



N° 75 | 2020

Didactique du français pour les adultes natifs : de l'illettrisme aux compétences clés

---

## Des corpus pour développer les compétences métalinguistiques des adultes francophones natifs en insécurité langagière

*Virginie André*

---

### Édition électronique :

#### URL :

<https://revue-tdfle.fr/articles/revue-75/1509-des-corpus-pour-developper-les-competences-metalinguistiques-des-adultes-francophones-natifs-en-insecurite-langagiere>

**DOI :** 10.34745/numerev\_1308

**ISSN :** 2553-5994

**Date de publication :** 08/07/2020

Cette publication est **sous licence CC-BY-NC-ND** (Creative Commons 2.0 - Attribution - Pas d'Utilisation Commerciale - Pas de Modification).

---

Pour **citer cette publication** : André, V. (2020). Des corpus pour développer les compétences métalinguistiques des adultes francophones natifs en insécurité langagière. *Revue TDFLE*, (75).

[https://doi.org/https://doi.org/10.34745/numerev\\_1308](https://doi.org/https://doi.org/10.34745/numerev_1308)

**Mots-clefs :**

Didactique, Illettrisme, Apprentissage sur corpus/data-driven learning, Français écrit

---

# Des corpus pour développer les compétences métalinguistiques des adultes francophones natifs en insécurité langagière

Virginie André - ATILF - Université de Lorraine et CNRS

[virginie.andre@univ-lorraine.fr](mailto:virginie.andre@univ-lorraine.fr)

**Résumé :**

Cet article propose d'exploiter la démarche du *data-driven learning*, traduit en français par apprentissage sur corpus (ASC), habituellement mise en œuvre pour enseigner et apprendre une langue étrangère, afin de développer des compétences métalinguistiques chez des adultes francophones natifs en insécurité langagière à l'écrit. Plusieurs études, qui portent notamment sur le fonctionnement cognitif et linguistique des personnes illetrées, sont unanimes aujourd'hui pour affirmer que les problèmes de maîtrise de la langue écrite sont liés à leurs faibles compétences métalinguistiques.

Cet article propose une présentation et une analyse d'expérimentations explorant la démarche de l'ASC avec ce type de public. Cette démarche permet un travail sur la métalangue et l'analyse de la langue, indispensable pour entrer dans l'écrit. Ce travail montre que des adultes en insécurité langagière réussissent à développer des compétences analytiques métalinguistiques qui les aident à entrer dans l'écrit.

**Mots-clés :** Didactique, apprentissage sur corpus/data-driven learning, français écrit, illettrisme.

**Abstract :**

Since the seventies, studies have shown an inability for poorly educated speakers to mobilize the linguistic means necessary to make themselves understood unambiguously

in the spoken word. However, even today, the training offered to these adults in the framework of social and professional reintegration schemes mainly focuses on improving their written skills. The objective of this study is to propose a methodology that would allow the implementation of remedial activities in oral communication and that could respond to a certain number of language difficulties encountered by these speakers so that they can intervene efficiently in all communication situations encountered at the professional level.

**Keywords** : illiteracy, oral, didactic, low schooling, proficiency

## Introduction

L'intrication des recherches sur corpus et de la didactique des langues semble aujourd'hui une perspective intéressante. D'un côté les analyses (socio)linguistiques proposent davantage de connaissances sur le fonctionnement de la langue en situation. Les résultats de ces analyses, transposés à des fins didactiques, accroissent la qualité de l'enseignement et de l'apprentissage des langues. D'un autre côté les méthodologies qui sont associées à ces analyses suscitent des applications didactiques innovantes. La linguistique outillée ou les analyses distributionnelles, par exemple, peuvent, parfois contre toute attente, trouver un accueil enthousiaste chez des apprenants de langue. Dans cette perspective, je propose des exploitations didactiques de l'analyse sociolinguistique des interactions verbales ou des études sur le français parlé en interaction (Mondada 2002, Kerbrat-Orecchioni 2005, Traverso 2016). L'exploitation de corpus oraux et multimodaux pour enseigner et apprendre à interagir à l'oral fait désormais partie des méthodologies didactiques actuelles (André 2018, 2019 ; Etienne, Jouin 2019). Elles sont expérimentées par de nombreux enseignants et apprenants de Français Langue Étrangère (FLE) en France et dans le monde.

Les résultats positifs de ces expérimentations exploitant des corpus avec différents types de public et pour atteindre des objectifs didactiques variés (voir par exemple Kerr 2013, Di Vito 2013, André 2018) m'ont poussée à détourner les corpus oraux de leurs objectifs d'apprentissage de l'oral afin de tester leur exploitation à des fins d'apprentissage de l'écrit. Les adultes francophones natifs qui participent à ces nouvelles recherches sont intégrés dans des formations de remise à niveau en français. Ils sont tous en situation d'illettrisme à différents degrés ou, autrement dit, en insécurité langagière à l'écrit (Adami, André 2014). Plusieurs études, qui portent notamment sur le fonctionnement cognitif et cognitivo-linguistique des personnes illettrées (Emé *et al.* 2009, Emé *et al.* 2011), sont unanimes aujourd'hui pour affirmer que les problèmes de maîtrise de la langue écrite sont liés à leurs faibles compétences métalinguistiques. Cet article propose une présentation et une analyse d'expérimentations explorant la démarche du *data-driven learning* (Johns 1991) ou, en français, de l'apprentissage sur corpus (Boulton, Tyne 2014) et permettant un travail sur la métalangue, indispensable pour entrer dans l'écrit (Lahire 1993, Adami 2001). Je rappellerai tout d'abord le contexte et les méthodologies de l'exploitation des corpus pour enseigner et apprendre

une langue étrangère, en insistant particulièrement sur les corpus oraux et multimodaux. Je présenterai ensuite une première tentative de perfectionnement du français avec des adultes natifs, à partir d'un corpus écrit et d'un corpus multimodal. J'expliquerai pourquoi cette tentative m'a orientée vers l'utilisation unique de ce même corpus multimodal. Enfin, je présenterai les résultats des expériences menées avec ce corpus dans l'objectif d'améliorer les compétences métalinguistiques d'adultes francophones natifs en insécurité langagière à l'écrit.

## 1. L'apprentissage des langues avec et sur corpus

L'exploitation de corpus écrits, oraux et multimodaux à des fins didactiques commence à intégrer et à compléter les méthodologies plus classiques d'enseignement et d'apprentissage des langues étrangères. L'utilité des corpus n'est plus à démontrer dans la mesure où ils proposent aux apprenants une exposition à la langue authentique, réelle et située. La notion d'authenticité a été largement discutée depuis plusieurs décennies (voir notamment Duda *et al.* 1972, Holec 1990, *Mélanges CRAPEL* 2009, Cortier 2019) et aujourd'hui aucune méthodologie sérieuse ne devrait proposer à des apprenants une exposition à une langue construite. C'est prendre le risque qu'elle ne corresponde à aucune réalité socio-interactionnelle. Si la didactique de l'écrit semble tentée de respecter ces intentions et ces pratiques, la didactique de l'oral souffre encore d'un retard important (Boulton, Tyne 2014). De nombreux manuels, même les plus récents, spécialisés dans l'enseignement du français parlé, proposent encore des énoncés construits voire des dialogues entiers inventés pour de soi-disant objectifs didactiques pertinents (Giroud, Surcouf 2016).

Pour comprendre ce qu'est un corpus, nous pouvons nous accorder sur la définition de 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* . Le traitement didactique des corpus consiste tout d'abord à exposer les apprenants à cet échantillon de langue. Ensuite, les types d'exposition et d'activités cognitives réalisés avec les corpus permettront aux apprenants de s'approprier les compétences sociolangagières visées. Les corpus oraux et multimodaux utilisés en didactique des langues peuvent être de différentes natures (André 2019). En effet, face aux manques de données langagières orales (au-delà de la radio, la télévision, le cinéma), différentes ressources audiovisuelles authentiques (non construites à des fins didactiques) peuvent être utilisées pour enseigner et apprendre une langue. Tout d'abord, il est possible d'utiliser des corpus existants, notamment ceux constitués par les (socio)linguistes qui cherchent à accroître les connaissances sur le fonctionnement de la langue en interaction. En ce qui concerne le français, je pense notamment aux corpus TCOF, CLAPI, ESLO, OFROM, PFC ou encore CFPP2000<sup>[1]</sup>. Ces corpus librement accessibles (au moins en partie pour certains) proposent différentes situations de communication (Hymes 1972) qui peuvent être utilisées comme des documents supports lors de formations linguistiques. Ensuite, il est possible de créer des petits corpus pour atteindre un objectif didactique spécifique précis. Comme l'exprime une enseignante du

département de FLE de l'Université de Lorraine : *on n'a jamais le bon document, le document idéal, celui qui va nous permettre d'atteindre l'objectif didactique que l'on s'est fixé !*. Dans ce cas, certains enseignants se lancent dans la constitution de leur propre corpus. Cette entreprise est extrêmement lourde, quiconque ayant déjà eu à faire ce type de travail peut en juger. Néanmoins, certains constituent des petits corpus spécifiques pour, par exemple, exposer les apprenants à des interactions de service dans des lieux culturels (Janot 2017). Enfin, entre ces deux types de corpus, il en existe d'autres tels que des corpus par domaine (professionnel ou universitaire comme FLEURON<sup>[2]</sup>), des sous-corpus élaborés et parfois didactisés pour leur intérêt et leur pertinence pour des apprenants (c'est le cas de CLAPI-FLE<sup>[3]</sup>) ou encore des corpus pour des publics particuliers, par exemple pour des migrants qui peuvent avoir besoin de comprendre des bandes sonores (à la gare, dans le métro, dans un supermarché) ou encore des répondeurs téléphoniques (demande d'aides sociales, d'aides à l'emploi, de renseignements de différentes natures) dans le cadre de leur intégration sociale et professionnelle (Adami, André 2013).

