



ISSN 1766-3059

ISSN en ligne 2260-7846

## Le corpus : un outil inductif pour l'enseignement-apprentissage de la grammaire

**Fanny Auzéau**

Université Paris-Est Créteil, France

fanny.auzeau@u-pec.fr

**Laure Abiad**

Université Paris 3 Sorbonne-Nouvelle, France

laure.abiad@univ-paris3.fr

Reçu le 15-03-2018 / Évalué le 09-08-2018 / Accepté le 08-10-2018

### Résumé

Cet article a pour objectif de mettre en évidence l'intérêt à utiliser un corpus en linguistique. Nous proposerons de revenir sur le concept de corpus en tant que tel, en linguistique et en didactique. A travers deux expériences, l'une en Français Langue Étrangère, l'autre en Français Langue Maternelle (ou considéré comme tel) nous analyserons les points forts et les faiblesses de cette pratique. Inscire l'apprenant dans une démarche collaborative de recherche nous semble important pour l'impliquer, le motiver et lui transmettre des techniques de chercheur.

**Mots-clés** : corpus, démarche inductive, analyse de la langue, dynamique motivationnelle

### The Corpus : An Inductive Tool for the Teaching and Learning of Grammar

### Abstract

This article aims to highlight interest in using a corpus in the fields of linguistic and grammar. We will summarize the concept of corpus in linguistics and in didactics. Through two experiences, one in French as Foreign Language and the other in French as L1, we will present the strengths and the weaknesses of this practice. In both cases, it engages learners to adopt a collaborative approach in their research, roots of motivation and researcher techniques.

**Keywords** : corpus, inductive reasoning, language analysis, motivation

De nombreuses critiques sont faites aux exercices de grammaire décontextualisés ou peu représentatifs de la langue à laquelle seront confrontés les apprenants en langue étrangère. Le choix du type d'activités, le vocabulaire et les tâches demandées demanderaient un réinvestissement concret, généralement effectué par l'enseignant en classe. Pour pallier le premier point, nous nous sommes intéressées

à l'analyse de corpus en classe de français langue étrangère et langue maternelle. Ce support généralement réservé aux spécialistes de l'analyse du discours se révèle être une aide précieuse en classe de langue, maternelle et seconde. L'intérêt du corpus en tant que ressource d'apprentissage se développant de manière croissante ces dernières années, nous avons souhaité l'utiliser pour diversifier nos pratiques et faire découvrir à nos étudiants un nouveau support. Dans cet article, nous ferons le point sur ce qu'est un corpus, ce qu'il apporte sur le plan de la linguistique appliquée, mais aussi nous verrons de plus près comment le travail sur corpus en français langue étrangère et en français langue maternelle est pertinent notamment par sa démarche inductive, et l'implication de l'apprenant dans ses activités. Malgré l'intérêt de cette démarche, force est de constater que ce qui fait l'intérêt même du travail sur corpus : sa diversité, son accessibilité, entre autres, présente, pour les mêmes raisons, des faiblesses. Enfin, nous présenterons deux utilisations concrètes de cette méthode en français langue étrangère et en français langue maternelle.

## 1. Le corpus

Un corpus donne des précisions sur un discours actualisé et situé. La transcription ou la retranscription ne sont pas vaines d'interprétation car « Transcrire de manière orthographique, c'est interpréter, comprendre, une situation langagière comme nous le faisons dans les interactions tous les jours. » (Boulton *et al.*, 2013 :18) bien que les chercheurs privilégient des recueils de données les plus authentiques possibles. (Gadet *et al.*, 2012 ; Tyne *et al.*, 2014). Cette interrogation renvoie également au dilemme établi depuis de nombreuses années sur la définition de « document authentique », de la limite de la didactisation d'un document pour l'exploiter en classe de langues étrangères. Cependant, ces supports sont riches sur les plans grammaticaux, segmentaux et suprasegmentaux également, car le linguiste, l'enseignant-chercheur ou l'étudiant ont donné des variables importantes grâce à des normes plus ou moins faciles d'accès pour un novice. Ce dernier peut avoir un regard sur une production « réelle », prendre du temps pour l'analyser et comprendre les éléments du discours.

