



Comment former les étudiants de Master FLE à l'utilisation pédagogique des corpus numériques ?

Cristelle Cavalla

► To cite this version:

Cristelle Cavalla. Comment former les étudiants de Master FLE à l'utilisation pédagogique des corpus numériques ?. Jan Goes; Luis Meneses-Lerin; Jean-Marc Mangiante; Françoise Olmo; Carmen Pineira-Tresmontant. Apports et limites des corpus numériques en analyse de discours et didactique des langues de spécialité, Editura Universitaria, pp.79-92, 2019, 978-606-14-1550-2. hal-02534091

HAL Id: hal-02534091

<https://hal.science/hal-02534091>

Submitted on 6 Apr 2020

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

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

Cavalla Cristelle (2019). Comment former les étudiants de Master FLE à l'utilisation pédagogique des corpus numériques. In J.Goes, L.Meneses-Lerin, J-M.Mangiante, F.Olmo, C.Pineira-Tresmontant (eds), *Apports et limites des corpus numériques en analyse de discours et didactique des langues de spécialité*, Craiova, Editura Universitaria, pp. 79-92.

Comment former les étudiants de Master FLE à l'utilisation pédagogique des corpus numériques ?

Cristelle CAVALLA

Université Sorbonne Nouvelle-Paris3

DILTEC EA2288

Dans cette contribution nous interrogeons la formation de futurs enseignants de langue à l'université – étudiants en 1^{re} année de Master didactique du FLE à l'Université Sorbonne Nouvelle-Paris 3 – à l'exploitation des corpus numériques pour leur formation et pour qu'ils se les approprient pour ensuite les intégrer à leur pratique d'enseignant. Cette contribution tentera de prendre en compte les réflexions autour de la complémentarité entre les outils électroniques que sont les dictionnaires en ligne et les corpus numériques afin de cerner ce qu'il faudrait retenir pour la formation initiale et continue des formateurs (Boulton et al., 2014). Les futurs enseignants formés vont finalement découvrir simultanément les corpus numériques, les formes figées spécifiques à certains discours et enfin le FOU, c'est-à-dire, le Français sur Objectif Universitaire (Mangiante et al., 2011). Les étudiants, futurs enseignants, sont réceptifs aux deux approches utilisées : bottom-up ou top-down¹ (Lundquist, 2013). Ils restent toutefois méfiants quant à l'utilisation des corpus numériques pour l'enseignement des langues. Nous allons expliquer l'apport des corpus numériques en classe de langue afin de comprendre comment aider les futurs enseignants de langue à intégrer ces outils à leur enseignement afin de contribuer à l'aide à l'apprentissage de leurs futurs apprenants.

Nous tenterons de cerner les compétences attendues pour l'utilisation de l'outil numérique et plus particulièrement pour celle des corpus numériques. Enfin, deux questions guideront notre réflexion : pourquoi former les enseignants à l'utilisation de ces corpus ? Comment les introduire en classe de langue ?

¹ Soit de l'élément linguistique rencontré en classe vers le FOU en passant par les corpus (bottom-up) ou à l'inverse top-down c'est-à-dire du FOU en passant par les corpus pour arriver à l'élément linguistique.

1. Compétences nécessaires pour l'utilisation d'outils numériques

Qu'est-ce qu'une compétence numérique ? Cette question trouve une réponse sur le site du MENESR² depuis 2015 à propos des formations au C2i (puis C2i2e), c'est-à-dire aux formations pour l'apprentissage de compétences à l'utilisation des outils numériques. Pour le MENESR, les compétences numériques correspondent

à la capacité à mobiliser les technologies numériques pour réaliser efficacement une tâche dans une situation donnée. Elles garantissent un usage sûr et critique des technologies. Elles ne se réduisent pas à la seule maîtrise d'outils numériques ou de bureautique. (MENESR, 2015)

Sur le site du MENESR, les auteurs précisent qu'il ne s'agit pas uniquement de savoir-faire numériques mais que les compétences afférentes sont des savoirs inhérents à l'utilisation à bon escient de ces outils. Par exemple, aider à développer la réflexion critique et argumentée des usages voire des technologies développées pour les outils accessibles en ligne, être au fait des évolutions des usages de ces outils et de la production de données ainsi que le partage des savoirs, enfin la dimension éthique véhiculée par leur usage est à interroger et discuter.

