

**Colloque international**  
***50 ans de linguistique sur corpus oraux: apports à l'étude de la variation***  
**14-16 novembre 2018**  
**Université d'Orléans**

Il y a cinquante ans, une équipe d'universitaires anglais prenait l'initiative de recueillir des enregistrements de français parlé, dans différentes situations de communication, auprès de locuteurs issus de toutes les classes sociales. L'Enquête Sociolinguistique à Orléans (ESLO - <http://eslo.huma-num.fr/>) est devenue à présent le témoignage le mieux documenté sur un état de langue dont les évolutions peuvent être appréhendées grâce à la poursuite d'une nouvelle collecte (ESLO2). La transformation en un objet numérique, son exploration à l'aide des outils issus du TAL, sa diffusion et son archivage pilotés par le LLL, sont représentatifs du bouleversement connu par la linguistique sur corpus, et au delà, par les sciences humaines et sociales ces dernières années.

A destination des chercheurs qui, quels que soient leur terrain et la langue qu'ils étudient, s'impliquent dans la sociolinguistique urbaine, les études sur l'oral et le changement en micro-diachronie, ce colloque se propose comme un moment privilégié d'échange d'expériences et d'analyses, ainsi qu'une occasion de réfléchir à la place des données variationnistes dans les développements de la linguistique sur corpus.

Les communications pourront s'inscrire dans l'un des trois axes du colloque et/ou relever des thématiques transversales de constitution, d'outillage, d'exploration et d'archivage des corpus oraux et multimodaux.

### **1. Sociolinguistique urbaine**

La sociolinguistique a entretenu dès ses origines, un rapport très étroit avec les pratiques langagières urbaines (Labov 1973 ; Millroy 1980 ; Calvet 1994 ; Gadet (dir) 2017).

Si les variations diatopiques ont été laissées à la dialectologie, la linguistique variationniste prend en compte les stratifications sociales et les mouvements de la société, l'hétérogénéité des locuteurs et la diversité des situations d'interaction dans la description et l'analyse linguistique.

La ville constitue un espace où la densité et la diversité sociale et culturelle des locuteurs et de leurs pratiques offre un point de vue privilégié sur la dynamique de la langue. Elle permet notamment d'observer les effets du contact de langues (des français avec de nombreuses langues) dans l'espace francophone et des situations de migrations qui contribuent à l'émergence de « vernaculaires urbains contemporains » (Rampton 2015), objet de nombreuses études dans le champ de la sociolinguistique urbaine.

Parce que la sociolinguistique s'est construite sur la question des données et leur relation à l'enquête, les pratiques linguistiques dans les villes ont été appréhendées à partir de corpus oraux (Corpus de Montréal Sankoff-Cedergren 1976 ; ESLO Blanc et Biggs 1971 ; ESLO2 2008 ; Corpus MPF Gadet 2013), construits pour appréhender les systèmes et structures linguistiques à partir des variations diastatique/diaphasique/diachronique socialement et géographiquement situées en milieu urbain.

Depuis les premières enquêtes sociolinguistiques, les villes ont subi des transformations considérables, elles sont plus que jamais le lieu où se croisent dynamiques sociales, dynamiques migratoires et dynamiques linguistiques. Reste à savoir dans quelle mesure la linguistique s'est elle-même développée, dans ses cadres théoriques, ses méthodes et ses outils.

Ainsi, seront considérées dans cette thématique toutes les propositions de communication qui s'interrogent sur la spécificité des pratiques langagières en milieu urbain, soit :

- parce qu'elles mettent en avant les précautions méthodologiques que ce terrain requiert pour y saisir les pratiques langagières ;
- parce qu'elles présentent des analyses montrant comment la particularité des interactions en milieux urbains conduit à l'actualisation de formes linguistiques si ce n'est innovantes, particulières ;
- parce qu'elles donnent à voir la place et l'influence de la/les norme/s dans les représentations et dans les pratiques langagières urbaines.

