

International symposium

50 years of oral corpus linguistics: Its contribution to the study of variation

14-16 November 2018

University of Orléans

Fifty years ago, a team of English academics took the initiative to collect recordings of spoken French in different communication situations with speakers from all social classes. The Sociolinguistic Survey in Orléans (ESLO – <http://eslo.huma-num.fr/>) has now become the best documented testimony of a state of language whose evolution can be addressed thanks to a follow-up new collection (ESLO2). The transformation into a digital object, its exploration using Natural Language Processing (NLP) tools, its dissemination and archiving piloted by the LLL, are representative of the upheaval experienced by corpus linguistics, and beyond, by the human and social sciences in recent years.

Intended for researchers who, whatever their field of study and whatever the language they study, are involved in urban sociolinguistics, oral studies and microdiachronic change, this symposium is intended as a privileged moment for the exchange of experiences and analyses, as well as an opportunity to reflect on the place of variationist data within the development of corpus linguistics.

The papers may fit into one of the three axes of the symposium and/or deal with the cross-cutting themes of constitution, tooling, exploration and archiving of oral and multimodal corpora.

1. Urban sociolinguistics

From the outset, sociolinguistics has had a very close relationship with urban language practices (Labov 1973, Millroy 1980, Calvet 1994, Gadet (dir) 2017).

While diatopic variations have been left to dialectology, variationist linguistics takes into account social stratifications and societal evolution, the heterogeneity of speakers and the diversity of interaction situations in linguistic description and analysis.

The city defines a space where the density and social and cultural diversity of speakers and their practices offers a privileged point of view on the dynamics of language. In particular, it makes it possible to shed light on the effects of language contact (of the French with many languages) in the French-speaking world and migration situations that contribute to the emergence of "contemporary urban vernaculars" (Rampton 2015), the subject of numerous studies in the field of urban sociolinguistics.

Sociolinguistics was founded on the issue of data and their relationship to the survey. Therefore, linguistic practices in cities were addressed thanks to oral corpora (Corpus de Montréal Sankoff-Cedergren 1976, ESLO Blanc et Biggs 1971, ESLO2 2008, Corpus MPF Gadet 2013), which had been built to investigate linguistic systems and structures via geographically and socially located diastatic/diaphasic variations.

Since the first sociolinguistic surveys, cities have undergone considerable transformations, and are more than ever the place where social, migration and linguistic dynamics intersect. The task of assessing to what extent linguistics has developed itself, in its theoretical frameworks, methods and tools, remains

Along these thematic lines, all presentations that question the specificity of language practices in urban context will be considered,

- whether they highlight the methodological precautions that this field requires to understand language practices;
- whether they feature analyses showing how the peculiarity of interactions in urban environments leads to the updating of linguistic forms, which are, if not innovative, particular;
- whether they underline the place and influence of the standard(s) in urban language representations and practices.

2. Linguistics of the oral language

This axis aims at the presentation of linguistic analyses of oral, qualitative or quantitative corpora implementing ecological approaches:

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

For half a century, numerous linguistic studies - in France in particular since the pioneering work of the Aix-en-Provence group (Blanche-Benveniste et al. 1987; Blanche-Benveniste 1990) - highlighted the profound differences between oral and written linguistic structure. In particular, the superimposition of utterance and text has a strong influence on oral construction: just as in writing, linearity is a constraint only in its constructional aspect and interlocutory dialogue is carried on with great stability during the elaboration process (typically by a monological discourse addressed to an absent recipient), in oral, discourse is constructed in real time and presence. In a tradition based on conversational analysis, studies have so far identified alternating "turn" mechanisms (Sacks et al. 1974; Schegloff 2007) and co-construction (Lerner 2002, Dausendschön et al. 2016) as well as the importance the coordination of the verbal dimension with other semiotic praxies and systems in the construction of oral meaning (Gumperz 1982; McNeill 1992; Givon, 2005). The treatment of real-time linguistic structures was addressed by the notion of "online" syntax (Auer 2000), where the projection of the structure is conceived as a trajectory whose instantiation is progressive, subjected to a process of formulation by "snatches" (Blanche-Benveniste 1990). In the absence of any external grammar that would fix its stability and contours, constructions themselves are captured as being emergent (Bybee 2007; Du Bois 2003), subjected to more or less important variabilities, within the temporalities of the present discourse and social variation.