La connaissance des corpus numériques pour la classe de langue, entre dans cette compétence numérique et doit être interrogée selon les entrées proposées comme « être au fait des évolutions » qui va par exemple conduire l'utilisateur à développer une compétence d'adaptation aux interfaces évolutives des corpus numériques. Il est également question des choix que l'enseignant peut faire en connaissant les corpus numériques disponibles en ligne ; il doit donc être au courant des derniers corpus développés. Pour la dimension éthique, il sera important de préciser les droits et les devoirs des auteurs des textes consultables en ligne via l'interface et expliquer pourquoi certains auteurs de corpus numériques exigent une inscription pour y avoir accès (le droit d'auteur est en jeu). Parmi les compétences autour des corpus numériques, celle en lien avec les outils connexes est importante car parfois il ne suffit pas de savoir manipuler le corpus, il est nécessaire de savoir manipuler les extractions une fois téléchargées dans un tableur ou un traitement de texte, ne serait-ce que pour en faire des exercices. Des logiciels tels que *Antconc* (Anthony, 2007) qui permettent de faire des statistiques sur des fichiers « .txt » peut être très utile pour gérer ses propres extractions, faire des comparaisons... Enfin, il existe des exemples d'exploitations des corpus, sur les sites des corpus (des exercices sont proposés dans

² Ministère de l'Éducation Nationale et de l'Enseignement Supérieur et de la Recherche.

*Lextutor*³ ou sur *IPFC*⁴ par exemple) et sur d'autres sites comme *Scientext – Écrits scientifiques et universitaires* (sur hypotheses.org). Nous résumons ces choix dans le tableau 1 ci-dessous.

Compétences	Connaitre les corpus accessibles en ligne	pour	Faire des choix
	S'adapter aux interfaces d'interrogation		Accéder rapidement aux données
	Connaitre les outils connexes (concordanciers par exemples)		Développer des exercices
	Trouver des exemples d'exploitation		Se servir de l'existant

Tableau 1 : Des compétences numériques pour l'utilisation des corpus numériques

2. Pourquoi et comment utiliser les corpus numériques en classe de langue ?

Les corpus numériques donnent accès à un grand échantillon de langue, et donc de variations des langues à plusieurs niveaux : les registres et les types de discours sont notamment très importants. Par exemple, la distinction des types de discours à l'écrit permet d'extraire des spécificités comme l'emploi très fréquent de l'expression « par exemple » dans le discours scientifique alors qu'elle est peu présente dans les discours de presse. Ces corpus – parfois spécialisés – sont désormais nombreux et souvent accessibles en ligne gratuitement.

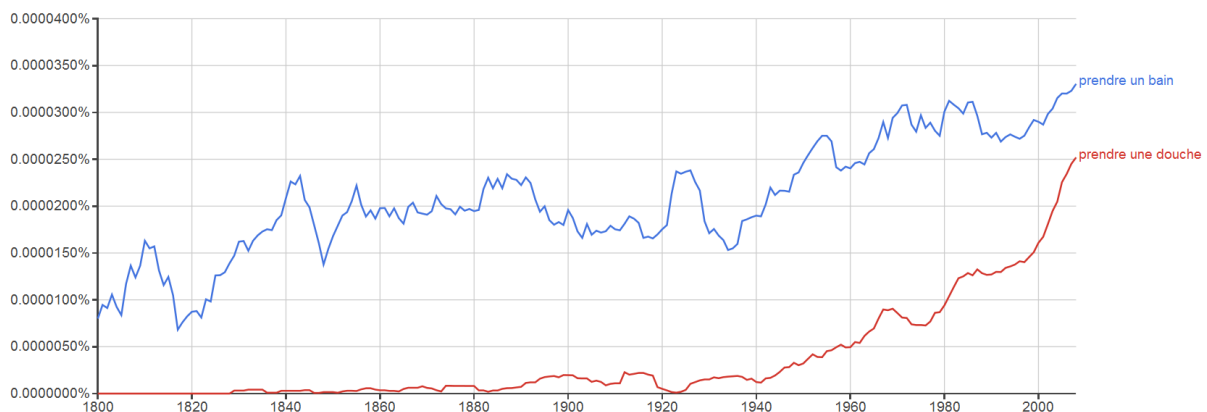
Les extractions des corpus numériques sont très nombreuses et impliquent une sélection des occurrences en fonction des objectifs didactiques. Enfin, l'utilisation des corpus vient parfois bouleverser ou du moins, modifier, les représentations ou les intuitions des francophones natifs et allophones. Il peut être surprenant de se rendre compte que l'expression « prendre un risque » dans le discours de presse, est davantage utilisée dans la finance qu'en politique (Cavalla et al., 2017). Le corpus de Google Book est utilisé pour donner des fréquences quant à l'utilisation à l'écrit de mots isolés ou d'expression dans l'interface *Ngram Viewer*. Le graphique 1 donne un aperçu de ces fréquences en comparant les expressions « prendre un bain » / « prendre une douche » :