Pour utiliser les corpus à des fins d'enseignement et d'apprentissage de la langue, et notamment en interaction, deux méthodologies didactiques se côtoient et se complètent (André 2018). La première consiste à utiliser les ressources du corpus une par une, en produisant une analyse interactionnelle conjointe - enseignant et apprenant. Dans ce cas, soit l'enseignant est formé à l'analyse des interactions, soit il bénéficie d'une transposition des résultats de l'analyse des interactions qui lui permet de décrire et de faire observer les mécanismes interactionnels et les pratiques langagières utilisées par les locuteurs en situation pour accomplir les activités visées (Alberdi *et al.* 2018). La seconde consiste à interroger un corpus avec un concordancier, selon la méthodologie du *data-driven learning* (Johns 1991) ou, en français, de l'apprentissage sur corpus (Boulton, Tyne 2014). Dans ce cas, l'apprenant peut observer et analyser les pratiques interactionnelles en cotexte et en contexte. Ces observations et analyses des occurrences permettent aux apprenants d'induire eux-mêmes les règles de fonctionnement des phénomènes recherchés dans le concordancier. Ces deux approches, avec et sur corpus, sont complémentaires et peuvent être mises en œuvre afin d'enseigner et apprendre à interagir. Ces méthodologies sont testées et expérimentées avec différents types de publics, notamment des apprenants-étudiants de FLE dans des centres de langues universitaires. L'apprentissage sur corpus a déjà fait ses preuves en ce qui concerne l'enseignement de l'anglais écrit par des étudiants (Boulton, Cobb 2017) et il commence à être mis en œuvre dans des formations linguistiques en FLE (André 2018). Les premiers résultats sont très encourageants et ces méthodologies méritent d'être déployées plus largement.

## **2. Des corpus au service de la maîtrise de la langue maternelle écrite**

L'exploitation de corpus à des fins didactiques suscite de nombreuses questions

quant à leur mise en œuvre, notamment avec un public de débutants (Boulton 2007). Après avoir réalisé différentes expérimentations auprès d'un public débutant et intermédiaire en FLE, ainsi qu'avec un public FLI (français langue d'intégration) débutant, j'ai souhaité expérimenter la méthodologie auprès d'un public d'adultes francophones natifs en insécurité langagière. Cette idée détourne complètement l'utilité première et l'origine de l'apprentissage sur corpus (désormais ASC) destiné à l'apprentissage de langues étrangères. Néanmoins, les expérimentations précédentes m'ont encouragée à tenter une telle méthodologie avec ce public. Si l'apprentissage sur corpus - de langues étrangères - avec des apprenants scolarisés dans leur langue première permet à ces derniers de mettre en œuvre des analyses métalinguistiques en étant guidé par les données, avec un public natif en insécurité langagière, il s'agit davantage de guider les apprenants pour qu'ils parviennent à réaliser ces analyses sans métalangage, sans conscience grammaticale et avec une faible compétence lexicosyntaxique. L'hypothèse émise dans le cadre de ce travail est que le développement de compétences métalinguistiques et de stratégies cognitives et métacognitives (Oxford 1999) va améliorer les compétences scripturales des adultes en insécurité langagière.

Les participants aux expérimentations ont été choisis parce qu'ils suivaient une formation linguistique visant à améliorer leurs compétences scripturales. Ils se trouvent tous en insécurité langagière à l'écrit, c'est-à-dire en situation d'illettrisme, avec des degrés de difficultés différents. A titre d'exemple, voici le profil social et scolaire de cinq

[4] des dix stagiaires\_\_\_ qui ont participé à la recherche :

Prénom	Sexe	Age	Niveau de scolarisation	Commentaires
Daniel	H	48	Jusque 3 <sup>ème</sup> Section d'Éducation Spécialisée	Fait partie des gens du voyage sédentarisés Recherche d'un emploi Ancien bucheron
Jordan	H	25	Jusqu'en 3 <sup>ème</sup> classique	Salarié d'une entreprise d'insertion (espaces verts). Souhaite passer le permis moto
Anthony	H	35	Section d'Éducation Spécialisée jusque 18 ans	Recherche un emploi
François	H	53	CAP agriculture/élevage	Salarié d'une entreprise d'insertion (espaces verts). Travailleur handicapé (physique)
Tristan	H	26	CAP hôtellerie/restauration	Salarié d'une entreprise d'insertion (espaces verts). Pris en charge par l'Agence Régionale de Santé (ARS)

Tableau 1 : Présentation de cinq adultes francophones natifs en insécurité langagière

L'objectif premier de cette expérimentation était de savoir si l'introduction des principes de l'ASC allait faciliter l'entrée dans l'écrit et permettre aux stagiaires de s'approprier plus aisément des règles de fonctionnement de la langue dans la mesure

où ils allaient les constater ou les dégager eux-mêmes. Ce test est également lié à l'hypothèse selon laquelle observer de façon répétée la forme d'un mot (ce que permet l'observation verticale des lignes de concordances), c'est-à-dire la façon dont il est orthographié, va faciliter son acquisition. Plus généralement, l'objectif est d'utiliser la démarche sur corpus pour développer une conscience langagière chez les participants. Ce fut un défi que j'ai souhaité relever tant les résultats avec des apprenants allophones, notamment avec les étudiants, étaient encourageants. Une grande part de l'enjeu réside dans le pari que ces personnes en insécurité langagière puissent réussir à endosser le rôle d'observateur, d'analyste et d'*apprenti-linguiste* décrit par l'ASC (Gavioli 2001, Bernardini 2004).

## 2.1. Une transcription pour tester le degré d'illettrisme

Afin de me rendre compte du degré d'illettrisme des stagiaires, j'ai commencé les séances expérimentales par un exercice de transcription. L'activité pédagogique de transcription est considérée comme une activité de découverte de la langue. Elle est utilisée dans le cadre de l'enseignement et de l'apprentissage du FLE mais aussi d'autres langues (Lebre-Peytard *et al.* 1977, Conrad 2004, Lynch 2007). J'ai utilisé une vidéo du corpus FLEURON (décrit en 3.1.) dans laquelle une employée présente quelques rayons d'une médiathèque. Voici une capture d'écran de la vidéo :



Figure 1 : Vidéo du corpus multimodal FLEURON

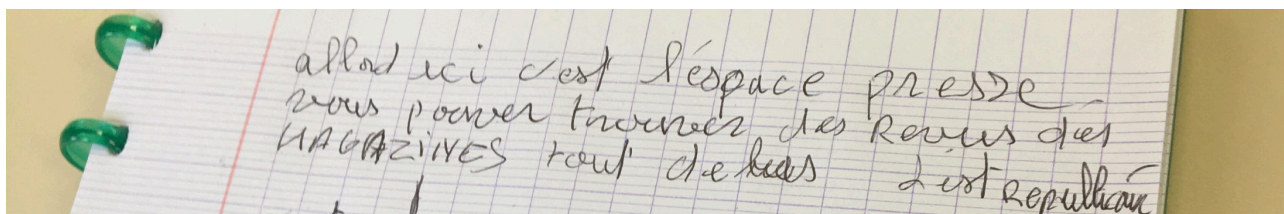
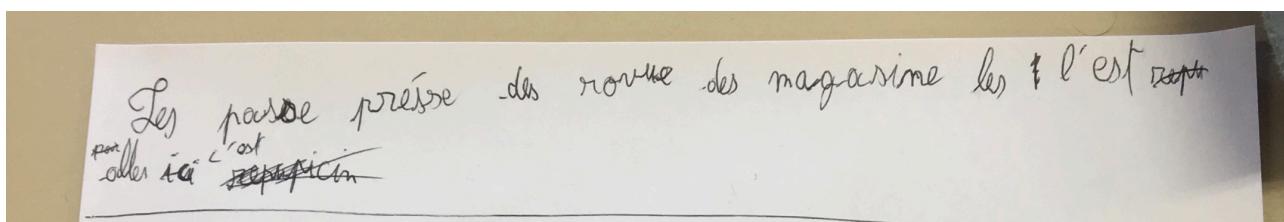
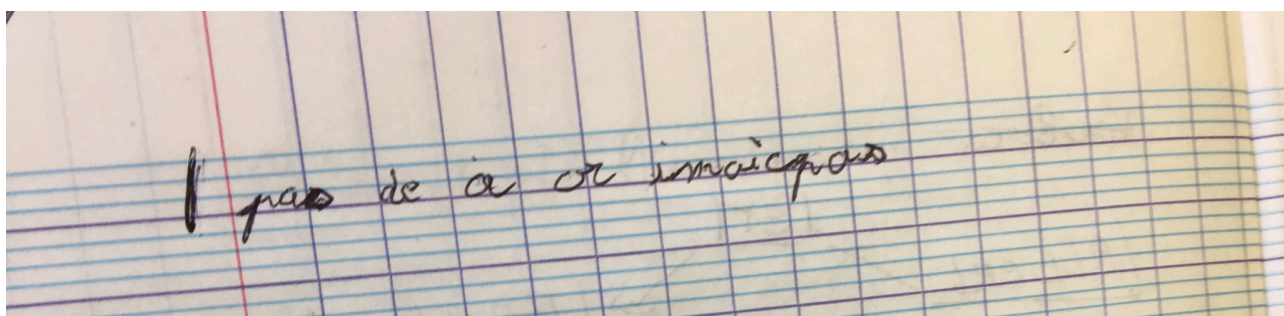


Le passage à transcrire concerne les informations relatives au rayon presse (ressource intitulée « Rayon presse de la médiathèque ») et dure onze secondes. Voici sa transcription :

*alors ici c'est l'espace presse vous pouvez trouver des revues des magazines au tout début ce sont les journaux donc l'actualité donc il y a l'Est Républicain notre journal régional*

Les participants ont pu visionner l'extrait filmé et l'écouter plusieurs fois, y compris par petits extraits.

