

L'apprentissage sur corpus : un concordancier au service de l'enseignement et de l'apprentissage de la syntaxe du français parlé en interaction

Cadre théorique et enjeux de la thèse

Cette thèse s'inscrit à l'interface entre la linguistique et la didactique des langues. Plus précisément, elle propose de faire des liens entre la linguistique de corpus et la didactique de l'oral en Français Langue Étrangère (FLE). L'exploitation de corpus oraux et multimodaux pour enseigner et apprendre à interagir fait actuellement partie des méthodologies didactiques qui sont expérimentées par de nombreux enseignants et apprenants de FLE selon différentes modalités (Ravazzolo, Etienne 2019 ; Etienne, Jouin 2019 ; André 2019a, 2018). Il existe plusieurs façons d'exploiter des corpus à des fins didactiques et plusieurs types de corpus peuvent faire l'objet de séquences pédagogiques en FLE (André 2019b). Cette confrontation des apprenants à des corpus en langue cible s'inscrit dans le prolongement de l'utilisation de documents authentiques en didactique (Holec 1990 ; Boulton 2009). Les recherches actuelles dans ce domaine s'accordent pour soutenir que l'exposition à la langue cible pour des apprenants est indispensable. Depuis les années 1970 et les travaux qui prônent l'utilisation des documents authentiques, de nouvelles formes d'exposition sont apparues notamment avec les progrès technologiques et l'accès au numérique.

La thèse s'intéressera à l'exploitation des corpus à des fins didactiques selon les principes du *data-driven learning* (Johns 1991, Aston 2001), traduit en français par l'Apprentissage Sur Corpus (ASC) (Boulton, Tyne 2014). La première mention de la possible exploitation outillée de corpus à des fins didactiques date du milieu des années 1980 avec la publication de Higgins et Johns (1984) et l'utilisation du terme *classroom concordancing*. Il s'agit d'utiliser un concordancier, un outil informatique créé pour faire de la linguistique de corpus qui permet de rechercher les occurrences d'un ou de plusieurs mots ou d'une chaîne de caractères dans une grande masse de données. Les résultats sont affichés verticalement, en alignant des lignes de concordances, comme c'est le cas dans la figure suivante :

Num	Fichier	Contexte Gauche	Occurrence	Contexte Droit
1	Aidesoignante...	des accompagnements à	pendant	un an j'ai découvert
2	Aidesoignante...	travaillé dans ce milieu là	pendant	quatorze ans et puis après
3	Aidesoignante...	contrat accompagnement à	pendant	mes deux ans et la
4	Aidesoignante...	suis repartie à l'école	pendant	six mois à «Verdun» j'
5	Algerie_AUD_17	école euh dans mon village	pendant	un ou deux ans et
6	Algerie_AUD_17	année du d-du bac	pendant	ce temps mon père était
7	Algerie_AUD_17	était cultivateur euh on rentrait	pendant	les grandes vacances et on
8	Algerie_AUD_17	des battages des des moissons	pendant	le pendant les grands travaux
9	Algerie_AUD_17	des des moissons pendant le	pendant	les grands travaux d'été
10	Algerie_AUD_17	les grands travaux d'été	pendant	l-on partait pas en
11	Algerie_AUD_17	on partait pas en vacances	pendant	les grandes vacances on restait
12	Algerie_AUD_17	et m-on a prié	pendant	tout le chemin à haute
13	Algerie_AUD_17	et moi je disais euh	pendant	que ma maman continuait de
14	anorexie_ban...	fais pas de l'anorexie	pendant	un moment j'ai arrêté
15	anorexie_ban...	était vraiment de l'anorexie	pendant	un moment j'y trouvais
16	anorexie_ban...	et un arrêt de travail	pendant	presque deux ans pendant
17	anorexie_ban...	travail pendant presque deux	pendant	deux ans je ne suis
18	anorexie_ban...	début il m'a suivi	pendant	trois mois il a commencé

Figure 1 : Extrait des résultats de la recherche de « pendant » dans le corpus TCOF (<https://www.cnrtl.fr/corpus/tcof/>) avec le concordancier JConc

Dans une perspective didactique, ces lignes sont analysées par des apprenants de langue en salle de classe. C'est le principe fondamental du *data-driven learning*, dont la première apparition du terme se trouve dans un article de Johns paru en 1990, repris un an plus tard dans l'ouvrage fondateur qu'il coédite avec Philip King (Johns, King 1991). Dans leur introduction, ils posent une première définition du *data-driven learning* qui peut être décrit comme :

an application of computers to language-learning that has come to be known as 'classroom concordancing' or 'data-driven learning' (DDL) – the use in the classroom of computer-generated concordances to get students to explore the regularities of patterning in the target language, and the development of activities and exercises based on concordance output. (Johns, King 1991 : iii).