³ URL : <https://www.lexutor.ca/conc/fr/>

⁴ Interphonologie du Français Contemporain, pfc.net/?option=com_content&view=article&id=172&Itemid=134

URL: <https://www.projet->

Graph these comma-separated phrases: ☐ case-insensitive
 between 1800 and 2008 from the corpus French (2009) with smoothing of 3 [Search lots of books](#)



Graphique 1 : Fréquence de « Prendre un bain » et « Prendre une douche » dans *Ngram Viewer* entre les années 1800 et 2008

L'utilisation des corpus est donc désormais facilitée par leur accès libre en ligne, par leurs interfaces d'interrogation plus ergonomiques qu'il y a quelques années et enfin, par les réflexions à leur propos pour des développements didactiques. En 1993, Fligelstone proposait déjà trois façons d'utiliser pédagogiquement les corpus numériques :

- « teaching about » : le corpus comme objet d'enseignement, soit l'enseignement de la linguistique de corpus ou de sa méthodologie ;
- « exploiting to teach » : le corpus comme support d'enseignement, soit l'enseignement de l'usage d'un corpus (quelles requêtes faire, dans quelle situation en fonction du système d'interrogation disponible ?) ;
- « teaching to exploit » : le corpus pour enseigner et apprendre la langue, soit l'exploitation des corpus et l'interprétation des résultats pour enseigner une langue.

Dans un article avec M. Loiseau (Cavalla et al., 2013 : 165), nous expliquions alors que seul le premier type d'exploitation n'entrait pas dans nos préoccupations pour l'enseignement des langues. Notons qu'il est parfois utilisé dans ce cadre, quand des enseignants demandent aux apprenants de constituer leur propre corpus à des fins d'apprentissage de la langue (Tyne, 2009). Nous n'avons pas choisi cette entrée mais les deux dernières qui deviennent complémentaires dans notre démarche dès l'instant où nous envisageons d'enseigner l'usage du corpus aux apprenants afin qu'ils soient autonomes dans leur recherche d'éléments langagiers. Dans cette perspective, nous devons adopter une démarche liée à la découverte des corpus en ligne.

3. Comment faire découvrir les corpus numériques ?

Le type de recherches que l'on peut faire sur corpus diffère beaucoup des recherches que l'on peut effectuer sur Internet ; en effet, le corpus et les outils d'interrogation de corpus permettent des interrogations beaucoup plus spécifiques et structurées, qui tiennent compte des genres textuels et des domaines de spécialité. (Kübler 2014: 40)