Voici cinq photos représentatives du niveau des dix participants aux expérimentations :





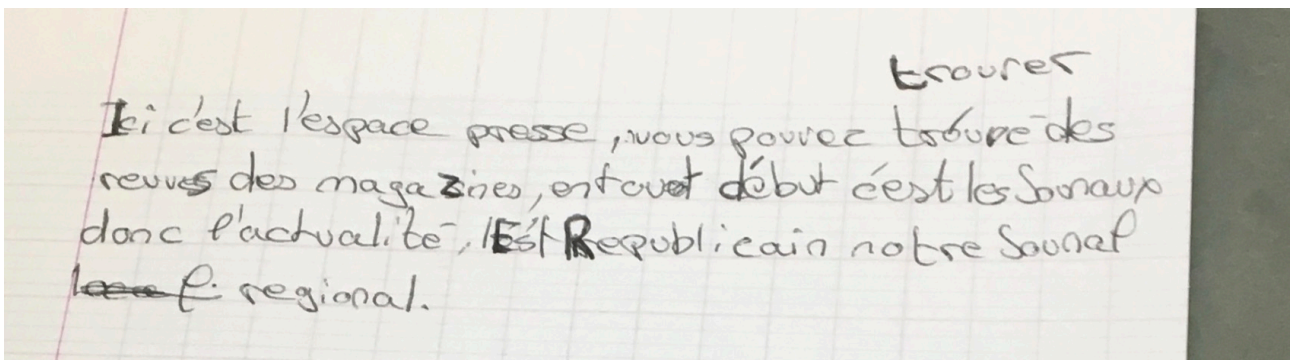
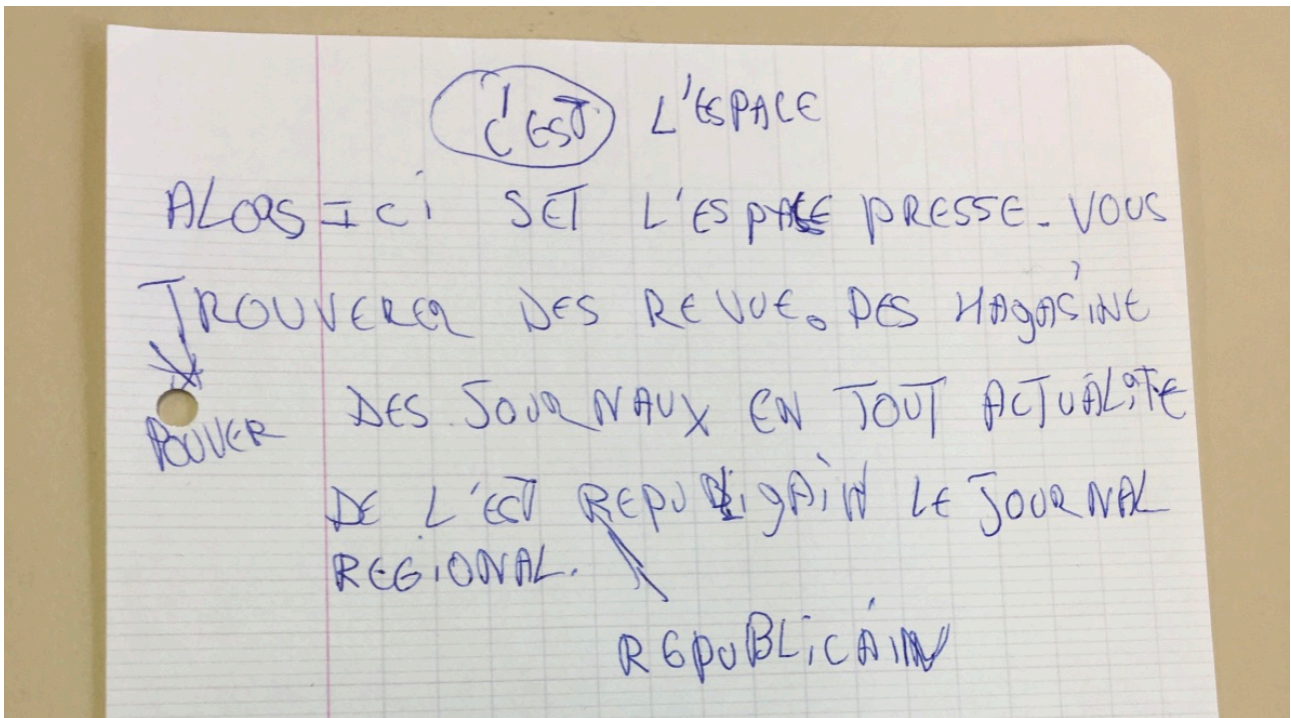


Figure 2 : Photos d'un exercice de transcription avec des personnes en insécurité langagière [5]

Ce premier exercice s'est révélé ludique pour les participants. Ils ont visionné le film puis ils ont tenté de transcrire les onze secondes proposées. Si les participants ont produit des textes différents, il est aisé de constater qu'ils ont tous des difficultés pour écrire. Les images de la figure 2 sont présentées dans l'ordre des difficultés des participants (de haut en bas, colonne de gauche puis de droite), des compétences scripturales les plus faibles aux plus élevées. On constate notamment des problèmes de transposition phonie-graphie et des problèmes de segmentation des mots. On voit également que certains ont recours à des pseudo-mots construits probablement à partir de mots déjà rencontrés. On remarque également des abandons, les participants renoncent plus ou moins rapidement à tenter d'écrire ce qu'ils entendent. Néanmoins, au-delà des difficultés rencontrées par les participants, avec la formatrice, nous avons pu constater que faire un exercice de transcription de français parlé avec ce type de public et ce type de document était bien accueilli par les participants. La visualisation de la situation et de la locutrice grâce à la vidéo a sans doute capté l'attention des participants. Le travail sur la transposition phonie-graphie n'est pas assimilé à une

dictée traditionnelle (énoncée par le formateur). Je ne vais pas détailler davantage cet exercice mais il aura été le point de départ de ces expérimentations consistant finalement, nous le verrons dans le dernier point, à détourner le dispositif FLEURON et l'apprentissage sur corpus.

## 2.2. Une séquence pédagogique à partir de corpus

La première partie de cette recherche a été consacrée à l'utilisation effective de corpus lors d'une séance de formation linguistique (remise à niveau prescrite par les employeurs des stagiaires). Pour mener à bien cette expérience, j'ai souhaité utiliser des corpus dont l'interface était la moins impressionnante pour un public qui ne maîtrise toujours pas les outils numériques (excepté l'usage de leur téléphone et la navigation sur certains sites internet spécifiques), même si je savais que j'allais moi-même

[6]

manipuler les corpus\_\_. J'ai choisi d'adapter une séquence pédagogique que j'avais déjà réalisée avec un public d'étudiants en FLE et qui avait bien fonctionné, tant du point de vue de la mise en place de la méthodologie de l'ASC que de l'apprentissage de la langue. J'ai adapté et simplifié une séquence d'observation de la négation en corpus (André à paraître 2020). Une formatrice en charge de ce public en insécurité langagière m'avait fait part du constat de la non-maîtrise de la négation en *ne... pas* de la part de ses stagiaires. Forte de ce constat et de mon expérience passée, j'ai planifié une séquence pédagogique sur la négation. Ce choix peut paraître surprenant dans la

[7]

mesure où cette difficulté peut sembler marginale pour ce public\_\_, mais il permettait une première séance d'expérimentation ainsi qu'une familiarisation avec les corpus.

J'ai utilisé uniquement deux corpus : l'Est Républicain, interrogeable *via* le site

[8]

ScienQuest\_\_ et OFROM (mentionné précédemment) qui possède son propre concordancier. Le premier corpus est un corpus écrit d'articles de presse et le second est un corpus oral de français parlé en Suisse romande et constitué à Neuchâtel. Cette première expérimentation a remporté un succès mitigé auprès des stagiaires. Je souhaite tout de même en rendre compte, dans la mesure où elle présente une première tentative d'exploitation de corpus à des fins didactiques et parce qu'elle a permis des ajustements didactiques et méthodologiques dans un domaine jamais expérimenté.

Tout d'abord, j'ai projeté le résultat de la recherche de « pas » dans le corpus de l'Est Républicain *via* le concordancier du site ScienQuest. Voici une capture d'écran :

1	Han par des candélabres en acier galvanisé noir n' a	pas	été prévu dans la délibération du 11 septembre 98 relative
2	Formule chère pour le cochon grillé Il n' en manquait	pas	« Après l' effort , le réconfort » , avait
3	5 décembre . Jacques Dodo , maire , n' a	pas	caché sa satisfaction de se trouver en présence d' une
4	une place de la mairie pourtant baignée de soleil ,	pas	la moindre ombre d' un bénévole , aussi bien du
5	un flash info déposé dans leur boîte aux lettres ,	pas	plus qu' un représentant du monde associatif prévenu lui par
6	bannière du « Printemps de l' environnement » n' a	pas	connu , à Fains-Véel , le succès que la municipalité
7	mettre un_peu de beurre dans ses épinards , n' hésite	pas	à sacrifier la sacro- sainte grasse matinée dominicale pour être
8	et l' Etat . « Bien_qu' il ne s' agisse	pas	d' un enterrement de première classe » , les agriculteurs
9	lors_de son séjour hivernal , le cirque Maximum n' a	pas	Depuis 4 ans , il sillonne les routes de France
10	bagages vers d'_autres cieux , la ville n' en fini	pas	Vendredi , le maximum sera une_fois encore réalisé pour séduire
11	une affinité plus marquée pour la pâtisserie , n' est	pas	loin_de regagner son fournil « parce_qu' il_y_a vraiment quelque_chose à
12	ne fait qu' une bouchée . « Ce n' est	pas	ça forcément qui nous ramène plus de clientèle , surtout
13	» , commente ainsi Nicole . « Je ne suis	pas	une très grande consommatrice de pain , mais je reste
14	l' accoutumée et sa réputation de sortie familiale n' a	pas	Le rallye a été préparé par l' équipe vainqueur l'
15	, boîtes , bouteilles et autres immondices très humains et	pas	naturels du tout , chaque équipe devait composer une statue
16	, Magali , l' épouse de Sébastien , ne cache	pas	sa surprise , car dans chacune des vitrines de la

Figure 3 : Extrait de la liste des occurrences de « pas » dans L'Est Républicain

Comme pour un public FLI avec lequel j'ai déjà exploité des corpus à des fins d'apprentissage de la langue, avec un public natif en insécurité langagière, peu scolarisé et peu habitué à manipuler des textes écrits, il faut expliquer le principe du concordancier. Cette première liste devait permettre aux stagiaires de se rendre compte de la présence de *ne* pour chacune des occurrences de *pas* (en excluant, avec des explications, les occurrences 4, 5 et 15 de la liste). Le constat a été fait, avec mon aide, et après de multiples incitations à l'observation du contexte gauche. Voici quelques extraits de mes explications <sup>[9]</sup> :

regardez bien regardez bien à gauche ((en montrant le tableau)) est-ce que vous repérez quelque chose un mot ou quelque chose qui est tout le temps là

(...)

regardez à gauche est-ce que vous voyez une forme qui revient tout le temps un peu avant pas une forme ou plusieurs formes qui se ressemblent ou même des lettres est-ce que vous réussissez à repérer des lettres à gauche qui sont présentes dans toutes les lignes

(...)

regardez avant pas même pas trop loin quand même il y a pas quelque chose qui revient + un mot + des lettres

Comme lors de l'explication du fonctionnement du concordancier, j'ai dû répéter les consignes d'observation en doutant encore aujourd'hui de leur totale compréhension par les stagiaires. Cependant, au bout de quelques allers-retours entre mes explications et les lignes de concordances, la présence de la première partie de la négation, prenant la forme de *ne* ou *n'*, a été remarquée. Un petit commentaire a été fait sur la différence entre les deux formes, selon la présence d'une voyelle ou d'une consonne. Cette remarque peut sembler anodine mais, pour un public qui ne réalise jamais ce type d'analyse métalinguistique, elle n'est pas si évidente.

