HAL
open science

Nouvelles actions didactiques : faire de la sociolinguistique de corpus pour enseigner et apprendre à interagir en français langue étrangère

Virginie André

To cite this version:

Virginie André. Nouvelles actions didactiques : faire de la sociolinguistique de corpus pour enseigner et apprendre à interagir en français langue étrangère. Action Didactique, 2018, 1 (71-88). hal-01862713

HAL Id: hal-01862713

https://hal.science/hal-01862713

Submitted on 27 Aug 2018

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.

Action Didactique
Revue internationale de didactique du français
http://univ-bejaia.dz/action-didactique

Virginie ANDRÉ
Université de Lorraine, CNRS, ATILF

Nouvelles actions didactiques : faire de la sociolinguistique de corpus pour enseigner et apprendre à interagir en français langue étrangère

Article reçu le 15.01.2018 / Modifié le 26.05.2018 / Accepté le 01.06.2018

Résumé

L'exploitation de corpus à des fins d'enseignement et d'apprentissage des langues apparaît comme une méthodologie relativement récente dans le paysage de la didactique du français langue étrangère (FLE). Elle apporte avec elle de nouvelles actions didactiques qui semblent faciliter l'appropriation du français en permettant aux apprenants d'observer et d'analyser le fonctionnement des interactions. Les apprenants ont un accès direct aux données, ils peuvent ainsi comprendre les pratiques langagières qui permettent de réaliser des activités langagières dans différentes situations de communication. Cet article propose de mettre en évidence l'importance des liens entre (socio)linguistique et didactique et de saisir les applications de ces liens. Le dispositif d'apprentissage FLEURON illustre cette nouvelle méthodologie d'apprentissage sur corpus et les actions didactiques qui sont associées, notamment une sociolinguistique de corpus réalisée par les apprenants.

Mots-clés : Apprentissage sur corpus, linguistique de corpus, sociolinguistique des interactions verbales, Français Langue Etrangère, corpus multimodaux

Abstract

The exploitation of corpus for language teaching and learning appears to be a new approach in the field of French as a Foreign Language (FLE). It brings with it new didactic actions which seem to facilitate the appropriation of French by allowing learners to observe and analyze the functioning of interactions. Learners have direct access to the data, so they can understand language practices that enable them to carry out language activities in different communication situations. This article proposes to show the importance of the links between (socio)linguistics and didactics and to identify the applications of these links. The FLEURON learning system illustrates this new data-driven learning methodology and its new didactic actions, in particular a corpus sociolinguistics carried out by learners.

Key-words: Data driven learning, corpus linguistics, sociolinguistics of verbal interactions, French as a Foreign Language, Multimodal corpora

Pour citer cet article :

ANDRÉ, V. (2018). Nouvelles actions didactiques : faire de la sociolinguistique de corpus pour enseigner et apprendre à interagir en français langue étrangère. *Action Didactique*, 1, 71-88. http://univ-bejaia.dz/pdf/ad1/Andre.pdf

Pour citer le numéro :

Amokrane, S. et Cortier, C. (dir.). (2018). Oral et oralité: perspectives didactiques, anthropologiques ou littéraires [numéro thématique]. *Action Didactique*, 1. http://univ-bejaia.dz/ad1

Introduction

Toutes les recherches actuelles en didactique des langues s'accordent pour soutenir que l'exposition à la langue cible pour un apprenant est indispensable. Cependant, l'exposition au français parlé en interaction n'est pas toujours aisée. Les documents authentiques représentatifs de toutes les situations de communication auxquelles les apprenants sont susceptibles de participer ne sont pas toujours faciles à trouver ou à recueillir. De plus, il ne suffit pas d'avoir les documents, il faut également réussir à les analyser et à les exploiter pour que les apprenants acquièrent des compétences sociolangagières. C'est pourquoi les liens entre sociolinguistique, analyses de corpus et didactique méritent d'être renforcés. Les études sur corpus du fonctionnement de la langue, de ses pratiques et de ses usages, en situation permettent à la didactique d'orienter ses objectifs d'apprentissage et ses méthodologies.

Dans cet article, nous montrerons qu'il est possible d'exploiter des corpus oraux et multimodaux, y compris ceux créés par et/ou pour des linguistes, à des fins didactiques. Nous exploiterons les analyses sociolinguistiques des interactions verbales pour aider la didactique à proposer des objectifs d'apprentissage appropriés. Nous montrerons également que les corpus de données authentiques peuvent servir de documents supports lors de séance de formation. En effet, il existe aujourd'hui différentes façons d'utiliser des ressources, c'est-à-dire d'exposer ou de s'exposer à la langue cible. Nous présenterons ici des outils et des activités pédagogiques qui peuvent venir compléter les pratiques d'enseignement et d'apprentissage de l'oral en français langue étrangère (FLE), en utilisant des corpus oraux ou multimodaux. Parmi les nouvelles actions didactiques possibles à partir des corpus, nous nous intéresserons à l'exploration outillée de corpus, avec un concordancier, et sur ses conséquences sur l'appropriation de la langue.

1. Des corpus d'interactions authentiques

Depuis les années 1970, l'utilisation de documents authentiques est fortement préconisée dans le cadre de l'enseignement et de l'apprentissage des langues. Aucune approche didactique ne remettra en cause cette pratique, même si, dans les faits, nous pouvons constater de nombreuses et fréquentes utilisations de documents fabriqués ou inventés, lors de séances de formation ou en salle de classe. Le document authentique est « produit à des fins autres que l'apprentissage d'une langue seconde » (Abé, Carton, Cembalo et Régent, 1979, p. 2). En d'autres termes, le document authentique en FLE est une ressource constituée d'un prélèvement de la réalité, c'est un « message élaboré par des francophones pour des francophones à des fins de communication réelle » (Cuq, 2003, p. 29). Avant d'être une ressource

utilisée à des fins didactiques, le document authentique est une réelle situation de communication dans laquelle échangent des locuteurs. Ces derniers interagissent tel qu'ils le feraient s'ils n'étaient pas enregistrés ou filmés afin de conserver les échanges pour s'en servir comme supports didactiques. L'objectif est de montrer aux apprenants les usages réels de la langue et de ses variations en fonction des situations de communication. Nous ne nous attarderons pas ici sur le caractère plus ou moins authentique des documents (voir notamment Widdowson, 1998 ; Chambers, 2009 ; Adami, 2009). Nous soulignerons simplement que ces documents prélevés à leur réalité pour être utilisés en formation ou sous la forme d'un corpus perdent nécessairement une partie de leur authenticité. Néanmoins, ces documents reflètent les pratiques langagières réelles des locuteurs de la langue cible. Ces dernières sont produites afin d'accomplir des activités langagières [1] dans des situations de communication particulières (André, 2015). L'exposition à ces données sociolangagières, autrement dit aux usages et à leur variation, semble indispensable dans le processus d'acquisition d'une langue, quel que soit le niveau de l'apprenant.

