



CLARIN



MAISON DES SCIENCES DE L'HOMME  
CLERMONT-FERRAND

UCA UNIVERSITÉ  
Clermont Auvergne

## Journée d'étude

CLARIN User Involvement Event

# COREFO

Lier **recherche** et **formation**  
professionnelle :  
des **corpus** à l'interface

Organisée par l'axe CA2LI

Corpus, Acquisition et Apprentissage, Langues en Interactions

Laboratoire de Recherche sur le Langage (EA 999)

*Vendredi 16 novembre 2018 – 9h/16h30*

Amphithéâtre 219 - Maison des Sciences de l'Homme

Université Clermont Auvergne

4 rue Ledru, 63000 Clermont-Ferrand

<https://corefo2018.sciencesconf.org/>

# Pour commencer

- 1 journée d'étude intitulée « **Lier recherche et formation professionnelle : des corpus à l'interface** »
- organisée par l'axe « **Corpus, Acquisition et Apprentissage, Langues en Interactions** »
- Soutien de **CLARIN** et du **LRL**



- 1 proposition de **publication** à la revue *Recherche et Formation*

**RECHERCHE  
& FORMATION**

# Petit point logistique

**2 pauses-café :**

- entre 10h45 et 11h15
- entre 15h et 15h20

**Repas du midi :** de 12h à 13h15

## Salle 28

RDC MSH, en bas des escaliers, à droite

# Petit point logistique

Enregistrements des interventions de la journée

**2 canaux de diffusion**

- UCA : WebTV de l'Université Clermont Auvergne

<https://webtv.u-clermont1.fr>

- CLARIN : Videolectures.NET

<http://videolectures.net/clarin/?q=clarin#o=hot&t=evt>

**Nos demandes :**

**Votre accord dans 2 documents**

-contrat de cession de droit à l'image de l'UCA

-contrat de cession de droit à l'image "Publishing consent and authorship statement" de CLARIN

**Yotre diaporama** (par email ou lors de la JE)

# Cadrage de la journée

## Les corpus aujourd'hui

- Grande **diversité** des corpus recueillis : écrits, oraux, multimodaux, ...
  - Sciences du langage – didactique / sciences de l'Education
    - corpus de situations d'interactions pédagogiques filmées ou médiées par les technologies (Muller et al., 2016 ; Roche et Gal-Petitfaux, 2015 ; Veillard et Tiberghien, 2012 (Mulce, 2015 ; Guichon et al., 2014)
  - Sciences médicales ou psychologiques
    - corpus variés de questions de pathologies médicales (LPL, 2012) ; corpus spécialisés de textes médicaux (Maniez, 2011) ; corpus d'entretiens cliniques (DECLICS (Dispositif d'Etudes CLIniques sur les Corpus Santé) ; Tanguy et al., 2011 ; Vasile, 2011).
- Développement de **méthodologies** et **d'outils de standardisation** et d'**annotation** de ces corpus hétérogènes
- Mise en place d'**infrastructures nationales et internationales** pour le **partage** des corpus dans différentes communautés scientifiques (Ortolang, ...)



# Cadrage de la journée

**Un constat** = utilisation des corpus pour la formation de praticien.ne.s, de professionnel.le.s de terrain assez peu documentée

**Lier recherche et formation via corpus : des initiatives, oui !**

**PLATEFORME NÉOPASS@CTION**



Quelles **formes d'appropriation** de ces « produits » de la recherche par le champ de la formation professionnelle ?

# Du côté de l'acquisition du langage

# Former qui ?

- Professeur.e.s des écoles
- Professeur.e.s de FLE (bilinguisme précoce)
- Professionnel.le.s de la petite enfance
- Orthophonistes
- Pédiatres...