Ensuite, j'ai proposé l'observation d'un corpus oral pour que les stagiaires constatent des différences avec le corpus écrit. J'ai donc projeté les résultats de la recherche de *pas* avec le concordancier du corpus OFROM, voici ce que les stagiaires pouvaient observer :

Locuteur	Texte
unifr11-esb	des choses où on n'arrive <b>pas</b> quoi
unifr11-esb	personne n'est <b>pas</b> bon en quelque chose il suffit de travailler mais
unifr11-esb	y en a qui sont <b>pas</b> f/ <b>pas</b> doués en math et d'autres en orthographe
unifr11-esb	oui je pense aussi mais après euh y a aussi euh des gens qui sont <b>pas</b> forts en orthographe et pis ça ben
unifr11-esb	ils comprennent <b>pas</b> beaucoup et pis euh
unifr11-esb	ouais le mot il leur vient <b>pas</b> directement juste alors c'est ça le gros problème en fait
unifr11-esb	euh oui assez quand même y a <b>pas</b> beaucoup de gens qui aiment l'ortographe euh
unifr11-esb	ça les gens ils vont <b>pas</b> non plus aller chercher dans un dictionnaire donc y a quand même des erreurs mais
unifr11-esa	ouais je pense ça va euh je sais <b>pas</b> hein mais je pense
unifr11-esa	des natels on on leur parle et puis ils nous donnent tout de suite ce qu'on veut je sais <b>pas</b> des trucs euh un peu
unifr11-esa	j'arrive <b>pas</b> trop à m'imaginer futur parce que
unifr11-esa	ouais je pense pis en plus ils avaient <b>pas</b> la DS la Gameboy euh
unifr11-esa	ouais euh on a ils ont <b>pas</b> des très bonnes notes non plus c'est
unifr11-esa	ceux qui ont MSN je trouve ouais ça influence euh <b>pas</b> mal
unifr11-esa	enfin le <b>pas</b> tout juste mais j'essaie de faire des efforts
unifr11-esa	non <b>pas</b> jus/ <b>pas</b> tout seuls comme ça
unifr11-esa	euh là je vais <b>pas</b> non plus aller intervenir chez tous les enfants qui vont sur internet mais

Figure 4 : Extrait de la liste des occurrences de « pas » dans OFROM

Avec cette liste, l'objectif était que les stagiaires constatent qu'à l'oral, la négation

pouvait se réaliser sans *ne*, mais pas systématiquement. Cet exercice a plutôt été un échec. Les stagiaires n'ont pas réellement compris où je voulais en venir et n'ont pas compris la différence, malgré les nombreux allers-retours entre les deux corpus et toutes les précisions sur le genre de discours (dans OFROM, il y a des narrations, des conférences et des discussions et ces métadonnées sont facilement accessibles), les locuteurs, les objectifs de la communication, etc. Ils ne perçoivent pas la différence entre *ne... pas* et *pas*. Je suis passée par une simulation pour tenter de faire comprendre au moins la différence entre l'oral et l'écrit, c'est-à-dire entre la possibilité de ne pas produire la négation complète et l'obligation de la produire. Cette simulation a consisté, dans un premier temps, à écrire un texto à un employeur (tous les stagiaires ayant participé à l'expérimentation étaient en emploi ou en formation et avaient un téléphone portable) pour prévenir d'une absence. Dans un second temps, ils devaient laisser un message vocal par téléphone à un ami proche qui venait de les inviter à manger alors qu'ils n'avaient pas faim. Cet exercice a été plutôt bien réussi. Le *ne* a été utilisé avec l'employeur (même si des stratégies d'évitement ont parfois été utilisées) mais pas avec l'ami (quasiment tous ont produit : *j'ai pas faim*). La variation a été perçue et ils ont finalement compris ce que je souhaitais leur montrer avec les listes de concordances. Ils ont eux-mêmes exprimé la variation diamésique (écrit/oral) en précisant ensuite qu'ils observaient davantage de variations en fonction du genre de discours et de l'interlocuteur. Ces analyses ont représenté un très bon résultat dans la mesure où elles semblaient montrer l'émergence d'une conscience métalangagière. Cependant, lorsque je leur ai montré de nouveau les listes de concordances, ils ne saisissaient plus la différence. Je souhaitais aborder la variation à l'oral avec OFROM mais ce nouvel exercice risquait plutôt de les perturber. Ce travail d'analyse leur a demandé beaucoup d'efforts et de concentration, j'ai dû l'abandonner avant d'obtenir des résultats plus satisfaisants.

Cette première expérimentation a tout de même été utile pour cette recherche et a orienté les suivantes. Même si je n'ai pas totalement mené l'exercice à son terme, j'ai pu me rendre compte que l'utilisation des corpus suscitait de la curiosité et un grand intérêt de la part des stagiaires. Cette façon nouvelle de travailler, et peut-être l'aspect expérimental, les ont poussés à se concentrer davantage que lorsqu'ils réalisent des exercices traditionnels habituels. Ce constat a été réalisé par les stagiaires eux-mêmes ainsi que par la formatrice. Enfin, cette première approche sur corpus a permis de révéler l'importance des vidéos et de leurs sous-titres et transcriptions en termes d'intérêt porté au discours. Les extraits sonores transcrits diffusés aux stagiaires retiennent moins leur attention que les vidéos sous-titrées.

### **3. Le choix de FLEURON pour améliorer les compétences scripturales**

Après la transcription pour saisir le niveau de maîtrise de l'écrit des individus et après l'expérimentation sur la négation, j'ai décidé de détourner le dispositif FLEURON avec ce type de public. L'expérimentation plutôt réussie de la transcription à partir des vidéos de FLEURON m'a poussée à tenter de nouvelles expérimentations mais



exclusivement avec FLEURON dont l'interface simple et ergonomique semblait faciliter l'entrée dans le corpus. En outre, l'utilisation d'un corpus multimodal (de vidéos) semble préférable et préférée compte tenu du rapport à l'écrit de ce public. Bien que je propose de travailler les compétences scripturales, je peux faire l'hypothèse que le passage par l'oral semble stimuler favorablement l'appréhension de l'écrit. Pour rappel, l'objectif est d'observer si les stagiaires – des adultes francophones natifs en insécurité langagière – sont capables de se lancer dans un processus d'observation, d'analyse et de classification malgré leurs problèmes de maîtrise de la langue et malgré leur conscience métalinguistique peu développée, pour dégager et comprendre des règles de fonctionnement (ici, plutôt des règles orthographiques lexicales et syntaxiques) à partir des transcriptions du corpus.

### **3.1. FLEURON**

Le projet FLEURON (Français Langue Étrangère Universitaire : Ressources et Outils Numériques) a développé un site internet (<https://fleuron.atilf.fr/>) qui propose un dispositif d'apprentissage du français parlé en interaction (André 2016). Le cœur de ce dispositif est constitué d'un corpus multimodal et d'un concordancier qui permet d'interroger ce corpus. FLEURON est destiné à des étudiants étrangers qui souhaitent faire un séjour universitaire en France et leur permet de se préparer aux interactions auxquelles ils seront susceptibles d'être confrontés dès leur arrivée et pendant leur séjour. Il s'agit notamment d'interactions à la scolarité, à la bibliothèque, au restaurant universitaire, avec des personnels administratifs, d'autres étudiants ou des enseignants. Le corpus compte également des interactions se déroulant en dehors d'un campus, tels que des échanges dans des lieux culturels, à la SNCF ou encore à la piscine municipale. Le domaine de ce corpus ne correspond donc pas, pour la majorité de ses ressources, aux adultes natifs qui composent le panel des expérimentations présentées dans cet article. Cependant, ce corpus a été choisi parce que les interactions sont filmées. Elles retiennent davantage l'attention (comparativement à l'expérience présentée précédemment avec le corpus sonore OFROM) des participants. Ces derniers s'intéressent aux échanges notamment parce qu'ils sont situés visuellement. Ils cherchent à comprendre les objectifs des interlocuteurs dans la situation de communication à laquelle ils participent. Les participants font même des commentaires sur les situations et les interactions qu'ils visualisent (sur les ingrédients des sandwiches du food-truck par exemple, dans une vidéo qui montrent un étudiant passer une commande).

FLEURON peut être utilisé selon les deux méthodologies exposées précédemment (partie 1). En effet, les vidéos de FLEURON peuvent être utilisées comme ressources pour observer et analyser le déroulement d'une interaction située. Dans ce cas, l'apprenant (accompagné ou non d'un enseignant) peut observer et analyser la vidéo accompagnée de ses sous-titres et/ou de sa transcription. Il comprend alors comment ouvrir une interaction, prendre son tour de parole, poser une question, exprimer son accord ou son désaccord, utiliser des petits mots de l'oral, etc. Ensuite, le concordancier permet d'interroger toutes les transcriptions du corpus et de présenter les résultats sous la forme d'une liste d'occurrences (voir la figure 5). La particularité de ce



concordancier est qu'il est multimodal ou aligné texte/son/vidéo (André, Ciekanski 2018). Cela signifie que l'utilisateur peut cliquer sur une occurrence pour accéder à la ressource correspondante et au moment où l'occurrence est prononcée par le locuteur. L'apprenant a ainsi accès au cotexte, plus ou moins large, ainsi qu'au contexte, chaque vidéo étant décrite par un titre et un résumé (lorsque l'image ne se suffit pas). Le concordancier permet donc de mettre en œuvre l'ASC avec FLEURON. Pour rappel, les principes de l'ASC tels qu'ils sont définis par Johns (1991, 1997) reposent sur trois phases [10] :

1. Observation des données d'un corpus via un concordancier.
2. Classification des données.
3. Généralisation des règles de fonctionnement à partir de la classification effectuée.