Grâce à l'utilisation du numérique, le corpus devient un groupement de textes pouvant être interrogés à l'aide d'un logiciel de recherche (Gilquin & Gries, 2009). Il peut aussi bien relater des productions écrites qu'orales, de niveaux dits « standards » ou familiers. Le contact avec des exemples variés donne à l'apprenant la possibilité d'acquérir des connaissances implicites et explicites sur la langue cible. Ces contextes variés peuvent le pousser à trouver des régularités dans ces fragments langagiers, ce qui lui donne accès à une démarche inductive et à émettre

des hypothèses plus ou moins vérifiables. La multimédialité est un atout majeur de cet outil informatisé pour varier les sources de compréhension, d'analyse et de réflexion. Il peut proposer des formes oralisées ou écrites, des vidéos accompagnées parfois d'une transcription synchrone par exemple, ou d'enregistrements sonores produits dans des contextes plus ou moins facilitants. Cette pluralité des inputs et de l'usage du numérique est parfois remis en question, car il nécessite des compétences supplémentaires. Schaeffer-Lacroix (2014) contrebalance cette idée en s'appuyant sur les dires de Anthony (2013) :

*Les enseignants [...] ont besoin de pouvoir accéder de façon rapide et aisée au corpus, de pouvoir filtrer les occurrences afin d'obtenir l'affichage de seulement celles qui sont pertinentes, et ils doivent pouvoir afficher, sauvegarder, et éventuellement imprimer ces occurrences afin de pouvoir les réutiliser dans des matériaux d'apprentissage. De même, les étudiants à qui on propose un apprentissage basé sur des données empiriques [...] ont besoin d'un outil d'exploration qui leur permette d'accéder au corpus de façon rapide et intuitive. De plus, cet outil doit leur fournir les occurrences susceptibles de les aider à résoudre rapidement une tâche d'apprentissage [...].*

Cette disponibilité et cette rapidité d'exécution représentent des atouts majeurs selon nous. Braun (2005 :55) insiste sur le fait que l'apprenant devra s'appuyer sur des connaissances connexes, mais que l'enseignant devra faire preuve de rigueur dans le guidage de l'activité sans que cela n'altère la qualité du support.

### **- le corpus en linguistique**

La discipline que représente la linguistique de corpus prend de l'ampleur et acquière de la légitimité (Tognini-Bonelli, 2010). Dans le monde anglosaxon, cette pratique est présente depuis plus quelques décennies déjà. De nombreuses études (Hyland, 2002 ; Thompson, Tribble, 2001 ; Tribble, 2002) s'accordent sur le fait que cette utilisation améliore les compétences des pratiques du discours et des usages académiques de la langue. Toutefois, ce type d'activités n'est pas facile d'accès à tous les étudiants et est souvent chronophage. Comme l'évoque O'Sullivan (2010 : 22), les apprenants en langue en voient l'utilité dans l'organisation textuelle et en langue de spécialité, mais rarement à une échelle plus fine. Read & al. (2001 : 388) précisent même que les apprenants passent plus de temps à découvrir les règles, ce qui pour une appropriation inductive des règles de la structure de la langue présente plusieurs avantages : un temps de réflexion, d'échanges entre pairs et d'induction (Mouchot, 2003 : 27). Ces différentes étapes consolident les connaissances et rendent l'apprenant actif et peuvent servir de soutien à une analyse

en morphologie dérivationnelle ou flexionnelle (Buchs, 2008). Toutefois, la multiplicité des occurrences peut effrayer les étudiants ou confondre leurs analyses. Ce travail de sélection devra être pris en compte par l'enseignant en s'adaptant au niveau des apprenants. Exploiter cette méthode nécessite des bases théoriques de la linguistique de corpus qui pourront être transmises par la suite en classe. Le passage de la sphère linguistique à didactique nécessitera quelques étapes d'adaptation (Gavioli, 2005).

