

ICODOC 2015 : Icar Colloque DOctorants/DOcteurs

<http://icodoc2015.sciencesconf.org>

Corpus complexes et enjeux méthodologiques : de la collecte de données à leur analyse

18 & 19 mai 2015, ENS de Lyon

Le laboratoire ICAR (UMR 5191) est spécialisé dans l'analyse multidimensionnelle des usages de la langue dans l'interaction et dans le texte. Ces analyses sont appréhendées de manière outillée sur des corpus « complexes » de données orales interactives et/ou de données textuelles. On entend par corpus « complexe » un ensemble cohérent de données « multi-modales » - associant au choix de la vidéo, du son, des textes, des traces numériques, des images, etc. - sans nécessairement impliquer un « gros » volume de données.

Pour la première fois, le laboratoire ICAR organise un colloque international pluridisciplinaire de jeunes doctorants/docteurs, impliquant, entre autres, les Sciences du Langage et les Sciences de l'Éducation (les deux disciplines représentatives du laboratoire ICAR). Ce colloque a pour thématique :

« Corpus complexes et enjeux méthodologiques : de la collecte de données à leur analyse ».

Pour cette première édition 2015, nous souhaitons en effet nous concentrer sur la notion de corpus complexes, qu'il s'agisse de corpus oraux, de corpus écrits (textuels), de corpus issus du web (oral-écrit) et/ou de corpus pluri/polysémiotiques¹. Les visées, les outils et les techniques intervenant dans la constitution de corpus complexes à des fins d'analyse par des chercheurs ainsi que les questions méthodologiques intéressent ce colloque.

Les communications attendues devront apporter une contribution en lien avec le langage et la constitution de corpus complexes/multi-modaux, et dont les méthodes d'approche peuvent être, à titre d'exemples :

- la linguistique interactionnelle,
- l'analyse du discours,
- les approches pluridisciplinaires de l'interaction,
- le traitement automatique des corpus écrits et oraux,
- la didactique de l'enseignement / apprentissage des langues, des sciences, des mathématiques et des savoirs professionnels,
- l'approche en corpus d'un fait de langue, qu'il soit phonologique, morphosyntaxique, lexical, orthographique, ou sémantique.

¹ Au sens très large : au niveau des modes sémiotiques (l'écriture, le dessin, la parole, les gestes) et au niveau des média portant ces modes (le papier, l'écran, le son de la voix, qu'il soit naturel ou enregistré, le corps) (Lund, Bécu-Robinault, 2009).

Sont invités à proposer une communication les doctorants, post-doctorants et jeunes chercheurs dont les objets d'étude sur le langage impliquent la constitution de corpus complexes et s'intéressent aux enjeux méthodologiques du recueil de corpus.

Nous proposons quatre axes de réflexion autour de la posture du chercheur :

Axe 1

La problématique de terrain : modalités d'approches du terrain et de recueil de corpus

On pourra réfléchir aux exigences multiples qui se posent au chercheur à différentes étapes du recueil de données (quelles que soient leurs formes).

- Quelles peuvent-être les modalités d'approches du terrain (en termes de réseaux, contacts, forme de négociation, etc.) ? Comment prépare-t-on le recueil de données sur un terrain (autorisation, accès, etc.) ? Quels sont les choix à faire, quels sont les choix qui s'imposent ?
- Quelles sont les contraintes diverses auxquelles doit penser le chercheur (psychologiques, éthiques, matérielles, institutionnelles, temporelles, techniques, etc.) ?
- Quelles articulations et mises en cohérence peuvent s'opérer entre les objectifs de la recherche, les contingences du terrain et la constitution de corpus ?

Axe 2

La problématique de traitements : manuel et/ou informatique

On s'interrogera ici sur les choix opérés par le chercheur au moment de traiter ses données, d'un point de vue méthodologique tout en préservant sa/ses problématique(s) de recherche.

- Quels sont les enjeux épistémologiques liés aux choix de transcriptions et/ou d'annotations (verbales, multimodales, etc.) ?
- Quelles nouvelles formes de transcriptions, d'annotations (i.e. représentation graphique d'une interaction, d'énoncés, de discours, de mots, etc.) sont aujourd'hui proposées au chercheur ? Quels en sont les avantages et les inconvénients ?
- Quel impact peut avoir l'usage de logiciels d'annotation sur l'orientation méthodologique de sa recherche ?
- En quoi le travail de balisage des données peut-il venir enrichir ou alourdir le travail analytique du chercheur ?

Axe 3

La problématique de l'analyse : approches qualitatives, quantitatives et mixtes

Au niveau de l'exploitation des données, on pourra s'interroger également sur l'articulation possible entre les analyses à la fois qualitatives et quantitatives.