## **2. Linguistique de l'oral**

Cet axe vise à la présentation d'analyses linguistiques de corpus oraux, qualitatives ou quantitatives, mettant en œuvre des approches *écologiques* :

« *Grammar is viewed as lived behavior, whose form and meaning unfold in experienced interactional and historical time* » (Ochs et al. 1996 :38)

Depuis un demi-siècle, de nombreuses études linguistiques – en France en particulier depuis les travaux pionniers du groupe aixois (Blanche-Benveniste et al. 1987 ; Blanche-Benveniste 1990) – ont mis en exergue les différences profondes entre la structure linguistique à l'oral et à l'écrit. En particulier, la superposition énonciation/énoncé exerce une forte influence sur la construction à l'oral : si, à l'écrit, la linéarité n'est contraignante que sous son aspect constructionnel et que l'interlocutivité est construite avec une grande stabilité au cours de l'activité d'élaboration (typiquement par un discours monologal adressé à un destinataire absent), l'oral se construit en présence et en temps réel. Dans une tradition orientée sur l'analyse conversationnelle, les études ont dégagé les mécanismes d'alternance des tours (Sacks et al. 1974 ; Schegloff 2007) et de co-construction (Lerner 2002 ; Dausendschön et al. 2016) ainsi que l'importance de la coordination du verbal avec d'autres praxies et systèmes sémiotiques pour la construction du sens à l'oral (Gumperz 1982 ; McNeill 1992 ; Givon 2005). Le traitement des structures linguistiques en temps réel a été saisi par la notion de la syntaxe « *online* » (Auer 2000), où la projection de la structure est conçue comme une *trajectoire* dont l'instanciation est progressive, soumise à un processus de formulation par « bribes » (Blanche-Benveniste 1990). En l'absence d'une grammaire externe qui en détermine la stabilité et les contours, les constructions même sont saisies comme émergentes (Bybee 2007 ; Du Bois 2003), sujettes à des variabilités plus ou moins importantes, dans les temporalités du discours présent et dans la variation sociale.

Nous invitons en particulier des communications sur les trois problématiques suivantes :

1) l'élaboration de la cohésion structurelle sous la contrainte de la séquentialité : p.ex. ressources de formulation / reformulation, co-construction, multimodalité ;

2) la modélisation de la structure à l'oral : catégories d'analyse (p.ex. unités micro-/macrosyntaxiques), principes de segmentation, interface morphologie/syntaxe et prosodie (rythme, intonation) ;

3) la routinisation variable des constructions et la caractérisation de l'hétérogénéité structurelle : quantification de la variation constructionnelle, tendances caractéristiques de certains genres/types de discours oraux ; profils de constructions, figures ou formats, leur diffusion, dynamique de changement, en lien ou non avec leur complexité.

Les communications rendant compte des projets/expériences d'annotation et d'analyse linguistique soutenus par des outils de traitement automatique sont les bienvenues.

### **3. Diachronie et micro-diachronie de l'oral**

C'est depuis plusieurs décennies déjà que la linguistique de l'oral s'est constituée comme champ d'investigation à part entière, grâce, entre autres, aux évolutions techniques permettant la constitution de corpus oraux d'envergure croissante, au point que les données disponibles permettent désormais un recul diachronique inédit pour l'analyse de l'oral. Adossée aux questions de la patrimonialisation des langues nationales, cette approche diachronique des corpus oraux a été adoptée pour documenter différentes langues : l'anglais (*Diachronic Corpus of Present-Day Spoken English*, UCL Royaume-Uni), le japonais (*Corpus of spontaneous Japanese, Corpus of Historical Japanese*, NINJAL Tokyo), l'italien (*C-Oral Rom*, Florence ; *LIP – Lessico di frequenza dell'italiano parlato*, hébergé à l'Université de Graz), ou encore le finlandais (corpus diachronique du finlandais parlé de l'Université de Turku).

Si elle est centrale dans la constitution du corpus numérique ESLO, mise en œuvre au LLL depuis 2008 suite au travail de collecte d'universitaires britanniques dans les années 60, la dimension (micro-)diachronique des corpus oraux a été peu exploitée, pour le français notamment, alors même que la documentation de l'oral couvre désormais plusieurs décennies, et qu'elle fournit des données sociologiquement situées, ce qui les rend propices à l'observation du changement (cf. Labov 2001). Par ailleurs, si la question des diachronies courtes (sur une dizaine d'années seulement) a déjà été envisagée pour l'écrit (cf. Siouffi et al. 2012), elle semble peu prise en compte pour l'oral.