L'exposition à la langue, pour un apprenant, est donc un aspect fondamental dans la démarche d'apprentissage et « l'authenticité de données comme input est l'un des piliers de la démarche dominante en vigueur aujourd'hui » (Boulton et Tyne, 2014, p. 22). Cependant, comme nous l'avons déjà mentionné, il n'est pas rare de constater que des enseignants de FLE utilisent encore des dialogues construits, inventés, créés pour les manuels de langue. Ces dialogues sont éloignés de la réalité des pratiques langagières en usage et ne proposent pas une exposition à la langue cible aux apprenants. Ils proposent souvent des interactions qui ne sont pas naturelles, une langue qui n'est pas utilisée ou qui n'existe que dans les manuels de langue. Néanmoins, si l'utilisation de ce type de dialogues existe encore aujourd'hui alors que le plaidoyer pour l'utilisation des documents authentiques est fort depuis plusieurs décennies (Duda, Esch et Laurens, 1972), c'est notamment parce que ces derniers restent difficiles à trouver et à utiliser.

En effet, les enseignants de FLE se plaignent fréquemment du manque de ressources orales représentatives de situations de communication ordinaires ou adaptées aux objectifs des apprenants. Pour pallier ce manque, nous proposons d'utiliser des données recueillies et traitées par des linguistiques. Ces données sont constituées en corpus, défini en linguistique notamment par Sinclair (1996) ou Blanche-Benveniste (2000) comme une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques et extralinguistiques explicites pour servir d'échantillon

[1] Nous préférons « activités langagières » à « acte de parole ».

d'emplois déterminés d'une langue. C'est donc à cet échantillon de langue que nous proposons d'exposer les apprenants.

Parmi les corpus oraux et multimodaux pouvant être utilisés à des fins didactiques, nous pensons notamment à TCOF (http://www.cnrtl.fr/corpus/tcof/) constitué à Nancy, à CLAPI (http://clapi.ish-lyon.cnrs.fr/) constitué à Lyon, à ESLO (http://eslo.huma-num.fr/) constitué à Orléans ou encore au projet PFC (https://www.projet-pfc.net/). La plateforme ORTOLANG présente également une liste de ce type de corpus (https://www.ortolang.fr/). Ces corpus de français parlé ont été recueillis et traités pour accueillir des analyses (socio)linguistiques. Néanmoins, ils présentent des ressources précieuses pour l'enseignement et l'apprentissage du FLE dans la mesure où ils sont composés d'enregistrements (son et/ou vidéo) de situations de communications authentiques, ressources rares et recherchées en didactique.

2. Sociolinguistique des interactions verbales

L'utilisation de corpus de données authentiques à des fins didactiques nécessite de repenser la façon d'aborder les documents supports en séance de formation. Tout d'abord, il s'agit de savoir observer et analyser ces données afin de rendre possible l'appropriation de la langue. En effet, la didactique des langues bénéfice de l'avancée des progrès de la linguistique qui tente de se munir de corpus visant une représentativité du français parlé en interaction. L'objectif de la linguistique dite « de corpus » est de proposer une description fine des usages de la langue. La sociolinguistique des interactions verbales (André, 2015) propose une description des pratiques langagières qui permettent d'accomplir des activités langagières dans des situations de communication particulières. Plus précisément, cette approche cherche à analyser les pratiques langagières, en examinant leurs spécificités linguistiques, interactionnelles, pragmatiques et contextuelles. Ces pratiques permettent d'accomplir des activités langagières (c'est-à-dire des actions communicatives telles que saluer, s'excuser, demander, etc.). Celles-ci dépendent du genre de discours (conversation, réunion de travail, entretien, etc.) dans lequel elles sont produites. En résumé, cette approche exige de saisir les multiples interinfluences entre les productions ou pratiques langagières et leur contexte d'apparition.

Les analyses proposées, à partir de corpus d'interactions authentiques, sont indiscutablement utiles, voire indispensables, à la didactique des langues. Elles permettent d'apprendre et d'enseigner à interagir de façon appropriée. De plus, les compétences sociolangagières orales sont généralement moins travaillées que celles de l'écrit, justement parce qu'elles sont moins décrites.

Les analyses sociolinguistiques des interactions verbales mettent en lumière un certain nombre de points :

- il n'y a pas congruence entre une activité langagière et une pratique langagière (c'est-à-dire que la même activité peut être réalisée différemment) ;
- il y a un nombre, presque infini, de pratiques langagières différentes qui peuvent réaliser une même activité langagière ;
- une même pratique langagière (parfois un seul mot) peut accomplir plusieurs activités langagières (par exemple plaisanter et reprocher) ;
- une activité langagière peut être réalisée avec un seul mot, une interjection, voire un seul geste ;
- une activité langagière peut être réalisée conjointement par plusieurs locuteurs.

Ces résultats vont généralement à l'encontre de ce que présentent les grammaires traditionnelles et les manuels de FLE ou alors ne sont simplement pas mentionnés. Cependant, il nous paraît important de les prendre en compte en didactique des langues. Nous allons rapidement illustrer cette approche avec l'analyse de quelques exemples de pratiques langagières.