### - le corpus en didactique

Ce support numérisé donne une autonomie dans le parcours de formation de l'étudiant (Kraif, 2003). Cette méthode est moins classique. Brun *et al.* (2002: 224) apportent une précision importante pour le contexte dans lesquels nous avons mis en place ces expérimentations : « Pour l'enseignement des langues, il ne s'agit donc pas de mettre aujourd'hui un cours en ligne qui permette d'apprendre une langue parfaitement mais plutôt de réfléchir et de construire des systèmes qui aident tout un chacun à apprendre ce dont il a besoin au moment où il en a besoin, à approfondir ses acquis, à s'entraîner pratiquement à l'infini sur les points où il a des difficultés. » Nous espérons ainsi que les étudiants ne se cantonneront pas à une utilisation guidée et/ou scolaire de cet outil informatisé, mais de prendre l'habitude de se tourner vers ce support en cas d'interrogation sur l'emploi d'une forme nécessitant une clarification, comme ils le font et/ou faisaient en utilisant un dictionnaire comme le Trésor de la Langue Française en ligne peut également le faire. Cette démarche métalinguistique dépasse les murs de la salle de classe. Cela n'est pas sans rappeler le concept de "*Data-driven learning*" (Johns, 1997) qui permet à l'enseignant d'avoir un support pédagogique exploitable et à l'apprenant de mener un auto-apprentissage. McEnery & Xiao (2010) découpent en trois étapes cette approche des « i » : tout d'abord l'illustration d'un phénomène en contexte réel, puis l'interaction pendant la phase de travail d'analyse et enfin l'induction. Nous pourrions également ajouter le terme d'intériorisation de ce concept pour les règles d'usage de la langue (Di Vito, 2013). Deux types d'utilisations sont préconisées par Tognini Bonelli (2001) la première *corpus-based* permet de valider une théorie existante, d'exemplifier un fait de langue ; la seconde *corpus-driven* sert de guide à l'élaboration d'une théorie. L'étudiant changera de rôles selon ces caractéristiques : facilitateur ou médiateur pour le premier, observateur ou acteur (Landure, Boulton, 2010 ; Tyne, 2012). Comme dans toutes pratiques, le résultat final (l'appropriation d'un savoir) n'est pas certain ; quels que soient la méthode ou méthodologie employée, l'outil ou le support (Ladure, Boulton, 2010 : 15).

## 2. L'utilité de cette pratique collaborative

Le travail sur corpus permet d'interroger « de manière directe le système linguistique que développe l'apprenant » (Boulton *et al.* 2013 : 14), démarche plus scientifique que de se fier à leur intuition. Travailler sur un corpus, collecter des données, c'est travailler sur un outil « palpable » (Boulton *et al.* 2013 : 19), même si celui-ci est hétérogène, peu varié, c'est encourager une démarche méthodologique motivante pour l'apprenant, l'incitant à découvrir la « règle », la théorie. Cette démarche inductive qui vise à construire l'acquisition du savoir par l'étudiant est particulièrement intéressante en grammaire.

Demander aux apprenants de faire des recherches ciblées dans un corpus, c'est également les soumettre à un défi et les inscrire ainsi dans une dynamique motivante. Comme le précise Viau (Viau, 2009) dans ses recherches sur la motivation dans le cadre scolaire et universitaire, une activité motivante doit être signifiante aux yeux de l'étudiant, authentique, et représenter un défi car cette condition influe sur la perception de ses compétences. Le travail sur corpus répond à ces exigences et l'apprenant crée ainsi une dynamique motivationnelle toujours plus vertueuse. En effet, une recherche de corpus exige un engagement cognitif de l'apprenant, en sollicitant ses stratégies d'apprentissage, en l'encourageant à faire des liens, en l'incitant à formuler des hypothèses, en les confrontant, ce qui aura pour conséquence de renforcer sa motivation et son engagement dans l'activité ; comme l'indique Brophy : « une activité ne doit pas seulement offrir aux élèves la possibilité d'être actif, elle doit les amener à penser » (Brophy, 2004 : 35).

Dans certaines activités, l'enseignant peut donner accès à un corpus, comme ce sera le cas dans l'activité en FLE, ce qui contribue à renforcer l'autonomie de l'apprenant prenant en charge la responsabilité de l'activité. Si les étudiants travaillent de manière collaborative sur des recherches de corpus, mais chacun dans un domaine spécifique, par exemple, les occurrences de la forme négative dans différents domaines (la presse et des médias, les corpus oraux, la littérature etc...) et qu'ils doivent aboutir à une mise en commun de leurs recherches, leurs ressources seront donc « interdépendantes », ce qui favorise leur travail collaboratif et participatif, dont l'intérêt n'est plus à démontrer (Johnson & Johnson, 1989). En effet, travailler sur un corpus de manière collaborative doit être dirigé vers une « interdépendance positive des buts » (Buchs, 2008 : 59).

