

JOURNÉE D'ÉTUDES

**Du recueil de données
à l'analyse des corpus
en didactique de l'anglais :
éthique, procédure, outils, moyens**



» Organisation

ARDA | www.ardaa.fr
ATILF | www.atilf.fr

» Contact

virginie.privas-breaute@univ-lorraine.fr

VENDEDI
24
JANVIER

2020

NANCY

Campus Lettres et Sciences Humaines
Bâtiment G | Salle G004

Pour le chercheur en didactique de l'anglais, le travail sur corpus est particulièrement enrichissant pour saisir la nature et les spécificités des productions et interactions dans la classe de langue. Le travail sur échantillon constitue une voie envisageable pour nourrir la réflexion autour de questions de recherche données. Pour ce faire, il est recommandé que la taille des échantillons d'appui soit « représentative »¹, pour satisfaire aux exigences de robustesse, validité et normes académiques habituelles de la sphère scientifique : enregistrer quelques heures d'interaction dans une salle de cours permet en effet de recueillir quelques éléments exploitables mais ne tient pas lieu d'échantillon représentatif de l'anglais parlé en situation réelle, par exemple. Les corpus peuvent donc prendre diverses formes en fonction de leur objet d'étude : des ensembles de productions écrites (scripturales ou numériques), de productions ou d'interactions orales, enregistrées en format audio ou vidéo, en situation écologique ou expérimentale. Le recueil peut alors être unique ou multiple, permettant ainsi une analyse longitudinale. Quant à la taille de ces corpus, elle peut être variable, allant de corpus écrits ou transcrits, comprenant plusieurs milliers ou centaines de milliers de mots, à des corpus plus réduits et éventuellement analysés d'après des approches plus qualitatives.

Construire un corpus s'inscrit dans une démarche scientifique globale (Dalbera, 2002) et implique une analyse spécifique en arrière-plan et le type de données que le.e chercheur.se sélectionne traduit une/des préoccupation.s sous-jacente.s précise.s, d'où la multiplicité des formes que peuvent adopter les corpus : corpus de textes, corpus d'unités de langue, ceux qui relèvent de l'écrit et ceux qui relèvent de l'oral, corpus conçus comme des échantillons de faits linguistiques, ceux qui se veulent exhaustifs dans un champ donné, etc. Dans une optique « responsable » de recherche (Narcy-Combes, 2005), il est donc essentiel pour le chercheur de penser son corpus et de le construire en fonction de ce qu'il souhaite analyser, étudier, démontrer.

Cette deuxième journée d'études organisée par l'association pour la recherche en didactique de l'anglais et acquisition (ARDAA) sera l'occasion de proposer un dialogue entre des chercheurs confirmés, des doctorants et des étudiants de Master MEEF anglais, sciences du langage et sciences de l'éducation français et internationaux, autour d'une variété de possibilités de recueil de données, de constitution de corpus, d'exploitation et d'analyse des données qu'ils contiennent.

Les échanges permettront aux jeunes chercheurs (étudiants de master et doctorants) d'identifier les principaux critères de classification de leurs corpus et leur exploitation : corpus de textes, corpus d'unités de langue, ceux qui relèvent de l'écrit et ceux qui relèvent de l'oral, corpus conçus comme des échantillons représentatifs des faits linguistiques et ceux qui se veulent exhaustifs dans un champ donné.

Les intervenants invités seront des spécialistes reconnus et incontournables des grandes notions de la journée d'étude. Ils.elles présenteront leurs corpus et les outils qu'ils.elles utilisent pour analyser leurs données et répondre aux questions de recherche qu'ils.elles poursuivent.

¹ Par exemple, en statistiques, les échantillons strictement supérieurs à 30 individus sont considérés comme de « grands échantillons » (Couty et al, 2004 : 95).