Par exemple, pour réaliser l'activité langagière « poser une question », différentes pratiques sont possibles. Les grammaires traditionnelles et les manuels de FLE présentent plusieurs de ces possibilités, généralement : inversion sujet-verbe, utilisation d'un mot interrogatif et ajout d'une intonation montante en fin d'énoncé. Cependant, l'analyse de corpus oraux montre qu'il existe d'autres possibilités, rarement voire jamais décrites (Debaisieux et Boulton (2007) puis Ravazzolo, Traverso, Jouin et Vigner (2015) évoquent déjà cet aspect). En analysant le corpus TCOF, nous pouvons remarquer d'autres façons de poser des questions dans des situations de communication particulières. Les exemples suivants présentent des énoncés « inachevés » qui peuvent réaliser cette activité langagière :

> *Exemple 1 : deux étudiants discutent de leurs études (Corpus TCOF – Conversation_ELI_13).*
> <|L2|> (…) bon attends je me concentre à faire ça parce que faut pas que je fasse de conneries
> <|L1|> c'est quoi
> <|L2|> un truc tout transférer la vidéo
> <|L1|> **sur euh un site ou**
> <|L2|> ben il faut que je transfère tout sur et sur le mac et sur une clé +
> <|L1|> ah ouais

Dans cet extrait, nous remarquons un phénomène assez fréquent en français parlé, L1 pose une question en produisant un énoncé inachevé, se terminant

par la conjonction « ou ». Il ne produit, en quelque sorte, que la première partie d'une question alternative, mais cette seule partie est suffisante d'un point de vue pragmatique. L2 comprend bien que son interlocuteur lui pose une question et répond immédiatement (*et sur le mac et sur une clé*). On constate le même procédé dans l'exemple suivant :

> *Exemple 2 : un étudiant raconte son séjour universitaire en Allemagne (Corpus TCOF - InterSDL_ALE_12).*
> <|L1|> **et les cours tu les fais ici a la fac ou**
> <|L3|> non c'était c'était en Allemagne à la fac on a deux heures obligatoires en licence mais c'est en dessous de mon niveau

Cette interaction se déroule en France, L1 interroge L2 sur des cours que ce dernier a suivis. Comme dans l'exemple précédent, une seule partie de l'alternative (*les cours tu les fais ici*) est évoquée et le locuteur ne semble pas avoir l'intention de produire la seconde partie. L'énoncé est inachevé d'un point de vue syntaxique mais achevé d'un point de vue pragmatique, il permet de poser une question. Ainsi, alors que cette façon de poser des questions est fréquente, elle n'est généralement jamais apprise par des apprenants de FLE et généralement jamais mentionnée par les enseignants. Un autre procédé, qui pourrait être utile pour des apprenants de langue, n'est également jamais enseigné. Il s'agit d'un procédé similaire mais qui consiste à produire une hésitation plutôt que la fin d'un énoncé. C'est le cas dans l'exemple suivant :

> *Exemple 3 : L1 raconte son mariage (Corpus TCOF – Mariage_ELO_14)*[2].
> <|L1|> d'accord et euh si vous auriez quelque chose à à changer dans cet événement **euh** +
> <|L3|> après si j'avais quelque < chose
> <|L1|> de mieux >
> <|L3|> à changer c'est que j'aimerais que mes parents justement euh soient (…)

Dans cet extrait, un énoncé syntaxiquement inachevé, suivi d'une hésitation (*euh*) et d'une pause (*+*) suffisent pour que L1 pose une question et pour que L3 comprenne qu'une réponse est attendue.

Ces énoncés inachevés produisent les effets pragmatiques escomptés, c'est-à-dire, pour les extraits analysés, qu'ils permettent aux locuteurs de poser une question et d'obtenir une réponse. Ces analyses mettent en lumière les spécificités des interactions orales et montrent que les productions langagières ne sont pas canoniques et peuvent s'éloigner de la norme. Pour des apprenants de FLE, ces façons de dire et d'agir peuvent être utilisées comme des stratégies de communication. Quel que soit le niveau de langue d'un apprenant, il peut observer et reproduire ce type d'énoncé. Ainsi, s'il ne

[2] Les conventions de transcription sont celles du projet TCOF : le + signale une pause et les chevrons <…> délimitent des paroles superposées.

sait pas produire la fin d'un énoncé ou l'alternative d'une question en deux parties, il peut tout de même essayer d'accomplir l'activité langagière visée avec ces procédés d'inachèvement.

D'autres analyses, réalisées avec les outils de la linguistique de corpus, permettent également de mettre en évidence des pratiques langagières souvent ignorées dans le cadre de l'apprentissage du français. Par exemple, une caractéristique du français parlé en interaction est l'importance des petits mots de l'oral que l'on peut également appeler marqueurs, ponctuants ou particules discursives. Les analyses sociolinguistiques des interactions verbales montrent à quel moment les locuteurs les utilisent, dans quels objectifs et dans quelles situations. Ces éléments peuvent remplir des objectifs interactionnels, comme c'est le cas dans l'exemple suivant :

> *Exemple 4 : deux étudiantes discutent du stage qu'elles doivent effectuer pendant leurs études (Corpus TCOF – Stage_SON_12).*
> <|L1|> **bon alors du coup** on fait quoi
> <|L2|> **alors** soit j'appelle ma mère je lui demande si elle veut bien signer
> <|L1|> soit +
> <|L2|> **et** elle me dira oui de toute façon et après on va les voir on leur dit que euh elle est d'accord parce qu'elle veut pas s'embarrasser avec le les histoires du rectorat et tout

Cet extrait illustre bien la fréquence des mots de l'oral, notamment pour commencer un tour de parole. Il est intéressant de noter que certains peuvent également être utilisés à l'écrit (c'est le cas de *alors* ici) mais avec des significations différentes. Les analyses outillées de corpus permettent de saisir le fonctionnement de ces éléments. En effet, la recherche d'occurrences par concordancier permet de décrire les usages les plus fréquents de ces marqueurs. Ces usages doivent être analysés à la lumière des métadonnées des corpus, c'est-à-dire des éléments de la situation de communication. En d'autres termes, c'est en analysant un grand nombre d'occurrences de ces éléments qu'il est possible de saisir leurs valeurs et leurs effets pragmatiques et interactionnels. Ces analyses méritent de trouver des applications en didactique du FLE voire d'entrer dans la constitution d'une méthodologie possible, accompagnée ou médiée par un enseignant, en séance de formation. Quel que soit le niveau de l'apprenant, celui-ci peut travailler avec des corpus, quelle que soit leur taille et la nature de l'échantillon de langue qu'ils représentent. Néanmoins, la façon de travailler avec ces corpus, c'est-à-dire les activités d'apprentissage mises en œuvre, ainsi que l'accompagnement pédagogique de l'enseignant vont alléger la charge cognitive que peut représenter un apprentissage sur corpus.