De plus, le travail entre pairs et la confrontation de leurs hypothèses dans le cadre d'un travail sur corpus est essentiel à l'acquisition de nouvelles notions. Les propos de Buchs (2008 : 61-62) nous semblent parfaitement éclairer notre position :

*[L]orsqu'elles sont gérées de manière constructive, ces confrontations augmentent la motivation à comprendre la position opposée, améliorent la compréhension des arguments et du raisonnement du partenaire et stimulent la prise en compte de nouvelles perspectives. On parle alors de régulation sociocognitive pour indiquer que les points de vue sont coordonnés et intégrés de manière à dépasser les confrontations initiales. L'intensité et le maintien des confrontations renforcent les effets positifs pour l'apprentissage.*

Outre l'intérêt acquisitionnel et interactionnel que représente l'analyse de corpus, il va de soi que, même dans le cadre d'activités grammaticales, le corpus confronte non seulement l'apprenant à des occurrences nombreuses et variées, mais également à l'aspect culturel de la langue cible, quand les apprenants-étudiants observent et relèvent les occurrences dans leur co-texte et leur contexte.

### 3. Les enjeux de l'utilisation de corpus

Ce travail sur corpus, généralement associé au travail de linguiste, ne peut être envisagé sans entrevoir les difficultés qu'il peut poser. La première question qu'il peut soulever, et à laquelle l'enseignant se doit de répondre, c'est la légitimité de l'activité, autrement dit, comment justifier auprès de son public l'intérêt de travailler sur un corpus d'énoncés ou de rechercher un corpus, plutôt que sur un texte authentique plus standard, rédigé, qui met en exergue des procédés grammaticaux, textuels visibles et reconnaissables directement par l'apprenant ? Il convient alors de devancer ces questionnements, en justifiant le choix de cette activité de recherche, qui met l'étudiant dans une position de « détective » (Johns, 1997 : 106), essentielle à toute démarche d'apprentissage et d'acquisition d'une langue maternelle ou seconde. Il est nécessaire également de penser l'activité corpus, davantage comme une activité de découverte, même si ce travail peut être approfondi : « Les corpus permettent un traitement en profondeur de l'*input*, en combinant la gestion du sens et le repérage conscient des données langagières par l'apprenant, lui permettant de faire évoluer son interlangue. » (Ciekanski, 2014 : 5). Par ailleurs, il convient de prendre en compte que cet apprentissage inductif ne convient pas à tous les profils d'apprenants, question soulevée par Boulton & Tyne (2008). Certains profils restent convaincus par une approche plus traditionnelle et déductive de la grammaire, c'est le cas de certains étudiants en FLE, dont la culture d'apprentissage est proche d'un enseignement magistral transmissif.

En outre, une autre problématique qui peut surgir lors d'un travail sur corpus est liée aux corpus oraux. Ceux-ci posent une difficulté méthodologique supplémentaire « la collecte, transcription et mise en forme de données orales sont coûteuses et la

composante orale des corpus existants se limite bien souvent à des transcriptions d'émissions de radio ou de télévision » (Boulton *et al*, 2013 : 12). La transcription ne s'avère pas totalement fiable et la collecte de données extrêmement chronophage. Une activité liée au corpus ne peut pas correctement fonctionner si celle-ci n'est pas suffisamment préparée par l'enseignant, il s'agit d'adapter le corpus en fonction de la problématique à travailler en particulier sur des corpus oraux. Manier l'outil peut s'avérer complexe et chronophage, les consignes et étapes de l'activité devront clairement être préparées en amont et clarifiées par l'enseignant : « si les buts escomptés sont clairs, les résultats concrets en termes d'apprentissage sont difficilement vérifiables, dépendant en large partie de la manière dont l'étudiant entre dans la tâche, et n'étant visibles qu'à travers des comptes-rendus d'apprentissage et des sondages d'attitudes et d'opinions. » (Tyne, 2009 : 108). Mais cette difficulté ne se limite pas aux corpus oraux, en effet, le travail sur corpus pose la question de la quantité de données collectées. Mettre un apprenant face à une représentation abondante de données, aussi pertinentes soient-elles, et même en lui demandant d'adopter une position de chercheur, peut paraître décourageant dans un premier temps. Cette difficulté peut être surmontée en proposant systématiquement un travail sur corpus collaboratif. En outre, rappelons ici également la nécessité de travailler sur des corpus variés, représentatifs de la grande variété des usages de la langue, bien qu'une difficulté de ce travail soit précisément de mesurer la variété des échantillons (Tomasello, Stahl, 2004). Ciekanski précise, à juste titre que « la plupart des corpus sont éphémères, notamment parce que la description des contextes d'extraction permettant de les réutiliser dans un autre contexte d'apprentissage n'est pas fournie. » Il est nécessaire alors, avant de débiter un travail sur corpus de sensibiliser l'apprenant, non seulement sur la pertinence de cette démarche didactique, mais aussi de le faire réfléchir en amont, sur la définition d'un corpus et sur l'intérêt qu'il apporte dans le fonctionnement d'une langue ou de sa langue. A ce titre, Rastier (2004) rappelle qu'« aucun corpus ne représente la langue : ni la langue fonctionnelle qui fait l'objet de la description linguistique, ni la langue historique, qui comprend l'ensemble des documents disponibles dans une langue. En revanche, un corpus est adéquat ou non à une tâche en fonction de laquelle on peut déterminer les critères de sa représentativité et de son homogénéité.» L'enseignant se doit donc de prendre en compte cette « subjectivité » du corpus en proposant des activités d'analyse de corpus et doit également en informer son apprenant.