# Avec quoi ?

- Corpus transversaux vs longitudinaux
- Corpus spontané / écologique vs contrôlé
- Corpus d'interactions (parents-enfants, enseignant-enfants, apprenant-apprenant)

# Quelles ressources ?

- Quantitatives : inventaire phonologique, morpho-syntaxique, lexical
- Qualitatives : structures de systèmes en cours de développement (système phonologique, syntaxique, réseau sémantique)

# Les corpus : quels intérêts pour le praticien ?

- Dégager des régularités du système et du bagage linguistique de l'enfant ( à un instant « t » ou sur une période)
- Etalonner / avoir des points de repères sur le développement « normal » vs « pathologique »
- Construire des profils d'enfants
- Connaître les périodes propices à certains types d'apprentissage
- Dédurre des scénarii d'acquisition

# Quelles contraintes ?

- L'outil informatique et ses spécificités techniques
- L'API ou autres langues de spécialistes
- L'austérité des formats / données présentées

# Se former à chercher quoi ?

Bilanter un enfant : c'ad savoir dans un corpus

- relever son stock lexical
- relever sa MLU
- relever son système phonologique (quels phonèmes acquis / non acquis ? Quelles variantes ?)
- relever son input et le mettre en relation avec ses acquisitions

# Quelles tâches à construire pour une telle formation ?

- Travail sur les données
  - Coder
  - Catégoriser (degré de granularité)
  - Comparer, croiser (entre sujets, entre enseignant/parent et sujets)
  - Mesurer (pourcentage de présence, corrélations)
  - Déduire, en lien avec la recherche actuelle !

**Du côté de la formation  
aux langues étrangères et  
de la formation des  
futur.e.s enseignant.e.s de  
langues étrangères**

# Quel(s) corpus utilisés ?

**Pour répondre aux besoins de formation :**

- À la langue étrangère
- À la traduction / l'interprétariat en langue(s) étrangère(s)
- À l'enseignement/apprentissage d'une langue étrangère

**3 grands types :**

- Des corpus de ressources linguistiques authentiques (textes, genres différents, ...)
- Des corpus d'interactions en situation d'enseignement/apprentissage (de la langue-cible ou d'autres disciplines)
- Des corpus de productions d'apprenant.e.s de langue étrangère (orales, écrites, multimodales)

# Quels usages des corpus ?

## Usage 1 : Le corpus comme « ressource »

- pour attester de phénomènes langagiers particuliers (approche *corpus-based*)
  - pour permettre l'exploration, la découverte de régularités langagières (approche *corpus-driven*)
- 
- Utilisation de logiciels de type concordanciers
  - Génération d'exercices automatiques à partir de concordances/contextes/fréquences
- 
- Plutôt pour la formation à la langue étrangère
  - Plutôt pour des apprenants avancés et/ou spécialistes
  - Plutôt sur des ressources authentiques (de natifs, d'experts) éventuellement multilingues et/ou spécialisés

# Quels usages des corpus ?

## Usage 2 : le corpus comme « référentiel »

- Pour bâtir des programmes de cours
  - Pour construire des supports de formation (modèles de productions, canevas de simulations, de jeux de rôle, etc.)
- Plutôt pour l'apprentissage de langue sur Objectifs Spécifiques (langue de spécialité, langue professionnelle, langue universitaire)
- Plutôt sur des corpus de textes spécialisés et/ou d'interactions langagières en situation professionnelle

# Quels usages des corpus ?

## Usage 3 : le corpus comme « mise en situation »

- Pour l'exploration, l'analyse des pratiques enseignantes contextualisées
- Pour la connaissance des publics apprenants
- Plutôt pour la formation des enseignant.e.s (de langue étrangère ou d'autres disciplines)
- Avec des corpus de productions d'apprenant.e.s
- Avec des corpus d'interactions pédagogiques en situation

# Quels problèmes ?

Malgré facilitation accès aux corpus → Utilisation encore restreinte quelque soit l'approche

## **Problèmes identifiés depuis longtemps**

-technicité / manque de convivialité des outils d'interrogation des corpus

-activité chronophage

-peu (pas ?) de recherche sur l'efficacité dans l'apprentissage et la formation de ces approches corpus

-raisons plus « culturelles » en lien avec l'organisation des formations d'enseignant.e.s, avec les conceptions de l'apprentissage d'une langue étrangère



# Des apports tout de même !

- Accroissement de l'autonomie langagière
- Développement de la sensibilité au sens contextuel, à la variation, etc.
- Facilitation du travail interlinguistique
- Lien avec les pratiques informelles des apprenants (qd apprenants font leurs propres corpus)
- Facilitation de la déconstruction des représentations des enseignant.e.s sur le métier
- Préparation à l'entrée sur le terrain professionnel
- Etc.