3. Corpus et apprentissage du français parlé en interaction

L'apprentissage de la langue grâce à l'observation et l'exploitation de données linguistiques (*data-driven learning*), notamment écrites, est bien exposé par Johns (1991) ou Aston (2001). L'approche sur corpus en didactique (Boulton et Tyne, 2014) s'appuie sur la possibilité pour un apprenant d'accéder et d'observer un grand échantillon de données langagières. Plusieurs types de corpus peuvent être utilisés ou constitués à des fins didactiques. Nous avons déjà mentionné les corpus constitués par et pour les linguistes mais d'autres types de corpus peuvent coexister en didactique des langues. En outre, plusieurs façons d'aborder et d'exploiter ces corpus à des fins d'apprentissage peuvent également coexister. Nous présenterons, dans un premier temps, un usage assez classique de l'utilisation des données authentiques et, dans un deuxième temps, l'exploration sociolinguistique des corpus avec un concordancier, une nouvelle pratique en didactique des langues.

Tout d'abord, à côté des grands corpus de linguistes (mentionnés dans le premier point et sur lesquels nous reviendrons) de plus petits corpus peuvent être utilisés en didactique. Par exemple, des sous-corpus de linguistes peuvent être constitués afin de réduire l'exposition des apprenants à une trop grande masse de données. Ainsi, il est possible, par exemple dans TCOF, de ne sélectionner que les conversations entre amis ou que les réunions de travail entre collègues afin de proposer aux apprenants des situations de communication qui vont correspondre à leurs objectifs d'apprentissage. Dans CLAPI, un sous-corpus d'interactions dans les commerces peut représenter une ressource intéressante pour les apprenants qui souhaitent venir s'installer en France. Par ailleurs, ces sous-corpus peuvent être didactisés. C'est le travail entrepris à Lyon avec CLAPI-FLE (http://clapi.ish-lyon.cnrs.fr/FLE/accueil.php). A partir du grand corpus d'interactions CLAPI, certaines ressources sont sélectionnées et des exploitations didactiques sont proposées (voir dans ce même numéro).

Ensuite, les enseignants peuvent constitués des corpus spécifiques. Dans ce cas, le corpus peut rassembler un nombre plus restreint de données langagières mais représentatives des objectifs pragmatiques visés voire de points particuliers de la langue. Certains enseignants utilisent le terme de *corpus* pour désigner un ensemble de données représentatives, par exemple, des demandes d'informations, des explications d'itinéraires, des présentations, des salutations, des remerciements, etc. La constitution de ce type de corpus est réalisée par les enseignants lorsque les données authentiques manquent. Par exemple, Janot (2017) a constitué un corpus multimédia de locuteurs donnant des informations (horaires, tarifs, visites, etc.) sur des lieux culturels (musées, bibliothèques, médiathèques, cinéma,

etc.). Ce corpus est né d'une demande d'apprenants. Lors d'une formation linguistique pour des migrants, ces derniers ont souhaité apprendre à demander et à apprendre à comprendre ce type d'informations afin de pouvoir se rendre dans ces lieux de leur ville. L'expérimentation a été concluante, les apprenants ont acquis les compétences sociolangagières visées. En outre, il nous paraît intéressant de noter que ce corpus constitué dans la ville des apprenants a davantage suscité d'intérêts pour les apprenants que le même type d'interactions ayant lieu dans des endroits où ils n'ont pas l'intention de se rendre.

Enfin, certains corpus peuvent présenter des spécificités en fonction du domaine, du public ou d'un objectif d'apprentissage. La spécification en domaines peut s'apparenter aux sous-corpus mentionnés précédemment mais elle peut également correspondre à des corpus de français universitaire ou de français des métiers de la propreté, du bâtiment ou encore du tourisme. Ces corpus par domaine sont des grands corpus mais représentatif d'un secteur d'activités particulier. Il s'agit de recueillir un ensemble de données représentant des situations de communication susceptibles d'être rencontrées par les apprenants évoluant ou se préparant à évoluer dans ce secteur. Les formations à visée professionnelle (Mourlhon-Dallies, 2008) peuvent utiliser ce type de corpus. Dans la dernière partie de cet article, nous montrerons de quelle façon un corpus de situations de communication universitaires peut être utilisé à des fins d'enseignement et d'apprentissage.

La didactique du FLE se spécialise en fonction des objectifs, des styles d'apprentissage, des contextes de formation mais également en fonction des publics. Les méthodologies s'adaptent aux apprenants. La formation linguistique des migrants est un bon exemple. La situation d'apprentissage du français par les migrants est tout à fait particulière. Ces derniers doivent réussir à s'intégrer dans la société d'accueil en participant aux interactions quotidiennes et en effectuant les démarches nécessaires à leur vie en France. Le français langue d'intégration (FLI) vise donc l'autonomisation des apprenants. Les formations en FLI s'appuient sur des corpus (Adami et André, 2013) représentant l'ensemble des réalités sociolangagières auxquelles les migrants sont susceptibles de participer pour organiser leur vie en France. Ces corpus sont rares, ils sont donc souvent constitués par des formateurs ou par des concepteurs pédagogiques. Ils proposent par exemples, des interactions : chez le médecin, à la Caisse d'Allocations Familiales, à la mairie, à la poste, à la banque, à l'école, à la préfecture, à la gare, à la pharmacie, etc.

