



HAL
open science

Faire de la linguistique de corpus avec des apprenants de français langue étrangère

Virginie André

► To cite this version:

Virginie André. Faire de la linguistique de corpus avec des apprenants de français langue étrangère. Pierre Larrivée, Florence Lefeuve. La didactisation du français vernaculaire, Presses Universitaires de Caen, pp.37-66, 2020. hal-02993341

HAL Id: hal-02993341

<https://hal.science/hal-02993341>

Submitted on 6 Nov 2020

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.

Faire de la linguistique de corpus avec des apprenants de français langue étrangère

Virginie ANDRÉ

ATILF (Analyse et traitement informatique de la langue française – UMR 7118)

Université de Lorraine

1. Introduction

Faire de la linguistique de corpus pour enseigner et apprendre le français langue étrangère (FLE) signifie donner un accès direct aux données langagières à des apprenants pour qu'ils puissent les observer, les analyser et s'approprier la langue. En d'autres termes, les apprenants se laissent conduire par les données ou se prêtent au *data-driven learning* (Johns 1990). Dans le cadre de cette étude, nous aborderons l'apprentissage de l'oral à partir de corpus oraux et multimodaux. Lors d'une expérimentation visant à faire utiliser des corpus par des apprenants de FLE, l'un d'entre eux a déclaré : « C'est une façon sympa de faire de la grammaire ! ». C'est cette façon de faire de la grammaire et, plus généralement, de s'approprier la langue que nous allons définir, expliquer et illustrer dans cette contribution. Nous nous inscrivons à l'interface entre linguistique et didactique et, plus précisément, nous étudions de quelles façons il est possible d'exploiter les résultats de l'analyse des interactions verbales à des fins d'enseignement et d'apprentissage du FLE.

Dans un premier temps, nous examinerons la place des corpus en didactique des langues. Nous verrons que les enseignants et les apprenants se trouvent face à un réel manque de données authentiques, malgré les approches didactiques des quarante dernières années, la préconisation de la production et de l'utilisation de « documents authentiques » (Holec 1990) et les technologies numériques actuelles. Ce manque de données nous conduit à orienter les enseignants et les apprenants de FLE vers l'utilisation de corpus constitués à des fins linguistiques et non didactiques. Cependant, leur exploitation à des fins didactiques n'est ni évidente ni aisée. Nous verrons de quelles façons les outils de la linguistique de corpus peuvent être détournés pour faciliter l'entrée dans les données par les apprenants qui deviennent rapidement des apprentis chercheurs en linguistique. Dans un deuxième temps, nous donnerons des exemples d'utilisations de corpus par des apprenants. Ces exemples sont extraits de séquences d'expérimentation avec des apprenants étrangers en France suivant des cours de FLE à l'université.

Nous présenterons trois corpus différents utilisés et utilisables à la fois par les enseignants et les apprenants. Enfin, dans la dernière partie, nous présenterons une séquence pédagogique réalisée afin de didactiser les fiches descriptives du projet FRACOV¹. Nous avons choisi la fiche descriptive sur le *ne* de négation et construit une séquence, à partir de l'observation de corpus oraux et multimodaux, pour que les apprenants se sensibilisent à la variation syntaxique, sémantique et situationnelle.

2. Corpus et didactique

En France, l'introduction des corpus en didactique des langues apparaît comme une nouvelle méthodologie qui s'inscrit dans une approche particulière, à la fois de la langue et de la didactique (Hymes 1972; Johns 1990; Aston 2001; Boulton & Tyne 2014). Ainsi, la conception de la langue, notamment, en ce qui concerne notre étude, du français parlé en interaction (Blanche-Benveniste & Jeanjean 1987; Mondada 2002; Kerbrat-Orecchioni 2005), et de son apprentissage, auxquels s'ajoutent les progrès technologiques, font des corpus outillés de nouvelles ressources pour la didactique. Les corpus qui peuvent être utilisés à des fins didactiques sont de natures différentes, notamment parce qu'ils sont constitués soit par des linguistes soit, plus rarement, par des enseignants de français, soit par une collaboration des deux. Quels que soient les corpus utilisés ainsi que la façon de les utiliser, leur introduction dans une salle de classe ou lors d'un apprentissage plus informel reste encore aujourd'hui un défi pour la didactique des langues (Debaisieux 2009).

2.1. Vers les corpus oraux et multimodaux

Notre approche, qui vise à utiliser des corpus à des fins didactiques, s'inscrit dans le prolongement de l'utilisation des documents authentiques en formation linguistique (Holec 1990). Quels que soient les objectifs d'apprentissage fixés par les enseignants, par les programmes ou par les apprenants, il est impensable aujourd'hui de les atteindre sans être confronté à la langue cible par l'intermédiaire d'un document support authentique, c'est-à-dire non inventé ou construit à des fins didactiques. Cette démarche est prônée depuis les années 1970 et personne ne semble la remettre en question, même si elle n'est pas toujours mise en œuvre, notamment pour des questions liées à la difficulté d'obtenir

1. FRACOV : Français contemporain vernaculaire, <http://www.univ-paris3.fr/fracov-227156.kjsp>.

des documents authentiques. Nous pouvons rapidement rappeler que le document authentique est « produit à des fins autres que l'apprentissage d'une langue seconde » (Abé *et al.* 1979 : 2). Ainsi, 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 : 29)². L'utilisation des documents authentiques est liée au souhait d'exposer et de confronter les apprenants à la langue cible. Ces documents peuvent être écrits : des coupures de presse, des journaux spécialisés, des correspondances, des mails, des textos, des pages Web, des romans, des petites annonces, des documents administratifs, et tout type d'écrits ; ils peuvent être oraux : des émissions de radio, des documentaires, des films, des séries, et tout type d'enregistrements audio et vidéo. En fonction des objectifs d'apprentissage, le document authentique approprié sera choisi, par l'apprenant ou par l'enseignant.

Aujourd'hui, il semble facile d'être exposé à la langue, notamment avec les nouvelles technologies, les nouveaux médias et Internet. Les enseignants et les apprenants de FLE peuvent trouver de très nombreux documents, écrits, oraux, vidéo, en français sur le Web. Néanmoins, les méthodologies didactiques d'exploitation de ces documents tout-venant ne sont pas toujours aisées à mettre en œuvre. L'authenticité ne fait pas tout. Il faut savoir faire ou faire faire des activités d'appropriation de la langue pour que les documents entrent dans une démarche didactique. Les méthodologies associées à l'utilisation des documents authentiques proposent des pistes d'exploitation de ces documents et des activités pédagogiques possibles. Plusieurs pistes et méthodologies peuvent coexister. Ces dernières ont évolué au fil du temps, au fur et à mesure de l'évolution des disciplines qui s'y intéressent ou qui sont connexes. Nous pouvons mentionner les liens entre la conception de la langue et différentes méthodologies. L'approche communicative, par exemple, va naître d'une conception fonctionnelle de la langue, permettant aux locuteurs d'échanger, d'accomplir des actions et d'interagir (Widdowson 1978 ; Brumfit & Johnson 1979).

Les évolutions méthodologiques sont également liées aux évolutions technologiques. L'utilisation de corpus en didactique va être impulsée par les progrès en informatique ainsi que par les avancées en linguistique de corpus. Ainsi, depuis quelques années, le *data-driven learning*

2. Nous ne débattons pas ici du caractère plus ou moins authentique de certains documents. Ce débat a notamment été mené dans le numéro 31 de la revue *Mélanges CRAPEL* (2009).

(Johns 1991), ou l'apprentissage sur corpus (Boulton & Tyne 2014), apporte de nouvelles méthodologies didactiques. Les travaux anglo-saxons sont les plus avancés dans ce domaine, notamment en ce qui concerne l'utilisation de corpus écrits. Dans le cadre de notre étude, dans la mesure où nous nous intéressons à l'analyse des interactions verbales et à l'exploitation de ces analyses à des fins d'apprentissage de la langue, nous nous intéressons uniquement aux corpus oraux et multimodaux.

L'exploitation de corpus à des fins didactiques (André 2017) présente donc de nouvelles opportunités et de nouvelles perspectives pour aborder de nombreuses questions liées à l'appropriation de la langue. Tout d'abord, les résultats de la linguistique de corpus, et plus précisément de l'analyse sociolinguistique des interactions verbales, ont des conséquences sur l'enseignement et l'apprentissage du FLE. Comme nous l'avons déjà mentionné, la conception de la langue et de son fonctionnement a nécessairement une implication dans la façon de l'enseigner ou de l'apprendre. Il est important de réduire l'écart entre ce que nous apprennent les analyses de corpus sur le français parlé en interaction (Bilger & Cappeau 2013) et la façon dont la langue est présentée en formation (Sinclair 2004).

Avant d'examiner de quelle façon enseigner et apprendre avec des corpus, nous allons brièvement préciser ce que nous entendons ici par corpus. En didactique, un ensemble de documents utilisé par un apprenant peut être appelé corpus. Cependant, nous utiliserons la définition linguistique de corpus notamment fournie par Sinclair (1996: 4): « *a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language* ». Il existe plusieurs corpus oraux en français qui sont des échantillons plus ou moins grands de la langue, sélectionnés en fonction de critères extralinguistiques. Nous pouvons par exemple citer les corpus nancéen TCOF (Traitement de corpus oraux en français), lyonnais CLAPI (Corpus de langue parlée en interaction), orléanais ESLO1 et ESLO2 (Enquêtes sociolinguistiques à Orléans), parisien CFPP2000 (Discours sur la ville. Corpus de français parlé parisien des années 2000), neuchâtelois OFROM (Corpus oral de français de Suisse romande) ou encore le projet PFC (Phonologie du français contemporain) (voir les références de ces corpus en bibliographie). Les analyses faites à partir de ces corpus permettent de comprendre le fonctionnement de la langue parlée en interaction.

D'un côté, la taille et la représentativité des corpus de données orales augmentent assez rapidement depuis une dizaine d'années en France, notamment avec la poursuite de l'enrichissement des corpus précédemment cités. D'un autre côté, les documents authentiques

prônés par toutes les approches didactiques actuelles manquent. En effet, les enseignants de FLE se plaignent régulièrement du manque de données authentiques, notamment de situations quotidiennes, malgré la prolifération de documents en tout genre sur le Web. La rencontre entre les corpus et la didactique semble donc incontournable. Donner accès aux corpus paraît un moyen d'exposer les apprenants à la langue cible. Néanmoins, cet accès mérite une réflexion sur les pratiques pédagogiques qu'il implique. L'exposition à la langue, et à de très nombreuses données langagières, grâce aux corpus, ne suffit pas pour garantir l'apprentissage de la langue.

