

Journées d'étude
Le sens des données

Statut du corpus et herméneutique à l'aune des humanités numériques

organisées par le Centre de Recherches Sémiotiques (CeReS) EA 3648

les 29 et 30 septembre 2016

à l'Université de Limoges

Comité d'organisation : Sophie Anquetil, Carine Duteil-Mougel, Vivien Lloveria

Cette journée s'inscrit dans le cadre d'activités menées par le Centre de Recherches Sémiotiques (CeReS) de Limoges et vise à rassembler sémioticiens et linguistes autour de la question du corpus numérique. Depuis l'avènement des technologies informatiques, des outils d'exploration offrent aux Sciences du langage et aux Sciences de l'information et de la communication la possibilité de constituer et de traiter des (bases de) données, qui engagent un nouveau rapport à l'empirique et donnent accès à de nouveaux observables.

De nombreux travaux consacrés aux corpus questionnent leur statut et s'intéressent aux modalités de leur traitement ; notre journée entend s'en démarquer en proposant une réflexion conjointe de la sémiotique et de la linguistique sur les corpus numériques contemporains afin de se demander si, et par quels modes d'exploitation, les données qui en résultent peuvent objectivement faire sens. Nous chercherons notamment à rendre compte d'une tension constatée entre, d'une part, l'injonction de scientificité à laquelle répondrait le corpus et, d'autre part, le caractère perméable et instable des corpus multimédiatiques numériques qui peuvent être le fruit de co-voire de multi-énonciations.

Pour organiser les débats nous proposons trois axes principaux et des pistes de questionnement :

Axe 1 : Qu'est-ce qu'un corpus numérique aujourd'hui ?

1. Comment une banque de données de textes et d'images devient-elle un corpus ?
2. Sur quels critères se fonde-t-on pour attribuer le statut de corpus ?
3. Quels traitements impose le statut de corpus ?
4. Suivant le mode d'exploration adopté, traite-t-on des réalisations effectives ou seulement des données perçues par le locuteur ?

Axe 2 : Quelles méthodologies du corpus aujourd'hui ?

1. Comment évaluer les critères de représentativité, d'exhaustivité et d'homogénéité du corpus dans un univers linguistique et sémiotique en mutation ?
2. Comment penser le corpus collaboratif ? Quels sont les retours d'expériences de co-constitution de corpus ?
3. Comment traiter des corpus polysémiotiques ?
4. Dans une perspective visuelle du corpus, quelle autonomie accorder à l'image vis-à-vis des textes qui l'accompagnent (critiques d'art, écrits biographiques, historiques) ?

Axe 3 : Quelles relations au corpus aujourd'hui ?

1. Comment appréhender cette récente "interaction homme-données" (Haddadi, H., Mortier, R., et al., 2013) qui concurrencerait l'interaction hommes-textes ?
2. Qu'entend-on par "humanités numériques" ?
3. Doit-on "laisser dire" le corpus ou le "faire dire" ? Quelle est l'épistémologie dominante ? Le corpus est-il un objet comportant ses propres lois ou doit-il être appréhendé comme une construction subjective ?
4. Peut-on parler d'une idéologie du corpus ? Se déclarer pro- ou anti- corpus a-t-il un sens ?

Les propositions de communication sont à envoyer conjointement à sophie.anquetil@unilim.fr, carine.duteil@ensil.unilim.fr, et vivien.lloveria@unilim.fr.

Elles comporteront le titre de la communication, un résumé de 1500-2000 signes (espaces non compris) accompagné d'une bibliographie indicative ainsi que de 5 mots-clés. Merci de préciser dans le corps du message le nom et l'appartenance institutionnelle de l'auteur ou des auteurs, et l'adresse de correspondance.

La date limite de soumission des propositions de communication est fixée au **15 mai 2016** et les notifications d'acceptation seront transmises le **15 juin 2016**.

Bibliographie indicative

Augustyn M., & Tutin A., 2009, "Constitution d'un corpus annoté autour du lexique des émotions : collocations et fonctions lexicales". In Beck D., Gerdes K., Milićević J. and Polguère A. *Actes de la Quatrième conférence internationale sur la Théorie Sens-Texte (MTT' 09)*, OLST (pp. 25-35). Disponible à l'adresse : <http://meaningtext.net/mtt2009/03-MAugustynATutin.pdf>

Beyaert-Geslin, A., Renoue, M., 2013, "L'image artistique à l'ère de la reproduction numérique : sémiotique visuelle et interfaces", *Interfaces numériques*, Vol. 2, n°2, pp.171-177.

Bommier-Pincemin, B., 1999, *Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents*, Thèse de l'Université Paris IV, Sous la direction de F. Rastier, 805 p.

