

Réunion COMERE 26 juin 2014

Présents (11) : Georges Antoniadis, Thierry Chanier, Linda Hriba, Kun Jin, Gudrun Ledegen, Julien Longhi, Paul Lotin, Camille Paloque, Céline Poudat, Benoît Sagot, Ciara Wigham,

Excusé : Achille Falaise

La réunion début à 10h au CNAM, Paris, avec l'ordre du jour suivant :

1) Point sur Corpus-écrits, formations, finance GT7 *** membres du COP (Céline Poudat, Benoit Sagot, Thierry Chanier) Cf. ppt de présentation sur le site

La formation aura lieu les 19 et 20 novembre au Grand Moulin, avant l'AG annuelle de Corpus-Ecrits le vendredi 21 novembre 2014 à Paris. 4 demi-journées pour 20 à 25 personnes. Elle constitue la continuation des premières journées de formation, et pourra s'ouvrir à d'autres participants en fonction des places disponibles.

2) Phase 1.1 (achevée) de CoMeRe : Application du modèle Interaction Space (TEI-CMC) aux 7 corpus (Textos, Tweets, Forum, Blogues, courriel, clavardage) *** Thierry Chanier Cf. ppt de présentation sur le site

3) Phase 1.1 (achevée) de CoMeRe : Travail du groupe Qualité *** Membres du groupe Qualité (Julien Longhi et Ciara Wigham) : présentation du parcours de vérification de la qualité des 7 corpus pour qu'ils accèdent à la version tei-v1 (prêts à entre dans Ortolang). Cf. ppt de présentation sur le site

4) Phase 1.1 (en cours) de CoMeRe : Phase de transfert des corpus CoMeRe vers Ortolang *** Paul Lotin Cf. ppt de présentation sur le site

5) Coopération européenne : groupe de travail TEI-CMC *** Thierry Chanier Cf. ppt de présentation sur le site

6) Phase 1.2 (en cours) de CoMeRe : Acquisition, organisation et vérifications corpus discussions Wikipedia (Wikiconflits) *** Cf. ppt de présentation sur le site

- Rappel objectifs : Céline Poudat

- travail accompli : Kun Jin

- travail de vérification, reconstitutions à faire : discussion (entre autres) avec membres du groupe nouvelles acquisitions concernés par cette sous-phase

7) Phase 2 (commencée) de CoMeRe : étiquetage morpho-syntaxique

- présentation de la phase pilote sur deux corpus : Benoit Sagot
- possibilité d'extension aux autres corpus : Georges Antoniadis, Benoit Sagot
- dernière mise en forme et vérification TEI avant passage dans Ortolang sur ces travaux

8) perspectives de collaborations / travaux scientifiques *** animé par Ciara Wigham

Le projet CoMeRe a comme premier objectif de permettre aux chercheurs collecteurs des données primaires de pouvoir ensuite lancer des programmes de recherche sur les corpus nouvellement organisés.

- publications : en cours (soumission) et perspectives
 - Revue TAL, Varia
 - Revue Corpus (Nice) : solliciter les collègues allemands et italiens
 - NLP-CMC
 - Language Research Evaluation (sur Qualité)
- conférences et ateliers : perspectives
 - NLP4CMC : 3 soumission (date limite 15/7)
 - Corpus Linguistics Lancaster (22-24/7/2015) (bi-annuel)
 - JEP-TALN, janvier 2015
 - TAL 2015
 - Conférence ELREC dans 2 ans
 - ACFAS, Canada
- projets : ANR, projets européens
 - ANR à démarrer à la rentrée pour mise en place en septembre 2015 autour de la notion de 'conflit' (à étudier sur plusieurs genres : wikipédia, tweets, sms, clavardage, ...)
 - ANR franco-allemand à mettre en route, maintenant que les corpus sont fonctionnels : rencontre à l'occasion de la conférence à Rome (18/9/2014) et organiser une rencontre élargie début 2015 (G. Ledegen recherchera les modalités de financements européens pour l'organisation d'une telle rencontre).

9) rappel dernières étapes avant fin du projet CoMeRE en 2014 : Thierry Chanier