

RECHERCHES  
EN DIDACTIQUE  
DES LANGUES  
ET DES CULTURES

## Recherches en didactique des langues et des cultures

Les cahiers de l'Acedle

11-1 | 2014

Notions en questions - Corpus

---

# Les corpus : de nouvelles perspectives pour l'apprentissage des langues en autonomie ?

Maud Ciekanski

---



### Édition électronique

URL : <http://journals.openedition.org/rdlc/1710>

DOI : 10.4000/rdlc.1710

ISSN : 1958-5772

### Éditeur

ACEDLE

### Référence électronique

Maud Ciekanski, « Les corpus : de nouvelles perspectives pour l'apprentissage des langues en autonomie ? », *Recherches en didactique des langues et des cultures* [En ligne], 11-1 | 2014, mis en ligne le 07 janvier 2014, consulté le 19 avril 2019. URL : <http://journals.openedition.org/rdlc/1710> ; DOI : 10.4000/rdlc.1710

---

Ce document a été généré automatiquement le 19 avril 2019.



*Recherches en didactique des langues et des cultures* is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

---

# Les corpus : de nouvelles perspectives pour l'apprentissage des langues en autonomie ?

Maud Ciekanski

---

## Introduction

- 1 Les corpus, en tant que ressources d'apprentissage, connaissent un intérêt grandissant en didactique des langues étrangères, comme en témoignent les nombreuses publications récentes sur le sujet. Cet intérêt fait suite à l'essor théorique et technique que la linguistique de corpus a connu ces dernières années. La pertinence des corpus écrits ou oraux pour l'apprentissage des langues est maintenant largement démontrée (Boulton & Tyne, 2014 ; Debaisieux, 2009 ; Frankenberg-Garcia, ce volume ; O'Keeffe *et al.*, 2007). Néanmoins, le recours aux corpus est encore souvent assimilé à un "défi" pour les enseignants (Debaisieux, 2009). Il subsiste en effet un écart entre le consensus théorique qui se dégage des recherches menées sur l'utilisation effective des corpus et leur intégration dans les pratiques enseignantes en contexte d'apprentissage. Les raisons invoquées sont diverses et pointent des difficultés en termes d'accès, de gratuité, d'ergonomie (Braun & Chambers, 2006 ; Debaisieux, 2009, souligne notamment la dispersion des "petits" corpus pour le français), en termes de compétences langagières requises (les corpus s'adressent plutôt à des publics avancés, selon Lüdeling & Walter, 2009), en termes de choix de langue cible (déficit de corpus pour les langues autres que l'anglais, selon Schaeffer-Lacroix, 2011) ou encore en termes de variété linguistique retenue (Boulton, 2008, qui souligne la prégnance de la norme écrite dans la constitution et la diffusion des corpus de langue française). On pourrait ajouter à toutes ces raisons le manque de formation et de familiarité à l'utilisation des corpus chez une majorité d'enseignants qui ne les ont pas utilisés quand ils étaient eux-mêmes apprenants - les corpus conservant dans les représentations de beaucoup l'image d'une ressource faite *par* et *pour* les linguistes.