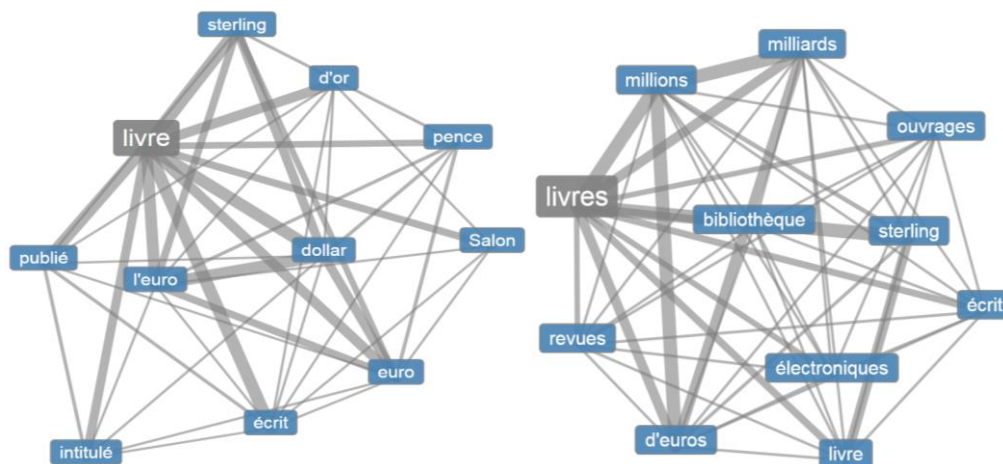
Les raisons évoquées par Kübler à propos des spécificités de la recherche sur corpus, nous invitent à privilégier les acquis antérieurs des apprenants pour leur faire découvrir les corpus numériques. Boulton et Tyne (2014) préconisent aussi cette entrée malgré les différences évoquées entre les interfaces des corpus numériques et les requêtes sur des moteurs de recherches courants. Pour reprendre les arguments de Kübler, les « outils d'interrogations de corpus » aident à mieux cibler les recherches et font émerger les variations possibles et les spécificités des discours spécialisés. Ces spécificités nous intéressent car le public que nous visons est souvent en FOU et a des besoins liés aux attentes universitaires (à l'écrit dans notre cas). Ainsi, le fait de savoir formuler une requête en ligne est utile pour avancer dans le type de requête possible sur des corpus numériques.

Une fois les acquis antérieurs vérifiés, les apprenants sont invités à se poser des questions sur les réponses à la requête : d'abord sur un moteur de recherche courant, puis sur l'interface d'interrogation d'un corpus. Enfin, les différences mises au jour, l'apprenant pourra interroger le corpus de façon précise en fonction de ses propres besoins.

Toutefois, avant d'en arriver à cette autonomie dans l'interrogation d'un corpus numérique, il est nécessaire d'aider l'apprenant à entrer dans ces outils quelque peu déroutants au premier abord. L'entrée choisie est totalement chiffrée à l'aide des fréquences. Il est apparu au fil des expériences d'enseignement, que la vue des fréquences d'utilisation des mots dans les corpus écrits, joue le rôle d'un déclencheur d'une prise de conscience d'éléments langagiers spécifiques. Comme évoqué précédemment, l'interrogation des corpus peut bouleverser les représentations des locuteurs. De ce fait, l'interface de *Ngram Viewer* déjà citée, ou celle du *Corpus Français de Leipzig*⁵, donnent à voir en premier des fréquences. Les apprenants sont alors frappés par les chiffres qu'ils découvrent, puis par les caractéristiques des sens d'un mot. Par exemple, dans le *Corpus Français de Leipzig*, la lexie « livre » apparaît le plus fréquemment avec « sterling » et les cooccurrents les plus proches sont liés au champ sémantique de la finance

⁵ URL : <http://corpora.uni-leipzig.de/en>

(« dollar », « euro »). En revanche, la lexie « livres » (au pluriel) apparaît plus proche des lexies « revues » et « bibliothèque » (même si « sterling » reste le premier). Le champ sémantique de « l'ouvrage à lire » (et pas de la finance) est alors plus présent et cela donne à voir les proxémies suivantes :



Graphique 2 : Proxémies de « livre » et « livres » dans le *Corpus Français de Leipzig*

Ces deux acceptions du substantif « livre » au singulier et au pluriel, révèlent deux contextes d'utilisation différents et surtout les deux sens de ce signifiant. Nous avons à faire à deux substantifs, donc deux signifiés, qui apparaissent dans toutes leurs diversités dans leur emploi en nombre (singulier vs pluriel). Il est intéressant de constater que dans la représentation proxémique du TLFi⁶ cette distinction n'apparaît pas malgré la mention des deux substantifs dans ce dictionnaire :



Graphique 3 : Clusters de la proxémie du TLFi pour « livre1 » (ouvrage) et « livre2 » (poids, monnaie)