#### 4. Les protocoles

Dans nos expérimentations, nous nous sommes intéressées à l'exploitation de corpus dans un contexte de Français Langue Étrangère et Langue maternelle en contexte universitaire en France. Nous avons centré nos pratiques sur l'observation de la langue ce qui donne un effet de loupe sur les usages en contexte. Cependant, comme il est courant d'observer dans les parcours en sciences du langage, le fait de créer un corpus ou de retranscrire un extrait de dialogue nous semblaient plus délicat pour le public en langue étrangère, non coutumier de ce type d'activité, et chronophage pour des étudiants de langue maternelle.

##### - en Français Langue Étrangère

Nous avons mené une expérience d'utilisation de corpus dans un cours d'argumentation destiné à des étudiants étrangers en classe de Français Langue Étrangère de niveau B1 (d'après le Cadre Européen Commun de Référence pour les Langues). Ces apprenants ont pour objectif de poursuivre leurs études au niveau Master dans des universités françaises, pour la majorité d'entre eux, dans les domaines du marketing, de la gestion et de la traduction. Ce groupe se compose de 18 étudiants issus du continent asiatique (Chine, Corée, Vietnam, Taïwan) ou du continent européen (Espagne, Italie, Turquie). Leur formation de 20 heures d'enseignement obligatoire inclut les quatre compétences (compréhensions écrite et orale, productions écrite et orale) avec un objectif universitaire de type FOU (Français sur Objectif Universitaire) pendant 13 semaines.

Le cours d'argumentation a pour vocation d'apprendre à structurer ses idées, les organiser de manière logique et pertinente pour convaincre et persuader à l'oral comme à l'écrit. Les attentes à ce niveau selon le CECR sont les suivantes à l'oral dans le cadre d'un monologue suivi : « Peut développer une argumentation suffisamment bien pour être compris sans difficulté la plupart du temps. Peut donner brièvement raisons et explications relatives à des opinions, projets et actions. » (2001 : 50) et dans une production écrite « Peut écrire des textes articulés simplement sur une gamme de sujets variés dans son domaine en liant une série d'éléments discrets en une séquence linéaire. » (*ibid* : 51).

Les apprenants avaient déjà été sensibilisés lors des séances précédentes à différencier les arguments et des exemples dans un paragraphe organisé, à découvrir la structure d'un texte argumentatif et d'en définir le point de vue. Nous souhaitons maintenant améliorer leurs compétences dans la maîtrise de structures verbales et grammaticales suivies de l'indicatif et du subjonctif. Ce dernier est introduit à ce niveau (Beacco *et al* :54 et suivantes). Comme nous l'avons évoqué dans la partie précédente, nous souhaitons mettre en place une démarche inductive, rendre les

apprenants acteurs dans leur enseignement/apprentissage et promouvoir l'entraide et la production d'une tâche coopérative.

Nous avons utilisé la plateforme Fleuron qui nous permettait d'avoir un accès libre aux productions écrites et orales grâce à un concordancier (Foucou, Kübler, 2000). Cet outil permet aux apprenants de retrouver les items recherchés dans un corpus écrit puis de les mener à la production orale en contexte. Cette plateforme traite de situations connues par les apprenants de FLE puisqu'elles traitent des situations d'inscription, de choix de cours, de dialogues entre étudiants. Cette adéquation avec leurs attentes nous a paru importante, car, comme nous l'avons mentionné : maîtriser l'outil qu'est le corpus en ligne demande déjà une certaine compétence.