Depuis une trentaine d'année, plusieurs travaux anglo-saxons montrent que le *data-driven learning* (Johns 1991, Aston 2001) apparaît comme une méthodologie efficace (Boulton, Cobb 2017). L'apprenant observe les données, les catégorise et dégage des règles de fonctionnement. Il se comporte en apprenti-linguiste ou en *Sherlock Holmes* (Johns 1997, Bernardini 2004) afin de saisir le fonctionnement de la langue en contexte. Cependant, même dans les pays anglo-saxons qui mettent en œuvre cette méthodologie, jusqu'à présent elle est déployée quasiment uniquement pour l'enseignement et l'apprentissage de l'anglais écrit (Timmis 2015). En France le retard est considérable. Quelques expérimentations sont menées, également souvent pour travailler des compétences scripturales en anglais (Landure 2011). Pour le français, les premières expérimentations de l'ASC sont mises en œuvre pour acquérir des compétences socio-interactionnelles à l'oral. Elles sont réalisées seulement depuis quelques années, notamment dans le cadre du projet FLEURON (<https://fleuron.atilf.fr/>) qui propose un corpus multimodal outillé d'un concordancier (André 2016). Ces expérimentations sont très encourageantes (André 2017) dans la mesure où elles montrent que les apprenants mettent en œuvre des activités métalangagières et des stratégies métacognitives (Gombert 1996) qui leur permettent de s'appropriier plus facilement, et de façon plus pertinente, les règles de fonctionnement de la langue parce qu'ils les dégagent eux-mêmes, tout en étant accompagnés ou guidés par un enseignant (Kerr 2013).

Lors des expérimentations menées jusqu'à présent, il a été constaté que l'ASC permettait aux enseignants et aux apprenants d'aborder la syntaxe d'une façon tout à fait innovante. La question de l'enseignement de la grammaire est récurrente en didactique du FLE, soit elle est au centre de l'enseignement, soit elle est évacuée pour laisser la place aux objectifs communicatifs (voir notamment Hamez, Pereiro 2018). Cependant, entre ces deux positions, de nombreux apprenants posent des questions auxquelles les enseignants ne savent pas ou ne veulent pas répondre, notamment parce que les descriptions du français parlé en interaction ne sont référencées dans aucun manuel et aucune grammaire (voir par exemple Blanche-Beneviste, Jeanjean 1987 ; Mondada 2002 ; Kerbrat-Orecchioni 2005 ; Traverso 2016 ; Giroud, Surcouf 2016). Ces questions sont légitimes et peuvent être traitées avec l'ASC qui semble permettre de développer une conscience métalangagière et un comportement d'apprentissage plus efficace que la grammaire explicite, celle qui est prescrite par l'enseignant (Lin 2019, Liu 2011, Miangah 2011). C'est ce que cette thèse va tenter de montrer.

Méthodologie et terrain de recherche

Dans un premier temps, le ou la doctorant.e devra repérer les phénomènes linguistiques qui posent le plus de difficultés pour des apprenants de FLE dans leur appropriation du français parlé. Pour cela, il ou elle pourra s'appuyer sur le réseau international d'enseignants constitué dans le cadre de plusieurs projets de recherche coordonnés par l'encadrante et notamment du projet FLEURON. Une sélection de phénomènes linguistiques sera opérée pour faire partie de l'étude. Dans un second temps, le ou la doctorant.e devra mettre en place les expérimentations longitudinales nécessaires, avec des groupes tests et des groupes témoins, afin d'évaluer les effets de l'ASC quant à l'appropriation de la langue et de ses usages en situation. Ces expérimentations pourront notamment avoir lieu au Département de FLE (DÉFLE) de l'université de Lorraine et au sein des organismes de formation pour adultes, partenaires du groupe de recherche Langage, Travail et Formation (LTF) de l'ATILF. Enfin, l'analyse de ces expérimentations et les ajustements de la méthodologie permettront de proposer une nouvelle approche didactique de la syntaxe du français parlé en interaction.

Retombées de la thèse

Les manuels de langue et les grammaires traditionnelles ne présentent pas la langue réelle en usage. L'absence de description du fonctionnement du français parlé en interaction dans ces documents de référence conduit les apprenants à s'approprier une langue prescrite, qui n'existe pas. Cet apprentissage nuit à l'intégration langagière des apprenants, qu'ils soient étudiants, professionnels ou migrants. La mise au jour d'une nouvelle méthodologie et la démonstration de son efficacité à long terme permettra de compléter les méthodologies actuelles et de développer chez les apprenants les compétences nécessaires pour interagir de façon appropriée dans les situations de communication qu'ils rencontrent.