2.2. Que faire avec les corpus ?

L'apprentissage « conduit par les données » ou le *data-driven learning* notamment développé par Johns (1990, 1991) apparaît comme une nouvelle approche théorique et méthodologique en didactique des langues, notamment dans le monde anglo-saxon (Leńko-Szymańska & Boulton 2015). Cette approche vise à prendre comme point de départ à l'apprentissage d'une langue la confrontation directe avec des données langagières utilisées par les scripteurs ou les locuteurs de la langue cible. Les apprenants sont alors amenés à observer, analyser, comprendre le sens, découvrir les usages, la variation, en consultant des corpus. L'outil privilégié pour accéder aux données est le concordancier. Il permet d'interroger des formes du corpus et de donner à observer aux apprenants des occurrences listées verticalement (voir les figures 1 et 2 ci-après).

Si l'accès aux corpus permet une exposition presque illimitée à la langue cible, cette masse de données peut décontenancer le plus motivé des apprenants. Les grands corpus, tels que ceux que nous avons cités précédemment en exemple, sont interrogeables avec des outils d'exploration. Dans le cadre de cette étude, nous utiliserons des concordanciers pour consulter, sélectionner, trier les données. Les avancées en linguistique de corpus et en informatique permettent aujourd'hui d'explorer des corpus de façon assez fine. Néanmoins, l'apprenant de langue n'est pas *a priori* un chercheur en linguistique. Il nous faut donc transposer certaines méthodologies et certaines analyses de données pour les rendre accessibles aux apprenants.

Nous allons examiner ce qu'il est possible de faire avec des corpus et les outils d'interrogation qui leur sont associés. Nous mentionnerons tout d'abord quatre objectifs que les apprenants peuvent atteindre avec l'apprentissage sur corpus, quel que soit leur niveau. Nous montrerons, ensuite, des exemples qui illustrent ces objectifs.

2.2.1. Observer des pratiques langagières

Le premier objectif que des apprenants peuvent se fixer en exploitant des corpus est l'observation de pratiques langagières, c'est-à-dire des productions verbales réelles des locuteurs, leurs spécificités lexicales, syntaxiques, sémantiques, pragmatiques, interactionnelles et situationnelles. En effet, un apprenant peut observer les productions verbales des locuteurs et saisir leurs objectifs communicatifs. Par exemple, il est possible de repérer dans les corpus tous les énoncés qui permettent de saluer un interlocuteur, tous les énoncés qui servent à dire qu'on est d'accord, ceux qui permettent d'interrompre son interlocuteur, de raconter des faits passés, etc. Pour certains de ces exemples, nous comptons beaucoup sur les progrès en termes d'annotation de corpus, qu'elle soit (morpho)syntaxique, en partie du discours ou pragmatique.

Avec les corpus, les apprenants peuvent également vérifier des usages qu'ils maîtrisent peu. Par exemple, lorsqu'un apprenant souhaite observer s'il existe des différences entre *parce que* et *puisque*, il peut rechercher les occurrences de ces deux mots grâce à un concordancier. Les figures ci-contre montrent le résultat de ces recherches dans le corpus TCOF interrogé avec le concordancier JConc.

En fournissant un grand nombre d'exemples à un apprenant, il peut se comporter en apprenti analyste, apprenant-chercheur (Gavioli 2001), ou en détective, voire en Sherlock Holmes selon Johns (1997), et faire de la linguistique de corpus sans le savoir. En effet, nous avons constaté lors de nombreuses expérimentations³ que les apprenants de FLE, quel que soit leur niveau, se prêtent tout à fait facilement à l'analyse des listes de résultats qu'ils obtiennent à l'issue de leurs requêtes. Les apprenants accèdent plus facilement au sens en analysant les occurrences en contexte. De plus, comme les apprenants dégagent eux-mêmes les règles, syntaxiques notamment, des faits de langue qu'ils observent, ils les comprennent mieux et surtout se les approprient mieux. Les apprenants sont amenés (souvent guidés par l'enseignant, au moins au début) à suivre les trois étapes du *data-driven learning*: observation, classification des données et généralisation des principes de fonctionnement à partir des observations et des classifications effectuées.

3. Les expérimentations mentionnées dans cette contribution ont été menées au département de français langue étrangère (DéFLE) de Nancy (université de Lorraine). Elles ont été effectuées avec Florence Poncet, enseignante de FLE, directrice du DéFLE et membre de l'ATILF. Pour plus d'informations: <https://defle-nancy.jimdo.com>.

Concordancier JConc - ** nouvelle session

puisque Mot RegExp Min/Maj Nb occurrences : 26 / 26
Nb fichiers : 12 / 12
Nb locuteurs : 3 / 3

Ajouter... Supprimer Vidier

39_45_EVA_14.trs
Accident_CAT_14.trs
Admission_14.trs
Cancer_LIS_14.trs
Coree_GHU_14.trs
Dos_COU_14.trs
Forum_GOM_14.trs
Forum_GOM_14.trs
Forum_GOM_14.trs

General Contextes

Caract. sép. [t\|n|r /[*+...?0@!~<>\'`-#=#]

Encodage ISO-8859-1

Nb mots à gauche 5

Nb mots à droite 5

Restreindre aux locuteurs :

Select. Locuteur Couleur

Num	Fichier	Contexte Gauche	Occurrence	Contexte Droit
1	39_45_EVA_14	les chevaux étaient très	puisque	les voitures ne circulaient pas
2	39_45_EVA_14	et elle venait en vélo	puisque	c'était toujours pareil on
3	Accident_CAT_...	rupture d'accord et euh	puisque	euh comme on dit une
4	Accident_CAT_...	d'accord j'avais pas	puisque	j'avais je ne regardais
5	Accident_CAT_...	euh par voies terrestres mais	puisque	moi j'ai mon cas
6	Accident_CAT_...	disais euh l-donc euh	puisque	j'ai été en hélicoptère
7	Accident_CAT_...	pouvoir les consolider et puis	puisque	maintenant elles sont elles
8	Admission_14	de psychologie en l'occurrence	puisque	c'est mon domaine euh
9	Admission_14	acceptée j'aurais jamais lâché	puisque	c'est quelque chose qui
10	HOW_PRO_14	conventions et puis	puisque	j'ai donné des cours
11	HOW_PRO_14	chose euh > bah oui	puisque	mon ex conjoint euh était
12	Marriage_ELO_14	bah je leur ai dit	puisque	c'est comme ça bah
13	Nonno_CRZ_14	on savait pas on ignorait	puisque	d'accord donc euh vu
14	Nonno_CRZ_14	c'était pas un problème	puisque	les deux mois que j'
15	Nonno_CRZ_14	le train à l'aveuglette	puisque	hum hum > Metz c'
16	Nonno_CRZ_14	ai dit non eh bah	puisque	c'est comme ça qu'
17	Papi_MAR_14	eu envie de le fêter	puisque	euh je c'était trop
18	Paternel_GOS	réussis à apprendre parler	puisque	forcément n'aurait fait

<[L2]> la fin du commencement d'une autre aventure donc je pourrais dire que professionnellement parlant j'ai eu trois vies j'ai été analyste financier donc salariée hein d'un établissement

<[L1]> ouais

<[L2]> financier américain j'ai créé ma société de formation en langues pour intervenir donc en entreprise donc j'ai été et directrice pour tout ce qui était contrats conventions et puis formations

<[L1]> j'ai donné des cours par la suite et puis par hasard je suis rentrée à l'université pour un remplacement aussi à l'UT de d'Epinal et puis on m'a rappelé et depuis j'y suis restée

<[L1]> d'accord

<[L2]> voilà

<[L1]> donc niveau professionnel ça fait un peu un quatre en un quoi tu as eu quatre événements marquants < trois événements trois événements marquants

<[L2]> trois < trois événements marquants enfin je veux dire > professionnel et puis après la grande aventure euh des études de la recherche etcetera

Figure 1. Recherche de *puisque* dans TCOF via le concordancier JConc

parce que Mot RegExp Min/Maj Nb occurrences : 475 / 475
Nb fichiers : 25 / 25
Nb locuteurs : 3 / 3

Ajouter... Supprimer Vidier

39_45_EVA_14.trs
Accident_CAT_14.trs
Admission_14.trs
Cancer_LIS_14.trs
Coree_GHU_14.trs
Dos_COU_14.trs
Forum_GOM_14.trs
Forum_GOM_14.trs
Forum_GOM_14.trs

General Contextes

Caract. sép. [t\|n|r /[*+...?0@!~<>\'`-#=#]

Encodage ISO-8859-1

Nb mots à gauche 5

Nb mots à droite 5

Restreindre aux locuteurs :

Select. Locuteur Couleur

Num	Fichier	Contexte Gauche	Occurrence	Contexte Droit
1	39_45_EVA_14	euh exploitée par les	parce que	qu'ils ont réquisitionné notre
2	39_45_EVA_14	douze kilomètres de ses	parce que	ses parents étaient aussi
3	39_45_EVA_14	m'avait gardé avec elle	parce que	elle voulait pas rester
4	39_45_EVA_14	y avait plus de ponts	parce que	les allemands avaient fait
5	39_45_EVA_14	toute seule sur la barque	parce que	qu'elle avait son vélo
6	39_45_EVA_14	à plusieurs et donc quand	parce que	quand ma ma sœur
7	39_45_EVA_14	ma ma sœur était là	parce que	quelque fois ma mère
8	39_45_EVA_14	que je ne bouge pas	parce que	je les enfants ça
9	39_45_EVA_14	on a pas été blessées	parce que	bien sûr c'est
10	39_45_EVA_14	monsieur qui ferait les	parce que	qu'à ce moment là
11	39_45_EVA_14	le frère de mon père	parce que	mon père avait un
12	39_45_EVA_14	courais je courais je courais	parce que	je disais le père
13	39_45_EVA_14	guère arrangé avec le temps	parce que	il y a ça
14	39_45_EVA_14	a on a toujours euh	parce que	quand on quand on
15	39_45_EVA_14	on avait terminé le repas	parce que	qu'on était nombreux a
16	39_45_EVA_14	restes puis elle lui dit	parce que	qu'on avait un chien
17	39_45_EVA_14	lui il avait bien mangé	parce que	nous mes parents maman
18	39_45_EVA_14	faire le vivane Verdun-Nancv	narrer	que c'était des locomotives

dans le village ou ma mère euh travaillait euh avec avec les allemands dans sur la ferme et dans cette école là je parlais tous les jours à l'école à pied et je p- je faisais plus d'un kilomètre à pied et je c'est un village qui était sur une grande longueur et quand je traversais ce village là je suis saisi devant un marchand-ferrant c'était un monsieur qui ferait les chevaux parce qu'à ce moment là les chevaux étaient très importants puisque les voitures ne circulaient pas les chevaux nous de -> ça nous servait de moteur c'est eux qui nous tr- ils faisaient tous les chevaux alors donc ce monsieur là ferait les pieds des chevaux et il venait le soir quelquefois chez ma mère pour voir le frère de mon père

que mon père avait un frère qui avait dix-sept ans qui n'av- qui n'était pas parti à la guerre et il venait pour blaguer avec lui et puis il me disait tu sais tous les jours quand tu vas à l'école que tu passes devant chez moi devant mon mon devant son son atelier il avait un atelier et bien il y a le père Fouettard qui est chez moi comme par hasard et il te regarde ah

