Rendre un grand corpus oral accessible pour la didactique du FLE : le projet ESLO-FLEU

Marie Skrovec, Chloé Sarah Tahar, Flora Badin, Britta Thörle

▶ **To cite this version:**

Marie Skrovec, Chloé Sarah Tahar, Flora Badin, Britta Thörle. Rendre un grand corpus oral accessible pour la didactique du FLE : le projet ESLO-FLEU. 11es Journées internationales de Linguistique de Corpus, Jul 2023, Grenoble, France. pp.319-325. halshs-04478301

HAL Id: halshs-04478301
https://shs.hal.science/halshs-04478301

Submitted on 26 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Rendre un grand corpus oral accessible pour la didactique du FLE : le projet ESLO-FLEU

Marie Skrovec [1], Chloé Tahar [1], Flora Badin [1], Britta Thörle [2]

[1] Laboratoire Ligérien de Linguistique, Université d'Orléans
[2] Département de Romanistik, Université de Siegen

marie.skrovec@univ-orleans.fr, chloe.tahar@univ-orleans.fr, flora.badin@univ-orleans.fr, thoerle@romanistik.uni-siegen.de

Introduction

Le corpus ESLO (Enquêtes SocioLinguistiques à Orléans), l'un des plus grands corpus oraux disponibles pour le français avec ses 600h d'enregistrements audio transcrits (422h en libre accès), constitue un important terrain d'observation du français parlé contemporain pour les linguistes. Structuré selon plusieurs axes, ce « portrait sonore » d'Orléans par ses habitants permet d'analyser le français parlé dans les années 60 et aujourd'hui, en accédant à une diversité de locuteurs dans des situations de communication variées. Le projet ESLO-FLEU (ESLO pour le Fle et la Linguistique dans l'Enseignement Universitaire), qui repose sur une collaboration entre le LLL à l'université d'Orléans et le département des langues romanes à l'université de Siegen, vise à rendre le corpus ESLO accessible à une communauté plus large d'étudiants et d'enseignants dans le domaine du FLE universitaire. Nous montrons comment nous avons sélectionné et structuré des données issues de ce « réservoir », dans une approche de pré-didactisation. Nous nous concentrons ici sur l'étape de sélection et d'indexation d'extraits, préalable à l'élaboration d'une ressource numérique. Cette ressource, constituée de plusieurs modules correspondant aux objectifs de formation des spécialistes du français en contexte universitaire à l'étranger (départements de français et d'études françaises et/ou romanes), est testée en 2022-2023 dans le cadre de cours de spécialité en sciences du langage, linguistique sur corpus et civilisation à l'université de Siegen. Nous expliquons également les choix techniques liés à un double objectif, celui de veiller à la réutilisabilité de ce travail d'enrichissement des données en linguistique (cumulativité), tout en garantissant l'accessibilité des données pour les non linguistes.

État de l'art

La question de l'utilisation des corpus des linguistes pour l'enseignement des langues étrangères est de grande actualité. Depuis plus d'une dizaine d'années, la méthode d'apprentissage sur corpus a été expérimentée, notamment pour l'anglais, domaine précurseur en la matière, à partir de corpus écrits ou oraux (Sinclair 2004, Boulton & Tyne 2014). Parallèlement, on note, à la suite de l'intérêt porté en linguistique pour le français parlé (Blanche-Benveniste 2010, Blanche-Benveniste & Martin 2010) un intérêt en didactique du FLE pour l'oralité et les spécificités du français parlé contemporain (Weber 2013). Parmi les linguistes, nombreux sont ceux à collaborer avec des didacticiens en vue de proposer des ressources pour la didactique du FLE (cf. les différents projets décrits dans Detey et al. 2011, André 2016, Etienne & Jouin 2019, Surcouf & Ausoni 2018).

