



Les corpus d'oral en interaction : outils, méthodes et objectifs

Vendredi 18 octobre

Unicaen, Campus 1, Bâtiment L, Amphi Laplace
Esplanade de la Paix Caen 14000

Atelier organisé par Thierry Ruchot et Natasha Romanova
(Laboratoire CRISCO)

Programme

- 9:00-9:30** **Thierry Ruchot**
Les corpus d'oral en interaction :
pourquoi sont-ils importants et comment les obtenir ?
- 9:30-10:30** **Benjamin Laillier, Selma Turalic, Amélie Guesnon**
Les enjeux juridiques de la collecte, la mise à disposition et l'archivage des données orales
- 10:45-11:00** **Thierry Ruchot**
Types de transcription
- 11:00-11:15** **Rayan Ziane**
Transcription automatique : problèmes et perspectives
- 11:15-12:15** **Yanka Bezinska**
Logiciel Elan pour la transcription des corpus oraux
- 12:15-14:00** **Pause repas**
- 14:00-16:00** **Philippe Martin**
Outils pour l'analyse prosodique : Praat et WinPitch.
- 16:00-17:00** **Table ronde**

Atelier : Les corpus d'oral en interaction, outils, méthodes et objectifs

Cet atelier s'adresse à des linguistes en activité, des doctorants ou des étudiants de master ayant un intérêt pour la linguistique de l'oral, notamment l'oral en interaction, et pour la création de corpus de ce type.

La création de corpus d'oral représente aujourd'hui un véritable défi aussi bien du point de vue théorique, que du point de vue pratique : pour le traitement automatique de données et le dialogue homme-machine, pour la linguistique clinique, qui a besoin d'observer du langage typique comme du langage pathologique, ou encore pour l'acquisition et la didactique des langues. Alors que l'on dispose déjà d'un nombre important de corpus écrits informatisés de toutes sortes : diachroniques ou synchroniques, unilingues, parallèles ou comparables, généraux, littéraires, journalistiques, techniques, de Web, de réseaux sociaux et de SMS, de productions écrites d'enfants et d'adolescents ou d'étrangers apprenant une langue seconde ou étrangère, de scripts de films et séries, etc., les corpus d'oral occupent encore une place assez modeste. De plus, si les corpus écrits sont mis à disposition essentiellement sous forme de concordanciers, qui ne donnent pas accès à l'intégralité des textes inclus dans le corpus, pour l'oral il est utile d'avoir un double accès : soit par concordancier, pour des recherches sur des points particuliers et l'établissement de calculs de fréquences, soit à l'intégralité des interactions transcrites, éventuellement avec le support enregistré pour pouvoir réaliser une étude qualitative de l'interaction et essayer d'en comprendre les mécanismes. Si l'on peut recueillir de nombreuses données orales sur une plateforme comme YouTube, outre les questions de droits, ces données ne représenteront que certains genres spécifiques : bulletins d'information et commentaires, tutoriels, conférences, cours, interviews, débats, mais très peu les interactions quotidiennes en lieux de services : consultations, réunions, entretiens professionnels, et encore moins les conversations les plus quotidiennes, celles qui constituent l'essentiel de notre pratique orale quotidienne, pour peu que nous ayons un minimum de lien social.

Beaucoup de chercheurs hésitent à aborder l'oral, parce qu'ils se posent de nombreuses questions : quels types de corpus d'oral peut-on développer et à quelles fins ? dans quels domaines sont-ils utiles ? comment procéder ? comment obtenir des données ? quelles obligations légales et éthiques s'imposent lorsqu'on réalise des enregistrements d'oral en interaction, à l'heure où la question de la protection des données personnelles est centrale ? comment transcrire les corpus et pour quels objectifs ? quel rôle accorder à la prosodie, notamment à l'intonation ? que faire de la multimodalité ? comment analyser l'oral en interaction ?

Il y a plusieurs raisons à cela. Tout d'abord la difficulté d'accès aux données de l'oral : au préalable il faut trouver des personnes volontaires pour être enregistrées et réaliser les enregistrements dans de bonnes conditions, en essayant de créer les conditions pour avoir une interaction la plus naturelle que possible. Puis, il faudra transcrire les paroles et, éventuellement, les gestes et les mimiques qui accompagnent la production verbale. Il faudra également réaliser une analyse de la prosodie en utilisant des programmes que beaucoup de chercheurs ont des appréhensions à utiliser. Ensuite, il faudra analyser ces données en utilisant des modèles linguistiques adaptés, sachant que, la plupart du temps, les modèles développées jusqu'ici se sont concentrés beaucoup plus sur l'écrit, et présentent des défauts et lacunes à l'heure d'aborder l'oral, particulièrement l'oral en interaction avec alternance de tours de parole.

Cet atelier visera à donner des pistes de réponse à plusieurs de ces questions, notamment les questions légales et méthodologiques, sans lesquels aucune tâche d'analyse des données ne peut être envisagée. La visée de cet atelier sur l'interaction est également... interactive, dans la mesure où l'échange sera favorisé durant tout l'atelier et où il se clora sur une table ronde, dans laquelle chacun sera invité à présenter ses impressions, ses projets, déjà achevés, en cours, bourgeonnants ou naissants, ses idées et intérêts, même encore vagues.