? Au final, ces questions convergent vers celle de la définition des unités (minimales, maximales) retenues pour l'annotation.

13h 00 – 14h 00 - DEJEUNER

Après-midi : Ateliers, restitutions et discussions

14h 00 – 15h 30

- Atelier Métadonnées - (Animateurs et sujets identiques à ceux de la matinée)
Christophe Benzitoun et Carole Etienne. Cet atelier se propose de discuter d'un jeu commun de métadonnées qui donnerait les informations nécessaires sur les corpus oraux pour faciliter leur réutilisation par des chercheurs ou des enseignants. En évitant de reprendre les débats sur les catégorisations ou de remettre en cause les descriptions existantes, cette étude cherche à définir les bonnes pratiques en matière de métadonnées. Cette réflexion s'inscrit dans la continuité des téléconférences en février et de la table ronde consacrée aux formats communs en juin et bénéficiera de nombreux exemples parmi les corpus du projet Orféo.
- Atelier Archivages - (Animateurs et sujets identiques à ceux de la matinée)
Olivier Baude (pour Cocoon), Audrey Viault (département Archives et Manuscrits, BnF), Jean-Marie Pierrel et Jérôme Blanchard (ORTOLANG/SLDR). Cet atelier présentera les étapes de l'archivage et de la valorisation de corpus oraux et multimodaux à travers quelques réalisations qui illustrent les spécificités de chaque institution. Les participants pourront proposer des corpus pour une simulation du processus d'archivage adapté à leur cas.
- Atelier Annotations - (Animateurs et sujets identiques à ceux de la matinée)
Olivier Baude (pour Cocoon), Audrey Viault (département Archives et Manuscrits, BnF), Jean-Marie Pierrel et Jérôme Blanchard (ORTOLANG/SLDR). Cet atelier présentera les étapes de l'archivage et de la valorisation de corpus oraux et multimodaux à travers quelques réalisations qui illustrent les spécificités de chaque institution. Les participants pourront proposer des corpus pour une simulation du processus d'archivage adapté à leur cas.

15h 30 – 15h 45 - PAUSE

15h 45 – 17h 00

- Restitutions et discussions