- Quelles sont les potentialités et les limites d'une analyse quantitative par rapport à une problématique de recherche ?
- Face à des données (nombreuses et complexes), quels choix (méthodologiques, théoriques, analytiques, etc.) se présentent au chercheur pour exploiter au mieux son corpus (en terme d'efficacité, de représentativité, de pertinence, d'outils d'exploitation, etc.) ?
- Quel impact l'usage des bases de données, des plateformes outillées dédiées à la recherche peut-il avoir aujourd'hui sur l'analyse quantitative et qualitative de ses propres corpus ?

Axe 4

La problématique dans la diffusion et la mutualisation des corpus : bases de données, plateformes outillées, etc.

Il y a actuellement différentes plateformes en France dédiées à la mutualisation des données. Un bref panorama informatif du paysage actuel sur cette question sera envisagé en introduction de cet axe. On pourra ainsi développer une réflexion sur la place et l'usage des bases et plateformes outillées aujourd'hui dans la recherche.

- Comment augmenter la visibilité des travaux et des ressources produites par les chercheurs à partir des bases et des plateformes outillées ?
- Comment diffuser et mutualiser des corpus - dans toute leur complexité - au sein des bases et des plateformes malgré les contraintes (techniques, juridiques, humaines, etc.) que cela peut parfois représenter ?
- Quelle est la place, le rôle du chercheur dans ce processus de diffusion et de mutualisation sur des bases de données et des plateformes ?

Les questions abordées autour de ces quatre axes permettront d'enrichir les discussions actuelles sur la notion de corpus « complexes » à laquelle est fréquemment confronté le jeune chercheur.

Ce colloque permettra aux jeunes doctorants/docteurs de présenter leurs recherches et de confronter leurs réflexions avec celles de collègues venus de France ou de l'étranger, mais aussi de créer un réseau avec d'autres équipes/laboratoires de linguistique ou de didactique², ainsi que de renforcer les réseaux existants (i.e. avec Montpellier, Paris, Toulouse, Grenoble, Orléans, Suisse, Belgique, Maghreb, Canada, etc.)

Publications prévues

Les contributions retenues (orales ou poster) donneront lieu à deux formes de publication possibles :

- Une version allégée des actes du colloque en ligne (résumé long, 5 pages maximum) avant le colloque.
- Une version papier sous forme d'articles dans un ouvrage collectif à l'issue du colloque. Nous sommes en discussion avec trois maisons d'éditions pour au final n'en retenir qu'une seule.

Format des contributions

Nous acceptons au choix deux formes de présentations :

- communication orale
- poster

Les propositions de contribution devront comprendre un résumé d'**au maximum** 5000 signes (espaces compris), et une courte bibliographie (non comprise dans le nombre total de signes).

Chaque participant ne peut déposer qu'une seule proposition (communication ou poster). L'évaluation des contributions se fera en double-aveugle. Le comité scientifique et le comité

² i.e. les acteurs de la banque de données ViSA - réseau spécialisé dans les questions méthodologiques - qui travaille notamment sur la question de la diffusion/mutualisation des corpus.

d'organisation d'ICODOC se réservent la possibilité d'orienter les soumissions vers l'une ou l'autre des deux formes après leurs évaluations.

Pour les communications orales (en français ou en anglais), il sera prévu vingt minutes de présentation et dix minutes d'échanges avec la salle.

Les posters seront au format A0, en portrait, en français ou en anglais. Chaque participant disposera de trois à cinq minutes (selon le nombre de présentations retenues) pour présenter oralement son poster (en français ou en anglais), lors d'une session qui sera prévue dans le programme du colloque.

