

Où en sont les corpus de français parlé ?

Rencontre internationale

03 et 04 septembre 2024

Université de Neuchâtel,
UniMail, salle E026 Millénium

Suivie d'un atelier dédié au logiciel TXM

Organisation :

Mathieu Avanzi (P.O.) & Laure Anne Johnsen (Prof. tit.)
avec la coll. de Yannick Emery (coll. scientifique)
& Céline Rumpf (ass. doc.)

Entrée libre

unine[®]
Université de Neuchâtel

Où en sont les corpus de français parlé ?

Description de la rencontre

Ce colloque fait suite à un précédent congrès dont l'intitulé était « Corpus de français parlés et français des corpus », qui s'était déroulé dans les murs de l'université de Neuchâtel les 08 et 09 mai 2014. Il est organisé dans le cadre du projet AGORA « Dis voir », et en lien avec d'autres projets de recherche actuels qui portent sur la description des variétés de français parlées en Suisse romande, dont le corpus OFROM, développé et hébergé à l'université de Neuchâtel.

Le but de la rencontre est de faire le point, près de 10 ans après la rencontre susmentionnée, sur l'état des corpus oraux dans la francophonie. Quand nous avons organisé le précédent colloque en 2014, nous soulignons le fait qu'en France, de plus en plus de chercheurs œuvraient à la mise au point de grands corpus. Nous avons alors eu l'idée de réunir quelques-uns des acteurs ayant contribué de façon significative à la mise au point de corpus ou d'archives de français parlés. En 2024, il nous a semblé que le temps était venu d'organiser une nouvelle rencontre du même type, en mettant cette fois-ci l'accent sur la francophonie.

Tous les invités disposeront d'une plage de parole de 45 minutes (discussion incluse), et s'efforceront de respecter lors de leur communication le format suivant :

- présentation succincte du corpus (bref historique du corpus/du projet ; nature des données traitées ; format des transcriptions et des annotations) ;
- présentation d'une ou de plusieurs études menées sur la base d'une partie au moins du corpus dans une perspective variationnelle : dia-variation (stylistique, régionale, sociale, etc.) et/ou variation -émique (morphosyntaxe, phonologique, lexicale, etc.)

La dernière demi-journée de cette rencontre consistera en un atelier TXM animé par Cyrielle Montrichard.

Programme prévisionnel de la rencontre et de l'atelier

Mardi 03 septembre

- 10.00-10.30** Accueil des participants et mot de bienvenue
- 10.30-11.15** Marie-Hélène CÔTÉ (U. de Lausanne)
Phonologie du français contemporain au Québec : traitement et exploitation
- 11.15-12.00** Wim REMYSEN (U. de Sherbrooke)
Les corpus oraux du français parlé au Québec : un patrimoine à pérenniser et à valoriser
- 12.00-14.00** Repas
- 14.00-14.45** Mathieu AVANZI¹, Gilles CORMINBOEUF², Laure Anne JOHNSEN¹,
14.45-15.30 Yannick EMERY¹ & Matthieu MONNEY²
(¹U. de Neuchâtel, ²U. de Fribourg)
OFROM+ : passé, présent et futur du corpus de français parlé en Suisse romande et dans la francophonie
- 15.30-16.00** Pause café
- 16.00-16.45** Audrey ROIG¹, Bertrand LUDES², Frédéric BARBUT³, Nizha CHATAR-MOUMNI⁴, Tania DELABARDE², Marianne DOURY¹, Cristina FIGUEIREDO¹, Lisa FRIEDLANDER⁵ & Caroline LACHET¹
(¹EDA, URP 4071 U. Paris Cité ; ²BABEL, UMR 8045 U. Paris Cité – CNRS ; ³3PHM, Inserm UMR-S 1139 U. Paris Cité ; ⁴MoDyCo, UMR 7114 U. Paris Cité – CNRS ; ⁵ECEVE, Inserm UMR-S 1123 U. Paris Cité)
Le corpus AS2-HP : paroles de soignants de l'AP-HP au terme de la 1^{re} année de covid-19