# Objectifs de la journée

- Une journée **exploratoire** et **pluridisciplinaire**
- Domaines de la **santé**, de la **rééducation du langage** ou de la **formation des enseignant.e.s** de différentes disciplines
- Etablir un 1<sup>er</sup> **état des lieux** des pratiques de l'utilisation des corpus notamment multimodaux en formation professionnelle
- Identifier **besoins** des formateurs-trices / **freins** en matière d'utilisation des corpus

# Quelques pistes de réflexion

- « **Corpus** » pour les chercheurs / « **corpus** » pour les formateurs ?
- **Transformation** des corpus en « **objets de formation** » ?
- **Interrogation** des corpus par les formateurs ?  
Quelle(s) interface(s)s ?
- Conception de **tâches d'apprentissage** à partir des corpus ?

# Programme de la journée

**9h30-10h - Damien Chabanal & Anne-Laure Foucher** (*Laboratoire de Recherche sur le Langage, Université Clermont Auvergne*) : Utilisation innovante des corpus : cadrage et objectifs de la journée

**10h-10h45 - Lidia Lebas-Fraczak** (*Laboratoire de Recherche sur le Langage, Université Clermont Auvergne*) : Ce que font les enfants et ce que les animateurs ne perçoivent pas toujours dans les discussions à visée philosophique : apport d'une analyse de corpus.

**10h45-11h15 - Pause-café**

**11h15-12h - Lionel Roche** (*Laboratoire Activité, Connaissance, Transmission, Education, Université Clermont Auvergne*) : L'analyse de l'activité des enseignants comme point de départ de la conception d'un dispositif multimodal de formation.

**12h-13h15 - Pause-déjeuner**

**13h15-13h30 - Présentation de l'Infrastructure Européenne de Recherche CLARIN.**

**13h30-14h15 - Mylène Blasco, Océane Advocat & Aline Delsart** (*Laboratoire de Recherche sur le Langage, Université Clermont Auvergne*) : Un corpus oral d'entretiens en santé : de la construction du corpus aux répercussions de l'analyse linguistique dans la formation

**14h15-15h - Eva Schaeffer-Lacroix** (*Centre de Linguistique en Sorbonne, Université Paris Sorbonne*) : Se servir des corpus d'apprenants et des corpus d'experts pour former en langue étrangère

**15h-15h20 - Pause-café**

**15h20-16h05 - Christophe dos Santos** (*Inserm U1253 « Imagerie et Cerveau » (iBrain), Équipe Psychiatrie Neurofonctionnelle, Université de Tours*) : Enseigner la création et l'utilisation de corpus langagiers à des étudiant-e-s en orthophonie : partage d'expérience pédagogique.

**16h05-16h30 - Clôture de la journée** : Prolongements de la journée autour de quelques questions-clés



CLARIN



MAISON DES SCIENCES DE L'HOMME  
CLERMONT-FERRAND

UNIVERSITÉ  
Clermont Auvergne

## Journée d'étude

CLARIN User Involvement Event

# COREFO

Lier **re**cherche et **fo**rmation  
professionnelle :  
des **co**rpus à l'**in**terface

Organisée par l'axe CA2LI

Corpus, Acquisition et Apprentissage, Langues en Interactions

Laboratoire de Recherche sur le Langage (EA 999)

*Vendredi 16 novembre 2018 – 9h/16h30*

Amphithéâtre 219 - Maison des Sciences de l'Homme

Université Clermont Auvergne

4 rue Ledru, 63000 Clermont-Ferrand

<https://corefo2018.sciencesconf.org/>

# Bonne journée à toutes et à tous !