Ainsi, l'apprenant aperçoit toute la complémentarité du dictionnaire et du corpus grâce à de tels exemples. Le dictionnaire apporte une précision sémantique figée qui permet de stabiliser un sens inhérent de base. Le corpus donne accès à la mise en contexte de ces lexies, à leur usage qui est obligatoirement influencé par l'entour lexico-sémantique proche et élargit. La « livre

⁶ Trésor de la Langue Française informatisé, URL: <http://atilf.atilf.fr/>

Sterling » apparaît dans la définition de « livre 2 » du TLFi car le signifié est en lien avec celui de la livre comme mesure de poids. L'apprenant accède à l'usage grâce au corpus et donc à tout une architecture lexicale, syntaxique et sémantique qui se construit autour de la base lexicale choisie par le locuteur. Grâce au dictionnaire, l'apprenant vérifie le signifié initial et peut envisager d'autres usages de la lexie en découvrant, peut-être, d'autres sens encore.

Ainsi, avec la découverte des chiffres des fréquences d'utilisation des lexies et celle des sens en usage de ces mêmes lexies, les apprenants prennent conscience de l'intérêt des corpus numériques pour leur apprentissage. L'enseignant peut alors les aider à organiser l'utilisation des outils à leur disposition notamment en poursuivant la réflexion autour de leur complémentarité. Pour ce faire, il est nécessaire d'aborder de façon progressive les différentes interfaces d'interrogation des corpus numériques à travers leurs aspects techniques puis linguistiques.

3.1. Aspects techniques

L'interrogation des corpus numériques est d'un abord technique dans un premier temps. Parmi les corpus numériques qui existent pour le français, plusieurs interfaces existent et leurs interfaces étant différentes, il est apparu indispensable de les présenter dans un ordre spécifique aux apprenants.

En effet, comme nous avons vu que les chiffres aident à une certaine prise de conscience vis-à-vis de la langue, nous avons opté pour présenter en premier, le *Corpus Français de Leipzig* déjà cité, et de travailler à partir des fréquences et des cooccurents. L'exemple de livre (graphique 2), permet à l'apprenant d'avoir simultanément des chiffres et des extraits. Une première approche peut alors être développée autour de la lexie en s'attardant sur ses cooccurents proches et donc ses constructions syntaxiques possibles à vérifier à l'aide des extraits.

Le deuxième corpus présenté, est le *lexicoscope*⁷ (Kraif, 2016). Les concepteurs ont fait le pari de développer un outil à la fois simple d'accès et performant quant à ses compétences d'extraction. Le *lexicoscope* est désormais le site que nous privilégions car nous n'avons pas à faire à des linguistes mais à des didacticiens dont l'objectif est d'extraire facilement des données pour passer davantage de temps à la réflexion pour leur didactisation. Le *lexicoscope* propose une seule case (comme les moteurs de recherche courants) et donne la possibilité de choisir par exemple le sens d'un mot, si une ambiguïté apparaît pour la machine, grâce à des exemples associés à des schémas syntaxiques qui seront la base de l'extraction (cf. Copie d'écran 1). Le fait d'ajouter des exemples aux schémas syntaxiques arborés, permet l'abord de ces schémas

⁷ URL: <http://phraseotext.u-grenoble3.fr/lexicoscope/index.php?errorAccess>

par des non spécialistes. Les didacticiens du FLE n'étant pas tous linguistes, de tels schémas ont parfois pour effet d'entraver l'entrée dans de tels outils de recherche. Ensuite, l'utilisateur clique sur le schéma voulu et choisi d'extraire un concordancier afin d'avoir des exemples d'utilisation de la lexie choisie. Enfin, dans la case unique, l'utilisateur peut taper une lexie isolée ou une expression (plusieurs lexies liées, cf. Copie d'écran 1) et l'outil cherchera les différentes acceptions par exemple les formes conjuguées d'un verbe, les formes plurielles et singulier d'un nom ou d'un déterminant (on parle de lemmatisation des formes lexicales).