Ainsi, les apprenants observent tout d'abord les items qu'ils ont recherchés ou qu'un enseignant leur a suggéré ou demandé d'observer. Ils analysent leur distribution, c'est-à-dire qu'ils comparent leurs conditions d'apparition et leurs usages. Ensuite, ils identifient les items qui fonctionnent de la même façon, qui sont toujours précédés ou suivis des mêmes éléments, qui sont placés à tel ou tel endroit dans un énoncé ou dans une interaction, etc. et font des catégories. Enfin, ils induisent les règles de fonctionnement des items recherchés. Ainsi, les apprenants deviennent des détectives de la langue, des Sherlock Holmes voire des apprentis-chercheurs. Gavioli (2001) et Bernardini (2004) parlent de *learner as researcher* pour évoquer les activités dans lesquelles s'investissent les apprenants en analysant des lignes de concordances. En détournant FLEURON avec un public natif, je souhaitais découvrir si ce dernier allait parvenir à accomplir ces différentes phases, avec des objectifs didactiques différents. Dans ce cas, la transcription du corpus représente l'outil didactique de travail. Elle devra permettre d'observer la transposition phonie-graphie, la segmentation des mots et plus généralement la mise en orthographe des mots prononcés par les locuteurs du corpus ainsi que leur rôle syntaxique. La faiblesse lexico-syntaxique ainsi que l'absence de prise en compte du contexte pourraient être une des causes de l'illettrisme (Emé et al. 2011). Ce travail d'observation et de réflexion sur la langue doit amener les stagiaires à développer des compétences métalinguistiques, indispensables pour entrer dans l'écrit.

### **3.2. Les consignes et les explications préalables**

La présentation de FLEURON, de ses principes et de ses outils a été réalisée dès la première séance puis à plusieurs reprises ensuite. Voici des extraits de transcription d'explications et de consignes qui permettent de saisir la démarche méthodologique proposée aux stagiaires :

*Exemple 1 : VA réexplique le fonctionnement de FLEURON*

VA alors vous vous souvenez un petit peu de mes jeux

S1 euh

VA mais c'est pas grave je vais vous réexpliquer tout ça (...) l'idée c'est toujours on va essayer de regarder donc ça c'est un site Internet euh qui est je sais pas si vous vous souvenez peut-être que vous vous aviez pas eu la présentation c'est un site Internet qui est fait pour des étudiants qui veulent voir un petit peu comment ça se passe quand on arrive dans une université en France mais en fait il y a plein de choses sur ce site qu'on peut utiliser différemment et en fait sur ce site il y a essentiellement des vidéos et il y a par exemple même des vidéos de personnes qui vont acheter qui vont chercher euh un sandwich sur le campus par exemple je vous montre une petite vidéo et ici vous voyez on a un jeune homme qui va acheter un sandwich ((diffusion de la vidéo)) alors **ce qui va nous intéresser nous ici l'avantage de ces petites vidéos c'est que vous voyez elles sont complètement sous-titrées**

S2 hum

VA donc en fait nous **ce qui va nous intéresser c'est de voir comment on écrit ce qui est dit**

Les formations pour adultes permettent de nombreuses entrées et sorties permanentes de la part des stagiaires. Cela signifie que le début et la fin de la formation d'un stagiaire sont indépendants du début de la session de formation, les stagiaires peuvent intégrer la formation (et en sortir) à n'importe quel moment de l'année. En outre, le taux d'absences des stagiaires est souvent élevé. La répétition des consignes est, par conséquent, souvent nécessaire. Dans ce premier exemple, j'attire l'attention des stagiaires sur ce qu'est la transcription et sur le fait qu'avec eux, nous allons nous intéresser à « comment on écrit ce qui est dit ». D'autres formulations sont possibles, l'exemple suivant est prononcé par une stagiaire de Master 2 FLE de l'université de Lorraine :

*Exemple 2 : Une formatrice (Ej) en stage qui participe aux expérimentations donne des explications.*

Ej on a vu la semaine dernière qu'on avait plusieurs mots comme ça qu'on prononce pareil mais à l'écrit ils sont différents souvent c'est ça qui pose problème c'est quand on va les écrire les reconnaître et **choisir le bon avec la bonne orthographe**

Avec cette consigne, la formatrice insiste sur l'orthographe. C'est effectivement une amélioration de l'orthographe qui est visée par ces expérimentations, notamment de l'orthographe des homophones.

Avant de commencer, et tout au long des expérimentations, un travail de reformulation a été effectué. Les deux exemples suivants illustrent d'autres façons d'expliquer aux stagiaires l'objectif de nos séances avec FLEURON. L'exemple 3 précise que les vidéos et leur transcription vont servir de supports :

*Exemple 3 : Des explications sur la transcription.*

VA avec FLEURON FLEURON c'est le nom du site Internet **tout ce qui est dit est aussi écrit donc on va essayer de voir comment ça s'écrit** justement on va travailler nous l'écrivons ensemble à partir de ce genre de vidéo

Si la transcription est assez simple à définir (en faisant souvent référence au principe du karaoké), saisir le fonctionnement du concordancier est bien plus compliqué. L'exemple 4 propose une tentative d'explication :

*Exemple 4 : Des explications du concordancier avec la recherche de « c'est » en exemple.*

VA alors pourquoi j'utilise ce truc là ce site Internet là parce qu'il y a un truc magique dans ce site Internet si je reviens au début on a ici un petit outil ici bon ça s'appelle un concordancier peu importe c'est comme un moteur de recherche quand vous tapez quelque chose dans Google par exemple **et il va nous chercher tous les** c'est comme ça écrit comme ça ((en montrant le concordancier sur le site)) et il va aller nous chercher comme tout est vous avez vu on écrit tout ce qui est prononcé donc **ça va chercher dans tout ce qui est prononcé** et nous alors on va regarder + **pour comprendre un petit peu comment ça fonctionne on va regarder à partir de celui du haut là donc on va regarder plusieurs c'est et on va essayer comment ça fonctionne c'est (...)** si vous vous souvenez on peut aller voir dans la vidéo au moment où elle prononce ça

Les activités métalangagières qui vont être réalisées lors de l'utilisation du concordancier nécessitent des explications simples et répétées pour que les participants comprennent ce qu'il leur sera demandé de faire. Les explications sont orientées par l'objectif de l'exercice. Il s'agit de faire comprendre aux stagiaires que les paroles sont transcrites et que l'observation des transcriptions va leur permettre de voir comment les mots sont orthographiés. Je présente explicitement ce travail aux stagiaires en leur expliquant que je cherche de nouvelles méthodologies et de nouveaux exercices pour les aider à améliorer leur écriture. Les stagiaires sont donc partie prenante et sont intégrés à l'expérimentation, ils peuvent ainsi plus facilement dire ce qu'ils en pensent et proposer éventuellement des pistes d'amélioration.

### 3.3. Les analyses réalisées par les stagiaires

Les activités proposées à partir du concordancier de FLEURON ont été choisies selon les difficultés habituellement rencontrées avec ce type de public (voir 2.1.) et en collaboration avec la formatrice en charge d'un groupe-expérimentateur. Pour cette étude, j'ai choisi de travailler les homophones. La première séance a été consacrée à la différence entre *c'est* et *sait*. Ce couple fait partie des nombreux homophones qui posent des problèmes à l'écrit (pour les adultes en insécurité langagière - voir figure 2, image 4 - comme pour les enfants en cours d'apprentissage de l'écriture, de nombreux travaux sont consacrés à ce sujet, voir notamment David *et al.* 2006, Leray et Tyne 2016). J'ai choisi délibérément des mots éloignés linguistiquement (des points de vue morphologique, lexical, syntaxique et sémantique) pour faciliter la réussite de l'exercice, le but n'étant pas de piéger les stagiaires déjà en grandes difficultés.

La première étape de ce travail sur corpus a été de faire écouter dix extraits d'interaction aux stagiaires, comprenant l'une ou l'autre des occurrences (*c'est* ou *sait*), en leur expliquant pour chaque énoncé le contexte nécessaire à la compréhension. Les participants devaient cocher dans un tableau (dessiné préalablement au tableau) s'ils pensaient qu'il s'agissait de *c'est* ou de *sait*. La majorité de leurs réponses a été choisie au hasard, ils ont avoué qu'ils voyaient parfois qu'il y avait une différence mais ne savaient pas laquelle. Ce premier exercice a permis de valider le choix des occurrences à analyser avec eux. Ce type d'exercice a été réalisé lors de la première séance d'expérimentation de FLEURON avec ce public. Il a permis d'entrer en contact avec les participants en leur proposant un exercice qui pouvait ressembler à des exercices traditionnels ou connus et déjà réalisés (cocher des réponses dans un tableau) avant de les lancer dans l'analyse distributionnelle d'une liste d'occurrences affichée grâce à un concordancier. Cependant, cet exercice n'a pas été renouvelé.

La seconde étape a donc été l'interrogation du corpus multimédia de FLEURON grâce à son concordancier. Voici des extraits des résultats du concordancier pour les recherches de *c'est* et *sait* :

A gauche du mot	Mot recherché	A droite du mot
L2 :	c'est	polyuniversitaire européen
L2 : qui font le même travail que moi en fait	c'est	c'est au niveau de c'est une politique des sites
L2 : qui font le même travail que moi en fait c'est	c'est	au niveau de c'est une politique des sites
L2 : qui font le même travail que moi en fait c'est c'est au niveau de	c'est	une politique des sites
L2 : euh lorraine donc	c'est	c'est un peu ça
L2 : euh lorraine donc c'est	c'est	un peu ça
L2 : et donc	c'est	uniquement pour les étudiants et nous on est là
L2 :	c'est	-à-dire tout ce qui est euh
A : mais	c'est	bien vous avez déjà pris de l'avance
A : j'ai je vois votre contrat d'études	c'est	bien
A :	c'est	madame XX Dalida
A : donc	c'est	bon
E : mais	c'est	le niveau débutant
A : alors là	c'est	différent hein pour les cours d'espagnol je vais vous donner le nom du professeur E : XX
A : si	c'est	pas assez clair si j'ai mal écrit E : hum hum

Figure 5 : Extrait de la liste des occurrences de « c'est » dans FLEURON

A gauche du mot	Mot recherché	A droite du mot
E2 : étudier on	sait	pas parce qu'elle peut rester que 3 mois E : étudier aussi
A : garde-le on	sait	jamais hein E : oui bah je le garde dans mes affaires
A : euh ouais non mais on est au courant du problème pour vous on	sait	que bon
A : vous inquiétez pas bon on	sait	que de toute façon vous devez avoir votre année d'acquise mais euh voilà
A1: en période d'examens et cetera on	sait	que les étudiants sont pas trop disponibles
A1: c'est bon à savoir oui et puis ça se	sait	pas forcément
A : j'essaierai de ne pas vous téléphoner euh mais on	sait	jamais
A: le C deux l c'est connu on le	sait	le contenu hein de de C deux l
A: donc quand vous dites que vous avez le C2i on	sait	