Vendredi 24 Janvier

9h30 **Accueil des participants | Café**

9h55 **Ouverture de la journée d'études | Virginie Privas-Bréauté** (Nancy)

10h00 – 11h00 **Conférence plénière | Alex Boulton** (Nancy)
Trente ans de recherches en apprentissage sur corpus

Présentations | modérateur **Elise Ouvrard, Caen**

11h00 – 11h45 **Clive Hamilton** (Paris)
Exploiter les corpus d'écrits universitaires : le poids de la discipline et le niveau d'expertise

11h45 – 12h30 **Anne-Marie Voise** (Paris) et **Marie Potapushkina-Delfosse** (Rouen)
Corpus en classe de langues à l'école maternelle

12h30 – 14h00 **Pause déjeuner**

Ateliers | modérateur **Joséphine Rémon, Lyon**

14h00 – 14h45 **Atelier 1 | Pauline Beaupoil et Justine Paris** (Paris)
Transcrire, coder, annoter et analyser des données vidéos : tour d'horizon de quelques outils

14h45 – 15h30 **Atelier 2 | Pascale Manoïlov, Agnès Leroux** (Paris)
Constituer et analyser un corpus d'interactions orales : d'un objectif à l'autre.

15h30 – 15h45 **Pause café**

15h45 – 16h30 **Atelier 3 | Stéphane Soulain** (Montpellier)
Apprendre par corps : défis méthodologiques d'une étude de transposition didactique auprès d'enseignants de langues

16h30 – 17h15 **Atelier 4 | Margot Kuligowska Esnault** (Nantes)
Productions écrites de type poétique – entre analyse qualitative et quantitative

17h15 – 17h45 **Clôture par les membres du bureau de l'ARDA**

Trente ans de recherches en apprentissage sur corpus

Les outils et techniques de la linguistique de corpus ne se limitent pas aux seules analyses linguistiques. Ils peuvent aussi améliorer les ressources disponibles (dictionnaires, grammaires, manuels...) et même être utilisés par les enseignants et apprenants de langue étrangère ou seconde. Cette conférence présente un recueil rigoureux de 351 publications en anglais (principalement des articles, chapitres ou actes de colloque) de 1989 à 2018 qui visent à évaluer différents aspects du *data-driven learning*, ou apprentissage sur corpus (ASC). L'analyse permet de contraster une vision historique de l'ensemble avec un sous-corpus composé de 46 articles publiés dans des revues à haut impact ces 5 dernières années. De nombreuses synthèses se limitent à des revues majeures, mais passent ainsi à côté de grandes quantités de travaux ; dans le cas présent, par exemple, seules 3 études en Iran représentent les 62 études de l'ASC dans les pays du Moyen Orient et les 23 en Turquie sont entièrement absentes.

Le codage des publications permet un certain nombre de constats sur le contexte des études, les participants, les corpus et outils, la méthodologie, etc. Le plus important est l'immense variation à tous les niveaux – des participants et leurs contextes, des corpus et des outils, des objectifs pédagogiques et linguistiques, des activités et procédures, des instruments et méthodologies. Ce constant appuie l'affirmation qu'il n'y a pas une seule forme universelle de l'ASC mais que celui-ci s'ouvre sur de multiples dimensions. Néanmoins, il reste de nombreuses pistes encore peu explorées, dont l'ASC mobile, pour d'autres langues que l'anglais, pour l'oral, ou encore en-dehors de la salle de classe universitaire. Les pratiques méthodologiques ainsi que les descriptifs fournis sont également à améliorer. Une dernière partie concerne une analyse de corpus de ces 351 publications pour un total de 1,8 millions d'occurrences. Une analyse des mots les plus fréquents et les plus significatifs par rapport à un corpus d'anglais général, ainsi que des groupes de mots récurrents, permet de mieux saisir le sujet à travers les yeux des chercheurs qui publient dans ce domaine ; d'autres analyses sont également possibles en comparant les premières recherches contre les plus récentes, ou encore entre types de supports, pays d'analyse, etc. Ensuite, on continue en explorant certains éléments saillants, des concepts et théories jusqu'aux auteurs les plus cités. La conclusion est que le domaine se porte bien avec un nombre croissant de publications à travers le monde, mais que si des méta-analyses montrent un effet important de l'approche, les recherches sont encore à poursuivre dans plusieurs domaines.

Corpus en classe de langues à l'école maternelle

Les corpus de nos recherches action en didactique des langues à l'école sont majoritairement composés d'enregistrements audio et vidéo des productions d'élèves, d'interactions élèves-enseignants et de transcriptions des entretiens menés en individuel ou en petits groupes auprès d'élèves.