Corpus et méthodologie

Corpus

Le corpus ESLO présente la particularité d'avoir été constitué en 2 temps (ESLO1 1968-1971, ESLO2 depuis 2008) selon une méthodologie similaire. Le projet trouve son origine à la fin des années 1960 lorsqu'une équipe franco-britannique, composée notamment d'enseignants de français, se fixe pour objectif de créer un corpus afin de constituer un ensemble de documents authentiques pour l'enseignement du français qui aboutit à la publication d'un manuel, *Les Orléanais ont la parole* (1976). L'équipe a également des objectifs de diffusion et d'analyse sociolinguistique (Blanc et Biggs, 1971) et réalise ainsi le « portrait sonore d'une ville » en tenant compte de « l'identité sociale de chaque locuteur et de la situation de communication dans chaque cas » (Lonergan et al., 1974 : 4). Depuis 2008, le laboratoire CORAL devenu LLL a entrepris de constituer un 2e volet, sous forme d'un échantillon comparable à ESLO1 basé sur la situation de l'entretien sociolinguistique, qui permet d'accéder aux représentations des Orléanais sur leur ville, la langue, leur rapport à la norme, etc. En outre, l'accent est mis dans ESLO2 sur la diversité des situations dans un contexte urbain, de manière à « répondre à un objectif de couverture étendue de la diversité des pratiques linguistiques dans la vie quotidienne » (Baude 2015 : 124). Parallèlement, un travail sur l'élaboration d'une typologie des situations de communication a été mené (Baude & Guerin 2014, Baude 2015), basée sur l'idée de continuum variationnel proposée par les romanistes allemands Söll (1974) puis Koch & Oesterreicher (2001) et relayée en France par Gadet (2003). Si plusieurs chercheurs du LLL s'intéressent à la question des transferts didactiques de l'accessibilité du corpus ESLO pour l'enseignement (Skrovec 2019, 2020 ; Hamma 2019), force est de constater qu'en l'absence d'un travail de pré-didactisation de ces données, la ressource reste peu accessible aux non spécialistes en linguistique de corpus en raison de la taille du corpus et de la technicité des outils d'exploration.

Méthodologie : sélection et indexation thématique

L'exploration du corpus a été guidée par la recherche d'extraits présentant un intérêt à plusieurs niveaux. Il s'agissait de :

• valoriser la dimension patrimoniale du corpus en sélectionnant des sujets de conversation saillants, des passages représentatifs du grand corpus ;

• rendre compte de la diversité des locuteurs et des situations d'interaction représentées dans le corpus ;

• identifier des extraits utilisables en contexte pédagogique (FLE universitaire en premier lieu, mais aussi FLE généraliste, voire dans d'autres disciplines), de par leur longueur et contenu. Les frontières ont été définies de manière à circonscrire des séquences d'une durée de 2 à 4 minutes maximum présentant une cohérence thématique et conversationnelle.

Le travail d'exploration-sélection a permis de faire émerger des regroupements thématiques, selon un processus prévoyant plusieurs étapes : recherche de tout type d'extrait d'intérêt, constitution de collections par regroupement thématique (recherche de similarités thématiques entre extraits), suppression d'extraits au thème marginal. L'intégralité des 150 extraits a fait l'objet d'une double indexation au minimum : chaque extrait appartient à un grand ensemble thématique (8 valeurs), et reçoit une sous-spécification thématique (43 valeurs). Parmi ces extraits, 76 intègrent un des 5 modules de formation (livrables du projet) et reçoivent une étiquette supplémentaire (5 valeurs).

L'indexation, pensée pour un public non-expert, a été réalisée en privilégiant le caractère intuitif des étiquettes. Elle vise d'une part (i) à identifier des catégories culturelles de sens commun, qui reflètent plusieurs aspects de la vie quotidienne des Orléanais (Métiers, Loisirs, Société, Vivre à Orléans) ; d'autre part (ii) à identifier des thèmes saillants émergeant dans certains dynamiques conversationnelles (Récits, Conversations, Savoirs, Témoignages). Par exemple, la sous-spécification thématique de l'ensemble des Récits rend compte du fait que les récits conversationnels sont le locus privilégié de l'élaboration de sujets conversationnels prototypiques (Récits familiaux, Expériences personnelles, Premiers emplois, Projets d'avenir, Remémorations d'histoires). La pertinence de l'annotation thématique a été évaluée en fonction de l'accord intersubjectif des deux chercheuses impliquées dans ce travail.