Figure 2. Recherche de *parce que* dans TCOF via le concordancier JConc

En outre, l'analyse sociolinguistique des interactions verbales a mis en lumière un certain nombre de pratiques langagières spécifiques, des points de vue linguistique et interactionnel. Ces spécificités sont notamment liées au déroulement temporel ou séquentiel des interactions ainsi qu'à leur coconstruction (voir par exemple Traverso 1999;

Mondada 2002 ; André 2014). Parmi ces spécificités, les apprenants peuvent observer l'apparition, le fonctionnement et le sens des marqueurs du discours ou des petits mots de l'oral, tels que *bon, ben, quoi, tu sais, voilà*, etc. L'interrogation des corpus présente un grand intérêt pour ces observations, notamment en compréhension orale d'abord, en expression ensuite (Bruxelles *et al.* 2015). En effet, les apprenants repèrent la place micro- ou macro-syntaxique des marqueurs, leurs sens en contexte, leur rôle pragmatique, leur fonction interactionnelle, leur fréquence et leur variation situationnelle. Au-delà de ces exemples précis, toutes les pratiques langagières sont observables en corpus. Les apprenants mettent en œuvre une démarche inductive au cours de laquelle ils effectuent un travail réflexif à partir des données des corpus.

2.2.2. Observer la variation

La question de la variation en didactique des langues est souvent problématique mais sa prise en compte est indispensable à la compétence sociolinguistique des apprenants (Tyne 2012). La place accordée à la variation dans les approches didactiques traditionnelles est assez mince. De plus, lorsque la variation est abordée, elle est fréquemment réduite à une opposition lexicale entre écrit et oral ou entre formel et informel. L'observation de pratiques langagières en contexte permet de sensibiliser les apprenants aux différents types de variations : diachronique, diaphasique, diastratique, diatopique (Conrad 2004 ; Gadet 2007).

Les corpus sont des ensembles à la fois de données primaires, c'est-à-dire des enregistrements, et de données secondaires qui accompagnent les premières (la transcription) et les décrivent (les métadonnées). Les métadonnées présentent les enregistrements et donnent des informations sur les locuteurs (sexe, âge, niveau d'études, lieu de naissance, etc.) et leurs relations (amicales, professionnelles, familiales, etc.), sur les conditions d'enregistrement (lieu et moment), sur le genre de discours auquel participent les interlocuteurs, etc. Ces informations sont indispensables pour comprendre les interactions verbales. Lors de l'observation des pratiques langagières produites par les locuteurs, les apprenants consultent les métadonnées pour examiner quelles sont les pratiques langagières appropriées à la situation de communication. Les figures 3 et 4 sont des extraits de la liste des métadonnées du corpus TCOF. La première concerne des informations générales sur l'enregistrement et la seconde présente des données sur un locuteur.

Un apprenant qui consulte l'enregistrement correspondant à ces métadonnées pourra ainsi s'apercevoir que les pratiques langagières observées sont produites par une femme assez âgée, peu scolarisée, lors d'un récit de vie non préparé.

Type de corpus	entre adultes
Modalités de recueil des données	Sollicitées par un chercheur
Canal de communication	Face à face
Cadre situationnel	Privé
Degré d'interactivité	Peu interactif
Genre de discours	Entretien
Typologie textuelle dominante	Narratif
Thème abordé / résumé	Une dame de 78 ans raconte les souvenirs qu'elle a de sa vie à l'époque de la seconde guerre mondiale.

Figure 3. Métadonnées générales d'un enregistrement dans TOCF

Étiquette	L2
Nombre de tours de parole	1
Age du locuteur	78
Sexe du locuteur	F
Niveau d'étude	Elémentaire
Profession(s) actuelle(s)	Retraitée
Profession(s) antérieure(s)	Agricultrice
Rôle dans l'interaction	interviewé
Degré de planification	Non planifié
Statut du français	Langue maternelle
Lieu de naissance	Verdun
Lieu de résidence (ville)	Verdun

Figure 4. Métadonnées pour un locuteur dans TCOF

L'observation des métadonnées permet aux apprenants de faire de la sociolinguistique de corpus. En effet, ils examinent des pratiques langagières en contexte, c'est-à-dire qu'ils effectuent un lien entre certaines pratiques et certains éléments de la situation de communication. La grande diversité des corpus, en termes de locuteurs, d'activités langagières, de genre de discours, et plus généralement de situations de communication (Hymes 1972), expose les apprenants à une diversité langagière qui n'existe pas dans les manuels ou dans des documents construits.

2.2.3. Choisir des objectifs d'apprentissage

Tous les apprenants sont différents et ont des objectifs d'apprentissage qui leur sont propres. En utilisant des corpus de situations de communication variées, l'apprenant peut se fixer des objectifs individuels. En d'autres termes, l'apprenant construit son interlangue, il n'imité pas une façon de faire mais choisit parmi des façons de faire. Il peut sélectionner les compétences langagières qu'il souhaite acquérir. Par

exemple, des apprenants nous ont avoué qu'ils souhaitaient apprendre à parler comme des jeunes. Dans ce cas, lors de travaux pratiques en classe, ils ont sélectionné (grâce aux métadonnées) les corpus dans lesquels les locuteurs avaient moins de vingt-cinq ans. D'autres apprenants ont sélectionné les corpus de conversations pour apprendre à interagir avec des amis francophones. D'un point de vue plus syntaxique, certains apprenants font également des choix. Par exemple, ils choisissent d'apprendre à utiliser les clivées en *c'est X qui* ou en *il y a Y qui* dans la mesure où ceux qui sont en immersion les entendent fréquemment. Un concordancier qui permet de rechercher des mots suivis d'autres mots permet de trier les données, comme c'est le cas dans JConc (fig. 5).

Ainsi, les apprenants peuvent choisir ce qu'ils ont envie de comprendre et ce qu'ils ont envie d'apprendre. La démarche propose une approche implicite et inductive de la grammaire, comme nous l'avons vu précédemment lors de l'observation et de l'analyse des pratiques langagières. L'apprenant est donc actif et acteur de son apprentissage alors que la grande majorité des approches sont explicites et déductives avec les traditionnels exercices d'application des règles transmises par l'enseignant (Fougerouse 2001).

Au-delà de constructions syntaxiques spécifiques, les apprenants peuvent choisir d'apprendre à réaliser certaines activités langagières particulières. En effet, certains apprenants souhaitent apprendre le français pour travailler dans le tourisme, pour faire des études, pour organiser leur vie en France, pour visiter des musées, pour comprendre les médias, etc. En choisissant des corpus ou des sous-corpus (grâce aux métadonnées), les apprenants peuvent orienter leur apprentissage. Ils peuvent ainsi observer des réalisations linguistiques communicativement situées pour découvrir ce qu'il est important d'apprendre. Ce travail peut être accompagné ou médié par un enseignant ou un guide, ou bien se faire seul, en autonomie.

2.2.4. Aller vers l'autonomie

Une des caractéristiques de l'utilisation des corpus à des fins d'apprentissage est la possibilité pour les apprenants de travailler en autonomie (Aston 2001). En effet, les corpus sont utilisables dans et en dehors de la salle de classe. Néanmoins, les expérimentations que nous avons menées montrent qu'il a toujours fallu une phase de prise en main avant que les apprenants commencent à travailler seul. De plus, cette prise en main a été transmise par des enseignants maîtrisant eux-mêmes les possibilités offertes par les corpus. Ainsi, l'utilisation des corpus à fins d'enseignement et d'apprentissage de la langue nécessite une

Search results for 'c'est' followed by 'qui' in the TCOF corpus using JConc:

Num	Fichier	Contexte Gauche	Occurrence	Contexte Droit
67	professeures...	mais ouais c'est sûr	c'est	pas ça qui va faire
68	professeures...	euh c'est attiré euh	c'est	la MDPH qui ou alors
69	professeures...	AVS quand c'est euh	c'est	euh l'inspection qui euh
70	professeures...	de solidarité hum hum donc	c'est	ce qui est proposé euh
71	professeures...	c'est pas vous	c'est	tout le monde qui le
72	professeures...	euh ben quand euh il	c'est	les enfants qui viennent nous
73	professeures...	autre vient nous voir donc	c'est	nous qui régions c'est
74	professeures...	régions c'est à dire	c'est	l'enseignant qui a eu
75	professeures...	qu'il faut faire attention	c'est	surtout les enfants qui n'
76	Projets_DUP_16	bon le le master MEEF	c'est	celui qui te prépare le
77	Projets_DUP_16	c'est vrai que ça	c'est	quelque chose qui m'a
78	Projets_DUP_16	cetera donc euh moi --	c'est	ça qui m'a beaucoup
79	Projets_DUP_16	des expériences	c'est	c'est quelque chose qui
80	Projets_DUP_16	enrichissantes euh hum c'est	c'est	quelque chose qui plaît surtout
81	Projets_DUP_16	a pas mal d'activités	c'est	ça aussi qui m'a
82	Projets_DUP_16	euh c'est vrai que	c'est	c'est ça qui est
83	Projets_DUP_16	est vrai que c'est	c'est	ça qui est pratique je
84	Projets_DUP_16	ça c'est vrai oui	c'est	un truc qui me lui

Options: ET OU

Suivi de :

qui

Occurrences in corpus:

- <L11> c'est pas vous personnellement tout le monde qui le règle en fait euh
- <L21> pour être euh
- <L22> ben quand euh il c'est les enfants qui viennent nous v- enfin soit nous on voit la bagarre soit ils viennent
- <L11> hum
- <L21> mais la bagarre que l'on ne voit pas en générale
- <L22> c'est parce que il y a quelqu'un qui a été on l'a pincé et caetera
- <L11> oui
- <L21> ça été fait dans un coin

Figure 5. Recherche de *c'est* suivi de *qui* dans le corpus TCOF avec JConc

formation des enseignants, notamment lorsqu'il s'agit d'utiliser des corpus de linguistes constitués pour des linguistes. Une fois formés, les enseignants montrent aux apprenants comment utiliser les corpus, observer, sélectionner ou encore interroger les données. La prise en main d'un corpus par un apprenant se fait rarement seul, elle est souvent médiée par les enseignants (Kerr 2013), même lors de l'utilisation de corpus spécifiques ou didactisés (nous en présenterons dans le point suivant). Cependant, une fois qu'un enseignant (ou un conseiller / un guide) a sensibilisé les apprenants à ces nouvelles pratiques, ces derniers développent une réelle forme d'autonomie. Les apprenants déploient de nouvelles compétences alliant observation et analyse de données langagières. Johns (1990) a bien expliqué que, même si l'on peut penser que les apprenants les plus autonomes ont davantage de facilités à travailler avec des corpus, le travail à partir de corpus rend les apprenants de plus en plus autonomes. Les dispositifs en ligne d'interrogation de corpus favorisent l'autonomie. Ils permettent aux apprenants de choisir les moments et les rythmes de travail ainsi que, comme nous l'avons expliqué précédemment, leurs objectifs d'apprentissage (Gremmo & Riley 1997). Comme le montrent Rivens Mompean & Eisenbeis (2009) dans le cadre de leurs dispositifs hybrides, le processus d'autonomisation est impulsé par le fait que l'apprentissage est motivé par la recherche conduite par les apprenants.