Figure 6 : Extrait de la liste des occurrences de « sait » dans FLEURON

Tout comme avec les adultes migrants faiblement scolarisés, le travail d'analyse doit être accompagné, médié par un formateur ou un enseignant. Par exemple, je lis moi-même à voix haute toutes les lignes de concordances en donnant parfois des données cotextuelles et situationnelles pour faciliter la compréhension des occurrences. Ce procédé permet aux apprenants de comprendre les lignes sans avoir besoin de les lire puisque la lecture est problématique pour la majorité d'entre eux. Pour assurer la compréhension des extraits, le recours aux vidéos est possible. En cliquant sur l'occurrence choisie, la ressource s'ouvre, comme l'illustre la figure suivante qui correspond à la dernière occurrence de la liste de la figure 5 :

Retour Dossier d'inscription d'une étudiante polonaise

Media Description Conseils

Masquer les sous-titres

A : donc c'est bon

Concordancier

c'est

Options de recherche :

A : oui oui oui elle est polonaise elle est d'origine polonaise oui oui  
 A : vous avez reconnu d'après le prénom  
 E : oui  
 A : voilà donc je je ne précise pas professeur de polonais hein vous avez deviné  
 E : hum hum  
 A : et puis pour hum  
 A : littérature  
 E : non non non  
 A : non linguistique non  
 A : donc XX  
 A : donc **c'est** bon  
 E : sinon je peux faire un cours d'italien ou d'espagnol  
 E : mais **c'est** le niveau débutant  
 A : alors là **c'est** différent hein pour les cours d'espagnol je vais vous donner le nom du professeur  
 E : XX  
 A : hein je vais je vais vous note je note  
 E : d'accord  
 A : Marie XX  
 A : professeur d'espagnol hein donc vous pouvez lui envoyer un petit mail aussi

Figure 7 : Visualisation de l'occurrence dans la ressource vidéo de FLEURON

La vidéo est calée au moment où *c'est* est prononcé. Elle est accompagnée du sous-titre et de la transcription dans laquelle les occurrences de *c'est* apparaissent en gras. Dans la mesure où l'objectif est d'essayer de faire comprendre aux stagiaires l'orthographe de *c'est* et de *sait*, je ne m'intéresse pas à la variation des usages à l'oral et à l'écrit.

Malgré l'absence de métalangage, de conscience grammaticale, et malgré les problèmes de maîtrise de la langue de ce public, il parvient à analyser les occurrences et à induire des règles de fonctionnement. Dans l'encadré suivant, j'ai reproduit neuf extraits de la transcription d'une séance collective qui illustrent le processus d'analyse mis en œuvre. Les énoncés sont prononcés par les stagiaires à la suite d'une aide à l'analyse des données extraites par le concordancier :

1. *c'est* avec une suite ou plutôt ça sert à terminer + on le trouve à la fin comme avec *c'est* magnifique
2. *sait* + *c'est* le verbe savoir dans on *sait* jamais
3. *c'est* encore ouvert + ça veut dire il est encore ouvert le bureau
4. *c'est* on peut le remplacer par cela est



5. *on le sait pas forcément + ça veut dire on n'est pas au courant*
6. *c'est on peut le remplacer par ceci est*
7. *c'est bon on dit ça quand on est invité à manger on peut le remplacer par c'était bon*
8. *pour faire la différence on peut changer de temps + mettre au passé*
9. *je sais si c'est bien euh c'est compliqué comment faire il y a un autre exemple je sais que c'est difficile euh je sais c'est je suis au courant que le travail est difficile avec être ça marche mais là il y a les deux alors faut plus de temps pour réfléchir*

Ces neuf étapes sont extraites d'un long processus d'analyse que j'ai accompagné en pointant certains éléments, en faisant des allers-retours dans les ressources afin de montrer un cotexte plus large et de donner des informations situationnelles, en posant des questions, etc. Certains énoncés (2 par exemple) montrent que le stagiaire qui l'a prononcé possède un métalangage, même si celui-ci est restreint au seul terme *verbe*. A la fin de la séance, les participants ont fait un résumé :

*en fait quand on peut dire c'était on met un c et quand on sait quelque chose quand ça veut dire qu'on est au courant on met sait avec s a i t ((mot épilé))*

Cette conclusion montre que les participants ont induit leurs propres « règles », celles qui leur permettent d'écrire correctement les mots. On constate qu'ils parviennent petit à petit à développer une capacité d'analyse et de réflexion sur la langue (voir l'encadré ci-dessus) qui leur permet de comprendre l'écriture des mots. Les participants ont également développé leur compétence pragmatique, c'est-à-dire leur capacité à prendre en compte le contexte des énoncés. Cette compétence est souvent lacunaire chez les personnes en situation d'illettrisme (Emé et al. 2009). A la séance suivante, à laquelle je n'ai pas assisté, la formatrice a proposé un exercice de systématisation <sup>[11]</sup> Les stagiaires devaient écrire *c'est* ou *sait* dans des énoncés (authentiques) à trous. Ils ont réinvesti les règles induites en remplaçant les mots recherchés par : *cela est*, *c'était* et *on est au courant*. Les résultats ont été positifs, à la surprise de tous, des stagiaires et de la formatrice.

Forte de ces premières expérimentations, d'autres ont suivi, sans activité traditionnelle préalable, en travaillant directement avec les lignes de concordances. Je vais rendre compte d'un autre exercice réalisé quelques semaines après l'analyse de *c'est* et *sait* (entre temps d'autres analyses du même type ont été effectuées par les stagiaires) mais qui s'inscrit dans la continuité. J'ai proposé d'analyser la différence entre *c'est* et *s'est*. Là encore, je projette les résultats des recherches. Voici celui de

s'est :

A gauche du mot	Mot recherché	A droite du mot
E1: tu me disais... la dernière qu'on	s'est	vues
E : je ne sais pas ce qu'il	s'est	passé cette année
E1: mais maintenant je crois avec heu la création de l'université de Lorraine ça	s'est	étendu
E1: alors du coup Mélissa est-ce que tu peux me m'expliquer comment ça	s'est	passé ta recherche de directeur

Figure 8 : Liste des occurrences de « s'est » dans FLEURON

Tout d'abord, il est intéressant de mentionner que l'un des participants a rappelé le principe dégagé ultérieurement qui consiste à remplacer *c'est* par *c'était* pour savoir comment écrire le son entendu. Même si je ne m'intéresse pas ici à l'efficacité de l'ASC sur le long terme, je peux constater que les règles de fonctionnement induites sont retenues par certains participants. Ensuite, lors de cette analyse comparative, un nouveau fonctionnement a été identifié pour *c'est*, celui de présentatif qui n'avait pas été analysé aussi précisément lors de la séance précédente (avec *c'est madame XX Dalila*, figure 5). Par ailleurs, l'analyse de *s'est* est facilitée par les seules quatre

[12]

occurrences du corpus\_\_\_\_. Un vrai travail d'analyse peut être lancé sans que les participants se sentent noyés par les données. Dans ces situations de formation, la sélection de l'input présentée aux stagiaires peut être conseillée. La séance de travail dure une heure avec de nombreux allers-retours entre les deux listes d'occurrences (même si 45 minutes au total sont consacrées à l'analyse de *s'est*) et vers les ressources (et leur contextualisation). Au début de l'analyse de *s'est*, les participants avouent ne pas connaître cette forme et ne l'avoir jamais écrite. Néanmoins, après les différentes analyses réalisées des occurrences en contexte, les stagiaires parviennent à induire des principes et des astuces pour réussir à choisir entre les différentes orthographes. Voici en substance ce qui a été dégagé, progressivement pendant la séance, à propos de *s'est* :

- c'est le verbe être
- après il y a aussi un verbe ou des actions
- le s apostrophe s'écrit s parce que c'est comme dans *se voir*
- au passé on peut dire *on s'est vus*
- devant il y a *on/il/ça* il y a un sujet ou quelque chose avant
- *se voir* c'est *voir* mais ensemble mutuellement

- c'est faire une action au passé

Cet encadré ne reproduit pas les paroles exactes prononcées par les stagiaires mais l'aboutissement des analyses réalisées conjointement entre les stagiaires et moi-même. Ainsi, la première observation réalisée est l'emploi du verbe *être*, puis celle de l'enchaînement avec une action, puis la présence du *s*, etc. Ensuite, les stagiaires réalisent spontanément une activité de mise en pratique pour montrer qu'ils ont compris. Les apprenants de FLE et de FLI qui ne réussissent pas à verbaliser les règles de fonctionnement, même sans métalangage et même dans leur langue première, utilisent souvent cette stratégie afin de montrer qu'ils ont compris le phénomène langagier observé. Finalement, les stagiaires concluent en s'exprimant sur les avantages des listes de concordances. L'un d'entre eux explique : *c'est facile quand on regarde les exemples on peut trouver des astuces pour écrire*. Cette remarque correspond exactement aux objectifs fixés par cette étude. Le travail avec le concordancier permet aux stagiaires de les aider et les amener à réfléchir sur la langue. Un autre stagiaire reprend en donnant les exemples suivants : *on se fait une soirée* et *on s'est fait une soirée* et en expliquant que le second exemple signifie que la soirée est déjà passée. Cette même construction est reprise par un autre stagiaire avec *se faire des crêpes*. Ainsi, les stagiaires trouvent leurs astuces et les résumant en produisant la formule suivante :

quand on a *on/elle* ou une personne + la possibilité de mettre au passé = *s'est*

Les différents exemples que nous avons analysés avec les stagiaires ont ensuite conduit à l'analyse de la différence entre *on* et *ont*. Enfin, les analyses distributionnelles de *on* et *ont* nous ont amenés à revenir sur la distinction entre *peux/peut/peu* travaillée lors d'une séance précédente que je vais résumer ci-dessous. Un des avantages de travailler de cette façon avec un corpus est la possibilité de répondre aux besoins des apprenants au fur et à mesure qu'ils émergent pendant les séances de formation (Bernardini 2000). Ils ne sont pas obligés d'attendre la leçon consacrée à tel ou tel phénomène langagier pour le travailler, ils peuvent immédiatement le chercher dans le concordancier (ou le faire chercher par le formateur ou la formatrice).