Méthodologie : choix des outils d'annotation et de consultation

La ressource est élaborée sous deux formes :

- Une version TXM (Heiden 2010) qui permet le balisage et l'annotation dans l'intégralité du corpus (correspondant à la partie accessible en ligne), pour l'utilisateur linguiste expert formé à l'annotation et l'interrogation avancée.
- Une version HTML sous forme de fichiers consultables sur un navigateur, pour une utilisation facilitée sans système de requête : consultation des extraits alignés au son grâce à une barre de lecture, ainsi qu'aux annotations, pour l'utilisateur linguiste non formé à l'annotation, l'apprenti linguiste ou le non linguiste.

Résultats

L'exploration a fait émerger 8 thématiques, répertoriées ci-après (figure 1).

Thème	Nombre d'extraits	Pourcentage
Conversations	15	10%
Loisirs	22	14%
Métiers	14	9%
Récits	24	15%
Savoirs	14	9%
Société	17	11%
Témoignages	25	16%
Vivre à Orléans	24	15%
Total	155	100%

figure . 1 Regroupement des extraits par thème

Les extraits sont également indexés selon un système de sous-spécification thématique (figure 2).

Thèmes	Sous-thèmes	Nombre d'extraits
Vivre à Orléans		**24**
	Équipements collectifs	4
	Fêtes orléanaises	5
	Les Orléanais	3
	Logement	4
	Modes de vie	5
	Sociolecte	3

figure . 2 Exemple des sous-spécifications thématiques du thème « Vivre à Orléans »

Après ce premier travail, une deuxième sélection de 76 extraits parmi les 150 de la sélection initiale a été effectuée en vue de constituer 5 modules pour la formation universitaire dans le cadre de cours de spécialité (linguistique interactionnelle, linguistique de corpus, sociolinguistique, didactique du FLE, civilisation). Visant une meilleure connaissance du français tel qu'il se parle, trois de ces modules permettent l'observation, la conceptualisation et la systématisation d'objets linguistiques caractérisant les interactions orales en français : « Module 1 : D'une situation à l'autre » (14 extraits ; variation diaphasique, liaison et constructions interrogatives), « Module 2 : Le parler jeune » (18 extraits ; variétés jeunes et usages des marqueurs discursifs, lexique familier, dispositifs syntaxiques spécifiques : dislocations), « Module 3 : Récits en interaction » (13 ; émergence des récits en interaction, discours rapporté et dispositifs syntaxiques spécifiques : clivées et présentatives complexes). Deux autres modules, à finalité socio-culturelle, « Voix variées » (11 extraits) et « Parole d'Orléanais » (20 extraits), rendent hommage au travail réalisé par les pionniers du corpus ESLO. « Parole d'Orléanais » allie thématiques contemporaines (fêtes de Jeanne d'Arc, sociolecte orléanais, centre-ville et quartiers, perception des Orléanais par eux-mêmes) et sujets historiques (mai 1968 et lutte des classes, Trente Glorieuses, pieds noirs d'Algérie), constituant une collection intéressante pour une exploitation dans un cours à thématique culturelle (Civilisation française, Culture et société, etc.).

Conclusion

Le projet ESLO-FLEU s'inscrit dans une démarche de transfert des recherches en linguistique et d'échanges avec le domaine de la didactique, au niveau universitaire et au-delà. Il s'agit ainsi de valoriser les corpus oraux au-delà de la communauté des linguistes, pour répondre à des orientations formulées en didactique des langues aujourd'hui (importance des compétences interactionnelles, exposition des apprenants à des conversation spontanées en contexte naturel). Prévus pour la formation universitaire des futurs enseignants de FLE dans le cadre de ce projet, les modules visant des objectifs linguistiques précis reposent sur un travail d'annotation des données issu des méthodes actuelles en linguistique. Mais cette ressource, conçue également pour un usage non expert, peut être utilisée par des non linguistes : dans la mesure où elle constitue une collection d'extraits sonores illustrant les usages du français parlé hier et aujourd'hui à partir de contenus socio-culturels divers, elle est susceptible d'intéresser les enseignants de FLE.

Références bibliographiques

André Virginie, 2016, FLEURON : Français Langue Étrangère Universitaire – Ressources et Outils Numériques. Origine, démarches et perspectives. *Mélanges Crapel 37*, p.69-92.