3. Des exemples d'exploitation de corpus à des fins didactiques

Pour illustrer les principes et méthodologies exposés dans la partie précédente, nous allons présenter des exemples d'exploitations de différents corpus à des fins d'enseignement et d'apprentissage du FLE. Il existe plusieurs types de corpus pouvant être utilisés par des apprenants. Nous présenterons ici trois corpus et trois façons de les exploiter.

3.1. Un corpus de linguistes

Parmi les différents corpus exploitables par des apprenants de FLE, il y a les corpus constitués par des linguistes pour des linguistes. Ces corpus permettent de décrire le fonctionnement du français parlé en interaction. Leur fonction première n'est donc pas d'être exploités à des fins didactiques. Néanmoins, la grande quantité de données qu'ils contiennent fait d'eux de bons moyens d'exposer les apprenants à la langue cible. Le corpus TCOF, que nous avons déjà mentionné précédemment, compte actuellement 309 enregistrements de différents genres de discours (conversations, entretiens, débats, réunions de travail, etc.) et qui présentent une grande variété de locuteurs (âge, niveau d'étude, lieu de résidence, etc.). Ce corpus peut être utilisé avec des apprenants qui s'interrogent sur le fonctionnement de la langue. Nous avons montré des exemples de requêtes (*parce que, puisque, c'est X qui, il y a Y qui*) faites grâce à un concordancier par des apprenants dans les points précédents. Tous les niveaux de la langue – lexicaux, syntaxiques, sémantiques, pragmatiques, interactionnels – peuvent être observés et analysés en corpus même s'ils ne peuvent pas tous être observés par un concordancier.

3.2. Un corpus spécialisé

Le projet FLEURON (Français langue étrangère universitaire: ressources et outils numériques) qui se concrétise par un dispositif d'apprentissage en ligne⁴ s'inscrit dans le cadre de l'exploitation outillée de corpus à des fins didactiques (André 2016). Le dispositif met à disposition des apprenants un corpus multimodal et un concordancier aligné texte / son / vidéo. Ce corpus est spécialisé dans le domaine universitaire, il présente des interactions auxquelles les étudiants sont susceptibles de participer pour organiser et gérer leur séjour d'études en France. Ce

4. <https://fleuron.atilf.fr>

dispositif est conçu pour des apprenants, le concordancier est donc plus ergonomique et plus simple d'utilisation que ceux de la linguistique de corpus, « détournés » à des fins didactiques.

Pour illustrer l'utilisation du corpus et du concordancier, nous allons relater la démarche d'un apprenant, étranger en séjour universitaire en France, qui a souhaité comprendre le fonctionnement de *même*. Son questionnement, dont il a fait part à un de ses enseignants à la fin d'un cours, est né de son incompréhension de ce mot alors qu'il l'entendait souvent lors d'interactions quotidiennes et universitaires. Nous avons donc suggéré à cet apprenant de rechercher les occurrences de *même* dans le concordancier de FLEURON. La figure 6 montre un extrait des résultats.

mais je viens quand	même	la première journée le vingt et un
mais je peux quand	même	venir à la rentrée le vingt et un
il vous faut quand	même	prendre hein
ça risque quand	même	il y a des chances que ça puisse
donc nous avons quand	même	un sa- plus de mille personnes
mais bon je privilégie quand	même	les événements qui se passent ici à la au campus
c'est normal hein parce qu'il y a quand	même	une correction les profs doivent vous corriger doivent saisir les notes donc
d'accord? mais attendez les résultats quand	même	avant d'aller le voir parce que...
oui mais vous allez en cours quand	même	hein
	même	si vous ne connaissez pas vos résultats
c'est dur d'aller en cours quand	même	
XX en	même	temps?
trois choses en en	même	temps alors
euh mais vous savez un petit peu quand	même	dans quelle discipline vous voulez étudier
donc qui voudraient faire du FLE ou	même	qui veulent apprendre une langue seule
donc une fois que tu as tes côtes tu vas te servir toi	même	en rayon
si possible une carte étudiante quand	même	pour prouver que vous êtes étudiant
euh c'était euh c'était quand	même	une différence une différence entre les deux pays
venir en France spécialement à Nancy la faculté est quand	même	assez bien réputée
il y a quand	même	une liste de de documents à fournir quand

Figure 6. Extrait de la liste des occurrences de *même* dans FLEURON

L'avantage du concordancier de FLEURON est qu'il est aligné texte / son / vidéo. En cliquant sur une occurrence dans la liste, le son ou la vidéo se lance au moment où l'occurrence est prononcée, comme nous pouvons le voir sur la figure 7 ci-après.

Une fois que la liste apparaît, l'apprenant se lance dans une analyse distributionnelle de *même*, il devient un apprenti chercheur en linguistique. En d'autres termes, il fait de la linguistique de corpus sans le savoir. Il analyse toutes les occurrences de *même*, leur place dans les énoncés, leur rôle syntaxique, leur sens, les collocations, etc. L'apprenant fait la liste de tous les fonctionnements différents de *même* en analysant chaque occurrence et en ayant recours à la ressource lorsque cela lui a paru nécessaire. En visionnant la ressource, l'apprenant a accès à un cotexte plus large, à la multimodalité et aux métadonnées. En croisant

ces données, verbales, non verbales et situationnelles, l'apprenant réussit à saisir les différents emplois et sens de *même*.

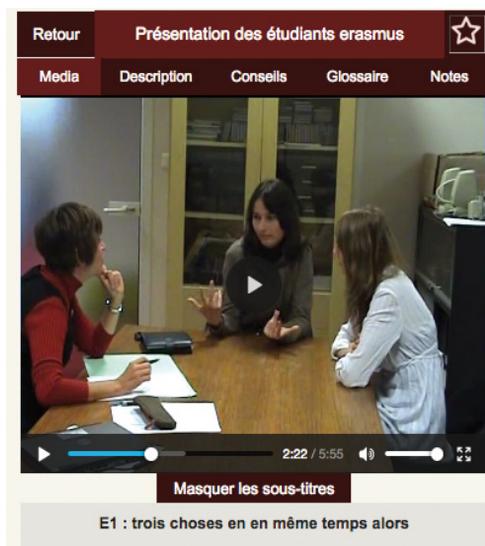


Figure 7. Vidéo correspondant à la 13^e occurrence de *même* dans la liste de la figure 6

3.3. Un corpus didactisé

Le corpus CLAPI est également un corpus de linguistes. À partir de ce corpus est né le site CLAPI-FLE⁵ qui présente des ressources didactisées du corpus CLAPI. Aujourd'hui, quarante extraits de CLAPI sont didactisés et mis à disposition des enseignants de FLE. Chaque extrait est décrit en termes de thèmes de la conversation, d'objectifs communicatifs et d'actions langagières. Pour chaque extrait, des exploitations didactiques sont proposées ainsi que des données quantitatives sur le lexique.

La plateforme CLAPI-FLE propose également des collections de faits langagiers et interactionnels illustrés par des extraits du corpus CLAPI. Par exemple, sont illustrées les constructions commençant par *le truc c'est que*, par *c'est vrai (que)* et les dislocations. Les deux extraits ci-après, une vidéo et un extrait audio, illustrent la construction en *le truc c'est que*:

5. <http://clapi.ish-lyon.cnrs.fr/FLE>

Exemple 1.1 - Corpus Repas - Kiwi

La jeune fille préférerait ne pas rester à l'université le vendredi où elle n'a pas cours d'analyse conversationnelle, mais elle y est plus ou moins contrainte (pour rendre service à son amie).

Vidéo (mp4)

Transcription (docx)



1 ELI et le vendredi normalement j'ai une fois sur deux tu sais

2 int-- euh hm comm[ent];

3 BEA _____[analyse conversationnelle (0.4)]

4 ELI exact mais euh le truc c'est que du coup même si: euh, quand

5 je l'ai pas je reste en fait parce que Laurence vu qu'elle

6 est en formation:

Exemple 1.2 - Corpus Invitation téléphonique - Proposition de déjeuner

La jeune fille (Florence) est ennuyée par la requête de son amie (Anaëlle) qui lui demande de déjeuner avec elle hors de l'université. Elle formule ce qui la gêne: elle n'a pas de monnaie

Audio (wav)

Transcription (docx)



1 Anaëlle ça va dis je voulais te demander: demain vu que euh on a qu'une

2 heure pour manger euh ça te dit que on aille manger dehors:

3 (1.3)

4 Florence euh hm pf je sais pas en fait parce que le truc c'est que j'ai

5 encore PAS de monnaie (0.5)

6 Anaëlle ouais

7 Florence donc du coup ça voudrait dire que on mangerait genre encore un

8 sandwich non: (0.8)

9 Florence ouais (0.3) ouais c'est c'est vrai c'est pas trop

10 Anaëlle c'est pas [trop le top

11 Florence _____[dans ce cas enfin (1.5)]

Figure 8. Extraits de CLAPI pour illustrer le fonctionnement de *le truc c'est que*

Les extraits sont précédés d'une petite fiche explicative du phénomène observé et sont classés selon les catégories déterminées dans la fiche. L'objectif du projet CLAPI-FLE est de transposer les résultats de l'analyse interactionnelle vers la didactique (Ravazzolo *et al.* 2015).