Concordances et profils combinatoires (cooccurrences)

Requête libre **Requête avancée** **Requête multi-pivots**

raconter des histoires

Concordances
Cooccurrences

p. ex. "considération" ou "prendre en considération"

```

graph TD
    A([raconter_VERB]) -- OBJ --> B([histoire_NOUN])
    
```

[Editer comme requête avancée](#)

Nombre d'occurrences :

101 Le touriste qui a voulu y vérifier si la télévision s'était émancipée de la littérature, si les écrivains étaient toujours les mieux placés pour **raconter** des **histoires** et si le roman inspirait encore les gens d'images, n'est pas reparti bredouille.

411 Contrairement aux opinions bien établies, il serait donc indispensable de " se **raconter** des **histoires** ".

49450 Dans *Malone meurt*, de Samuel Beckett, le narrateur essaie de se **raconter** des **histoires** mais elles s'effondrent sur elles-mêmes, il dit " Non, je ne peux pas ", c'est trop, ça n'a pas de sens.

49453 La maladie en tant que telle m'apparaît donc plutôt comme un défi redoutable posé à notre capacité à nous **raconter** des **histoires**.

79254 " On continue à se **raconter** de belles **histoires** ", se désole, un brin désabusé, Hervé Mariton, député de la Drôme.

```

graph TD
    A([raconter_VERB]) -- VMOD_POSIT1 --> B([histoire_NOUN])
    
```

[Editer comme requête avancée](#)

Nombre d'occurrences :

394 Ne pas se **raconter** d'**histoires**, c'est d'habitude un signe de la lucidité.

Cliquer sur le schéma pour sélectionner l'arbre lemmatisé

Copie d'écran 1 : Requête pour « raconter des histoires » dans le *Lexicoscope*

Pour un premier abord de la langue uniquement, sans passer par des chiffres, l'utilisateur n'est généralement pas perdu car sa familiarité avec ce type d'interface a pour effet de rassurer l'entrée dans la langue. Un tel effet serait à mesurer plus précisément auprès d'enseignants de FLE même si cela est corroboré par les études sur l'utilisation du « déjà-là » de l'apprenant (Meirieu 2012). Sans entrer dans les détails des extractions, le *lexicoscope* offre des possibilités de tri intéressantes pour l'enseignant et l'apprenant et la diversité des corpus à disposition donne accès à des discours de spécialité fort appréciés des enseignants de langues.

À ce stade de l'exploration, d'autres corpus peuvent être présentés et notamment *Lextutor* malgré son interface peu conviviale. Les outils offerts par le site plongent l'utilisateur dans des dédales et T. Cobb (2007) réussit à nous faire oublier que toutes les considérations ergonomiques du site ont été évacuées. L'aspect technique retenu pour la formation de futurs enseignants, est double pour ce corpus : 1/la possibilité d'explorer des petits corpus (ce qui est de moins en moins vrai car l'auteur augmente son corpus) et 2/l'interrogation simple.

Dans ces deux exemples d'interfaces, des chiffres sont encore présents notamment en ce qui concerne les fréquences des lexies ou des expressions. L'entrée dans la langue se fait donc par étapes qui associent à la fois les spécificités des corpus et les attentes pour extraire des éléments utiles à la classe de langue.

3.2. Aspects linguistiques

L'introduction des corpus pour la recherche d'information linguistique permet à l'utilisateur de prendre conscience du contexte de l'usage d'une lexie. Nous l'avons vu précédemment avec les lexies « livre » et « livres » (cf. Graphique 2). Les corpus sont désormais construits autour de discours spécifiques et de genre. La notion de genre telle que nous l'envisageons ici est issue de la théorie de la sémantique interprétative de Rastier (1987). Ainsi, les requêtes sur les corpus offrent plusieurs façons d'entrer dans la langue, ce qui est fort pertinent pour l'enseignement. Soit l'utilisateur entre par la lexie et oriente donc la recherche dans le corpus vers cette lexie, il s'agit de l'approche « corpus-based » explicitée par Leech (1992), soit il veut explorer ce que le corpus contient et peut lui apporter et il l'abordera alors selon l'approche « corpus-driven » (ou DDL : Data Driven Learning) développée, entre autres par la suite, par Firth (1957) et Sinclair (1987). Ces deux approches ne sont pas exclusives, au contraire, il est important de les associer afin d'extraire à la fois des éléments qui confirment des hypothèses et d'autres qui viennent les compléter car le corpus révèle des éléments non envisagés au départ.

