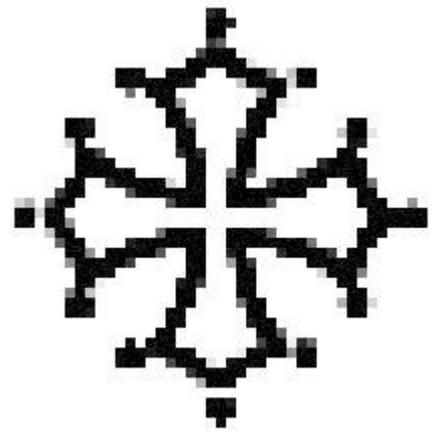


OCC-Haton 1

L'annotation linguistique
de corpus occitans en
diachronie longue



Vendredi 4 avril, Maison de la recherche, 28 rue Serpente
Organisation : Gilles Couffignal, Sorbonne Université
Partenaires : INRIA, AIEO

Le principe du OCC-haton est de dédier une journée de travaux à la réalisation de tâches computationnelles (hackhaton) sur des données en occitan.

Cette première journée est consacrée à l'annotation linguistique des corpus occitans, depuis les premiers textes jusqu'à l'époque contemporaine.

L'objectif est de mieux connaître et comprendre les différents projets existants mais également de permettre une harmonisation des pratiques afin de mettre en commun des données d'entraînement et d'ouvrir la perspective à des outils de lemmatisation généralistes et performants.

Présentation des bases existantes

08h45 Accueil des participants

09h00-09h15 Les enjeux des référentiels d'annotation (G. Couffignal, SU)

09h15-09h45 DocLing (M. Glessgen, Zürich)

09h45-09h00 Tolosa TreeBank (M. Bras, Toulouse 2)

10h15-10h45 CorAG (N. Romanova, B. Francioni, Caen)

10h45-11h15 Old Gallo-Romance corpus (Th. Rainsford, Stuttgart)

11h15-11h45 Transfert de technologie : Latin et ancien français (T. Clérice, L. Ing, O. Nédey, INRIA)

Ateliers

14h00-14h20 Table ronde sur les corpus en projet et la question de la segmentation

14h20-14h40 Proposition et discussion d'un référentiel basé sur Levy et LoFlòc

15h00-16h30 Prise en main de l'interface de post-correction et de contrôle des référentiels Pyrrha (L. Ing)

Contact :

gilles.couffignal@sorbonne-universite.fr