4. L'exemple de la négation – le projet FRACOV

Le projet FRACOV (Français contemporain vernaculaire) s'est donné deux objectifs. Le premier vise à décrire la pratique spontanée, non surveillée, du français. Pour atteindre cet objectif, plusieurs fiches descriptives sont élaborées à partir de corpus oraux. Ces fiches décrivent par exemple le fonctionnement de l'exclamation et de l'interjection, de l'interrogation, de la négation, du passif, des propositions relatives, du futur, etc. Le second objectif est la didactisation de ces fiches. En d'autres termes, l'objectif est d'examiner de quelles façons ces fiches décrivant certains phénomènes morphosyntaxiques du français peuvent être utiles et utilisées par des enseignants de FLE. Concrètement, il s'agit de concevoir des séquences pédagogiques permettant un travail d'appropriation de ces phénomènes langagiers. Pour harmoniser le travail des membres du projet, les séquences pédagogiques créées visent un public d'apprenants niveau B2 et durent environ deux heures.

4.1. La fiche descriptive Ne

Dans le cadre de cette didactisation du français vernaculaire, nous avons choisi de travailler à partir de la fiche descriptive portant sur le *ne* de négation (reproduite en annexe). Nous avons choisi de réfléchir à la

didactisation du fonctionnement du *ne* de négation tout d'abord parce que son actualisation est bien différente à l'écrit et à l'oral, et que l'oral reste encore le parent pauvre de la didactique. En outre, notre choix a été influencé par le fait que de très nombreux travaux en français parlé ont traité de ce sujet (dès les années 1980 en France, voir notamment Blanche-Benventiste & Jeanjean 1987). C'est devenu notamment l'exemple classique des différences entre la grammaire de l'écrit et la grammaire de l'oral. Ainsi, comme nous l'avons expliqué précédemment, transposer les résultats d'analyses linguistiques nous semble être une démarche importante pour la didactique. Enfin, nous avons choisi d'élaborer cette séquence sur *ne* parce que c'est un point compliqué en didactique du FLE à l'oral. Nous avons même rencontré des enseignants de FLE en Chine qui nous ont avoué simplifier leur enseignement et expliquer à leurs apprenants que *ne* n'est jamais actualisé à l'oral. Cependant, les étudiants étrangers en France se rendent compte que les choses sont plus complexes. Plusieurs nous ont signalé des problèmes d'utilisation de la négation. Le plus souvent, contrairement à ce qu'on peut leur enseigner, les apprenants de FLE actualisent presque systématiquement le *ne* de négation, en référence à la norme. Néanmoins, ils observent que cette pratique n'est pas toujours appropriée (Donaldson 2017).

Notre objectif est de voir si les apprenants réussissent à se construire une fiche descriptive du fonctionnement de *ne* à l'oral proche de la fiche du projet FRACOV, sans le métalangage propre au linguiste. Notre méthodologie d'apprentissage sur corpus implique les apprenants dans une démarche inductive. Plutôt que d'expliquer la fiche descriptive de FRACOV, ce qui pourrait avoir lieu dans un cours de FLE avec une approche explicite de la grammaire, nous préférons guider les apprenants vers une analyse personnelle, comme nous l'avons expliqué avec le corpus FLEURON et les occurrences de *même*.

4.2. La séquence pédagogique

La séquence pédagogique que nous proposons ici et qui a été testée avec des apprenants s'inscrit dans le cadre de l'apprentissage sur corpus. Elle a pour prérequis un travail sur la négation et ses différentes formes. Elle vise comme objectif communicatif : savoir utiliser et reconnaître une négation. Plus précisément, en compréhension écrite et orale, il s'agit de comprendre que quelque chose n'est pas possible, et, en expression écrite et orale, il s'agit de dire que quelque chose n'est pas possible.

La séquence se déroule en plusieurs étapes :

1. observations à l'écrit ;
2. systématisation à l'écrit ;

3. observations à l'oral;
4. vérification à partir des transcriptions des corpus oraux;
5. mise en situation à l'oral.

Nous allons détailler chacune de ces étapes.

La première étape est une phase d'observation de la négation dans des corpus écrits. Cette phase est brève, elle consiste à demander aux apprenants de rechercher *pas* dans un concordancier afin de comprendre le fonctionnement de la négation en *ne... pas* en français écrit. L'enseignant précise qu'il y a des exemples d'énoncés construits avec *pas* qui ont une polarité négative mais qui ne sont pas des négations en *ne... pas* (par exemple : *pas de problème, pas mal*). Il peut attirer l'attention des apprenants sur quelques exemples. Les corpus proposés aux apprenants sont : Scientext⁶, un corpus de textes scientifiques ; *Le Monde* de 1998⁷, un corpus d'articles de presse ; et Frantext⁸, un corpus de textes littéraires.

Lors de cette première étape, les apprenants recherchent donc *pas* dans le concordancier du corpus Scientext, en ayant la possibilité de choisir la discipline et le genre de discours qu'ils souhaitent interroger, comme nous pouvons le voir sur l'interface suivante :

The screenshot shows the Scientext search interface. At the top, there is a navigation bar with tabs: Corpus, Textes, Recherche, Résultats, Historique, Connexion, and À propos. Below this, there are sub-tabs for Critères and Liste. The main content area displays the current search selection: "Actuellement votre sélection comporte 5 063 315 mots (205 textes), sur les 5 063 315 mots (205 textes) du corpus. Cliquez pour voir les détails de la sélection...". Two filter panels are visible: "Discipline" and "Genre".

Discipline

- Sciences expérimentales
 - Biologie (744 534 mots, 28 textes)
 - Médecine (97 139 mots, 8 textes)
- Sciences humaines
 - Linguistique (1 633 579 mots, 67 textes)
 - Psychologie (528 554 mots, 17 textes)
 - Sciences de l'éducation (1 280 515 mots, 60 textes)
 - TAL (753 025 mots, 17 textes)
- Sciences appliquées
 - Mécanique (139 912 mots, 3 textes)
 - Électronique (419 234 mots, 5 textes)

Tout Rien

Genre

- HDR (517 025 mots, 7 textes)
- Article (348 168 mots, 45 textes)
- Communication (532 333 mots, 112 textes)
- Thèse (4 198 966 mots, 41 textes)

Tout Rien

Figure 9. Interface de choix des métadonnées de Scientext

6. <https://corpora.aiakide.net/scientext20>
7. Ce corpus est directement interrogeable sur le site Lextutor : <https://www.lexutor.ca/conc/fr>.
8. <https://www.frantext.fr>

Les apprenants obtiennent ces résultats⁹:

Corpus	Textes	Recherche	Résultats	Historique	Connexion	À propos	Aide
			Sémantique Lire Avancé				
<input checked="" type="checkbox"/>	1		l'enseignement secondaire, l'enseignement spécialisé qui n'est	pas	celui de son origine. En effet, la TSD a	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	2		que doit observer le chercheur en sciences humaines pour ne	pas	devenir prisonnier d'un modèle particulier ni surtout de l'	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	3		fait que la science et l'objectivité scientifique ne résultent	pas	des tentatives d'un savant individuel pour être " objectif	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	4		le projet scientifique en sciences sociales il faut faire un	pas	de plus, dont les sciences de la nature peuvent	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	5		didactique construit des modèles qui, tout en n'étant	pas	mathématiques, se doivent d'être, non exacts,	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	6		Bien des théories, on le pressent, ne satisfont	pas	à ce critère: ce sont les théories trop générales	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	7		théories. Notons qu'une telle méthodologie globale n'existe	pas	dans ce champ de recherche, ainsi qu'il a été	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	8		L'usage de méthodologies locales ou diverses n'est ainsi	pas	à confondre avec un opportunisme des méthodes suivant le phénomène	Cliquer pour afficher le contexte large et l'arbre syntaxique #1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	9		série de questions ou de phénomènes: ce n'est	pas	à la théorie choisie de déterminer les questions pertinentes sur le champ	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	10		réalité didactique étudiée. En dernière analyse, la théorie ne fonctionne	pas	pour son propre développement mais pour l'étude de l'	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	11		me sert-il à voir quelque chose que je ne verrais	pas	sans cet outil? Par exemple l'idée de processus /	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	12		,de les classer...? Ou ne qualifie-t-on	pas	de concept ce qui n'est qu'une métaphore?	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	13		et un volet didactique. Qui dit réalité ne dit	pas	" monde réel " toute science reconstruit une réalité observable	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	14		, " Certes la didactique des mathématiques n'en vient	pas	à un système axiomatique, et Gonseth parle ici de	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	15		et N., 1988), il n'y a	pas	, dans la théorie, de réflexion sur les ostensifs	#1 - HDR - Sciences de l'éducation - Introduction	
<input checked="" type="checkbox"/>	16		l'on souhaiterait entendre; mais ils ne le sont	pas	moins pour comprendre l'interprétation des signes mathématiques et des	#1 - HDR - Sciences de l'éducation - Introduction	

Figure 10. Extrait de la liste des résultats de la recherche de pas dans le corpus Scientext

Il est possible d'obtenir un contexte plus large et d'accéder aux textes, comme nous le voyons ci-dessous :

<input checked="" type="checkbox"/>	10		fonctionne	pas	pour son propre développement mais pour l'étude de l'	Introduction
<input checked="" type="checkbox"/>	11		me sert-il à voir quelque chose que je ne verrais	pas	sans cet outil? Par exemple l'idée de processus /	#1 - HDR - Sciences de l'éducation - Introduction
<input checked="" type="checkbox"/>	12		,de les classer...? Ou ne qualifie-t-on	pas	de concept ce qui n'est qu'une métaphore?	#1 - HDR - Sciences de l'éducation - Introduction
		Afficher <input checked="" type="radio"/> Contexte <input type="radio"/> Arbre syntaxique		définition qui me permet de qualifier les faits qui en relèvent, de les repérer, de les étiqueter, de les classer...? Ou ne qualifie-t-on pas de concept ce qui n'est qu'une métaphore? Ou, pire, est-on victime d'une idéologie de l'éducation? Ainsi que le dit Bourdieu pour la sociologie mais en l'occurrence on peut remplacer sociologie par éducation il s'agit d'un champ ou : " le monde se sert en droit de dire la vision légitime du monde éducatif, dans lequel le didacticien intervient aussi, mais avec une ambition tout à fait spéciale, que l'on accorde sans problème à tous les autres savants, et qui, dans son cas, tend à paraître monstrueuse : dire la vérité ou pire, définir les conditions dans lesquelles on peut dire la vérité " (Bourdieu, 2001, p. ... Dans notre principal paradigme de recherche la Théorie des Situations Didactiques il sera donc nécessaire d'éprouver les concepts utilisés ou introduits, à la fois du côté théorique de la didactique, et par confrontation à la contingence, c'est-à-dire à la réalité de l'enseignement, dans les dimensions où nous prétendons les observer: l'enseignement spécialisé et l'enseignement des notions de l'analyse à l'articulation secondaire / supérieur. Cette		Isabelle BLOCH QUELQUES APPORTS DE LA THÉORIE DES SITUATIONS À LA DIDACTIQUE DES MATHÉMATIQUES DANS L'ENSEIGNEMENT SECONDAIRE ET SUPÉRIEUR Université Paris 7 2005
<input checked="" type="checkbox"/>	13		et un volet didactique. Qui dit réalité ne dit	pas	" monde réel " toute science reconstruit une réalité observable	#1 - HDR - Sciences de l'éducation - Introduction

