



Consortium « Corpus Écrits »

ILF, Paris, 29 octobre 2012

Chères Collègues, Chers Collègues,

Le consortium « Corpus écrits » (Corpus-IR) organise sa réunion plénière annuelle le vendredi 24 novembre 2012, de 9h30 à 18h, au Campus des Cordeliers (15, rue de l'École de Médecine, 75006, Paris). Cette réunion sera consacrée à la **présentation des activités des différents groupes de travail du consortium** :

1. Usage des corpus et droits d'auteurs ou d'éditeurs (aspects juridiques)
2. Corpus d'états anciens de la langue (numérisation, codage)
3. Numérisation (OCR, saisie), correction
4. Pluralité de systèmes d'écriture
5. Corpus multilingues (parallèles, comparables...)
6. Description de corpus collaborative - métadonnées
7. Corpus d'écrits modernes et prise en compte de nouveaux modes de communication (SMS, mail, blog, etc.)
8. Annotation de plus haut niveau : syntaxe, sémantique, référence (annotations collaboratives)
9. Annotation de surface : segmentation lexicale, description morphosyntaxique, chunking, lemmatisation, entités nommées, etc.
10. Exploration de corpus (méthodes, outils)
11. Qualité scientifique et accessibilité des corpus (place des corpus dans l'évaluation de la production scientifique des UR)

Chaque présentation sera suivie d'une discussion d'une vingtaine de minutes.

Cette journée, très importante pour l'avancée de la réflexion commune sur les corpus écrits, sera suivie, le lendemain, **samedi 24 novembre**, d'une **journée d'information et d'échanges sur les aspects juridiques de la propriété et de l'archivage des corpus**, dont le programme sera communiqué très prochainement.

La participation de toutes les personnes intéressées par ces journées est vivement encouragée par le comité de pilotage, qu'elles soient ou non inscrites à un groupe de travail.

Si la participation à ces journées est libre, **l'inscription est obligatoire**. Vous trouverez en pièce jointe le **formulaire d'inscription** à retourner **au plus tard le 12 novembre 2012** au secrétariat de l'Institut de Linguistique Française, institution gestionnaire du Consortium, qui, le cas échéant, prendra contact avec vous pour organiser votre mission.

Le consortium contribuera au financement des missions des participants actifs des groupes de travail.

Au plaisir de vous accueillir nombreux les **23 et 24 novembre prochains**.

Pour le comité de pilotage du Consortium « Corpus écrits »
Franck Neveu, Directeur de l'ILF

Le comité de pilotage du Consortium « Corpus écrits » :

Franck Neveu pour l'ILF, FR 2393 – **porteur du consortium**

Sylvie Archambault (suppléant **Bernard Colombat**) pour HTL – UMR 7597 - Université Denis Diderot - Paris 7

Benoît Sagot pour ALPAGE – INRIA- Université Denis Diderot - Paris 7

Serge Heiden pour ICAR - UMR 5191 - Université Lumière Lyon 2

Damon Mayaffre (Suppléante **Mahé Ben Hamed**) pour BCL - UMR 6039 - Université Nice Sophia Antipolis

Jean-Marie Pierrel pour l'ATILF - UMR 7118 – Nancy - Université

Clément Plancq (suppléant **Olivier Bonami**) pour le LLF - UMR 7110 - Université Denis Diderot - Paris 7

Céline Poudat pour le LDI - UMR 7187 – Université de Paris 13

Catherine Schnedecker (suppléante **Amalia Todirascu**) pour LILPA – EA 1339 – Université de Strasbourg

Agnès Tutin (suppléante **Marie- Paule Jacques**) pour le LIDILEM – EA 609 – Université Grenoble 3