Ainsi, plusieurs types de corpus peuvent être utilisés à des fins didactiques. Que les corpus soient de très grands corpus de français parlé tentant d'être

représentatifs de tous locuteurs du français dans un maximum de situations de communication ou qu'ils soient représentatifs d'une activité langagière, ils permettent une alternative à l'utilisation des manuels de langue. Il existe notamment deux façons de les exploiter à des fins d'enseignement et d'apprentissage du français. La première convoque les connaissances issues de l'analyse sociolinguistique des interactions verbales. Elle observe les interactions en les analysant dans leur séquentialité. En d'autres termes, il est possible d'utiliser chacune des ressources d'un corpus comme un document authentique unique dont l'analyse montrera à l'apprenant le fonctionnement d'une interaction dans son entièreté. Il peut revenir à l'enseignant d'accompagner l'apprenant dans sa découverte de ce fonctionnement et des pratiques langagières mises en œuvre selon les objectifs communicatifs et la situation de communication. Par exemple, il est possible de pointer les pratiques langagières permettant de poser des questions, comme nous l'avons montré dans le point précédent.

La seconde façon d'exploiter ces corpus est d'utiliser les outils de la linguistique de corpus, et notamment les concordanciers. En effet, en cherchant les occurrences de certaines pratiques langagières ou de certains mots, les apprenants peuvent observer et comprendre le fonctionnement de la langue. Par exemple, lors de séances de formation, des apprenants ont déclaré ne pas comprendre l'apparition fréquente de « du coup » dans les interactions auxquelles ils participent. Nous avons interrogé le corpus TCOF avec les apprenants pour les aider à comprendre. Voici un extrait du résultat de la recherche :

470	Ang_jul_07	je préfèrais mh mh et	du coup	elle s'est sacrifiée elle
471	Ang_jul_07	ah oui d'accord et	du coup	euh et euh il y
472	Assemblee_di...	le blocage et et et	du coup	ça ça fournit un peu
473	Assemblee_m...	euh oui donc	du coup	encore une fois je suis
474	Automobile_gu...	sûre que si tu veux	du coup	il y a plus de
475	Boi_m1_09	même temps ben alors du	du coup	on en parle ça c'
476	Boi_m1_09	matos > et comme ça	du coup	on lui dit euh ben
477	Boi_m1_09	que tu étais passé et	du coup	ben oui ça l'a
478	Cha_hey_07	vois ouais et là euh	du coup	c'était une journée très
479	Comb_bach_08	euh leur enfant et donc	du coup	après nous avec les bénévoles
480	Comb_bach_08	alors en fait c'est	du coup	le l'enfant euh parce
481	Comb_bach_08	que > l'enfant lui	du coup	parce s'il y a
482	Comb_bach_08	plus vieux que lui donc	du coup	il arrive plus facilement à
483	Comb_bach_08	se projeter voilà et donc	du coup	de voir qu'un é–
484	Comb_bach_08	accessible quoi en fait donc	du coup	les enfants vu qu'ils
485	Comb_bach_08	ben l'étudiant il arrête	du coup	après les parents ils se
486	Comb_bach_08	comme ça quoi et donc	du coup	ils font même plus l'
487	Comb_bach_08	bien sûr > parce que	du coup	le jeune il se dit

Figure 1 : Extrait de la liste des résultats de la recherche de « du coup » dans TCOF avec le concordancier JConc

L'apprenant peut observer et analyser l'apparition des occurrences, leur emplacement dans les interactions, leur rôle et leur sens. En cas de besoin,

l'apprenant peut recourir à un contexte plus large et écouter l'enregistrement. Le corpus TOCF étant aligné texte-son, si un apprenant clique sur une occurrence, une fenêtre s'ouvre avec un contexte plus large. La figure ci-dessous montre la transcription correspondant à la deuxième occurrence de « du coup » dans la liste précédente :

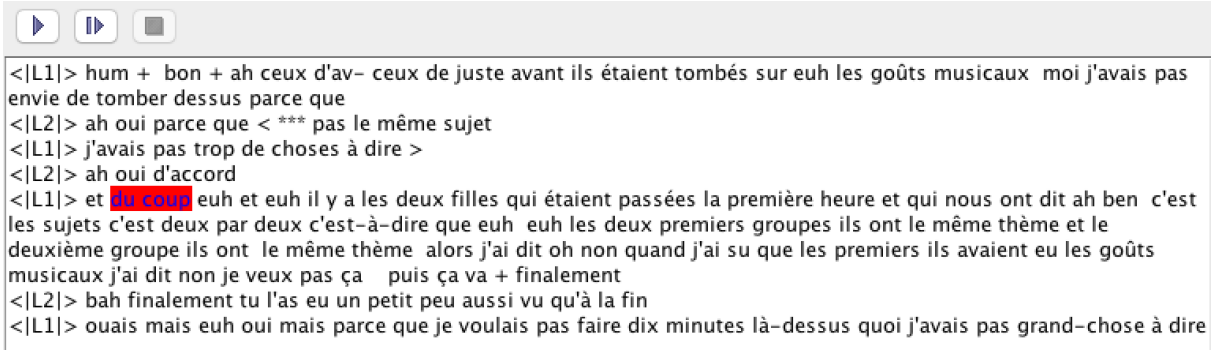

Figure 2 : Extrait de la transcription contenant « du coup » avec possibilité d'écouter l'enregistrement

Ainsi, grâce à la présentation en liste des occurrences et à l'accès au corpus (transcription et son), les apprenants réussissent à comprendre le fonctionnement de « du coup » ainsi que son sens et sa valeur pragmatique en interaction. Cette compréhension n'implique pas nécessairement celle de l'énoncé tout entier ou celle de l'interaction entre les locuteurs. Néanmoins, l'observation de plusieurs occurrences de « du coup » permet de saisir le rôle interactionnel de ce marqueur. Le travail réalisé à partir des résultats du concordancier peut être modulable selon le niveau des apprenants. Par exemple, les débutants peuvent repérer les mécanismes de prise de parole et les plus avancés pourront aller jusqu'à l'analyse des arguments introduits par le marqueur.