Seront considérées dans cette thématique des communications autour de la constitution et de l'exploitation de corpus oraux dans une perspective diachronique. Il pourra s'agir :

- de faire un état des lieux sur les types de données existantes, sur leur empan diachronique et leur échantillonnage, en s'interrogeant sur la question de leur représentativité et de leur accessibilité en tant que données orales diachroniques, et de leur éventuelle patrimonialisation ;
- d'adopter une approche réflexive sur les méthodes mises en œuvre pour saisir la variation dans toutes ses dimensions, depuis la constitution de sous-corpus sociologiquement situés, l'échantillonnage et le choix des coupes synchroniques, au type d'analyse, qu'il soit quantitatif (relevés de fréquence, phénomènes rares) ou qualitatif (examen des contextes d'emplois, annotation outillée, etc.) ;
- d'illustrer par des études ciblées les types d'analyses pouvant être conduites, et ce, dans tous les domaines (phonologie, morphosyntaxe, sémantique, lexicale, discours

et interaction, etc.), en interrogeant tant le rapport entre l'objet et la temporalité considérée, que la nature de la variation, entre variabilité des usages, effets de mode, et changement systémique.

### **Thématiques transversales**

Au moment où l'Enquête Socio-Linguistique à Orléans était constituée comme une archive sonore, indexée, cataloguée et diffusée pour des analyses dans tous les domaines des sciences humaines et sociales, le corpus de Montréal connaissait les premières tentatives d'un traitement informatique. Depuis ces deux exemples, les corpus oraux et multimodaux ont été au cœur des transformations technologiques mais aussi méthodologiques et théoriques de la linguistique sur corpus numériques et du domaine de la conservation de documents sonores. Les outils et instruments de transcription, d'annotations, de traitement du signal, de textométrie, de visualisation, et plus généralement tous les outils du TAL et du traitement de données, les plateformes de conservation et de diffusion de corpus, les initiatives visant l'interopérabilité des données ont un effet sur les analyses et sur les opérations de constitution et d'exploitation de corpus.

Enfin, à l'heure du web de données, les questions posées par l'archivage et par la réutilisation de corpus, tout comme les projets de sciences contributives dépassent le domaine de la linguistique bien que concerné en premier lieu. Dans ce vaste mouvement qui nécessite des opérations de normalisation et de formatage, la place de données hétérogènes permettant l'étude de la variation reste à questionner.

Les travaux de recherche sur toute langue seront les bienvenus.

Les langues du colloque seront le français et l'anglais.

### **Bibliographie**

- Auer P. 2000, « Projection in interaction and projection in grammar », Communication on the EURESCO Conference on Interactional Linguistics, Spa. *Text*, 25/1, 7-36.
- Baude, O., Dugua, C., (2016) « Les ESLO, du portrait sonore au paysage digital », *Corpus 15*, Corpus de français parlé et français parlé des corpus, Nice, 29-56
- Bergounioux, G. Dir, (1992). Enquête, corpus et témoin, *Langue française*, n°93, Paris, Larousse.
- Blanc, M. Biggs, P., 1972 « L'enquête socio-linguistique sur le français parlé à Orléans » (1971) *Le français dans le monde* n°85, Paris, 16-25
- Blanche-Benveniste C. 1990, *Le français parlé. Etudes grammaticales*. Paris : CNRS (coll. : sciences du langage).
- Blanche-Benveniste C. & C. Jeanjean 1987, *Le français parlé. Transcription et édition*. Paris : CNRS / INALF.
- Bybee J. 2007, *Frequency of use and the organization of language*. Oxford : Oxford University Press.
- Calvet L.-J. 1994, *Les Voix de la ville, Introduction à la sociolinguistique urbaine*, Paris : Payot.
- Dausendschön-Gay U. & E. Gülich & U. Krafft (eds.), 2016, *Ko-Konstruktionen in der Interaktion*, Bielefeld : transcript-Verlag.
- Du Bois J.W. 2003, « Discourse and grammar », in : Tomasello, Michael (ed.), *The new psychology of language*, Mahwah, NY : Laurence Erlbaum Associates, vol.2, 47-87.
- Eshkol-Taravella I., Baude O., Maurel D., Hriba L., Dugua C., Tellier I., (2012) Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012. in *Ressources linguistiques libres, TAL*. Volume 52 – n° 3/2011, 17-46.