Anne-Marie Voise exposera quelques dispositifs d'élaboration de différents types de corpus oraux et multimodaux, leurs fondements, leurs traitements et leurs usages dans le cadre de recherches menées au tout début de l'apprentissage des langues en contexte scolaire.

Marie Potapushkina illustrera ce propos à l'aide de quelques exemples d'extraits de corpus qu'elle analysera à la lumière des questions qui guident les investigations des deux chercheuses: en quoi l'immersion dans un monde imaginaire, l'identification aux personnages de celui-ci, la gestuelle iconique et rythmique, la coopération avec les pairs, etc. permettent à de très jeunes élèves d'accéder à une nouvelle langue.

Quelques références bibliographiques

Potapushkina-Delfosse, M., 2019. Le sens tactile dans l'initiation à une langue vivante à l'école maternelle. *Methodal – Méthodologie de l'enseignement/apprentissage des langues*, n°3. En ligne <https://methodal.net/Le-sens-tactile-dans-l-initiation-a-une-langue-vivante-a-l-ecole-maternelle>

Potapushkina-Delfosse, M., 2019. Langues, art et empathie à l'école: communauté et singularité d'expérience. *TIPA – travaux interdisciplinaires sur la Parole et le Langage*, 35/2019. En ligne <http://journals.openedition.org/tipa/3375>

Voise Anne-Marie., 2018. Corps – accords : une approche holistique et transdisciplinaire des langues à l'école maternelle, *Cahiers de l'APLIUT*, Vol. 37 N°2 | 2018: Corps, voix et langues dans l'enseignement supérieur <https://journals.openedition.org/apliut/6345>

Voise Anne-Marie., 2018. Délier- relier les langues en maternelle. Orizons, Université de Haute Alsace.

Exploiter les corpus d'écrits universitaires : le poids de la discipline et le niveau d'expertise

Selon Braud et al. (2015), 60% des universités françaises proposent des cours de langues de spécialité dès la première année de licence. Ces cours sont souvent limités, cependant, à l'apprentissage du lexique, de la grammaire et accessoirement de la terminologie (Chaplier, 2016). La question de la syntaxe spécialisée, des éléments lexicaux ou discursifs inhérents à ces discours de spécialité n'est pas abordée ou prioritaire dans ces formations. Face à ce constat, ce travail de réflexion met en avant, l'importance de la prise en compte, non seulement des domaines de spécialité, mais aussi du genre textuel dans la conception des cours de langues de spécialité d'une part et propose d'analyser trois éléments, a priori, caractéristiques du discours scientifique anglais, d'autre part.

Nous nous appuyons sur deux corpus d'écrits scientifiques : l'un composé d'écrits d'étudiants anglophones en première année d'université (corpus BAWE), et l'autre composé d'articles scientifiques rédigés par des chercheurs anglophones (corpus EIIDA). Chacun des corpus comporte une discipline issue des sciences humaines et une autre des sciences naturelles. Les résultats principaux indiquent une grande variabilité d'emploi de certains traits caractéristiques, à la fois au sein d'une même discipline, qu'entre les disciplines elles-mêmes ou entre les deux groupes de locuteurs natifs.

Bibliographie

Braud, V., Millot, P., Sarré, C., & Wozniak, S. 2015b. « You say you want a revolution... » Contribution à la réflexion pour une politique des langues adaptée au secteur LANSAD, Recherche et pratiques pédagogiques en langues de spécialité, Vol. XXXIV N° 1 | URL : <http://apliut.revues.org/5020> ; DOI : 10.4000/apliut.5020

Chaplier, C. (2016). Anglais des sciences, variété d'anglais de spécialité : réflexions sur la formation des enseignants en France. *Journal of Languages for Specific Purposes*, pp.83-93

Du recueil de données à l'analyse des corpus en didactique de l'anglais : éthique, procédure, outils, moyens

Transcrire, coder, annoter et analyser des données vidéos : tour d'horizon de quelques outils

L'objectif de cette intervention sera de faire un tour d'horizon de quelques outils les plus utilisés en matière de transcription et codage de données dans le champ de l'acquisition du langage et des langues secondes. Nous nous intéresserons tout d'abord au concept de transcription et de codage des données : pourquoi transcrire, combien de temps faut-il compter pour transcrire des données, quelles sont les implications théoriques, quelle distinction faire entre codage et analyse, pourquoi coder des données dans une grille ?