Mercredi 04 septembre

- 10.30-11.15** Julie GLIKMAN¹, Nicolas MAZZIOTTA² & Christophe BENZITOUN¹
(¹.U de Lorraine, ATILF ; ²U. de Liège, Traverses)
Les corpus de français non standard au 21^e siècle : enjeux du projet Les Vocaux
- 11.15-12.00** Anne DISTER (UCLouvain – Saint-Louis – Bruxelles)
Documenter le français parlé à Bruxelles : comment ? pourquoi ? pour qui ?
- 12.00-14.00** Repas
- 14.00-15.30** Cyrielle MONTRICHARD (U. Paris-Est Créteil)
Atelier TXM (partie 1/2)
- 15.30-16.00** Pause café
- 16.00-17.30** Cyrielle MONTRICHARD (U. Paris-Est Créteil)
Atelier TXM (partie 2/2)

Résumés des interventions

Phonologie du français contemporain au Québec : traitement et exploitation

Marie-Hélène CÔTÉ (Université de Lausanne)

Dans le corpus Phonologie du français contemporain (Durand et al. 2002, 2009 ; Detey et al. 2016), le Québec est largement représenté, par 440 locuteurs provenant de 32 points d'enquête répartis sur le territoire (Côté 2014 ; Côté et Saint-Amant Lamy 2023). Si la partie québécoise adopte l'ensemble du protocole de base PFC, elle se distingue par des tâches et des procédures de traitement supplémentaires visant à documenter les particularités des variétés québécoises, mais également à bonifier l'exploitation des données. Nous présenterons ici le corpus, les raisons et les avantages de ses spécificités de traitement, ainsi que quelques exemples d'exploitation.

Les corpus oraux du français parlé au Québec : un patrimoine à pérenniser et à valoriser

Wim REMYSEN (Université de Sherbrooke)

Le Québec s'est doté de plusieurs corpus destinés à étudier les particularités de sa variété de français à partir des années 1960 et 1970. Les corpus oraux prennent une place importante dans cet ensemble documentaire qui comprend également quelques corpus textuels d'envergure. Bien plus que de simples pièces d'anthologie, les corpus de langue orale ouvrent la possibilité d'étudier l'évolution récente du français parlé au Québec, notamment à la lumière de données réunies à date plus récente. Or la préservation des données et leur accessibilité demeurent des enjeux majeurs. Dans le cadre de cette communication, nous présenterons une plateforme numérique récemment créée par le Centre de recherche interuniversitaire sur le français en usage au Québec en vue de pérenniser et de mutualiser les corpus oraux québécois, tant patrimoniaux que contemporains. Cette plateforme, baptisée Fonds de données linguistiques du Québec, rend possible la consultation de différents corpus et l'interrogation des contenus.

OFROM+ : passé, présent et futur du corpus de français parlé en Suisse romande et dans la francophonie

Mathieu AVANZI¹, Gilles CORMINBOEUF², Laure ANNE JOHNSEN¹, Yannick EMERY¹ & Matthieu MONNEY²

(¹Université de Neuchâtel, ²Université de Fribourg)

L'archive de parole OFROM (corpus oral du français en Suisse romande) a été mise au point dans le but de décrire les pratiques linguistiques de francophones bien ancrés en Suisse romande. Avec la récolte initiée en 2007, puis l'ouverture du site au grand public quelques années plus tard, la base a non seulement permis de disposer du premier corpus de français oral de Suisse romande, mais aussi de rattraper, grâce aux 1,3 millions de mots alignés texte/son qu'elle contient actuellement, le retard pris par rapport à d'autres pays francophones. Dans cette présentation, nous ferons état des derniers développements qu'a connus la base (ouverture à la francophonie et à d'autres genres de parole, notamment des messages vocaux ; utilisation des données à des fins didactiques) et illustrerons avec des études de cas (*tu* générique) les potentiels de la base pour la recherche en linguistique française dans une perspective décentralisée.