L'approfondissement de l'exploration conduit alors l'utilisateur à se rendre compte que les genres peuvent modifier le sens des lexies ou que certaines lexies n'apparaissent que dans certains genres. Nous avons évoqué l'expression « par exemple » qui est spécifique du discours

scientifique et du genre « thèse » dans le corpus *Scientext*⁸. De telles remarques peuvent apparaître également dans un corpus diachronique (*Frantext*⁹ par exemple). Ainsi, l'exploration conduit-elle l'utilisateur à prendre conscience qu'il existe des genres et que ces derniers contribuent à la distinction des sens des lexies. Enfin, l'aspect diachronique est important car travailler sur un corpus de presse des années 2000 ne donne pas les mêmes résultats que l'exploration d'un corpus de presse des années 2010 (pourtant ici les différences ne sont pas si importantes). De même, puisque le choix du genre va influencer le choix des lexies et des sens diffusés, les utilisateurs prennent conscience des liens entre les mots et leur environnement. Des réflexions émergent et les futurs enseignants formés à ces outils ont alors ciblé davantage leurs requêtes afin d'introduire les corpus dans leur classes de langue avec des entrées précises pour leurs apprenants dans un premier temps.

4. Comment introduire les corpus en classe de langue ?

La question a déjà des réponses notamment grâce à Chambers (2005, 2011) qui a bien décrit deux approches qui varient selon les objectifs que l'enseignant fixe pour les apprenants :

- soit une approche indirecte, c'est-à-dire donner aux apprenants des concordanciers triés et les faire travailler sur ces éléments ;
- soit une approche directe où l'apprenant est placé devant un corpus (de préférence petit) et l'enseignant le guide dans l'interrogation en ligne.

La taille du corpus est importante dans l'approche directe car il ne faut pas effrayer l'apprenant qui extrait des milliers d'exemples qu'il faudrait trier. Toutefois, au vu de l'accroissement exponentiel du nombre de mots dans les corpus numériques développés actuellement, il est de plus en plus difficile de placer l'apprenant face à un corpus de petite taille. De fait, la pédagogie mise en place doit obéir à des règles de consignes strictes : guider l'apprenant c'est le faire pas à pas et dans un premier temps, ne pas lui donner la possibilité de dévier des consignes à suivre. Par exemple, taper la requête « livre » au singulier dans le *lexicoscope* et ne s'intéresser qu'aux 10 premières lignes de concordance.

L'introduction des corpus dans la classe de langue n'est plus vraiment une question pour les chercheurs surtout depuis l'ouvrage de Boulton et Tyne (2014) qui a démocratisé le corpus numérique en le hissant au rang des documents authentiques courants. Ainsi, les corpus numériques sont-ils des outils qui complètent l'usage des dictionnaires en classe de langue, mais aussi celui des manuels en confirmant ou ajoutant des exemples d'utilisation de lexies ou d'expressions. C'est une manne intarissable pour les langues de spécialités et notamment pour

⁸ URL : <http://scientext.msh-alpes.fr/scientext-site/spip.php?article1>

⁹ URL : <http://www.frantext.fr/>

le FOU qui était l'entrée de la formation prise pour exemple dans cet article. Enfin, quand cela est possible, le travail en groupe est toujours bénéfique à plusieurs niveaux : pour la découverte d'un nouvel outil (répartition des tâches entre ceux qui avancent plus rapidement sur les interfaces et les autres), pour les requêtes à formuler (répartition des tâches de tri entre ceux qui avancent plus rapidement sur l'analyse que les autres), pour les hypothèses à émettre et ensuite pour la didactisation des sorties.

5. Conclusion

Nous allons conclure en reprenant les questions de l'introduction, d'abord celle interrogeant l'utilité de la formation des enseignants de langues à l'utilisation des corpus numériques. Plusieurs arguments ont été évoqués, voici le rappel de quelques-uns : la multitude des exemples (à maîtriser), le choix possible, l'accès rapide, la complémentarité avec d'autres outils (les dictionnaires par exemple). Puis la question de leur introduction en classe de langue auprès des étudiants. Nous avons évoqué les approches pédagogiques « directe » et « indirecte », et nous avons opté pour une approche directe du corpus, à savoir, rendre autonome les apprenants face à ces outils et donc qu'ils travaillent directement sur les corpus.