Je terminerai donc par la présentation d'une expérimentation plus simple, pour les stagiaires, destinée à maîtriser les différences entre *peux*, *peut* et *peu*. Si l'on regarde les résultats du concordancier pour les apparitions de *peux*, *peut* et *peu*, le cotexte gauche semble aisément identifiable :

E : sinon je	peux	faire un cours d'italien ou d'espagnol
A : c'est ce que vous avez déjà apporté donc on peut mettre euh E : en fait est-ce que je	peux	le faire ici euh remplir ce document
E1: et le et le en bas tu	peux	garder ton pantalon E2: oui E1: oui d'accord
A : je	peux	écrire en minuscule votre adresse électronique
A : hein si je	peux	je l'envoie en fin d'après-midi
E : vous voulez je	peux	l'écrire
E : et demain je	peux	je viens A : vous pouvez venir le chercher oui si vous voulez E : bien
A : euh parce que là c'est vrai qu'effectivement là c'est hum ça va poser problème je	peux	pas vous faire euh
E : rien A : je	peux	pas vous faire l'inscription pour la carte parce que j'ai vraiment besoin de ce document
A : euh bah je	peux	

Figure 9 : Extrait de la liste des occurrences de « *peux* » dans FLEURON

Il est tout à fait envisageable de ne montrer que cette liste (extraite parmi les 111 occurrences de *peux*) aux stagiaires. La récurrence et la fréquence mises en évidence par le concordancier facilitent les analyses distributionnelles. Il est également possible de demander aux stagiaires de ne pas observer certaines occurrences en expliquant les raisons. C'est le cas pour la liste suivante qui présente des *peut-être* qui n'entrent pas dans les analyses demandées mais qui peuvent faire l'objet d'une parenthèse explicative.

A gauche du mot	Mot recherché	A droite du mot
L2 : alors il est important de savoir qu'on	peut	avoir des pôles universitaires
A : c'est ce que vous avez déjà apporté donc on	peut	mettre euh E : en fait est-ce que je peux le faire ici euh remplir ce document
E : hum hum A : parfois on	peut	vous demander hein on peut faire un petit contrôle et demander
E : hum hum A : parfois on peut vous demander hein on	peut	faire un petit contrôle et demander
E2: ouais ça	peut	ça ça prend toute la journée
A : pour euh dire que vous êtes bien arrivé en France	peut	-être E : oui oui
A : euh entre temps vous aurez	peut	-être vu monsieur Smith
A : est-ce qu'on	peut	prendre rendez-vous jeudi 14 heures?
A : vos parents peuvent	peut	-être appeler
A : hein ça	peut	se faire rapidement je pense
E : oui A : ça	peut	se faire rapidement je pense hein
A : bah vous avez dû avoir	peut	-être un numéro d'étudiant non?
E : ouais donc c'est un visa d'études A : oui une fois que la personne a le visa elle	peut	venir s'inscrire
E : l'année prochaine A : ah oui alors à ce moment-là il faudrait	peut	-être revenir euh
E : elle	peut	rentr- elle elle va rentrer en France elle peut rentrer en France parce qu'elle a un passeport X A : hum
E : elle peut rentr- elle elle va rentrer en France elle	peut	rentrer en France parce qu'elle a un passeport X A : hum

Figure 10 : Extrait de la liste des occurrences de « *peut* » dans FLEURON

La figure 11 présente également une partie des 154 occurrences de *peu*.

A gauche du mot	Mot recherché	A droite du mot
L1 : donc oui est-ce que vous pourriez déjà m'expliquer un petit	peu	quelle est votre fonction ici au pôle européen
L2 : euh lorraine donc c'est c'est un	peu	ça
A : toujours on vous laisse un petit	peu	de temps hein pour changer les cours hein E : hum hum super oui
A : euh vous pouvez oui si vous avez un petit	peu	de temps vous pouvez vous mettre euh
E1: d'accord oui pour faire les radios E2: oui c'était j'étais un	peu	mal à l'aise parce que
A : c'est bien donc attendez un petit	peu	et puis euh
A : mais si vous partez un petit	peu	avant je je vous ferai une attestation
A : donc Sarah donc je vais vous expliquer un petit	peu	la procédure hein
E : en français euh je connais un	peu	la langue parlée mais je sais pas très bien écrire en français en fait
A : parce que là c'est un	peu	tôt hein faut faire faut faire E : oui oui A : la demande quelques mois avant E : hum
A : vous pouvez attendre encore un petit	peu	euh E : hum hum A : il y a pas de problèmes euh
A : euh et on vous laisse un petit	peu	de temps pour faire le choix des cours
A : mais tu as encore un petit	peu	de temps hein pour choisir les cours euh
A : tu as encore un petit	peu	de temps on vous laisse le temps

Figure 11 : Extrait de la liste des occurrences de « peu » dans FLEURON

On peut remarquer que ce type d'analyse est tout à fait réalisable sans matériel informatique dans la salle de cours (les premières expériences de *data-driven learning* réalisées notamment par Johns (1991) ont été menées de cette façon). La séquence peut être préparée à l'avance et les listes de concordances peuvent être imprimées. Dans ce cas particulier d'analyse de la différence entre *peux/peut/peu*, les stagiaires peuvent réussir l'exercice sans davantage de contexte.

Au fur et à mesure de l'observation des résultats des recherches dans le concordancier, les stagiaires identifient rapidement que :

- *peux* est immédiatement précédé de *je* ou *tu*
- *peut* est immédiatement précédé de *on*, *elle* ou *ça*
- *peu* est immédiatement précédé de *un* ou de *un petit*
- *peu* se trouve dans l'expression *un petit peu de temps*, *un peu tôt*, *un peu mal*
- *peu* s'oppose à *beaucoup*

Une fois encore, en les amenant et en les aidant à réfléchir sur la langue, les stagiaires trouvent des astuces, voire des règles de fonctionnement, qui leur permettent d'écrire correctement le mot. La récurrence et la régularité semblent favoriser l'induction des règles voire l'acquisition des formes lexico-syntaxiques (Boulton 2017). Lors de cette

expérimentation avec les différentes formes de [pø], un stagiaire explicite parfaitement la démarche en disant que c'est la récurrence des formes identiques présentées en colonne, les unes en dessous des autres, qui l'a aidé à immédiatement comprendre pourquoi parfois *peu-* était suivi d'un *-x*, d'un *-t* ou de rien. Ainsi l'exposition contextualisée et volumineuse facilite l'appropriation des formes. L'objectif est le même que pour une langue étrangère, l'idée est de faire prendre le recul et la distance réflexive nécessaires aux apprenants pour saisir le fonctionnement de la langue. Les données auxquelles les stagiaires sont exposés sont analysées, caractérisées et généralisées, selon les préconisations de Johns (1991, 1997). Cette approche paraît plus pertinente qu'une approche explicite au cours de laquelle un formateur donne les règles de fonctionnement. Le public cible de ces expérimentations a été confronté à des règles à de nombreuses reprises, à l'école, en formation pour adultes, en formation d'insertion. Malgré ces règles énoncées et répétées, ces personnes restent en situation d'illettrisme, avec des gros problèmes pour orthographier les mots. Les stagiaires qui ont fait partie de cette recherche s'expriment sur l'utilité de la démarche. L'un d'entre eux m'a confié : *on a dû me répéter la règle des dizaines et des dizaines de fois quand j'étais gamin et dans toutes les formations que j'ai suivies mais j'ai jamais compris maintenant c'est bon je visualise et j'ai même compris*. Cette remarque, parmi d'autres, montre que la méthodologie de l'ASC permet aux adultes en insécurité langagière de développer des compétences métalinguistiques et métapragmatiques ainsi qu'une conscience lexico-syntaxique qui facilitent l'entrée dans l'écrit. L'extraction des données avec le concordancier, l'observation et l'analyse de ces données offrent de nouvelles perspectives pour de nouvelles pratiques didactiques.

## Conclusion

Parmi les défis que lance l'exploitation didactique des corpus (Debaisieux 2009), celui que j'ai tenté de relever avec des adultes francophones natifs en insécurité langagière semble ouvrir des perspectives intéressantes. Les applications présentées ici concernent essentiellement l'orthographe des mots, et plus particulièrement des homophones. Cela peut paraître un problème modeste dans une masse d'autres problèmes, parfois bien plus importants tels que celui de la cohérence d'un texte ou la capacité d'abstraction que nécessite le processus d'écriture, mais il me semble intéressant de le traiter d'une façon nouvelle. La réflexion que nécessite l'orthographe des homophones pose la question centrale des compétences métalinguistiques, indispensables à l'entrée dans l'écrit et qui peuvent être développées grâce à l'ASC. En outre, les problèmes d'orthographe permettent aussi d'aborder des éléments de syntaxe avec un public peu scolarisé en insécurité langagière à l'écrit. Adami (2018) tente également de répondre à cette question de la grammaire, en minimisant l'utilité de la grammaire explicite, avec un public suivant des formations d'intégration et d'insertion sociale, professionnelle et citoyenne. Avec nos expérimentations et l'utilisation du concordancier de FLEURON, la grammaire est bien abordée mais de façon tout à fait particulière, et surtout de façon implicite. En tentant de rapprocher lexique et grammaire en utilisant la démarche de l'ASC dans le cadre de la maîtrise de la langue écrite, il est possible d'envisager des perspectives intéressantes. Le problème de



l'homophonie que j'ai abordé avec des adultes en situation d'illettrisme a également été abordé par Leray et Tyne (2016) avec des enfants entre sept et douze ans du niveau CE2 à la sixième qui ont été soumis à l'ASC. Si le processus analytique et métalangagier mis en œuvre par les élèves n'est pas décrit, les résultats chiffrés de leur enquête sont encourageants. Ils soulignent par ailleurs que « la question de l'exposition aux formes écrites comme moyen d'avancer dans la maîtrise de l'orthographe semble suffisamment pertinente pour que l'on s'y intéresse par le biais de l'ASC » (*Ibid.* : 134). Certains de leurs résultats les amènent également à conclure que « l'ASC constituerait ici une base plus solide pour que l'enfant puisse ensuite progresser par lui-même » (*Ibid.* : 141). Je peux m'aventurer à faire le même constat avec un public adulte en insécurité langagière. Il s'agit de donner des habitudes analytiques métalangagières aux stagiaires et de leur donner quelques clés pour qu'ils identifient eux-mêmes les astuces qui leur seront utiles dans leurs futures activités de rédaction. Cette démarche peut sans aucun doute être mise à l'épreuve pour aller au-delà de ces astuces et pour faciliter des productions naturelles plus élaborées.