Le corpus AS2-HP : paroles de soignants de l'AP-HP au terme de la 1^{re} année de covid-19

Audrey ROIG¹, Bertrand LUDES², Frédéric BARBUT³, Nizha CHATAR-MOUMNI⁴, Tania DELABARDE², Marianne DOURY¹, Cristina FIGUEIREDO¹, Lisa FRIEDLANDER⁵ & Caroline LACHET¹

(¹EDA, URP 4071 U. Paris Cité ; ²BABEL, UMR 8045 U. Paris Cité – CNRS ;

³3PHM, Inserm UMR-S 1139 U. Paris Cité ; ⁴MoDyCo, UMR 7114 U. Paris Cité – CNRS ;

⁵ECEVE, Inserm UMR-S 1123 U. Paris Cité)

Fruit d'un projet interfacultaire SH-Santé soutenu par l'Université Paris Cité, le corpus « AS 2-HP » est constitué de paroles de 33 soignants recueillies juste un an après la 1^{re} vague de Covid-19, en mai et juin 2021, au terme de la 3^e vague de Covid en France. Travaillant dans les services de Réanimation, d'Urgences ou d'Odontologie de différents hôpitaux publics parisiens ou au sein de l'Institut médico-légal, les aides-soignants, infirmiers, internes et titulaires interrogés se sont exprimés sur les émotions ressenties au fil de l'année écoulée et sur leur impression d'avoir été acteurs ou spectateurs de la crise sanitaire. Semi-directifs, les entretiens, filmés puis retranscrits, ont couvert des thèmes variés tels que les moyens dont ces soignants ont disposé, les soutiens dont ils ont bénéficié ou encore la possible porte de sortie offerte par la vaccination au printemps 2021.

Les corpus de français non standard au 21^e siècle : enjeux du projet Les Vocaux

Julie GLIKMAN¹, Nicolas MAZZIOTTA² & Christophe BENZITOUN¹

(¹Université de Lorraine, ATILF ; ²Université de Liège, Traverses)

Ces dernières années ont vu se développer un nouveau mode de communication sous la forme de messages vocaux enregistrés, appelés « SMS vocaux », « notes vocales » ou « vocaux », terme que nous retiendrons ici, diffusés sous forme de messages, à destinataires uniques ou multiples, connus ou inconnus, via le téléphone portable. Véritables « sms vocaux/vidéos », ils se substituent à l'envoi de SMS, sans pour autant constituer un retour à la conversation téléphonique. Le corpus *Les Vocaux* développé dans le cadre du projet Oralidia (I dex Université de Strasbourg) a pour objectif d'étudier ces productions. L'objectif de notre communication est triple. Nous commencerons par une présentation globale du corpus, ce qui nous permettra d'interroger le statut de ces productions (1) et de montrer les enjeux et les contraintes scientifiques, méthodologiques, techniques et éthiques de la constitution d'un tel corpus, du recueil au format final de distribution (2). Nous terminerons par quelques exemples d'analyses sur le corpus (3).

Documenter le français parlé à Bruxelles : comment ? pourquoi ? pour qui ?

Anne DISTER (UCLouvain – Saint-Louis – Bruxelles)

Le *Corpus de français parlé à Bruxelles* (CFPB) a vu le jour en 2013 (Dister et Labeau 2017) ; il se veut le pendant bruxellois du *Corpus de français parlé parisien* (CFPP200, Branca-Rosoff et al. 2011). En effet, moyennant quelques adaptations, les intervieweurs utilisent le même guide d'entretien semi-dirigé dont le thème principal est celui de la relation des habitants à leur quartier, et plus généralement leur ville. Le CFPB vient assurément combler un vide. En effet, il n'existait pas jusqu'à sa mise en place de corpus académique, comme c'est le cas par exemple à Montréal (Sankoff et al. 1976), documentant le français parlé dans la capitale belge. Dans cette communication, nous verrons notamment le terrain particulier, du point de vue linguistique, que constitue Bruxelles, ville officiellement bilingue, mais aussi ville multiculturelle, capitale des institutions européennes. Nous présenterons également des résultats issus de deux analyses comparées, l'une avec le corpus parisien et l'autre avec des données montréalaises.