L'accès à de grands corpus modifie l'accès aux compétences sociolangagières par les apprenants. Ils observent, comparent, analysent le fonctionnement de la langue en cotexte et en contexte. Ils déduisent eux-mêmes les règles, les usages et les variations sociolinguistiques des pratiques langagières. Ces nouvelles actions didactiques facilitent l'appropriation de la langue. La possibilité d'avoir recours immédiatement aux données permet de répondre aux interrogations des apprenants. L'apprentissage se fait par recherches d'occurrences dont les résultats conduisent à de nouvelles recherches. L'apprentissage se fait donc par détours, par hypothèses, par tâtonnement, par essais-erreurs mais également par choix. En effet, certains apprenants choisissent d'acquérir certaines pratiques langagières et pas d'autres ou encore une variété particulière, par exemple celle des jeunes, celle des individus au travail, celle des commerçants ou encore celle des étudiants. Cette action didactique, qui consiste à faire ou à faire faire de la

(socio)linguistique de corpus à des apprenants de FLE, n'est pas liée à leur niveau de langue. Tous les apprenants peuvent utiliser l'observation de données en corpus à des fins d'apprentissage. Ce sont plutôt les pratiques langagières recherchées dans le concordancier qui vont varier en fonction du niveau des apprenants ainsi que le degré d'accompagnement de la part de l'enseignant.

4. L'exemple de FLEURON (Français Langue Etrangère Universitaire – Ressources et Outils Numériques).

FLEURON est un projet de recherche qui se situe à l'interface entre sociolinguistique et didactique. Il se concrétise par un site Internet, libre et gratuit, proposant un dispositif d'apprentissage du français (https://apps.atilf.fr/fleuron/). Ce dispositif est destiné aux étudiants qui souhaitent faire un séjour universitaire en France. Le cœur de FLEURON est constitué d'un corpus multimédia de situations de communication que les futurs étudiants sont susceptibles de rencontrer dès leur arrivée en France, et lors de l'organisation de leur vie étudiante, ainsi que d'un concordancier qui permet d'interroger ce corpus (André, 2016). Le corpus FLEURON appartient à un domaine particulier, celui dans lequel évoluent les étudiants. Ce corpus a été constitué, et ne cesse d'être enrichi, afin d'être un échantillon représentatif des interactions auxquelles les étudiants participent. Il est composé d'interactions authentiques entre des étudiants et :

- des personnels administratifs de l'université : scolarité, secrétariat, bibliothèque, santé, maison de l'étudiant, appariteur, service des relations internationales, centre de langue, service d'aide aux logements, service d'assistance sociale et médicale, service informatique, associations étudiantes, etc. ;
- des enseignants ;
- des personnels de services non universitaires mais indispensables à la vie étudiante : préfecture, transports en commun, lieux de culture, etc. ;
- d'autres étudiants, étrangers, qui témoignent de leurs problèmes, surprises ou découvertes.

Les interactions sont de trois natures différentes :

- des situations de communication écologiques, c'est-à-dire qu'elles ont eu lieu en milieu naturel et qu'elles auraient eu lieu même si nous ne les avions pas filmées ;
- des situations de communication sollicitées pour obtenir des explications, elles auraient pu exister sans notre intervention ;
- des témoignages d'étudiants étrangers.

Le corpus est interrogeable de différentes façons. Les ressources multimédia sont accessibles directement et sont classées en différentes catégories : procédure d'inscription, obtenir un document officiel, choisir ses cours, explications du système universitaire, comprendre la vie étudiante, utiliser les transports, témoignages, etc. Toutes les situations présentées par le site relèvent de la vie étudiante, sur un campus et en dehors d'un campus mais utiles à un étudiant (ouvrir un compte à la banque, louer un vélo, etc.). Si un apprenant choisit d'apprendre à interagir avec le personnel de l'administration, il peut visionner une ressource correspondante. Chaque ressource est décrite par son titre et par un court résumé. L'apprenant pourra alors également avoir accès aux sous-titres de la ressource et à la transcription, comme nous pouvons le voir ci-dessous :