Bibliographie

- André V. (2019a). Pourquoi faire de la sociolinguistique des interactions verbales avec des enseignants et des apprenants de Français Langue Étrangère. *LINX*, n°79. <https://journals.openedition.org/linx/3694>
- André V. (2019b). Des corpus oraux et multimodaux authentiques pour acquérir des compétences sociolangagières. In Gajo L., Luscher J.-M., Racine I., Zay F. (eds), *Variation, plurilinguisme et évaluation en français langue étrangère*. Bern : Peter Lang, p.209-223.
- André, V. (2018). Nouvelles actions didactiques : faire de la sociolinguistique de corpus pour enseigner et apprendre à interagir en français langue étrangère. *Action didactique*, n°1, 71-88. <http://univ-bejaia.dz/pdf/ad1/Andre.pdf>
- André V. (2017). Un corpus multimédia pour apprendre à interagir en situations universitaires en France. *Actes colloque international de l'ATPF « Enseigner le français : s'engager et innover »*, p.292-315.
- André V. (2016). FLEURON : Français Langue Étrangère Universitaire – Ressources et Outils Numériques. Origine, démarches et perspectives. *Mélanges Crapel*, 37, p.69-92. <http://www.atilf.fr/IMG/pdf/art3.pdf>
- Aston, G. (Ed.). (2001). *Learning with corpora*. Houston: Athelstan.
- Bernardini S. (2004). Corpora in the classroom: An overview and some reflexions on future developments. In Sinclair J. (ed.). *How to Use Corpora in Language Teaching*. Amsterdam: Johns Benjamins, p.15-36.
- Blanche-Benveniste C., Jeanjean C. (1987). *Le français parlé. Transcription et édition*. Paris, Didier.
- Boulton A. (dir.) (2009). *Des documents authentiques oraux aux corpus : questions d'apprentissage en didactique des langues*, *Mélanges CRAPEL*, 31.
- Boulton, A., Cobb, T. (2017). Corpus Use in Language Learning: A Meta-Analysis. *Language learning*. Volume 67, Issue 2, 348-393. <https://doi-org.bases-doc.univ-lorraine.fr/10.1111/lang.12224>
- Boulton, A., Tyne, H. (2014). *Des Documents Authentiques aux Corpus. Démarches pour l'Apprentissage des Langues*. Paris : Didier.
- Giroud A., Surcouf C. (2016). « De « Pierre, combien de membres avez-vous ? » à « Nous nous appelons Marc et Christian » : réflexions autour de l'authenticité dans les documents oraux des manuels de FLE pour débutants. *SHS Web of Conferences* 27. CMLF 2016. <https://doi.org/10.1051/shsconf/20162707017>.
- Gombert J.-É. (1996). Activités métalinguistiques et acquisition d'une langue. *Acquisition et interaction en langue étrangère*, 8, p.41-55.
- Hamez M.-P., Pereiro M. (eds) (2018). *Les langues Modernes. « Grammaire ? Vous avez dit grammaire ? Représentations et pratiques enseignantes »*, n°3.
- Johns T. (1991). Should you be persuaded: Two examples of data-driven learning. In Johns T., King P. (dir.), *Classroom Concordancing. English Language Research Journal*, 4, 1-16.
- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (dir.), *Teaching and Language Corpora*. Harlow : Addison Wesley Longman, 100-115.
- Johns, P. King P. (dir.) (1991). *Classroom Concordancing. English Language Research Journal*, 4.
- Kerbrat-Orecchioni C. (2005). *Le discours en interaction*. Paris : Armand Colin.
- Kerr B. (2013). Grammatical Description and Classroom Application. Theory and Practice in Data-Driven Learning. *Bulletin suisse de linguistique appliquée*, 97, p.17-39.
- Landure C. (2011). Data-Driven Learning : Apprendre et enseigner à contre-courant. *Mélanges CRAPEL*, 32, p.163-178.
- Lin M. H. (2019). Becoming a DDL teacher in English grammar classes: A pilot study. *The Journal of Language Learning and Teaching*, 9(1), p.70-82.
- Liu D. (2011). Making grammar instruction more empowering: An exploratory case of corpus use in the learning/teaching of grammar. *Research in the Teaching of English*, 45(4), p.353-377.
- Miangah T. M. (2011). The effect of exploiting corpora in TEFL classroom: A case study. *Theory and Practice in Language Studies*, 1(4), p.370-378.
- Mondada L. (2002). Pour une linguistique interactionnelle. In Santacroce M. (dir.). *Faits de langue - faits de discours. Données, processus et modèles. Qu'est-ce qu'un fait linguistique ?* Vol. 2. Paris : L'Harmattan, p. 95-136.
- Ravazzolo E., Etienne C. (2019). Nouvelles ressources pour le FLE à partir des études en interaction. *LINX*, n°79. <https://journals.openedition.org/linx/3454>
- Timmis I. (2015). *Corpus Linguistics for ELT. Research and Practice*. London: Routledge.
- Traverso V. (2016). *Décrire le français parlé en interaction*. Paris : Éditions Ophrys.