Dans un premier temps, en nous appuyant sur le principe du *corpus driven* (cf. *supra*). Notre choix s'est porté sur cette pratique, car il nous paraissait délicat de confronter les apprenants en langue étrangère à une quantité importante de données. Pour cela, nous avons préparé une grille afin que les apprenants soient guidés tout au long de cette tâche. La consigne leur a été donnée de compléter le tableau en observant le type de phrase (affirmative, négative), le verbe, la position de l'énoncé dans la phrase, les signes de ponctuation éventuels. En binômes, ils devaient rechercher les expressions suivies du mode indicatif ou subjonctif. L'utilisation de ces expressions et de ce métalangage n'a pas été effectué au préalable pour ne pas dévoiler le résultat final. Munis de leurs téléphones, les apprenants ont commencé leurs recherches. L'outil du concordancier a nécessité des explications supplémentaires. Les étudiants habituellement à l'aise dans les tâches dites « classiques » l'étaient également, ceux, plus en difficulté, ne comprenaient pas l'objectif à atteindre. Ces derniers lisaient, écoutaient les items trouvés sans activer leur esprit d'analyse. Après une vingtaine de minutes, nous avons procédé à la mise en commun (document projeté grâce à un vidéoprojecteur), les apprenants ont décelé de nombreux éléments concernant la syntaxe, la ponctuation et, du point de vue grammatical, ont soulevé le fait que certaines expressions étaient suivies de l'imparfait. Cette réponse, bien qu'erronée, a permis de découvrir les similitudes et différences des formes de l'imparfait et du subjonctif, mais également de revenir sur la concordance des temps.

Nous restons convaincues de l'utilité de cet outil pour son aspect inductif qui développe des stratégies d'auto-apprentissage et de réflexion métalinguistique, mais nous avons observé certaines limites, notamment le caractère chronophage de la mise en commun et la difficulté de distanciation. Cela est sans doute dû aux cultures scolaires des étudiants. Les étudiants issus des pays européens et asiatiques semblent plus enclins à cette démarche, tandis que les étudiants des

continents sud-américains et africains le sont moins. Toutefois, cette conclusion est à relativiser, car notre expérimentation a été testée sur un seul public et le nombre d'étudiants concerné (18) est trop faible également.

- **en Français Langue Maternelle**

Nous avons conduit une expérience de travail sur le corpus dans un cours d'énonciation destiné aux étudiants de L2 de Langues Etrangères Appliquées. L'objectif de ce cours est de doter les étudiants d'outils théoriques et méthodologiques afin qu'ils puissent analyser les différents genres de textes qu'ils pourront être amenés à rencontrer dans leur vie professionnelle de rédacteur ou de traducteur. L'étude de l'énonciation, de la subjectivité dans le discours, autrement dit les traces du sujet parlant ou écrivant dans son discours oral ou écrit, nous a conduits vers la deixis personnelle, et à nous interroger sur les valeurs du *Je*, *tu*, *on*, *nous*, et *vous*. Le pronom *on* ayant différentes valeurs (indéfini, générique, inclusif, exclusif...), il était donc intéressant de partir de leurs propres usages du pronom *On* pour établir une classification. Il leur a donc été demandé de partir d'un relevé d'énoncés, d'un corpus produit par les étudiants, qui devaient s'interroger et interroger leurs pairs sur tous les usages possibles du pronom *On* et les relever. L'étudiant devient alors enquêteur et responsable de la fabrication du corpus, ce qui encourage sa motivation et son sentiment d'*ownership* (Seidhfolher, 2002 : 222).

Ils devaient ensuite établir une classification des différents usages de *On* qui présente un intérêt majeur en linguistique, puisqu'il revêt différentes valeurs. Ce travail collaboratif entre pairs, leur a permis de confronter leurs énoncés, puis de faire des hypothèses sur le classement de ceux-ci. Lors de cette étape, ils ont eux-mêmes éliminer les doublons (deux usages du *on* générique par exemple : *quand il fait chaud, on doit boire pour s'hydrater / aujourd'hui, on dirait qu'il fait froid*) ce qui les a naturellement conduits à se questionner sur ces différentes valeurs. Leurs recherches ont donc abouti au classement en catégories suivantes :

1. *On* indéfini renvoyant à un ensemble vague (ex : *on dit que tu vas être renvoyé*)
2. *On* renvoyant à un ensemble de personnes identifiées (ex : *En France, on aime le pain*)
3. *On* ayant une valeur spécifique mais non précise (ex. *On nous écouter derrière la porte*) ou générique renvoyant à un sujet universel (ex. *On doit boire pour s'hydrater*)
4. Les valeurs du *On* exclusif, avec dans cette catégorie les différentes personnes désignées (ex. *On va m'opérer demain* où *on* désigne les médecins = *ils vont m'opérer demain*) mais *On* peut renvoyer également à *il* (ex. *On dit quoi ? on dit merci* ou *on = il (un enfant)* etc.