Figure 11. Visualisation du contexte plus large en cliquant sur une occurrence dans la liste des résultats

Ensuite, il est proposé aux apprenants de faire la même recherche dans le corpus Frantext. L'enseignant précise que certains textes sont anciens pour que les apprenants ne soient pas surpris par les énoncés. La figure suivante présente la façon dont les résultats sont présentés :

- Afin d'aider le lecteur à comprendre la démarche, nous avons réalisé des copies d'écran de certaines interfaces. Ces images sont réduites, elles permettent notamment de visualiser rapidement la présentation des données aux apprenants.

ma vieillesse, encore vigilante, n'alourdit pas mes yeux. *Agamemnon cette étoile brillante et s'élançe près des pléiades, vois. Je n'ai pas entendu gazouiller les oiseaux, et les vents pareille est douce. L'homme obscur, que n'a pas ébloui la gloire, vit heureux, ô vieillard, dans la bouche d'un roi. Non, non, tu ne fus pas engendré par ton père pour goûter tous les b e déplaisir d'*Hélène et de son lit n'éteint pas le désir, y nourrissant sans cesse une offre rbare, je trompais une mère et je n'hésitais pas d'apprêter à ma fille un indigne trépas. Moi uis abîmé ! Mais cours et hâte-toi ; ne cède pas à l'âge ; évite les ruisseaux et les sources e verdure opposée au soleil, et ne te laisse pas charmer par le sommeil. *Le *Vieillard sois peut-être. *Le *Vieillard non, tu ne devais pas , portant la main sur moi, connaître des secr sur moi, connaître des secrets qui n'étaient pas à toi. *Ménélas a ton tour, devais-tu te cha

Figure 12. Extrait de la liste des résultats de la recherche de *pas* dans Frantext

Et enfin, les apprenants recherchent *pas* dans le corpus constitué par les numéros du journal *Le Monde* de 1998. La recherche s'effectue avec Lextutor qui affiche les résultats de la façon suivante :

915. □ à 10 000 francs). Le prix de la communication n'a pas encore été fixé par Iridium, mais son concurrent
916. □ é son oeil de biche. Son rire est franc, elle n'a pas peur de découvrir ses dents. En scène, elle se
917. □ éron, deux jeunes artistes auxquels le public n'a pas ménagé ses encouragements. Juan Manuel Quintan
918. □ és. L'Allemagne (112 milliards de francs) qui n'a pas bénéficié de la croissance en 1997 devrait con
919. □ bus, la régie publicitaire de la RATE, n'a-t-elle pas affirmé, par la voix de son PDG, Gérard Cros,
920. □ de l'équipe de France", la société Hom n'a-t-elle pas demandé à trois joueurs d'exhiber leurs jambes
921. □ éme de la civilisation judéo-chrétienne, n'a-t-il pas droit au RMI de jouissance consistant à rencon
922. □ février. Les analystes estiment qu'il n'aboutira pas et qu'un éventuel troisième scrutin, qui pourr
923. □ dépolitise le corps social; tant qu'il n'aboutira pas à un véritable engagement dans un respect de l
924. □ Le père de Jamal, farouche antisémite, n'accepte pas cette liaison. Le cousin de Jamal, narrateur,
925. □ ir ses professionnels de santé. S'ils n'acceptent pas ces règles, on ne travaillera pas avec eux. Ma
926. □ en place. Les autorités américaines n'accepteront pas que les plans du gouvernement japonais soient
927. □ ion et des lois de l'Argentine. "Nous n'acceptons pas", a souligné le ministre argentin de la Justice
928. □ sait à qui attribuer dans un pays qui n'accueille pas de groupes anti-américains. Trois personnes so
929. □ tion et à la géométrie. Pour autant, il n'affirme pas encore une oeuvre singulière. Alors qu'il côto
930. □ M. Bouygues : "Je ne te cacherais pas que je n'ai pas trouvé le niveau d'information dont j'estime a
931. □ risque pour la sécurité des personnes". "Je n'ai pas envie de finir ma carrière et mon mandat dans
932. □ un index irrité vers les journalistes : "Je n'ai pas eu de relations sexuelles avec cette femme, Mo

Figure 13. Extrait de la liste des résultats de la recherche de *pas* via Lextutor dans *Le Monde*, 1998

Cette première étape est une étape rapide d'observation. Les apprenants constatent essentiellement que *ne/n'* précède *pas* dans les constructions syntaxiques à polarité négative. Ils constatent également quelques collocations particulières telles que *certainement pas*, *pas de problèmes*, ainsi qu'un éloignement plus ou moins important entre *ne* et *pas*. La deuxième étape de cette séquence pédagogique est une phase de systématisation. L'apprenant doit transformer des énoncés positifs, extraits d'un corpus, en énoncés négatifs avec *ne... pas*. Les énoncés à transformer présentent des constructions syntaxiques variées.

Avec la troisième étape, nous entrons au cœur de la séquence pédagogique qui s'intéresse à la didactisation du français vernaculaire. Cette troisième étape est une nouvelle phase d'observation, réalisée avec des corpus oraux et multimodaux. Nous avons tout d'abord choisi d'utiliser FLEURON pour faire écouter aux apprenants différents énoncés négatifs avec et sans *ne*. Nous avons recherché *pas* dans le concordancier et fait écouter les énoncés, sélectionnés de façon aléatoire. En visualisant les résultats, dont une partie est représentée par la figure 14 ci-après, l'enseignant voit quels sont les énoncés qui réalisent ou pas la double négation.

qui soient	pas	européens
et nous on transmet à la préfecture mais en aucun moment nous on ne remet	pas	les cartes de séjour
de ma carte d'identité mais en fait je n'ai	pas	euh rempli le formul- le formul-
je vais vous le donner oui	pas	de problème
et puis toujours demander l'accord à votre université si elle est d'accord nous il y a	pas	de problème
voilà donc je je ne précise	pas	professeur de polonais hein vous avez deviné
si vous comprenez	pas	si...
si c'est	pas	assez clair si j'ai mal écrit
donc on vous demande ce n'est	pas	très cher euh l'adhésion
je pense que les cours vont	pas	lander euh la rentrée c'est le 28
et puis là vous gardez pour vous ce papier ça indique que vous n'avez	pas	payé de de de frais vous n'avez rien payé
ah parce que tu n'avais	pas	pris ton carnet de santé

Figure 14. Extrait de la liste des résultats de la recherche de *pas* dans FLEURON

Nous avons demandé aux apprenants s'ils pouvaient constater quelque chose quant au fonctionnement de la négation à l'oral. La réponse attendue est évidemment que *ne* n'est pas toujours actualisé.

Nous avons élaboré une deuxième phase d'écoute pour tester la reconnaissance de la négation. Dans le corpus FLEURON, nous avons sélectionné dix extraits, avec et sans *ne*. Les apprenants doivent cocher dans un tableau à deux colonnes s'ils entendent *ne... pas* ou *pas*.

Séquence	<i>ne... pas</i>	<i>pas</i>
1		
2		
...		

Figure 15. Tableau pour les apprenants

Cette activité montre que les apprenants, quel que soit leur niveau, ont parfois du mal à déceler si la négation est en deux parties ou en une seule. Ils justifient leur difficulté en expliquant que certains sons rendent ardue la discrimination. Nous retrouvons ici des aspects phonétiques de la réalisation du « pronom atone contenant un schwa susceptible de réduction » (Larrivée & Ramasse 2013). Nous demandons aux apprenants de vérifier eux-mêmes leurs réponses en s'appuyant sur la transcription. C'est à partir de ce travail avec la transcription que les apprenants commencent à analyser les occurrences de *ne* et de *ne... pas* pour tenter de comprendre le fonctionnement de la négation.

Pour compléter cette phase d'observation et d'analyse, nous demandons aux apprenants de rechercher *pas* dans d'autres corpus, notamment dans TCOF, CLAPI, CFPP2000. Les apprenants constatent, une fois encore, que le *ne* de négation est parfois actualisé à l'oral et parfois pas. Ils observent les cotextes d'apparition et ne réussissent pas immédiatement à dégager de règles. Ils constatent néanmoins que, dans un même énoncé, un locuteur peut produire deux négations successives, une avec et une sans *ne*. Ces phases d'observation n'éclairent pas immédiatement les

apprenants sur le fonctionnement de la négation à l'oral. Cependant, la complexité du phénomène les amène à envisager différentes analyses.

Ainsi, parallèlement à ces constats, les apprenants réalisent des analyses pour tout de même essayer de comprendre les apparitions et les omissions. Tout d'abord, les apprenants constatent qu'avec certaines expressions telles que *il y a*, *c'est* ou *je sais* (très fréquentes dans les corpus), le *ne* de négation est rarement présent. Ce constat correspond à ce qui est décrit par Pierre Larrivée dans la fiche FRACOV (en annexe) et dans d'autres études (par exemple Moreau 1986; Gadet 1997). L'environnement phonétique explique cette réduction de la négation. Même si les apprenants n'ont pas le métalangage pour décrire leurs observations, ils réussissent à réaliser et à verbaliser des analyses tout à fait justes, concernant par exemple certaines règles phonologiques. Les observations de l'écrit ont permis aux apprenants de constater les places syntaxiques des deux éléments de la négation, une position clitique préverbale et une position postverbale. En observant l'oral, ils ont constaté que souvent seul l'élément postverbal subsiste. Évidemment, ils avaient déjà fait ce constat, soit lors de leur cours de FLE, soit en écoutant interagir des francophones, natifs ou non natifs. C'est justement les raisons de l'absence du *ne* que nous essayons de déterminer avec eux. D'autres analyses sont réalisées par les apprenants, notamment en ce qui concerne la valeur pragmatique de la négation. Ils constatent que, dans certains cas, il serait inapproprié d'utiliser la double négation et, inversement, l'effet pragmatique escompté ne serait pas atteint avec des énoncés sans *ne*. Les apprenants repèrent des valeurs d'emphase et d'intensité. Ces valeurs ont été identifiées notamment par Fonseca-Greber (2007) ou Van Compernelle (2009).