## Bibliographie

- Adami H. (2018). Existe-t-il une vie didactique sans la grammaire ? *Les langues Modernes*, 3, p.85-92.
- Adami H. (2001). L'oralité et la métalangue dans les rapports au langage des scripteurs/lecteurs en insécurité à l'écrit. *Mélanges*, 25. p.7-37.
- Adami H., André V. (2014). Les processus de sécurisation langagière des adultes : parcours sociaux et cursus d'apprentissage. *Revue française de linguistique appliquée*, Vol. XIX-2, p. 71-83.
- Adami H., André V. (2013). Corpus et Français Langue d'Intégration (FLI). *LINX*, n°68-69, p.135-158.
- Alberdi C., Etienne C., Jouin-Chardon E. (2018). Les apports des corpus d'interactions naturelles en situation de classe : enjeux et pratiques. *Action Didactique* 1, p.55-70. En ligne : <http://univ-bejaia.dz/action-didactique/pdf/ad1/Alberdi-Etienne-JouinChardon.pdf>
- André V. (à paraître 2020). Faire de la linguistique de corpus avec des apprenants de français langue étrangère. *Syntaxe et sémantique*.
- André V. (2019). Des corpus oraux et multimodaux authentiques pour acquérir des compétences sociolangagières. In Gajo L., Luscher J.-M., Racine I., Zay F. (eds), *Variation, plurilinguisme et évaluation en français langue étrangère*. Bern : Peter Lang, p.209-223.
- André V. (2018). Nouvelles actions didactiques : faire de la sociolinguistique de corpus pour enseigner et apprendre à interagir en français langue étrangère.

*Action didactique*, n°1, p.71-88. En ligne : <http://univ-bejaia.dz/pdf/ad1/Andre.pdf>

André V. (2016). FLEURON : Français Langue Étrangère Universitaire - Ressources et Outils Numériques. Origine, démarches et perspectives. *Mélanges Crapel*, 37, p.69-92. En ligne : <http://www.atilf.fr/IMG/pdf/art3.pdf>

André V., Ciekanski M. (2018). Apprendre à interagir à l'oral à partir d'un concordancier multimodal : effets sur le développement de la conscience langagière et sur l'autonomie de l'apprenant dans le dispositif FLEURON. In Dejean C., Mangenot F., Nissen E., Soubrié T. (coord.). *EPAL - Échanger Pour Apprendre en Ligne, Juin 2018, Grenoble, France*. En ligne : <https://hal.archives-ouvertes.fr/hal-01996477/document>

Bernardini S. (2004). Corpora in the classroom : An overview and some reflexions on future developments. In Sinclair J. (ed.). *How to Use Corpora in Language Teaching*. Amsterdam: Johns Benjamins, p.15-36.

Bernardini S. (2000). Systematising serendipity: proposals for concordancing large corpora with language learners. In Burnard L., McEnery T. (dir.). *Rethinking language pedagogy from a corpus perspective*. Frankfurt: Peter Lang. p.225-234.

Boulton A. (2017). Corpora in language teaching and learning, *Language Teaching* 50-4, 483-506.

Boulton A. (2007). DDL Is in the Details... and in the Big Themes. In Davies M., Rayson P., Hunston S., Danielsson P. (eds), *Proceedings of the Corpus Linguistics Conference: CL2007*. En ligne : <http://ucrel.lancs.ac.uk/publications/CL2007/>

Boulton A., Cobb T. (2017). Corpus use in language learning: a meta-analysis. *Language Teaching*, 67, p.348-393.

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

Conrad S. (2004). Corpus linguistics, language variation and language teaching. In Sinclair J. (dir.). *How to Use Corpora in Language Teaching*. Amsterdam: Johns Benjamins, p.67-85.

Cortier C. (2019). Des documents authentiques aux corpus oraux, vers une approche sociodidactique ? In Gajo L., Luscher J.-M., Racine I., Zay F. (eds), *Variation, plurilinguisme et évaluation en français langue étrangère*. Bern : Peter Lang, p.1997-208.

David J., Guyon O., Brissaud C. (2006). Apprendre à orthographier les verbes : le

- cas de l'homophonie des finales en /E/. *Langue française*, 151, p.109-126.
- Debaisieux J.-M. (2009). Des documents authentiques oraux aux corpus : un défi pour la didactique du FLE. *Mélanges CRAPEL*, 31, p.35-56.
- Di Vito S. (2013). L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE. *Linx*, 68-69, p.159-176. En ligne : <https://journals.openedition.org/linx/1519>
- Duda R., Esch E. et Laurens J. P. (1972). Documents non didactiques et formation en langues. *Mélanges pédagogiques*, p.1-48.
- Eme E. Chaminaud S., Bernicot J., Laval V. (2011). Capacités pragmatiques des adultes en situations d'illettrisme : compréhension du langage non littéral et connaissances métapragmatiques. *L'année psychologique*, vol. 111, p.3-39. En ligne : <https://www.cairn.info/revue-l-annee-psychologique1-2011-1-page-3.htm>
- Eme E., Reilly J., Almecija (2009). Compétences narratives et communicatives chez des personnes en situation d'illettrisme. *Revue européenne de psychologie appliquée*, 59, p.123-138.
- Etienne C., Jouin E. (2019). Constituer des ressources pédagogiques pour enseigner le français oral à partir des recherches menées en interaction. In Gajo L., Luscher J.-M., Racine I., Zay F. (eds), *Variation, plurilinguisme et évaluation en français langue étrangère*. Bern : Peter Lang, p.225-240.
- Gavioli L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In Aston G. (ed.). *Learning with corpora*, Houston: Athelstan, p.108-137.
- Giroud A., Surcouf C. (2016). De « Pierre, combien de membres avez-vous ? » à « Nous nous appelons Marc et Christian » : réflexions autour de l'authenticité dans les documents oraux des manuels de FLE pour débutants. *CMLF, SHS Web of Conferences* 27. DOI:10.1051/shsconf/20162707017.
- Holec H. (1990). Des documents authentiques, pour quoi faire. *Mélanges pédagogiques*, p.65-74.
- Hymes D. (1972). Models of the interaction of language and social life. In Gumperz J., Hymes D. (Eds), *Directions in Sociolinguistics. The Ethnography of Communication*. Oxford : Basil Blackwell, p.35-71.
- Janot S. (2017). *Constitution d'un corpus multimédia pour travailler la compréhension orale. Étude menée auprès d'un public migrant*. Mémoire de Master sous la direction de V. André, Université de Lorraine.
- Johns T. (1997). Contexts: The background, development and trialling of a

- concordance-based CALL program. In Wichmann A., Fligelstone S., McEnery T., Knowles G. (eds). *Teaching an Language Corpora*. Harlow: Addison Wesley Longman, p.100-115.
- Johns T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In Johns T., King P. (dir.). *Classroom Concordancing, English Language Research Journal*, 4, p.1-16.
- Kerbrat-Orecchioni C. (2005). *Le discours en interaction*. Paris : Armand Colin.
- Kerr B. (2013). Grammatical description and classroom application: Theory and practice in data-driven learning. *Bulletin suisse de linguistique appliquée*, 97, p.17-39.
- Lahire, B. (1993). *La raison des plus faibles : rapport au travail, écritures domestiques et lectures en milieux populaires*. Lille : Presses Universitaires de Lille.
- Lebre-Peytard M., Lieutaud S., Beacco J.-C., Malandain J.-L. (1977). *Le document oral brut dans la classe de français : la transcription de documents sonores authentiques*. Paris : Bureau d'études pour les langues et les cultures.
- Leray M., Tyne H. (2016). Homophonie et maîtrise du français écrit : apport de l'apprentissage sur corpus. *Linguistik online*, 78(4), p.131-150.
- Lynch T. (2007). Learning from the transcripts of an oral communication task. *ELT Journal*, 61/4, p.311-320.
- Mélanges CRAPEL* (2009). Des documents authentiques oraux aux corpus : questions d'apprentissage en didactique des langues, 31. En ligne : <http://www.atilf.fr/spip.php?rubrique584>
- Mondada L. (2002). Pour une linguistique interactionnelle. In Santacrose M. (Ed.). *Faits de langue - Faits de discours. Données, processus et modèles. Qu'est-ce qu'un fait linguistique ?* Vol 2. Paris : L'Harmattan, p.95-136.
- Oxford R. (1990). *Language Learning Strategies: What Every Teacher Should Know*. New York: Newbury House Publisher.
- Sinclair J. (1996). Preliminary recommendations on corpus typology. EAGLES Document TCWG-CTYP/P. En ligne : <http://www.ilc.pi.cnr.it/EAGLES/corpusTyp/corpusTyp.html>
- Traverso V. (2016). *Décrire le français parlé en interaction*. Paris : Éditions Ophrys.

---

[1]

\_\_\_ Les corpus sont respectivement disponibles aux adresses suivantes : <https://www.cnrtl.fr/corpus/tcof/>, <http://clapi.icar.cnrs.fr/>, <http://eslo.huma-num.fr/>, <http://www11.unine.ch/>, <https://www.projet-pfc.net/>, <http://cfpp2000.univ-paris3.fr/>.

[2]

\_\_\_ <https://fleuron.atilf.fr/>

[3]

\_\_\_ <http://clapi.icar.cnrs.fr/FLE/>

[4]

\_\_\_ Les adultes francophones natifs qui suivent des formations linguistiques sont appelés « stagiaires » dans le milieu professionnel, je conserve donc cette appellation.

[5]

\_\_\_ Certaines images comportent des éléments de correction apportés après l'exercice, tels que « c'est » entouré, « républicain » et « l'espace » (image 4) ou encore « trouver » (image 5).

[6]

\_\_\_ Au-delà de la non-maitrise de l'outil informatique de ce public, les associations et certains organismes de formation qui assurent les séances ne possèdent que très rarement des salles contenant des ordinateurs pour tous les stagiaires.

[7]

\_\_\_ Ce problème peut paraître également marginal pour des apprenants de FLE mais des enseignants ont exprimé leur impuissance pour trouver des explications pédagogiques, pourtant sollicitées par des apprenants.

[8]

\_\_\_ Disponible à l'adresse suivante : <https://corpora.aiakide.net/scientext20/>.

[9]

\_\_\_ Les extraits sont transcrits en respectant les conventions de transcription de TCOF (consultable ici : <https://www.cnrtl.fr/corpus/tcof/TCOFConventions.pdf>). Les + symbolisent des pauses courtes.

[10]

\_\_\_ Johns (1991 : 4) propose les trois termes : « *Identify - Classify - Generalise* ».

[11]

\_\_\_ L'apprentissage sur corpus étant une méthodologie particulière, elle n'exclut pas la mise en œuvre d'activités de systématisation. Ces dernières viennent compléter les activités de découverte que seraient

les activités d'observation et d'analyse réalisées avec le concordancier. Cependant, j'ai fait le choix de laisser ces activités à la discrétion des formateurs qui m'ont accueillie pour les séances d'expérimentation sur corpus.

[12]

\_\_\_ Au moment de ce travail, le corpus FLEURON ne comptait que quatre occurrences de *sait*. Le corpus étant sans cesse enrichi, ce chiffre a augmenté, il y a dix occurrences au moment de la rédaction de cet article.