- 2 Si la plupart des recherches-actions menées sur l'utilisation des corpus à des fins d'apprentissage de langues se situent en contexte d'hétéroformation (en classe de langue majoritairement), nombreux sont les chercheurs qui perçoivent les corpus comme une ressource particulièrement pertinente pour un apprentissage en autonomie (Boulton, 2011 ; Kettemann & Marko, 2011 ; Uzar, 2003), à l'instar d'Aston : "*Perhaps the greatest attraction of corpora for language pedagogy is their potential for autonomous learning*"<sup>1</sup> (Aston, 2001 : 41). Cette autonomie de l'apprenant liée à une situation d'apprentissage favorisant un accès direct aux ressources est pour beaucoup synonyme de la possibilité de travailler à son rythme sur les données rassemblées en corpus, en fonction de ses préférences et de son style d'apprentissage (ex. Kettemann & Marko, 2011). Chez la plupart des auteurs, les corpus sont sélectionnés ou créés par l'enseignant (Schaeffer-Lacroix, 2011). Toutefois, le fait que l'apprenant ait un accès direct aux corpus inviterait à modifier le rôle de l'enseignant, comme suggéré par Kettemann et Marko : "*the teacher consequently acts as a guide and as an advisor, not as a mediator of knowledge nor a judge of students' performance*"<sup>2</sup> (2011 : 20). En outre, le rôle de l'apprenant n'est pas seulement limité à l'exploitation de données sélectionnées en amont par l'enseignant. D'autres pratiques plus autonomisantes émergent des recherches sur l'utilisation des corpus en classe de langues : Landure et Boulton (2010) engagent l'apprenant dans un travail d'autocorrection à partir des corpus ; Tyne (2009) accompagne les étudiants dans un travail de constitution de corpus *ad hoc* à partir de données qu'ils ont eux-mêmes recueillies.
- 3 La démarche d'apprentissage sur corpus s'apparenterait aux pratiques d'apprentissage relevant du paradigme de l'autonomie (Albéro, 2000) ou, à tout le moins, impliquerait des pratiques d'apprentissage plutôt caractéristiques de l'autoformation (Tremblay, 2003). On parle d'un apprentissage par "serendipité", c'est-à-dire par hasard, (*serendipitous findings*) chez Bernardini (2000), d'un apprentissage "incidentel", c'est-à-dire non intentionnel, (*incidental learning*) chez Duda et Tyne (2010), mais aussi d'un apprentissage par découverte (Bernardini, 2002 ; Debaisieux, 2009) ou encore d'un apprentissage exploratoire (Kettemann & Marko, 2011). Ces différentes modalités d'apprentissage sont toutes caractéristiques de la stochastique, une situation d'apprentissage non déterminée qui peut être le fruit du hasard et qui est au cœur des démarches autodidactes ou autonomes (Tremblay, 2003 : 147).
- 4 Pour autant, Chambers (2009) rappelle que des expérimentations menées en contexte autodirigé ou lors d'apprentissage non formel seraient nécessaires pour établir les conditions de l'utilisation de corpus *sans* la médiation de l'enseignant :
- If corpus consultation... is to become a common activity for learners across the broad spectrum of language studies (general language learning, literary studies, languages for specific purposes, translation, etc.), it would seem necessary for developments to take place in a broader context... It is perhaps outside the classroom that the next important step in research in this area will take place. (Chambers, 2007 : 13)<sup>3</sup>
- 5 Paradoxalement, la question des corpus comme ressources d'apprentissage a suscité, à notre connaissance, peu d'échos dans le domaine de l'apprentissage des langues en autonomie. Un survol des principales revues scientifiques publiant sur l'apprentissage des langues en autonomie<sup>4</sup>, ainsi qu'un survol des ressources proposées dans les centres de ressources en langues (CRL) de quelques universités françaises<sup>5</sup>, permet de constater le peu de place occupée par les corpus parmi les ressources analysées ou recensées<sup>6</sup>. Pourquoi ce manque d'intérêt apparent pour les corpus dans le champ de l'autonomie ?

Quelles seraient les nouvelles perspectives apportées par les corpus pour cette situation d'apprentissage spécifique ?

- 6 Cet article vise à définir les potentialités des corpus pour un apprentissage en autonomie. Nous chercherons, d'une part, à mieux cerner la nature de cette ressource d'apprentissage : les corpus répondent-ils aux caractéristiques d'une ressource adaptée à l'apprentissage en autonomie ? Nous discuterons, d'autre part, des freins et des leviers liés à l'utilisation des corpus dans les apprentissages en autonomie, en questionnant plus particulièrement la nature de la médiation à proposer autour de l'accès aux corpus et à leur exploitation.

## De la pertinence des corpus pour l'apprentissage des langues en autonomie

- 7 Comme le rappellent Landure et Boulton (2010), un corpus ne renvoie pas à n'importe quel type de recueil de documents. Il s'agit, selon la définition de McEnery *et al.* (2006 : 5), "*d'un grand ensemble de textes<sup>7</sup> authentiques représentatif d'une variété donnée quelle qu'elle soit et exploitable sous forme électronique*". Un corpus se comprend comme un modèle explicatif de l'objet considéré (Chanier & Ciekanski, 2010), ce qui implique un recueil systématique de documents liés à l'objet d'étude. La couverture et la taille sont des indicateurs de la systématisme du recueil. Cette collection de données langagières, sélectionnées et organisées à partir de critères explicites, peut ainsi servir d' "*échantillon du langage*" pour Sinclair (1996 : 4). Pour autant, un corpus n'est pas une agrégation de données langagières. Il implique une structuration permettant l'exploitation des données recueillies.
- 8 Dans cette partie, nous chercherons à mieux comprendre l'intérêt des corpus comme ressource d'apprentissage des langues (section 2.1) et les démarches d'apprentissage favorisées par les corpus (section 2.2). Nous discuterons ensuite de la pertinence de l'utilisation des corpus en situation d'apprentissage en autonomie, à partir des trois caractéristiques proposées par Carette et Holec (1995) auxquelles doit répondre une ressource utilisable en autonomie : l'adaptabilité, l'auto-suffisance et l'accessibilité (section 2.3).