Baude, Olivier, 2015, *Observatologie : Vers une science de l'adéquation observationnelle en linguistique*. Linguistique. Thèse d'habilitation, Université Paris Ouest Nanterre la défense.

Baude, Olivier & Guerin, Emmanuelle, 2014, Pourquoi et comment dresser le portrait sonore d'une "grande ville"? L'exemple d'ESLO2. *Les métropoles francophones en temps de globalisation*, Juin 2014, Nanterre, France.

Blanc, Michel & Biggs, Patricia, 1971, L'enquête socio-linguistique sur le français parlé à Orléans. Le français dans le monde 85: 16-25.

Blanche-Benveniste, Claire, 2010, *Approches de la langue parlée en français*. Paris : Ophrys.

Blanche-Benveniste, Claire & Martin, Philippe, 2010, *Le français. Usages de la langue parlée*. Leuven/Paris : Peeters.

Boulton, Alex & Tyne, Henry, 2014, *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*, Paris : Didier.

Detey, Sylvain & Durand, Jacques & Laks, Bernard & Lyche, Chantal, 2011, *Les variétés du français parlé dans l'espace francophone. Ressources pour l'enseignement*. Paris : Ophrys.

Etienne Carole & Jouin Emilie, 2019, Constituer des ressources pédagogiques pour enseigner le français oral à partir des recherches menées en interaction. In Gajo L., Luscher J.-M., Racine I., Zay F. (eds), *Variation, plurilinguisme et évaluation en français langue étrangère*. Bern : Peter Lang, p.225-240.

Gadet, Françoise, 2003, *La variation sociale en français*. Paris : Ophrys.

Hamma, Badreddine, 2019, Quand l'interaction n'est pas là, la souris est mangée par le chat ! Remarques sur l'enseignement du passif en classe de français. In : *Linguistique interactionnelle, grammaire de l'oral et didactique du français*. Anne-Sophie Calinon, Badreddine Hamma, Katja Ploog et Marie Skrovec (éds.). Besançon : Presses Universitaires de Franche-Comté, 237-262.

Heiden, Serge & al., 2010, TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement, in *Actes du 10th International Conference on the Statistical Analysis of Textual Data*, Rome, Edizioni Universitarie di Lettere Economia Diritto 2, 1021-1032.

Koch, Peter & Oesterreicher, Wulf, 2001, Gesprochene Sprache und geschriebene Sprache / Langage parlé et langage écrit. In: G. Holtus, M. Metzeltin, Ch. Schmitt (éds), *Lexikon der Romanistischen Linguistik*, Bd. I/2, 584-627.

Lonergan J., Kay J. & Ross J. 1974, *Étude sociolinguistique sur Orléans, catalogue des enregistrements*. Colchester : Multigraphié.

Sinclair, John, 2004, *How to use corpora in language teaching*. Amsterdam/Philadelphia : John Benjamins.

Skrovec, Marie, 2019, « Grammaire et corpus oraux : réflexions pour un prolongement didactique de travaux réalisés sur ESLO », in : Calinon, A.-S., Hamma, B., Ploog, K. & Skrovec, M. (éds.) : *Linguistique interactionnelle, grammaire de l'oral et didactique du*

français, actes des Journées d'étude Linguistique et Didactique, Orléans, 23 mai et Besançon, 5 décembre 2014, Besançon : Presses Universitaires de Franche-Comté, 263-294.

Skrovec, Marie, 2020, « Le futur en français parlé, simple ou périphrastique ? Pour une adaptation des corpus oraux en didactique du FLE ». In : La didactisation du français vernaculaire, *Syntaxe et Sémantique*, 113-151.

Söll, Ludwig, 1974, *Gesprochenes und geschriebenes Französisch*. Berlin : E. Schmidt.

Surcouf Christian & Ausoni Alain, 2018, « Création d'un corpus de français parlé à des fins pédagogiques en FLE : la genèse du projet FLORALE », EDL (Études en didactique des langues), n°31, pp. 71-91.

Weber, Corinne, 2013, *Pour une didactique de l'oralité. Enseigner le français tel qu'il est parlé*, Paris : Didier.