Une dernière observation est proposée aux apprenants avec le corpus OFROM. Nous leur avons demandé de faire parallèlement deux recherches de *pas* dans le concordancier associé au corpus, une première en sélectionnant « conférence » dans le genre de parole et une seconde en sélectionnant « discussion ». La figure 16 ci-après montre ce concordancier, dans lequel « conférence » est sélectionné.

Seules les dix premières occurrences de chaque résultat ont été analysées par les apprenants. Ceux-ci constatent que, dans les résultats pour « conférence », le *ne* de négation est actualisé quatre fois et aucune fois dans les résultats pour « discussion ». Ils en déduisent que le genre de discours influence la fréquence d'utilisation de la négation en deux parties (Gadet 2000). Tout au long de la séquence, les apprenants sont accompagnés ou guidés par un enseignant dans leurs observations, leurs classifications et leurs généralisations. La dernière étape de cette séquence consiste à mettre en situation les apprenants. Nous leur avons

Figure 16. Recherche de *pas* dans les conférences du corpus OFROM

donc demandé d’imaginer des situations de communication dans lesquelles ils auraient à répondre négativement à des questions. En fonction des éléments de la situation qu’ils déterminent, les apprenants actualisent ou pas la négation en deux parties.

En conclusion, à l’issue de cette séquence pédagogique, grâce à une démarche inductive permise par l’observation de données en corpus, les apprenants ont réussi à déduire le fonctionnement syntaxique et sémantique décrit dans la fiche descriptive du projet FRACOV. En résumé, les apprenants ont constaté que *ne* est une forme « faible », qu’il est toujours associé à un verbe et qu’il est le premier du groupe des clitiques antéposés au verbe (constat déjà fait avec les observations de l’écrit). Ils ont constaté sa présence systématique à l’écrit et sa rareté à l’oral, tout en remarquant tout de même des actualisations. L’environnement phonétique a été repéré comme un élément pouvant influencer l’absence de *ne*, notamment lorsque des groupes consonantiques se forment, comme *je sais pas*, *je veux pas* ou *je peux pas*. D’ailleurs, les apprenants ont également remarqué que *ne* est plus absent lorsque le sujet est un clitique que lorsque le sujet est un mot lexical. Ils ont aussi analysé les apparitions de *ne* avec d’autres éléments négatifs tels que *point* ou *rien* et avec l’expletif *que*. Enfin, ils concluent que l’absence ou la présence du *ne* de négation n’est pas toujours expliquée et que des éléments extralinguistiques peuvent être une source d’influence. Au-delà de la complexité relevée par les apprenants, cette séquence leur a permis de faire de la sociolinguistique de corpus sans le savoir et d’analyser le fonctionnement syntaxique, sémantique et pragmatique de la langue.

5. Conclusion

L'exploitation de corpus à des fins d'enseignement et d'apprentissage du FLE offre de nouvelles perspectives pour la didactique. Donner un accès direct aux données langagières constituées en corpus à des apprenants permet non seulement de pallier le manque de données authentiques mais également de trouver une alternative aux dialogues construits des manuels de langue, encore utilisés en salle de classe aujourd'hui. En amenant les apprenants à travailler avec des corpus, ces derniers partent à la découverte des données, ils observent des faits de langue, analysent leurs actualisations, peuvent vérifier des hypothèses, des intuitions sur le fonctionnement de la langue et des interactions verbales, ils comprennent le sens et les usages en contexte. Nous avons constaté que la démarche inductive que nous proposons aux apprenants leur permet de dégager eux-mêmes des règles syntaxiques, de faciliter l'interprétation des énoncés, de saisir la valeur pragmatique des activités langagières ou encore de cerner les mécanismes de prise de parole. En d'autres termes, nous proposons une méthodologie qui fait faire de la linguistique de corpus à des apprenants de FLE. Nous pouvons même préciser qu'ils font de la sociolinguistique de corpus dans la mesure où la contextualisation des données facilite l'exposition à la langue cible et les analyses qu'ils peuvent en faire. Les apprenants acquièrent ainsi une conscience métalangagière. Cette méthodologie peut être médiée par un enseignant puis, éventuellement, être menée en autonomie.

L'approche sur corpus permet de prendre en compte la variation et, dans d'autres termes, d'enseigner et d'apprendre le français vernaculaire. En plus d'exposer à une langue utilisée par des locuteurs en situations naturelles, l'avantage est que cela permet également de décomplexer les apprenants. Ils s'aperçoivent que le discours se construit au fur et à mesure de l'interaction et que ce discours n'est jamais constitué des énoncés parfaitement normatifs qu'ils apprennent généralement dans les manuels. De plus, elle permet de repenser les façons de faire de la grammaire avec des apprenants de langue. La relation entre lexique et grammaire est également abordée d'une façon inductive. Les apprenants examinent, font des analyses distributionnelles, remarquent des collocations, apprennent des *chunks* ou des blocs de mots. En salle de classe, l'enseignant peut aider à l'analyse et orienter les apprenants. Toutes les pratiques langagières sont observables et analysables en corpus ainsi que tous les aspects de la langue – phonétiques, prosodiques, lexicaux, syntaxiques, sémantiques, pragmatiques, interactionnels.

Enfin, nous avons présenté la façon la plus simple d'introduire l'exploitation de corpus dans la formation linguistique des apprenants,

même si cette méthodologie n'en est qu'à ses premières expérimentations en France avec des corpus oraux et multimodaux. En plus des concordanciers (simples), d'autres outils d'exploitation peuvent être utilisés à des fins didactiques. Par exemple, il serait intéressant que les apprenants continuent leur chemin d'apprenti linguiste, pour s'approprier la langue, en recherchant un son particulier, un lemme, ou, plus précisément, les verbes au futur, les prépositions, les verbes à l'infinitif, les interrogations, etc. L'exploitation de corpus annotés, une fois qu'elle sera entièrement développée et accessible à des apprenants, offrira encore de nouvelles perspectives méthodologiques à la didactique des langues. Dans un contexte technologique en plein essor, les apprenants sont souvent réceptifs aux nouvelles technologies et aux méthodologies inductives, grâce auxquelles ils découvrent et font eux-mêmes avancer leurs apprentissages.

Références bibliographiques

Études

- ABÉ D., CARTON F., CEMBALO S. M., RÉGENT O. (1979), « Didactique et authentique: du document à la pédagogie », *Mélanges pédagogiques*, n° 10, p. 1-14.
- ANDRÉ V. (2014), « L'énonciation conjointe: trace et ressource de la construction collaborative du discours », *SHS Web of Conferences*, n° 8, 4^e congrès mondial de linguistique française – CMLF 2014, p. 1891-1904, en ligne : https://www.shs-conferences.org/articles/shsconf/pdf/2014/05/shsconf_cmlf14_01323.pdf.
- ANDRÉ V. (2016), « FLEURON: Français langue étrangère universitaire, ressources et outils numériques. Origine, démarches et perspectives », *Mélanges CRAPEL*, n° 37, p. 69-92.
- ANDRÉ V. (2017), « Un corpus multimédia pour apprendre à interagir en situations universitaires en France », in *Enseigner le français: s'engager et innover* (Actes du troisième colloque de l'ATPF, Bangkok, 19-20 octobre 2017), p. 292-315, en ligne : http://www.atpf-th.org/Actes_du_Colloque.pdf.
- ASHBY W. J. (1981), « The Loss of the Negative Particle *Ne* in French: A Syntactic Change in Progress », *Language*, vol. 57, n° 3, p. 674-687.
- ASTON G. (2001), « Learning with Corpora: An Overview », in *Learning with Corpora*, G. Aston (dir.), Houston, Athelstan, p. 1-45.
- BILGER M., CAPPEAU P. (2013), « Comment les données de corpus pourraient renouveler les manuels de grammaire? », *Linx*, n° 68-69, p. 177-199, en ligne : <https://journals.openedition.org/linx/1526>.

- BLANCHE-BENVENISTE C., JEANJEAN C. (1987), *Le français parlé. Transcription et édition*, Paris, Didier.
- BOULTON A., TYNE H. (2014), *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*, Paris, Didier.
- BRANCA-ROSOFF S., FLEURY S., LEFEUVRE F., PIRES M. (2012), « Discours sur la ville. Présentation du corpus de français parlé parisien des années 2000 », université Sorbonne Nouvelle – Paris 3, en ligne : <http://cfpp2000.univ-paris3.fr/CFPP2000.pdf>.
- BRUMFIT C. J., JOHNSON K. (dir.) (1979), *The Communicative Approach to Language Teaching*, Oxford, Oxford University Press.
- BRUXELLES S., JOUIN-CHARDON É., TRAVERSO V., 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 », in *Langues parlées, interactions sociales. Une variété d’usages pour l’apprentissage du français*, I. Guinamard, É. Jouin-Chardon, M. Rispaïl, V. Traverso, Trinh Duc Thai (dir.), Paris, L’Harmattan, p. 131-153.
- CONRAD S. (2004), « Corpus Variety: Corpus Linguistics, Language Variation, and Language Teaching », in *How to Use Corpora in Language Teaching*, J. M. Sinclair (dir.), Amsterdam – Philadelphie, J. Benjamins, p. 67-85.
- CULBERTSON J., LEGENDRE G. (2008), « Qu’en est-il des clitiques sujet en français oral contemporain ? », in *Congrès mondial de linguistique française – CMLF’08*, J. Durand, B. Habert, B. Laks (dir.), Paris, Institut de linguistique française, p. 2663-2674, en ligne : <https://www.linguistique-francaise.org/articles/cmlf/pdf/2008/01/cmlf08308.pdf>.
- CUQ J.-P. (dir.) (2003), *Dictionnaire de didactique du français langue étrangère et seconde*, Paris, CLE international.
- DEBAISIEUX J.-M. (2009), « Des documents authentiques oraux aux corpus : un défi pour la didactique du FLE », *Mélanges CRAPEL*, n° 31, *Des documents authentiques oraux aux corpus : questions d’apprentissage en didactique des langues*, p. 36-56.
- DONALDSON B. (2017), « Negation in Near-Native French: Variation and Sociolinguistic Competence », *Language Learning*, vol. 67, n° 1, p. 141-170.
- FONSECA-GREBER B. B. (2007), « The Emergence of Emphatic “Ne” in Conversational Swiss French », *Journal of French Language Studies*, vol. 7, n° 3, p. 249-275.
- FOUGEROUSE M.-C. (2001), « L’enseignement de la grammaire en classe de français langue étrangère », *Études de linguistique appliquée*, n° 122, p. 165-178, en ligne : <https://www.cairn.info/revue-ela-2001-2-page-165.htm>.
- GADET F. (1997), *Le français ordinaire*, 2^e éd., Paris, A. Colin.