5. Les valeurs du *On* inclusif, renvoyant à *nous* (ex. *On arrive dans deux minutes !*), etc. De leur corpus, un usage beaucoup plus actuel de *On* a émergé où *on* renvoie à *je* (ex. *ça se passe bien la fac pour toi ? ça va, on s'adapte*).

L'intérêt de cette activité ne résidait pas dans le fait d'être exhaustif mais de leur faire prendre conscience que le pronom *on* pouvait renvoyer à toute la gamme des pronoms personnels (certains dans des usages familiers, bien entendu), ce qui entraîne d'un point de vue grammatical, des changements d'accords, dont ils n'avaient pas forcément conscience auparavant. En outre, inscrire les étudiants dans cette activité collaborative, déductive, dans laquelle ils sont en position de chercheur, a été très moteur pour le reste des activités de ce cours, car elle a permis de susciter chez eux un questionnement constant sur les normes et usages de la langue. Cette démarche s'avère essentielle dans leur cursus universitaire en langues, car ce qui a été fait dans le cadre d'un cours sur l'énonciation peut être décliné dans les langues étudiées.

## 5. Conclusion et perspectives

Ces deux domaines d'utilisation et d'exploitation des corpus nous ont permis de mettre en exergue l'intérêt de ce type de pratiques dans l'enseignement-apprentissage d'une langue première ou étrangère. Nous avons également pu identifier les difficultés méthodologiques que cette approche pouvait soulever tant sur le plan de la manipulation de l'outil que sur la démarche, généralement nouvelle pour les étudiants concernés. Le balisage nécessaire dans toute tâche ou activité pédagogique est fondamental dans ce type d'approche, car la masse d'informations à laquelle peut être confrontée l'étudiant peut le rendre réticent à ce type de pratique. Néanmoins, nous restons persuadées que cette démarche inductive, qui met l'apprenant au cœur de la réflexion sur la langue, est essentielle pour mettre l'étudiant face à la variété des usages et normes et développer ainsi son esprit critique.

L'utilisation de corpus nous a ouvert de nouvelles pistes de réflexion et d'utilisation dans le domaine de la prosodie et de l'analyse des contours suprasegmentaux notamment. L'apport de la vidéo et du son nous permettent d'avoir accès à de nouvelles données ou indices de compréhension pour les étudiants en langue étrangère et d'observation des marqueurs de l'oral pour les étudiants en langue maternelle. Nous pensons qu'une tâche de découverte ou de réinvestissement permettrait à ces publics de lier la théorie à la pratique ou comme nous l'avons démontré tout au long de cet article de déduire les règles et usages de certains phénomènes d'accentuation par exemple.

