

Appel à contribution

Numéro spécial de la revue Corpus (<http://corpus.revues.org>)

Spécificités et contraintes des grands corpus de textes scolaires :

problèmes de transcription, d'annotation et de traitement

Coordonnateurs :

- Claire Doquet (EA 7345 Clesthia, Université de la Sorbonne Nouvelle)
- Jacques David (EA 1392 CRTF, Université de Cergy-Pontoise)
- Serge Fleury (EA 7345 Clesthia, Université de la Sorbonne Nouvelle).

Les écrits des élèves suscitent un intérêt grandissant chez de nombreux chercheurs appartenant à des domaines ou des paradigmes de recherche aussi variés que la linguistique, la psycholinguistique, la sociolinguistique et la didactique du français. Cet intérêt s'explique à la fois par la singularité de l'objet discursif qu'ils constituent et par la rareté des études empiriques appuyées sur des corpus de grande envergure. Si des corpus importants ont pu être publiés notamment dans de nombreuses thèses en sciences du langage, les initiatives de ce type demeurent relativement rares et limitées ; l'analyse des textes reste partielle et leur accès difficile.

La liste de l'université de Louvain recensant les bases de données de Learner Corpora (<http://www.uclouvain.be/en-cecl-lcworld.html>) ne fait apparaître que 5 bases en français écrit, dont 3 sont du FLE ou FLS. Le Learner Corpus French (Université de Genève, Yuji Kawaguchi Tokyo University of Foreign Studies), projet de grande ampleur, est en développement. En France, certaines équipes (en particulier à l'université de Cergy-Pontoise, cf. Elalouf & Boré, 2007) ont constitué des corpus mais butent sur des difficultés à les faire partager et surtout à les traiter. De fait, malgré les avancées considérables des outils informatiques d'analyse de textes et les méthodologies liées aux grands corpus oraux, le traitement quantitatif des données langagières émanant de scripteurs débutants ou en cours d'apprentissage est difficile du fait du caractère linguistiquement peu normé (ou autrement normé) de leurs productions.

Après la tenue d'une journée d'études sur ces questions le 18 mars dernier (<http://www.univ-paris3.fr/activites-ecriscol-300518.kjsp?RH=1416243625396>), nous souhaitons donner une visibilité à ces travaux en programmant un numéro spécial de la revue Corpus à paraître au printemps 2016. Ce numéro accueillera des articles proposant une synthèse des problèmes et/ou faisant état de travaux en cours dans le domaine. Plusieurs questionnements

peuvent être envisagés :

- Mode de présentation des données dans une base : numérisation des écrits manuscrits, transcription diplomatique vs linéaire, annotation pour traitement informatique, textométrie, etc.
- Norme, mise aux normes : quelle(s) interface(s) entre le matériau disponible, leur publication ou mise en ligne et les exigences des outils informatiques ?
- Aspects génétiques de l'écriture : la place des notes, l'analyse des biffures et ratures, la saisie des brouillons et de l'ensemble des avant-textes et péri-texte d'un écrit d'élève. Pour les copies d'élèves se posent de manière aiguë, en plus de la textualisation, tous les problèmes liés à l'acquisition du système linguistique, en particulier les unités de segmentation, la ponctuation, l'ensemble des données (ortho)graphiques, mais aussi les procédés sémantico-syntaxiques et d'organisation textuelle.
- Prise en compte des contextes didactiques : démarche, consignes, supports, modes de production et d'édition... en particulier dans le choix des métadonnées.

Dates à respecter (des indications complémentaires seront envoyées aux auteurs des articles retenus) :

- 20 septembre 2015 : réception des intentions de publication (un résumé d'une page et une bibliographie de référence).
- 10 octobre 2015 : notification de l'acceptation du projet d'article.
- 5 janvier 2016 : réception des articles.
- 31 janvier 2016 : retour des évaluations du comité de lecture.
- 25 février 2016 : réception des articles révisés.

Le comité de lecture du numéro est en cours de constitution.

Merci d'envoyer vos propositions d'articles aux trois coordonnateurs :

Claire Doquet : doquet.claire@orange.fr

Jacques David : jak.david@orange.fr

Serge Fleury : sergefleury@gmail.com