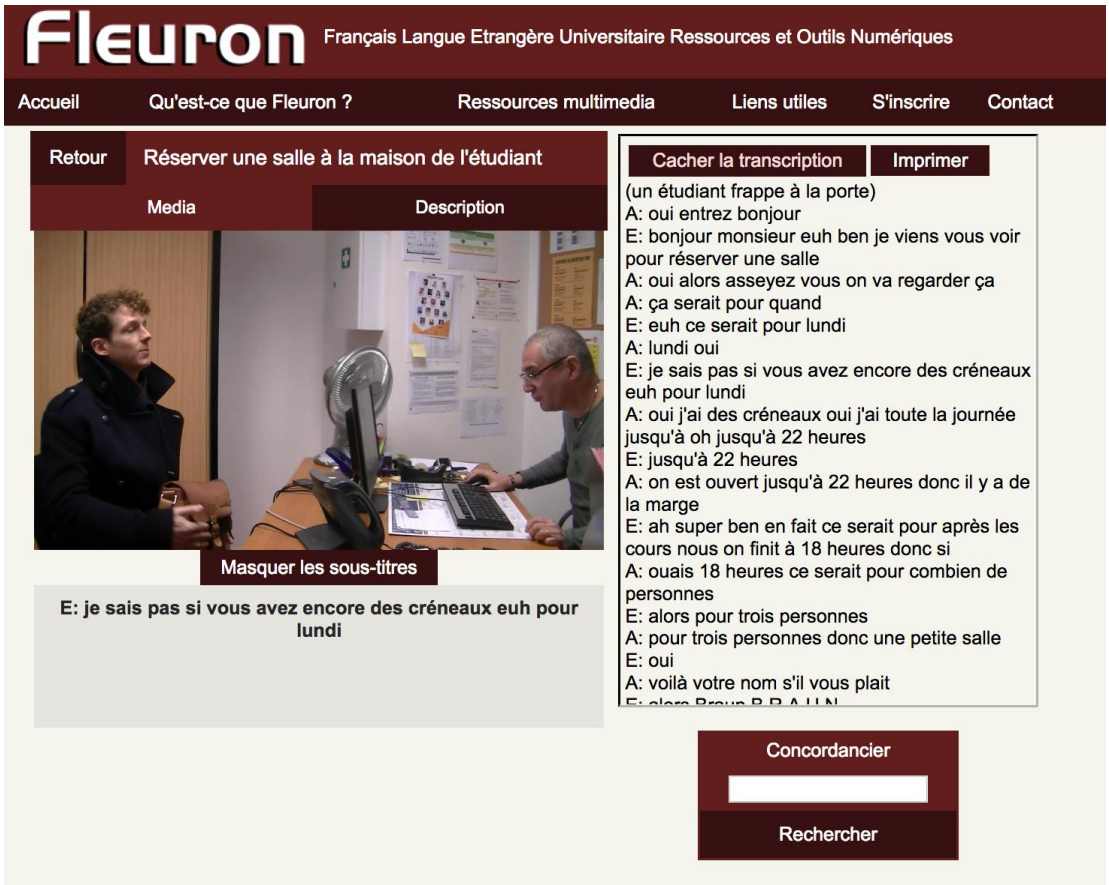

Figure 3 : Une ressource multimédia du site FLEURON avec transcription et sous-titres

L'apprenant peut choisir, en fonction de ses stratégies d'apprentissage, d'écoute et de compréhension, d'afficher ou pas les sous-titres et la transcription. Il peut analyser l'interaction et son déroulement, dès son ouverture jusqu'à sa clôture.

L'originalité de ce dispositif est que l'ensemble des ressources est également interrogeable par un concordancier simple, créé spécifiquement pour le site (voir en bas à droite de la figure 3). FLEURON est le seul dispositif d'apprentissage du FLE qui permette d'interroger un grand corpus multimédia d'interactions authentiques. Les apprenants peuvent saisir une chaine de caractères dans le concordancier et obtenir toutes les actualisations de cette chaine dans les interactions. Lors des expérimentations menées jusqu'à présent, en France ou à l'étranger, nous constatons que les apprenants recherchent fréquemment des mots de l'oral tels que *bon*, *ben*, *du coup*, *donc*, *alors*, *enfin*, etc. Le concordancier permet d'observer leur fonctionnement. La figure suivante présente un extrait de la liste des résultats de la recherche de « ben » :

A gauche du mot	Mot recherché	A droite du mot
non	ben	c'est
	ben	écoutez c'est parfait
bien	ben	vous pouvez aller chercher votre carte
donc vous euh	ben	vous irez une semaine hein
et puis euh bon	ben	pour l'instant vous pouvez s- donc suivre les cours de français et puis
	ben	alors ce qu'il y a c'est que le bureau en face est fermé
	ben	ils m'ont dit que c'était seulement un chèque
voilà	ben	c'est bon
	ben	je vous remercie
	ben	pour le remettre c'est aussi simple en fait
	ben	souvent voilà ça c'est le... L1 : c'est là où ça coince hein parce que
	ben	d'accord je vous remercie
et	ben	voilà
d'accord	ben	je vous remercie
et	ben	très bien
et	ben	je vous remercie bonne journée
t puis si vous avez d'autres questions	ben	n'hésitez pas comme d'habitude je te rends ta carte
ok ok d'accord	ben	merci beaucoup
oui oui oui évidemment c'est	ben	ça fait déjà cinq ans que je suis là euh
	ben	je dirais que c'est une expérience
e viens de visiter l'appart du dessus et du coup	ben	c'est bon il est pour moi donc je viens me présenter je suis la nouvelle voisine

Figure 4 : Extrait de la liste des résultats de la recherche de « ben » dans le corpus FLEURON

De plus, l'apprenant a la possibilité de visionner les ressources correspondant aux occurrences qu'ils souhaitent écouter. En cliquant sur l'occurrence choisie, la vidéo s'ouvre au moment où l'occurrence est prononcée (quelques secondes avant) :

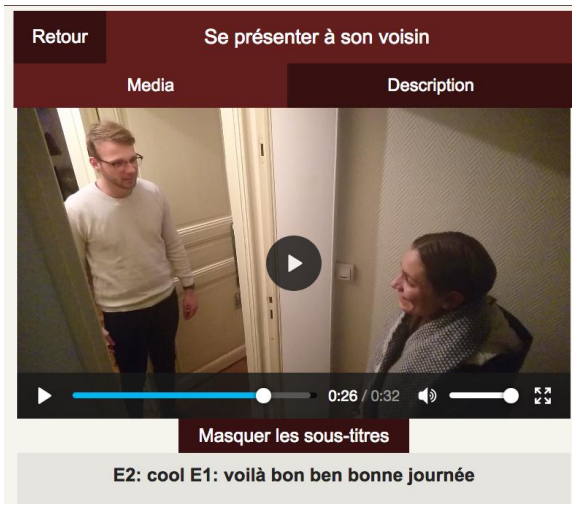

Figure 5 : Vidéo positionnée sur l'occurrence du « ben » choisi dans la liste des résultats

En exploitant les corpus multimodaux, soit ressource par ressource en analysant l'interaction dans sa séquentialité, soit avec une liste d'occurrences, les apprenants font de la sociolinguistique de corpus, sans le savoir. En effet, ils observent des pratiques langagières puis ils les analysent en tentant de saisir les multiples influences qui conditionnent leur apparition (locuteurs, lieux, objectifs communicatifs, règles interactionnelles, etc.). L'enseignant peut confirmer ou infirmer ces hypothèses, il peut en suggérer d'autres, expliquer le fonctionnement repéré, aider à l'analyser, mettre sur la voie, accompagner les démarches, etc. En procédant de cette façon, les apprenants, accompagnés d'enseignants formés aux liens entre sociolinguistique et didactique, acquièrent des compétences sociolangagières et apprennent à interagir de façon appropriée.