Bonaccorsi, J., 2013, "Pratiquer les images en Sciences de l'information et de la communication: semiose, eikones, montage". *Revue française des sciences de l'information et de la communication*, (3).

Brandhorst, J. P. J., 1993, "Quantifiability in iconography". *Knowledge organization*, 20(1), 12-19.

Chartrand, L., Chartier, J. F., & Meunier, J. G. « Des mots pour se retrouver : recherche d'information dans l'œuvre de Magritte à l'aide d'un corpus de descripteurs sémiotiques ».

Condamines A., 2005, « Linguistique de corpus et terminologie », *Langages*, vol. 39, n°157, pp.36-47.

Habert, B., Nazarenko, A. et Salem, A., 1997, *Les linguistiques de corpus*. Paris, Armand Colin — Masson.

Legallois, D., 2006, « Le texte et le problème de son et ses unités : propositions pour une déclinaison », *Langages*, vol. 40, n° 163, pp. 3-9.

Legallois, D. et François J. , 2011, « La Linguistique fondée sur l'usage : parcours critique », *Travaux de linguistique*, vol.1 n°62, pp. 7-33. DOI : 10.3917/tl.062.0007

Mayaffre, D., 2002, « Les corpus réflexifs : entre architextualité et intertextualité », *Corpus*, I, 1, pp. 51-70.

Mayaffre D., 2010, « Corpus et Web-corpus. Réflexion sur la corporalité numérique », *Cahiers de praxématique*, Publications de l'Université Paul Valéry, pp.233-248.

Mayaffre D., 2015, « Textes augmentés et discours numériques Pour une approche logométrique des digital corpora », *Text and discourse in confrontation in the European context: towards an epistemological and heuristic renewal*, Sep 2015, Metz, France.

- Haddadi, H., Mortier, R., McAuley, D., & Crowcroft, J., 2013, "Human-data interaction". *University of Cambridge*.
- Panofsky, E., 1967, "Essais d'iconologie: thèmes humanistes dans l'art de la Renaissance", trad. C. Herbette & B. Teyssèdre, Paris, Gallimard.
- Paveau M-A, 2014, « Quand les corps s'écrivent. Discours de femmes à l'ère du numérique », dans Éric Bidaud (dir.), *Recherches de visages. Une approche psychanalytique*, Paris, Hermann, p. 207-241.
- Paveau M-A, 2015, « Les textes numériques sont-ils des textes ? », *Itinéraires LTC*, 2015, Textualités numériques, 2014-1 2015, <https://itineraires.revues.org/2312>.
- Paveau M-A., 2015, « L'intégrité des corpus natifs en ligne. Une écologie postdualiste pour la théorie du discours », *Cahiers de praxématique* [En ligne], 59 | 2012, document 3, mis en ligne le 01 janvier 2015, consulté le 26 janvier 2016. URL : <http://praxematique.revues.org/3359>
- Paveau M-A, 2015, « En naviguant en écrivant. Réflexions sur les textualités numériques », in Adam J-M., *Faire texte. Frontières textuelles et opérations de textualisation*, Presses universitaires de Franche-Comté.
- Rastier F., « Sémiotique des sites racistes », *Mots. Les langages du politique* [En ligne], 80 | 2006, mis en ligne le 01 mars 2008, consulté le 25 janvier 2016. URL : <http://mots.revues.org/17113>
- Rastier F., 2010, « Web Semantics vs the Semantic Web : the Problem of Keyness », in Marina Bondi and Mike Scott, *Keyness in Texts*, Studies in Corpus Linguistics, 41, Amsterdam-Philadelphia, Benjamins, pp. 93-112
- Rastier F., 2011, *La Mesure et le Grain. Sémantique de corpus*, Paris, Champion.
- Stockinger, P., 2012, *Analyse des contenus audiovisuels : métalangage et modèles de description*, Lavoisier, Paris.
- Trudel, É., & Hébert, L., 2011, « Protocole d'analyse de la base de données "Magritte". Toutes les œuvres, tous les thèmes ».
- Valette M., et Egle E., 2015, « Une sémantique de corpus pour la fouille de textes », *La sémantique et ses interfaces. Actes du colloque 2013 de l'Association des Sciences du Langage*, textes réunis et présentés par A. Rabatel, A. Ferrara-Léturgie et A. Léturgie, éd., Lambert-Lucas, Limoges, pp. 205-224.
- Vanni L., Luong X., Mayaffre D., 2014, « Arbre et co-occurrences Nouvel outil logométrique sur le net. Application au discours de François Hollande », *JADT 2014*, Jun 2014, Paris, France.