## Corpus et apprentissage des langues

- 9 Pour Dubois *et al.* (2010), à propos de l'évolution des méthodologies en didactique des langues, le développement de nouvelles technologies (qu'elles soient numériques ou non) ou de nouvelles ressources, accompagne, voire provoque, souvent l'avènement de nouvelles méthodes d'enseignement/apprentissage. Les corpus sont porteurs de différentes transformations qui contribuent à modifier la conception de la langue-cible et la conception des ressources pour l'apprentissage des langues. Premièrement, le recours aux corpus propose une alternative aux documents de référence (ex. dictionnaire ou manuel) utilisés jusqu'alors. La nature des corpus permet notamment de sortir de l'exemplarité qui a longtemps servi de support à la compréhension du fonctionnement de la langue-cible en situation d'apprentissage des langues. Cette approche place les contextes d'usage et la variation de la langue au cœur de l'appréhension de la langue-cible. Ainsi, les corpus représentent "*des ressources importantes pour l'apprentissage des*

*langues : phénomène de collocation et phraséologie, microsyntaxe des entrées lexicales, étude des langues de spécialité, typologie des textes*" (Habert *et al.*, 1997 : 15). Comme le souligne Debaisieux (2009 : 43) au sujet des corpus oraux, les corpus apportent un éclairage nouveau sur certaines caractéristiques du discours masquées par des découpages de la langue artificiels opérés par les approches traditionnelles de la linguistique et qui ont nourri les manuels de langues. Le renouveau apporté par la linguistique de corpus a contribué à modifier les savoirs et savoir-faire liés à la description de la langue. Il contribue également à faire évoluer les cultures langagières des enseignants autant que celles des apprenants.

- 10 Deuxièmement, le recours aux corpus pour apprendre les langues hérite de toute la réflexion sur l'utilisation des documents authentiques à des fins d'apprentissage comme le rappelle Boulton (2009). Qu'ajoutent les corpus à l'utilisation de documents authentiques ? Ils apportent trois dimensions importantes : la structuration, la massification et la multicontextualisation (Cobb, 1997) des données consultables ou observables. La confrontation à la complexité de la langue est facilitée par la mise en corpus qui permet la formulation de requêtes et soutient ainsi l'appropriation de la langue par l'apprenant. Pour McEnery *et al.* (2006 : 197, traduit par Debaisieux, 2009 : 51) :
- C'est précisément parce qu'ils n'offrent pas un modèle simple à imiter que les données de corpus présentent un intérêt pour les apprenants. Les données doivent être interprétées de façon subjective. Pour les apprenants, la réalité des données de corpus réside essentiellement dans la possibilité qu'elles leur offrent d'être interprétées et de créer leurs propres modèles.
- 11 Debaisieux (2009 : 44-51) souligne quatre plus-values liées à l'utilisation de corpus dans l'apprentissage des langues :
- l'accès aux régularités à travers un grand nombre d'occurrences ;
  - l'accès aux sens les plus fréquents (et qui n'est possible ni par les dictionnaires, ni par l'intuition chez les natifs) ;
  - l'accès à la variation ;
  - l'accès aux contextes d'utilisation.
- 12 Troisièmement, l'accès à des ressources abondantes tel que le proposent les corpus est-il redondant par rapport à l'utilisation massive d'Internet que l'on constate dans les pratiques des apprenants ces dernières années ? Ces millions de textes disponibles sur Internet forment en effet une "ressource" facile d'accès que la plupart des apprenants ont déjà l'habitude de consulter (Chambers, 2009). De plus, l'essor des technologies numériques a considérablement facilité les moyens d'effectuer des recherches d'information. Pour autant, Internet peut-il être considéré comme un corpus ? Cette question épineuse se pose notamment dans le cas des centres de ressources en langues qui recensent de plus en plus fréquemment des liens Internet renvoyant à des documents authentiques présélectionnés, en termes de contenu, pour une utilisation en autonomie. Toutefois, rappelons avec Rastier (2004 : n.p.) que la notion de critère linguistique est centrale pour pouvoir attribuer le statut de corpus à une collection de textes. Sans cette organisation, Internet n'est qu'une compilation de données langagières non structurées qui ne permet pas de requête systématique.
- 13 Les corpus offrent ainsi une alternative aux documents de référence en termes de contenu, mais aussi une alternative aux documents authentiques issus ou non d'Internet en termes d'utilisation.