Bibliographie non exhaustive sur la notion de corpus

- Adam, J.M., Viprey, J.M. (2009/8). Corpus de textes, textes en corpus. *Corpus* [En ligne], mis en ligne le 01 juillet 2010, consulté le 09 juillet 2014. URL : <http://corpus.revues.org/1670>
- Durand, J., Gut, U., Kristoffersen, G. *The Oxford Handbook of Corpus Phonology*, Oxford: Oxford University Press.
- Abeillé, A., Godard D. (2012/4). La Grande Grammaire du Français et la variété des données, *Langue française* n°176, p. 47-68.
- Banks, M. & J. Ruby (2011). *Made To Be Seen: Historical Perspectives on Visual Anthropology*. Chicago : University of Chicago Press.
- Baude O. (coord.), C. Blanche-Benveniste, M.-F. Calas, P. Cappeau, P. Cordereix, L. Goury, M. Jacobson, I. de Lamberterie, C. Marchello-Nizia et L. Mondada, (2006). *Corpus oraux - Guide des bonnes pratiques 2006*. CNRS Editions & Presses : Universitaire d'Orléans. 203 p.
- Bilger, M. (Ed.) (2008). *Données orales. Les enjeux de la transcription*. Perpignan : Presses Universitaires de Perpignan.
- Bruxelles S. et al. (éds) (2009), *Cahiers de Linguistique de Louvain n° 33 : Grands corpus de français parlé*.
- Erickson, F. (2006). Definition and analysis of data from videotape: Some research procedures and their rationales. In J. Green, J. Camilli, and P. Elmore (eds.) *Handbook of complementary methods in educational research*. (3rd ed.) American Educational Research Association.
- Derry, S. J. (Ed.) (2007). *Guidelines for Video Research in Education: Recommendations from an Expert Panel*. Prepared for the National Science Foundation, Interagency Education Research Initiative, and the Data Research and Development Center. Available at: <http://drdc.uchicago.edu/what/video-research.html>.
- Dister A. et al. (2009). Du corpus à la banque de données. Du son, des textes et des métadonnées. L'évolution de banque de données textuelles orales VALIBEL (1989-2009), *Cahiers de linguistique* 33 (2), p. 113-129.
- Garric, N., Longhi, J. (2012/3). L'analyse de corpus face à l'hétérogénéité des données, dans *Langages* n° 187, 152 p.
- Goldman, R., Pea, R., Barron, B. J. & S. Derry (Eds.). (2007). *Video research in the learning sciences* (pp. 239-254). Mahwah, NJ : Erlbaum.
- Groupe ICOR (Bert M., Bruxelles S., Etienne C., Mondada L., Teston S., Traverso V., Jouin-Chardon E., Justine L). (2010). Grands corpus et linguistique outillée pour l'étude du français en interaction (plateforme CLAPI et corpus CIEL), dans *Pratiques - Interactions et corpus oraux*, n°147-148, p. 17-34. <http://halshs.archives-ouvertes.fr/halshs-00622858>.
- Groupe ICOR (Bert M., Bruxelles S., Etienne C., Mondada L., Traverso V.). (2009). Exploitation de la plateforme Corpus de Langue Parlée en Interaction (CLAPI) : le cas de 'voilà' dans les chevauchements, dans *Cahiers de Linguistique* n°33, 2, p. 243-268. <http://halshs.archives-ouvertes.fr/halshs-00370875>.
- Guillot, C., Heiden, S., Lavrentiev, A., Marchello-Nizia, C. (Dir.) (2008). Constitution et exploitation des corpus d'ancien et de moyen français, *CORPUS 7*, Université de Nice- Sophia Antipolis.
- Habert, B., Nazarenko, A. & Salem, A. (1997). *Les linguistiques de corpus*, Paris : Armand Colin.

- Laks, B. (2011/1). La phonologie du français et les corpus, *Langue française* n°169, p. 3-17. <http://www.cairn.info/revue-langue-francaise-2011-1-page-3.htm>.
- Lund K., Bécu-Robinault K. (2009). La reformulation multimodale et polysémiotique comme aide à la compréhension de la physique, dans Alain Rabatel (Ed.), *Analyse sémiotique et didactique des reformulations*. <http://halshs.archives-ouvertes.fr/halshs-00376594>
- Mellet, S. (2002/1). Corpus et recherches linguistiques. *Corpus* [En ligne], mis en ligne le 15 décembre 2003, consulté le 08 juillet 2014. URL : <http://corpus.revues.org/49>
- Mondada, L. (2012). The Conversation Analytic Approach to Data Collection, dans J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis*, Oxford : Wiley-Blackwell, p. 32-56.
- Schmale G. (2010), *Pratiques* n° 147-148 : *Interactions et corpus oraux*, CRESEF.
- ten Have, P. (2007). *Doing Conversation Analysis: A Practical Guide*, London : Sage Publications.
- Traverso, V. (2012/3). Analyses interactionnelles : repères, questions saillantes et évolution, *Langue française* n°175, p. 3-17.
- Traverso, V. (2008). Analyser un corpus de langue parlée en interaction : questions méthodologiques, *Verbum* 30 (4), p. 313-328.
- Veillard, L., & Tiberghien, A. (2011). *Instrumentation de la recherche en Education. Le cas du développement d'une base de vidéos de situation d'enseignement et d'apprentissage ViSA*. Paris : Maison des Sciences de l'Homme.
- Veillard, L. (2013). Les méthodologies de constitution et d'analyse des enregistrements données vidéo dans les recherches en éducation, dans L. Veillard & A. Tiberghien (Eds.), *ViSA : Instrumentation de la recherche en éducation*. <http://books.openedition.org/editionsmsh/1930>