Nous présenterons ensuite trois outils de transcription et de codage de données qui nous paraissent centraux en linguistique et didactique des langues aujourd'hui pour mener à bien des analyses qualitatives et/ou quantitatives : CLAN, PRAAT et ELAN. Chacun de ces outils permet de proposer une transcription langagière (CLAN) ou des annotations prosodiques (PRAAT) et gestuelles (ELAN).

Enfin, dans l'optique de rendre ce panorama le plus concret possible, nous reviendrons sur des extraits choisis de nos données de thèses – collectées respectivement auprès d'enfants natifs anglophones et d'apprenants francophones de l'anglais langue seconde. Nous exposerons comment nous les avons transcrits, codés et analysés à la lueur de questions de recherche, cadres théoriques et procédés méthodologiques différents.

Constituer et analyser un corpus d'interactions orales : d'un objectif à l'autre

Cette communication présentera les modalités de construction d'un corpus d'interactions orales filmées en classe de LVE entre des apprenants de cycle 4 (classe de 3^{ème}).

À partir de questions de recherche, nous nous interrogerons sur les enjeux de la collecte des données en milieu écologique (modalité, type de collecte, format), leur traitement ainsi que les analyses qui peuvent en découler.

Nous montrerons comment un corpus construit à partir d'hypothèses et de questions de recherche, puis analysé selon des paramètres de pragmatique conversationnelle peut servir d'autres approches théoriques et répondre à d'autres problématiques.

Apprendre par corps : défis méthodologiques d'une étude de transposition didactique auprès d'enseignants de langues

La communication rend compte d'une recherche qui étudie les processus de transposition didactique mis en œuvre à l'issue d'une formation menée en 2017-18 auprès un groupe d'enseignants du premier degré et de formateurs en langues. Après une présentation des visées de la recherche, la première partie de la communication expose la méthodologie retenue pour cette recherche (l'observation des enseignants en situation professionnelle et la conduite d'Entretien d'Explicitation). La seconde partie propose quelques résultats, notamment :

1. La posture de l'enseignante et l'interaction avec les élèves.
2. Les activités corporelles et l'utilisation de différents types de gestes.
3. La boucle d'actions qui conçoit l'apprentissage comme une circularité entre réception-production orale et action.

Mots clés

Corps - énonciation - entretien d'explicitation - geste - transposition

Productions écrites de type poétique – entre analyse qualitative et quantitative

Cette communication s'appuie sur une partie d'une étude doctorale axée sur la poésie dans l'enseignement-apprentissage des langues. Il s'agira d'exposer et d'interroger les choix qui ont été retenus pour analyser un corpus de 32 textes poétiques en L2 anglais, produits par des élèves du secondaire en France. L'analyse de ces productions écrites était à la fois de type qualitatif (analyse de contenu, Bardin, 2013) et quantitatif (le logiciel d'analyse lexicale en ligne VocabProfile-Classic et le logiciel SPSS).

Biographies

Alex Boulton est professeur en anglais et didactique des langues à l'Université de Lorraine, et directeur du laboratoire *Analyse et Traitement Informatique de la Langue Française* (ATILF – UMR 7118 : CNRS & Université de Lorraine). Il enseigne principalement dans la première UFR Lansad (langues pour spécialistes d'autres disciplines) en France et dans d'autres départements avec des cours de langue anglaise, de linguistique générale, de didactique, de linguistique de corpus et de méthodologie de la recherche, et encadre des mémoires et des thèses dans les domaines de ses intérêts de recherche. Ceux-ci concernent au plan large le lexique, la didactique des langues, l'exploitation des TIC ; et plus précisément l'apport des corpus (*data-driven learning*) pour l'enseignement/apprentissage des langues. Il a de nombreuses publications dans ces domaines et fait partie de comités de différentes associations dont l'AFLA (vice-président), TaLC et EUROCALL ; et de revues dont ReCALL (éditeur), Alsic, ASp, CALL-EJ, Eurocall Review, IJCALLT, JALT-CALL Journal, Language Learning & Technology, et Al-Lisaniyyat.