- GADET F. (2000), « Des corpus pour (ne)... pas », in *Corpus: méthodologie et applications linguistiques*, M. Bilger (dir.), Paris, H. Champion, p. 156-167.
- GADET F. (2007), *La variation sociale en français*, Paris, Ophrys.
- GAVIOLI L. (2001), « The Learner as Researcher: Introducing Corpus Concordancing in the Classroom », in *Learning with Corpora*, G. Aston (dir.), Houston, Athelstan, p. 108-137.
- GREMMO M.-J., RILEY P. (1997), « Autonomie et apprentissage autodirigé: l'histoire d'une idée », *Mélanges CRAPEL*, n° 23, p. 81-107.
- HOLEC H. (1990), « Des documents authentiques pour quoi faire ? », *Mélanges CRAPEL*, p. 65-74.
- HYMES D. (1972), « Models of the Interaction of Language and Social Life », in *Directions in Sociolinguistics. The Ethnography of Communication*, J. J. Gumperz, D. Hymes (dir.), Oxford, B. Blackwell, p. 35-71.
- JOHNS T. (1990), « From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-Driven Learning », *CALL Austria*, n° 10, p. 14-34.
- JOHNS T. (1991), « Should You Be Persuaded: Two Examples of Data-Driven Learning Materials », *English Language Research Journal*, n° 4, *Classroom Concordancing*, T. Johns, P. King (dir.), p. 1-16.
- JOHNS T. (1997), « Contexts: The Background, Development and Trialling of a Concordance-Based CALL Program », in *Teaching and Language Corpora*, A. Wichmann, S. Fligelstone, T. McEnery, G. Knowles (dir.), Londres – New York, Longman, p. 100-115.
- KERBRAT-ORECCHIONI C. (2005), *Le discours en interaction*, Paris, A. Colin.
- KERR B. (2013), « Grammatical Description and Classroom Application. Theory and Practice in Data-Driven Learning », *Bulletin suisse de linguistique appliquée*, n° 97, p. 17-39.
- LARRIVÉE P. (2014), « Reanalysis of Negatives as Polarity Markers? The Last 400 Years of Decline of the French Preverbal Negative Clitic », *Lingua*, vol. 147, p. 40-49.
- LARRIVÉE P., RAMASSE D. (2013), « Dimensions stylistique et phonétique de la disparition de *ne* en français », *The Phonetician*, n° 107/108, p. 86-106.
- LEŃKO-SZYMAŃSKA A., BOULTON A. (dir.) (2015), *Multiple Affordances of Language Corpora for Data-Driven Learning*, Amsterdam – Philadelphie, J. Benjamins.
- MEISNER C., ROBERT-TISSOT A., STARK E. (2016), « L'absence et la présence du *ne* de négation », in *Encyclopédie grammaticale du français*, en ligne: www.encyclogram.fr/notx/008/008_Notice.php.

- Mélanges CRAPEL (2009), n° 31, *Des documents authentiques oraux aux corpus : questions d'apprentissage en didactique des langues*, en ligne : <http://www.atilf.fr/spip.php?rubrique584>.
- MONDADA L. (2002), « Pour une linguistique interactionnelle », in *Faits de langue - faits de discours. Données, processus et modèles. Qu'est-ce qu'un fait linguistique?*, M. Santacroce (dir.), Paris, L'Harmattan, vol. II, p. 95-136.
- MOREAU M.-L. (1986), « Les séquences préformées : entre les combinaisons libres et les idiomatismes. Le cas de la négation avec ou sans *ne* », *Le français moderne*, vol. 54, n° 3-4, p. 137-160.
- MULLER C. (1991), *La négation en français : syntaxe, sémantique et éléments de comparaison avec les autres langues romanes*, Genève, Droz.
- RAVAZZOLO E., TRAVERSO V., JOUIN-CHARDON É., VIGNER G. (2015), *Interactions, dialogues, conversations : l'oral en français langue étrangère*, Vanves, Hachette FLE.
- RECOURCÉ G. (1996), « Une double analyse de la particule *ne* en français », *Langages*, n° 122, p. 62-78.
- RIVENS MOMPEAN A., EISENBEIS M. (2009), « Autoformation en langues : quel guidage pour l'autonomisation ? », *Recherches en didactique des langues et des cultures. Les cahiers de l'Acedle*, vol. 6, n° 1, p. 221-244.
- SINCLAIR J. M. (1996), *Preliminary Recommendations on Corpus Typology*, EAGLES Document EAG-TCWG-CTYP/P.
- SINCLAIR J. M. (2004), « New Evidence, New Priorities, New Attitudes », in *How to Use Corpora in Language Teaching*, J. M. Sinclair (dir.), Amsterdam – Philadelphie, J. Benjamins, p. 271-299.
- TRAVERSO V. (1999), *L'analyse des conversations*, Paris, Nathan.
- TYNE H. (2012), « La variation dans l'enseignement. Apprentissage d'une langue 2 », *Le français d'aujourd'hui*, n° 176, p. 103-112.
- VAN COMPERNOLLE R. A. (2009), « Emphatic *Ne* in Informal Spoken French and Implications for Foreign Language Pedagogy », *International Journal of Applied Linguistics*, vol. 19, n° 1, p. 47-65.
- WIDDOWSON H. G. (1978), *Teaching Language as Communication*, Oxford, Oxford University Press.

Corpus

- CFPP2000 : Discours sur la ville. Corpus de français parlé parisien des années 2000, S. Branca-Rosoff, S. Fleury, F. Lefevre, M. Pires (dir.), en ligne : <http://cfpp2000.univ-paris3.fr>.

- CLAPI: Corpus de langue parlée en interaction, laboratoire Interactions, corpus, apprentissages, représentations (ICAR), université Lumière – Lyon 2, en ligne: <http://clapi.ish-lyon.cnrs.fr>.
- ESLO: Enquêtes sociolinguistiques à Orléans, Laboratoire ligérien de linguistique (LLL – UMR 7270), université d’Orléans, en ligne: <http://eslo.huma-num.fr>.
- FLEURON: Français langue étrangère universitaire: ressources et outils numériques, laboratoire Analyse et traitement informatique de la langue française (ATILF), université de Lorraine et CNRS, V. André (dir.), en ligne: <https://fleuron.atilf.fr>.
- OFROM: Corpus oral de français de Suisse romande, M. Avanzi, M.-J. Béguelin, F. Diémoz (dir.), en ligne: <http://www11.unine.ch>.
- PFC: Phonologie du français contemporain, M.-H. Côté, J. Durand, B. Laks, C. Lyche (dir.), en ligne: <https://www.projet-pfc.net>.
- TCOF: Traitement de corpus oraux en français, V. André, C. Benzitoun, E. Canut, J.-M. Debaisieux, C. Masson (dir.), en ligne: <http://www.cnrtl.fr/corpus/tcof>.

Annexe

Fiche descriptive du *ne* de négation du projet FRACOV

NE

Syntaxe

Le marqueur de négation *ne* est un clitique : c'est une forme « faible » face à *non*, qui n'entre en rapport qu'avec un verbe, et refuse d'entrer en relation syntaxique avec d'autres mots (expliquant l'impossibilité en réponse de **absolument ne* face à *absolument pas*) (Muller 1991 ; Recourcé 1996). Il est le premier du groupe de clitiques compléments antéposés au verbe.

Ainsi, on ne le retrouve pas avec les impératifs à clitiques compléments postposés.

1. pourquoi tu fais ça pourquoi tu fais ça mais fais-le pas (CFPP2000)
2. (* *ne*) fais-le pas
3. (*ne*) le fais pas

L'emploi de *ne* subsiste dans le vernaculaire à des taux d'environ 5 % des propositions négatives (Van Compernelle 2009), et à 0,5 % en français d'Amérique du Nord. Son identification peut être délicate, après un mot finissant par une consonne nasale comme *on*, et suppose un examen phonétique avancé (Larrivée & Ramasse 2013). Son absence est liée à une multitude de facteurs : l'environnement phonétique (selon qu'il est précédé d'une pause, ou apparaît entre consonnes), le caractère figé des séquences négatives comme *je sais pas* (Ashby 1981), le rôle pragmatique de l'emphase (Fonsecca-Greber 2007). La principale corrélation reste la nature du sujet, les clitiques amenant l'absence de *ne* à 94 % avec un sujet clitique et à 17 % avec un sujet lexical (Culbertson & Legendre 2008 ; voir le tableau synoptique de Meisner *et al.* 2016 : 19).

Interprétations

On reconnaît trois emplois de *ne* en français. Un sondage des 300 premières occurrences de *ne* non élidé dans le corpus CFPP2000 (Larrivée 2013) donne 92 % d'emplois avec une autre négation, et 6,7 % avec l'exceptif *que*. Les emplois de *ne* seul constituent 1,3 % des cas. L'emploi explétif qui lui est associé en français normé n'est pas attesté dans cet échantillon.

4. ils n'ont jamais rien eu (CFPP2000)
5. il n'y a que les gens du onzième qui adorent me dire tout ça (CFPP2000)
6. il exposait des écharpes il vendait les écharpes (mm) je ne sais Vlada (CFPP2000)

Ne marque ainsi la portée propositionnelle de la négation, et n'a jamais de portée de constituant¹⁰ (qu'a *rien* dans *il travaille pour rien*, où un travail est affirmé et son résultat nié).

Il est la source de la focalisation quand il s'emploie seul, et de la polarité négative, même si ces emplois que fait attendre le français normé restent à attester dans le vernaculaire.

Pierre LARRIVÉE

(21 décembre 2013, révisé le 21 juillet 2016)

10. On pourrait croire que puisqu'ils ne forment pas de proposition principale, l'infinitif et le participe présent ne permettent pas une négation à portée propositionnelle. C'est pourtant bien le cas comme le montre la paire minimale *ne travaillant pour rien au monde* face à *travaillant pour rien*, où dans le premier cas *ne* indique que c'est la proposition constituée autour de *travaillant* qui est concernée par la négation.