- Gadet F. (dir.) 2017, *Les parlers jeunes dans l'île de France multiculturelle*, Paris : Ophrys.
- Givón T. 2005, *Context as other minds : the pragmatics of sociality, cognition, and communication*. Amsterdam : Benjamins.
- Gumperz J. J. 1982, *Discourse Strategies*. Cambridge : Cambridge University Press.
- Labov W. 1973, *Sociolinguistic Patterns*, Philadelphia : University of Pennsylvania Press.
- Labov W. 2001, *Principles of Language Change I : Social factors*, Oxford, Blackwell.
- Lerner G. 2002, « Collaborative turn sequences » in : Gene H. Lerner (ed.), *Conversation Analysis. Studies from the first generation*. Amsterdam: Benjamins, S. 225-256.
- McNeil D. 1992, *Gesture and Thought*, Chicago : University of Chicago Press.
- Ochs E. & E. Schegloff & S. Thompson (eds.) 1996, *Interaction and grammar*. Cambridge : Cambridge University Press (Studies in Interactional Sociolinguistics).
- Milroy L. 1980, *Language and social networks*, London – Baltimore : Basil Blackwell - University Park Press.
- Rampton B. 2015, « Contemporary urban vernaculars » in J. Nortier and B. Svendsen (eds.), *Language, Youth and Identity in the 21st Century. Linguistic Practices across Urban Spaces*, Cambridge : Cambridge University Press.
- Sacks H. & E. Schegloff, G. Jefferson, 1974, « A simplest Systematics for organization of turn-taking for conversation », *Language* 50/4, 696-735.
- Schegloff E. 2007, *Sequence Organisation in Interaction. A Primer in Conversation Analysis*. Cambridge : Cambridge University Press.
- Siouffi G., Wionet C. & Steuckardt A. 2012, « Comment prendre en compte les diachronies courtes et contemporaines », Congrès Mondial de Linguistique Française – CMLF 2012, SHS Web of Conferences 1.

## Site du colloque

<https://anniveslo-50ans.sciencesconf.org/>

## Modalités de soumission

Les propositions doivent être envoyées **avant le 8 avril 2018** à l'adresse suivante :

*anniv-eslo.iii@univ-orleans.fr*

Elles prendront la forme d'un résumé anonymisé (en français ou en anglais) d'une longueur maximale de 500 mots (bibliographie non comprise, en Times New Roman 12). Dans le corps du message, doivent être indiqués le nom, prénom et l'affiliation de l'auteur.

## Calendrier

Soumission des propositions : 8 avril 2018

Notification aux auteurs : 1er juin 2018

Confirmation de participation : 15 juin 2018

Dates du colloque : 14-16 novembre 2018 - Université d'Orléans

## Comité scientifique

- Peter Auer (Universität Freiburg, Allemagne)
- Christophe Benzitoun (Université de Lorraine, France)
- Jan Blommaert (Université de Gand, Belgique)

- Jeanne-Marie Debaisieux (Université Paris 3 Sorbonne Nouvelle, France)
- Anne Dister (Université Saint-Louis, Bruxelles, Belgique)
- Gaëtane Dostie (Université de Sherbrook, Canada)
- Andreas Dufter (LMU, Munich, Allemagne)
- Françoise Gadet (Université Paris Nanterre, France)
- Marja-Liisa Helasvuo (Université de Turku, Finlande)
- Anne Lacheret (Université Paris Nanterre, France)
- Takehiko Maruyama (NINJAL, Tokyo, Japon)
- Alessandro Panunzi (Université de Florence, Italie)
- Stefan Pfänder (Universität Freiburg, Allemagne)
- Shana Poplack (Université d'Ottawa, Canada)
- Claus Pusch (Universität Freiburg, Allemagne)
- Wim Remysen (Université de Sherbrook, Canada)
- Nathalie Rossi-Gensane (Université Lumière Lyon 2, France)
- Thomas Schmidt (IDS Mannheim, Allemagne)
- Cyril Trimaille (Université Grenoble-Alpes, France)

### **Comité ESLO**

- Lotfi Abouda (Université d'Orléans)
- Olivier Baude (Université Paris Nanterre)
- Gabriel Bergounioux (Université d'Orléans)
- Caroline Cance (Université d'Orléans)
- Gilles Cloiseau (Université d'Orléans)
- Céline Dugua (Université d'Orléans)
- Emmanuelle Guerin (Université d'Orléans)
- Badreddine Hamma (Université d'Orléans)
- Loyal Kanaan-Caillol (Université d'Orléans)
- Paola Pietrandrea (Université de Tours)
- Katja Ploog (Université d'Orléans)
- Noëlle Serpollet (Université d'Orléans)
- Marie Skrovec (Université d'Orléans)

### **Comité d'organisation**

- Le *comité ESLO*
- Yossra Ben Ahmed
- Fatma Ben Barka
- François Delafontaine
- Athéna Dupont
- Jennifer Ganaye
- Luka Raickovic

### **Contact :**

*anniv-eslo.iii@univ-orleans.fr*