Marie Potapushkina-Delfosse est maître de conférences en didactique des langues à l'INSPE de l'Université de Rouen. Sa recherche porte sur l'enseignement/apprentissage énonctif des langues à l'école avec un intérêt particulier pour des médiations artistiques : théâtrale, picturale, dansée. En mettant régulièrement en œuvre des dispositifs didactiques expérimentaux qui relient les arts et les langues, auprès d'un public de scolaires, elle explore le rôle du corps mouvant, du geste globalement corporel et manuel ainsi que des émotions dans les processus d'appropriation langagière.

Pauline Beaupoil-Hourdel est linguiste, maître de conférences à Sorbonne Université, INSPE de l'académie de Paris, CeLiSo EA 7332. Son travail porte sur l'acquisition de la négation et des postures multimodales d'opposition dans les interactions adultes-enfants en milieu naturel et spontané. Formatrice à l'Institut National Supérieur du Professorat et l'Éducation de l'académie de Paris, elle s'intéresse également à la didactique des langues étrangères et aux processus d'apprentissage des langues dans le premier degré. Elle est porteuse du projet de recherche IMAAJEE (Interaction et Multimodalité pour l'Apprentissage de l'Anglais par les Jeunes Enfants à l'Ecole) financé par le GIS RREEFOR-INSPE qui s'appuie sur des données filmées d'interactions polyadiques en classe de maternelle. Dans le cadre de ses recherches, Pauline Beaupoil-Hourdel a développé une méthode de codage pour l'analyse multimodale des interactions adultes-enfants en associant approches quantitative et qualitative.

Justine Paris est maître de conférences en linguistique anglaise et didactique des langues à l'Université de Paris. Ses recherches portent sur le langage figuré en acquisition du langage L1/L2 ainsi que sur l'hybridation et la personnalisation des parcours dans le domaine l'apprentissage des langues. Elle s'intéresse ainsi à la transformation des pratiques pédagogiques via le numérique et à l'usage des TICE en classe de langue qui permettent des modalités d'apprentissage sur mesure. Elle porte également un grand intérêt à l'autonomisation de l'étudiant et aux interactions entre le Centre de Ressource en Langues (CRL) et la classe de langue, et par là même à la prise en compte des pratiques informelles pour le travail de classe.

Stéphane Soulain est agrégé d'anglais et docteur en sciences du langage - didactique des langues. Il est actuellement Maître de Conférences à l'Université de Montpellier - Inspe, responsable des formations en didactique des langues, auprès des étudiants futurs enseignants des 1^{er} et 2nd degrés. Ses recherches portent sur l'approche corporelle de l'enseignement de l'anglais qui associe les domaines de la linguistique (phonologie), la théorie du mouvement et les pratiques artistiques. L'ensemble des travaux s'inscrit dans le paradigme des savoirs incorporés. Les recherches actuelles sont orientées vers la production orale des élèves, la transposition didactique chez les enseignants en formation, ainsi que les processus d'interaction dans les actes de langage en langue étrangère.

Organisation



Contact

virginie.privas-breaute@univ-lorraine.fr

Web

ARDAA | www.ardaa.fr

ATILF | www.atilf.fr

Plan d'accès



-  : ENTRÉE CAMPUS
-  : POINT D'ACCUEIL / LOGE
- BU** : BIBLIOTHÈQUE UNIVERSITAIRE
-  : POINT DE RESTAURATION
-  : SOIP (Service Orientation Insertion Professionnelle)
-  : MAUL (Musée Archéologique de l'Université de Lorraine)
-  : AMPHITHÉÂTRE
- MDE** : MAISON DE L'ÉTUDIANT
- ATILF** : UNIVERSITÉ DE LORRAINE / CNRS
Laboratoire d'Analyse et Traitement Informatique de la Langue Française
Centre de documentation M. Dinet ● Site linguistique ● Site didactique des langues
-  : SERVICE DE SANTÉ
-  : SUAPS (Activités Physiques et Sportives)
- WC** : TOILETTES
-  : PORTE D'ACCÈS AUX BÂTIMENTS
-  : POINT DE RASSEMBLEMENT
-  : ASCENSEUR
-  : ENTRÉE POUR PERSONNES À MOBILITÉ RÉDUITE