Enfin, nous voulions juste évoquer l'exploitation des corpus numériques pour les classes de langue auprès d'enseignants peu familiarisés avec ces outils, mais qui, généralement, se forment rapidement. Nous souhaiterions alors savoir, dans une prochaine étude, comment les futurs enseignants formés à ces outils, les réinvestissent dans leurs classes de langue. Nous avons vu que les approches DDL et corpus-based se complètent, nos prochaines investigations se tourneront vers l'appropriation de ces approches par les enseignants et les modules développés par eux pour l'intégration de ces outils en classe de langue pour l'enseignement et l'apprentissage.

Références

- ANTHONY Laurence, 2007, « AntConc », [Online], <http://www.laurenceanthony.net/software.html> (consulté le 07.07.2017).
- BOULTON Alex et TYNES Henry, 2014, *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*, Paris, Didier.
- CAVALLA Cristelle et LOISEAU Mathieu, 2013, « Scientext comme corpus pour l'enseignement », in Agnès TUTIN & Francis GROSSMANN, (éd.), *L'écrit scientifique : du lexique au discours. Autour de Scientext*, Rennes, PUR, p. 163-182.
- CAVALLA Cristelle & SORBA Julie, 2017, « Prendre un bain, des risques ou la fuite : étude diachronique du figement », in Francis GROSSMANN, Salah MEJRI & Inès SFAR, (éd.), *Phraséologie : Sémantique, Syntaxe, Discours*, Paris, Champion, p. 187-201.
- CHAMBERS Angela, FARR Fiona et O'RIORDAN Stéphanie, 2011, « Language teachers with corpora in mind: from starting steps to walking tall », *The Language Learning Journal*, n° 39:1, pp. 85-104.

- CHAMBERS Angela, 2005, « Integrating Corpus consultation in Language studies », *Language Learning & Technology*, 9 - n°2, pp. 111-125.
- COBB Tom, 2007, « Why & How to Use Frequency Lists to Learn Words », [Online], <http://www.lex tutor.ca/research/> (consulté le 15.10.2017).
- FIRTH John Rupert, 1957, « A Synopsis of Linguistic Theory 1930–1955 », in John Rupert FIRTH, (éd.), *Studies in Linguistic Analysis*, Oxford: Basil Blackwell, p. 1-32.
- FLIGELSTONE Steve, 1993, « Some reflections on the question of teaching, from a corpus linguistics perspective », *ICAME journal*, n° 17, pp. 97-109.
- KRAIF Olivier, 2016, « Le lexicoscope : un outil d'extraction des séquences phraséologiques basé sur des corpus arboré », *Cahiers de lexicologie*, n° 108, pp. 91-106.
- KÜBLER Natalie, 2014, « Mettre en œuvre la linguistique de corpus à l'université - Vers une compétence utile pour l'enseignement/apprentissage des langues ? », *Les Cahiers de l'Acedle*, n° 11, pp. 37-77.
- LEECH Geoffrey, 1992, « Corpora and Theories of Linguistic Performance », in Jan SVARTVIK, (éd.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium, 4-8 August 1991*, Berlin-New York : Mouton de Gruyter, p. 105-122.
- LUNDQUIST Lita, 2013, *Lire un texte académique en français*, Paris, Ophrys.
- MEIRIEU Philippe, 2012 [1987], *Apprendre... oui, mais comment*, Issy les Moulineaux, ESF.
- MENESR, 2015, *C2i2e*, [Online], <https://c2i.enseignementsup-recherche.gouv.fr/enseignant/les-competences-numeriques> (consulté le 01/11/17).
- MANGIANTE Jean-Marc et PARPETTE Chantal, 2011, *Le français sur objectif universitaire*, Grenoble, PUG.
- RASTIER François, 1987, *Sémantique interprétative*, Paris, PUF.
- SINCLAIR John McHardy, 1987, *The nature of the evidence. Looking Up: an account of the COBUILD Project in lexical computing*, London and Glasgow, Collins.
- TYNE Henry, 2009, « Corpus oraux par et pour l'apprenant », *Mélanges CRAPEL*, n° 31, pp.91-111.