We particularly welcome papers on the three following issues:

- 1) the elaboration of structural cohesion under the constraint of sequentiality: e. g. formulation resources / reformulation, co-construction, multimodality;
- 2) Oral structure modeling: categories of analysis (e. g., micro/macro-syntactic units), segmentation/chunking principles, morphology/syntax interface and prosody (rhythm, intonation);
- 3) the variable routinization of constructions and the characterization of structural heterogeneity—quantification of constructional variation, characteristic trends of given genres/types of oral discourses; construction patterns, figures or formats, their delivery, dynamics of change, related or not to their complexity.

Communications reporting on annotation projects/experiences and linguistic analysis supported by automatic processing tools are welcome.

3. Diachrony and oral microdiachronics

It is now several decades since oral linguistics became a full-fledged field of investigation, thanks, among other things, to technical developments enabling the creation of oral corpora of increasing scope, to the point that the available data now allow for an unprecedented diachronic retrospective for oral analysis.

Confronted with the issues of the national heritage of languages, this diachronic approach of oral corpora has been adopted to document different languages: English (Corp. of Present-Day Spoken English, UK UCL), Japanese (Corpus of spontaneous Japanese), Corpus of Historical Japanese, NINJAL Tokyo), Italian (C-Oral Rom, Florence, LIP - Lessico di frequenza dell'italiano parlato, hosted at the University of Graz), or Finnish (diachronic corpus of Finnish spoken from the University of Turku).

The (micro) diachronic dimension of oral corpora, at the core of the development of the ESLO digital corpus, implemented at LLL since 2008 following the collection of British academics in the 1960s, has not been used widely, in particular for the study of French, even though spoken language documentation now covers several decades, and provides sociologically situated data, which is made suitable for observation of change (see Labov 2001). Moreover, while the issue of short-term diachronies (over a period of only ten years) has already been considered for written text (see Siouffi et al 2012), it seems to be hardly taken into account for its spoken counterpart.

Communications around the constitution and exploitation of oral corpora in a diachronic perspective will be considered under this thematic axis. These might:

- take stock of the existing types of data, their diachronic span and their sampling, by questioning the issue of their representativeness and their accessibility as diachronic oral data, and how they might be made national heritage;
- adopt a reflexive approach on the methods implemented to capture the variation in all its dimensions, from the constitution of sociologically located sub-corpora, the sampling and the choice of the synchronic sections, to the type of analysis, whether it is quantitative

(frequency records, rare phenomena) or qualitative (examination of employment contexts, annotated tools, etc.);

- illustrate via targeted studies the types of analyses that can be conducted in all domains (phonology, morphosyntax, semantics, lexicon, speech and interaction, etc.), by questioning both the relationship between the object and the temporality considered, as the nature of the variation, between variability of uses, effects of fashion, and systemic change.

Cross-cutting themes

At a time when the Socio-Linguistic Survey in Orléans was constituted as a sound archive, indexed, catalogued and disseminated for analysis in all fields of the humanities and social sciences, the Montreal corpus underwent the first attempts of computer processing. Since these two examples, oral and multimodal corpora have been at the heart of technological but also methodological and theoretical transformations in linguistics on digital corpus and in the field of preservation of sound documents. The tools and instruments of transcription, annotation, signal processing, textometry, visualization, and more generally all the tools of NLP and data processing, the platforms aiming at archiving and/or disseminating corpora, the initiatives aiming at data interoperability have an effect on the analyses and on the operations of constitution and exploitation of corpora.

Finally, at the time of the web of data, the questions raised by the archiving and reuse of corpora, similarly to contributory science projects, go beyond the domain of linguistics, although it is primarily concerned by those questions. In this vast movement which requires standardization and formatting operations, the place of heterogeneous data allowing the study of variation remains to be questioned.

Research on any language is welcome.

The conference will be held both in French and in English.

Bibliography

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 John 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 (2012) DOI 10.1051/shsconf/20120100214.

Site of the symposium

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

Submission procedures

Proposals should be sent to the following address:

anniv-eslo.iii@univ-orleans.fr

The submission deadline is April 8th 2018.

They will take the form of an anonymized abstract (in French or English) of up to 500 words in length (not including bibliography, in Times New Roman 12). The author's surname, first name and affiliation must be indicated in the body of the message.

Key Dates

Abstract submission deadline: 8th April 2018

Notification of acceptance/rejection to authors: 1st June 2018

Confirmation of participation: 15th June 2018

Dates of the symposium: 14th-16th November 2018 - University of Orléans

Scientific committee

- 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)

ESLO committee

- 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 Hama (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)

Organizing committee

ESLO Committee

Fatma Ben Barka

Yossra Ben Ahmed

François Delafontaine

Athéna Dupont

Jennifer Ganaye

Luka Raickovic

Contact e-mail

anniv-eslo.iii@univ-orleans.fr