Conclusion

L'exploitation des corpus d'interactions verbales à des fins didactiques représente une alternative à l'utilisation des manuels de langue. Si elle voit le jour dès la fin des années 1980 dans des études anglo-saxonnes (Johns, 1991 ; Aston, 2001), notamment avec l'exploitation de corpus écrits, elle n'en est qu'à ses débuts en France. Elle est néanmoins en phase de devenir une approche particulière en proposant de nouvelles actions didactiques. Ces dernières se développent avec les avancées de la sociolinguistique des interactions verbales et de l'analyse outillée de corpus (André, 2017). En effet, des expérimentations associant linguistes et didacticiens sont mises place afin d'expliciter et de tester ces nouvelles actions didactiques. Nous pensons notamment aux travaux que nous avons déjà mentionnés autour du corpus TCOF et du site FLEURON mais également aux travaux lyonnais qui font des propositions didactiques à partir d'analyses interactionnelles (voir

par exemple à propos de l'expression « du coup », Bruxelles, Jouin-Chardon, Traverso et Guinamard, 2015).

L'exploitation de corpus à des fins d'enseignement et d'apprentissage permet de confronter les apprenants à des pratiques langagières réelles produites par des locuteurs qui accomplissent des actions dans différents contextes. Les corpus d'interactions authentiques donnent à voir ces données et permettent aux apprenants de se les approprier en les observant et en faisant leur analyse sociolinguistique. Ces nouvelles actions didactiques ne vont pas de soi. Si elles permettent aux apprenants d'apprendre à apprendre et d'apprendre en autonomie, elles nécessitent également le plus souvent un accompagnement de la part des enseignants. En outre, ces derniers doivent également être formés à ces pratiques professionnelles et avoir le temps de s'adonner à la linguistique de corpus, en termes de constitution de données et d'analyses. Qu'ils aient l'habitude d'utiliser des méthodes traditionnelles ou qu'ils soient encore en formation universitaire, la mise en œuvre de ces actions didactiques n'est pas simple mais elle a pour l'instant enthousiasmé tous les enseignants et les apprenants qui l'ont expérimentée, essentiellement avec le dispositif proposé par le site FLEURON.

Références bibliographiques

Abé, D., Carton, F., Cembalo, S. M., et Régent, O. (1979). Didactique et authentique: du document à la pédagogie. *Mélanges pédagogiques*, *10*, 1-14.

Adami, H. (2009). Les documents authentiques dans la formation des adultes migrants: pratiques pédagogiques et contraintes institutionnelles. *Mélanges CRAPEL*, *31*, 159-172.

Adami, H., et André, V. (2013). Corpus et apprentissage du Français Langue d'Intégration (FLI). *Linx. Revue des linguistes de l'université Paris X Nanterre*, 68-69, 135-158.

André, V. (2017). Un corpus multimédia pour apprendre à interagir en situations universitaires en France. Récupéré de *http://www.atpf-th.org/Actes_du_Colloque.pdf*.

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

André, V. (2015). Sociolinguistique des interactions verbales : de l'analyse des situations de travail aux implications sociales. *Langage, Travail et Formation*, 1-10.

Aston, G. (Ed.). (2001). *Learning with corpora*. Houston : Athelstan.

Blanche-Benveniste, C. (2000). *Approches de la langue parlée en français*. Paris : Editions Ophrys.

Boulton, A., et Tyne, H. (2014). *Des Documents Authentiques aux Corpus. Démarches pour l'Apprentissage des Langues*. Paris : Didier.

Bruxelles, S., Jouin-Chardon, E., Traverso, V. et Guinamard, I. (2015). « Du coup » dans l'interaction orale en français : description de ses usages situés à partir d'une base de données multimédia, et considérations didactiques. Dans Guinamard I., Jouin-Chardon E., Traverso V., Thai T. D. (dir.), *Langues parlées, interactions sociales. Une variété d'usages pour l'apprentissage du français* (p. 131-153). Paris : L'Harmattan.

Chambers, A. (2009). Les corpus oraux en français langue étrangère : authenticité et pédagogie. *Mélanges Crapel*, *31*, 15-33.

Cuq, J. P. (2003). *Dictionnaire de didactique du français*. Paris : CLE international.

Debaisieux, J. M., et Boulton, A. (2007). Alors la question c'est...? Questions pragmatiques et annotation pédagogique des corpus. *Cahiers de l'AFLS.*, *13*(2), 31-59.

Duda, R., Esch, E. et Laurens, J. P. (1972). Documents non didactiques et formation en langues. *Mélanges pédagogiques.* Récupéré de *http://www.atilf.fr/spip.php?article3528*.

Janot, S. (2017). *Constitution d'un corpus multimédia pour travailler la compréhension orale. Etude menée auprès d'un public migrant*. Mémoire de Master sous la direction de Virginie André, Université de Lorraine.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *Classroom Concordancing, English Language Research Jounral*, 4, 1-16.

Mourlhon-Dallies, F. (2008). *Enseigner une langue à des fins professionnelles*. Paris : Didier.

Ravazzolo, E., Traverso, V., Jouin, E., et Vigner, G. (2015). *Interactions, dialogues, conversations: l'oral en français langue étrangère*. Paris : Hachette.

Sinclair, J. (1996). Preliminary recommendations on corpus typology. *EAGLES Document TCWG-CTYP/P*. Récupéré de *http://www. ilc. pi. cnr. it/EAGLES/corpustyp/corpustyp. html*.

Widdowson, H. G. (1998). Context, community, and authentic language. *TESOL quarterly*, *32* (4), 705-716.

AUTEUR

Virginie ANDRÉ est enseignante-chercheuse en Sciences du langage à l'Université de Lorraine, à Nancy, et membre du laboratoire CNRS ATILF (https://apps.atilf.fr/homepages/VirginieAndre/). Elle est spécialiste de l'analyse sociolinguistique des interactions verbales et de ses applications en didactique des langues. Ses recherches se situent à l'interface entre sociolinguistique de corpus et didactique. Ses travaux s'inscrivent également dans le domaine « Langage, Travail et Formation », nom du groupe de recherche qu'elle co-anime ainsi que du réseau international dont elle est membre fondateur (https://apps.atilf.fr/reseaultf/).