## Bibliographie

- Beacco, J.-C., Blin, B., Houles, E., Lepage, S., Riba, P. 2011. *Niveau B1 pour le français (utilisateur/apprenant indépendant) niveau seuil*, un référentiel, Paris : Didier.
- Boulton, A., Canut, E., Guerin, E., Parisse, C., Tyne, H. 2013. « Corpus et appropriation de L1 et L2 », *Revue Linx*, n° 68-69, p. 9-32.
- Brophy, J.-E. 2004. *Motivating students to learn*. Psychology Press, London : Routledge.
- Buchs, C. 2008. La distribution des informations dans des dispositifs d'apprentissage entre pairs. In : Rouiller, Y. Lehraus, K. (Eds) *Vers des apprentissages en coopération : rencontres et perspectives*, p.57-80, Bruxelles : Peter Lang.
- Ciekanski, M. 2014. « Les corpus : de nouvelles perspectives pour l'apprentissage des langues en autonomie ? », Les Cahiers de l'ACEDLE, *Recherches en didactique des langues et des cultures*, n° 11-1, p.1-19.
- Conseil de l'Europe. 2001. *Cadre Européen Commun de Référence pour les Langues. Apprendre, enseigner, évaluer*. Didier : Paris.
- Di Vito, S. 2013. « L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE », *Revue Linx*, n° 68-69, p.159-176.
- Foucou, P.-Y., Kübler, N. 2000. A web-based environment for teaching technical English. In : Burnard, L., McEnery, T. (dir.). *Rethinking language pedagogy from a corpus perspective*. Francfort : Peter Lang. p. 65-74.
- Gadet, F., Ludwig, R., Mondada, L., Pfänder, S., Simon, A.-C. 2012. « Un grand corpus de français parlé : choix épistémologiques et réalisations empiriques », *Revue française de linguistique appliquée*, n° 17/1, p. 39-54.
- Gavioli, L. 2005. *Exploring corpora for ESP learning*, Amsterdam : John Benjamins.
- Gilquin, G., Gries, S. 2009. « Corpora and experimental methods: A state-of-the-art review », *Corpus Linguistics and Linguistic Theory*, n°5/1, p. 1-26.
- Hyland, K. 2002. Activity and evaluation: reporting practices in academic writing. In : Flowerdew, J. (ed.), *Academic Discourse*, London, Pearson Education, p.115-130.
- Johns, T. 1997. Contexts : the Background, Development and Trialling of a Concordance-based CALL Program. In : Wichmann, Fligelstone, McEnery and Knowles (eds), *Teaching and Language corpora*, Longman, p. 100-115.
- Kerbrat-Orecchioni, C. 1980. *L'énonciation de la subjectivité dans le langage*. Paris : Armand Colin.
- Kraif, O. 2003. « Propositions pour l'intégration d'outils TAL aux dispositifs informatisés d'apprentissage des langues ». *Lidil*, n°28, p. 153-165.
- Landure, C., Boulton, A. 2010. « Corpus et autocorrection pour l'apprentissage des langues ». *ASp, La revue du GERAS*, n°57. p. 11-30.
- Mcenery, T., Xiao, R. 2010. What corpora can offer in language teaching and learning. In : E. Hinkel (éd.), *Handbook of Research in Second Language Teaching and Learning*. London : Routledge, p. 364-380.
- Mouchot, C. 2003. *Méthodologie de l'économie*. Paris : Seuil.
- O'Sullivan Í. « Using corpora to enhance learners' academic writing skills in French », *Revue française de linguistique appliquée*, 2010/2 (Vol. XV), p. 21-35.
- Rastier, F. 2004. « Enjeux épistémologiques de la linguistique de corpus ». *Revue Texto !* [En ligne] : [http://www.revue-texto.net/Inedits/Rastier/Rastier\\_Enjeux.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html) [consulté le 05 avril 2018].
- Read, B., Francis, B., Robson, J. 2001. « 'Playing safe': understanding essay writing and the presentation of the student 'voice' ». *British Journal of Sociology of Education*, 22(3), p. 387-399.
- Schaeffer-Lacroix, E. 2014 « Utiliser des corpus numériques avec un public Lansad ». *Alsic*, Vol. 17. p.36-54.

Seidlhofer, B. 2002. Pedagogy and local learner corpora: working with learner-driven data. In : S. Granger, J. Hung, S. Petch-Tyson (éds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam : John Benjamins, p. 213-234.

Thompson, P., Tribble, C. 2001. « Looking at citations : using corpora in English for academic purposes ». *Language Learning and Technology*, 5(3), p.91-105.

Tognini-Bonelli, E. 2001. *Corpus linguistics at work*. Amsterdam : John Benjamins.

Tognini-Bonelli, E. 2010. « Theoretical overview of the evolution of corpus linguistics », In : A. O'Keeffe & M. McCarthy (éds.), *Routledge Handbook of Corpus Linguistics*, London : Routledge, p. 14-27.

Tribble, C. 2002. Corpora and corpus analysis: new windows on academic writing. In : Flowerdew, J. (ed.), *Academic Discourse*, London, Pearson Education, p.131-149.

Tyne, H. 2009, « Corpus oraux par et pour l'apprenant ». *Mélanges CRAPEL*, n° 31, p. 91-111.

Tyne, H. 2012. Corpus work with ordinary teachers: data-driven learning activities. In : Thomas, J., Boulton, A. (dir.). *Input, process and product: developments in teaching and language corpora*. Brno : Masaryk University Press. p. 114-129.

Viau, R. 2009. *La motivation en contexte scolaire*. Bruxelles : Editions De Boeck.

### Sitographie

<http://atilf.atilf.fr> Trésor de la langue française informatisé [consulté le 07 avril 2018].

<https://apps.atilf.fr/fleuron/> Plateforme FLEURON [consulté le 05 avril 2018].