## Corpus et modèles d'apprentissage

- 14 Il existe deux grandes approches pour l'utilisation des corpus : *corpus-based* et *corpus-driven* (Tognini-Bonelli, 2001). La première approche fait référence à une démarche *déductive* à partir de catégories prédéfinies permettant l'exemplification du fonctionnement de la langue ; la seconde fait référence à une démarche *inductive*, sans catégories prédéfinies, les catégories de la langue émergeant des données analysées. Nous nous intéresserons plus particulièrement au "*data-driven learning*" (Johns, 1991) ou apprentissage "orienté par les données" qui renvoie à l'exploitation directe des données par l'apprenant, à partir d'une approche qui paraît plus favorable à l'apprentissage en autonomie.
- 15 En quoi consiste cette approche ? Pour Grosbois (2012 : 93), les corpus relèvent d'une "*dynamique actionnelle pleinement constructiviste*" dans laquelle l'activité de l'apprenant est le moteur de l'apprentissage. Cette activité est décrite par Ormrod de la façon suivante :
- An approach to instruction through which students interact with their environment – by exploring and manipulating objects, wrestling with questions and controversies, or performing experiments<sup>8</sup>. (Ormrod, 1995 : 442).
- 16 Dans le cadre de cette approche constructiviste, les chercheurs travaillant sur les corpus en tant que ressources d'apprentissage plébiscitent notamment les activités de découverte (Debaisieux, 2009 ; Bernardini, 2001, 2002), qui soutiennent en particulier l'acquisition du système de la langue, en invitant l'apprenant à adopter une posture réflexive sur le fonctionnement langagier (Kettemann & Marko, 2011). Les corpus permettent un traitement en profondeur de l'*input* (*deep processing*), en combinant la gestion du sens et le repérage conscient des données langagières par l'apprenant, lui permettant de faire évoluer son interlangue. Toutefois, comme le souligne Randall (2007 : 131), un input présélectionné faciliterait le repérage et donc l'apprentissage des apprenants. Ce constat qui se retrouve dans de nombreux travaux sur les applications pédagogiques des corpus questionne l'intérêt réel d'un accès direct aux données brutes en termes d'acquisition.
- 17 L'usage du concordancier associé aux corpus est également perçu comme une aide à l'exploration des corpus, mais aussi à la production langagière (Grosbois, 2012 : 105). Les concordanciers, en proposant des "*chunks*" (un stock de blocs lexicalisés dans lequel le locuteur peut puiser et dont Ellis, 2001, a démontré le rôle dans le processus d'acquisition des langues), donnent à l'apprenant la possibilité de décomposer plus facilement une séquence d'informations en éléments constitutifs de façon à réutiliser ensuite ces derniers de manière créative lors de la production. Les concordances permettent aux apprenants de se confronter à de l'input riche en sens et varié et, à partir de là, de tirer eux-mêmes des conclusions, à se forger des repères sur la langue. Toutefois, Chini (2004 : 80, citée par Grosbois, 2012 : 107) souligne :
- [...] la difficulté qu'il y a non seulement à sortir du schéma traditionnel de transmission d'un savoir prédéfini pour laisser à l'élève un espace de liberté où construire du sens, mais aussi à adapter les concepts et discours de la théorie pour en faire des outils utiles à l'apprentissage.
- 18 Ces difficultés posent la question de la médiation des activités à mettre en place pour soutenir l'apprentissage. Grosbois (2012 : 107) rappelle la complexité cognitive de la recherche d'informations (Rouet *et al.*, 2004). La recherche d'une information pertinente à

partir de données importantes, comme dans le cas des corpus, ne va pas de soi. On retrouve le même constat chez Kettemann et Marko (2011 : 45-46) :

we think that students' explorations would properly benefit from slightly more interventionism and feedback and slightly less autonomy... Emphasis should be laid on the difference between unique, singular phenomena and regular ordinary patterns<sup>9</sup>.

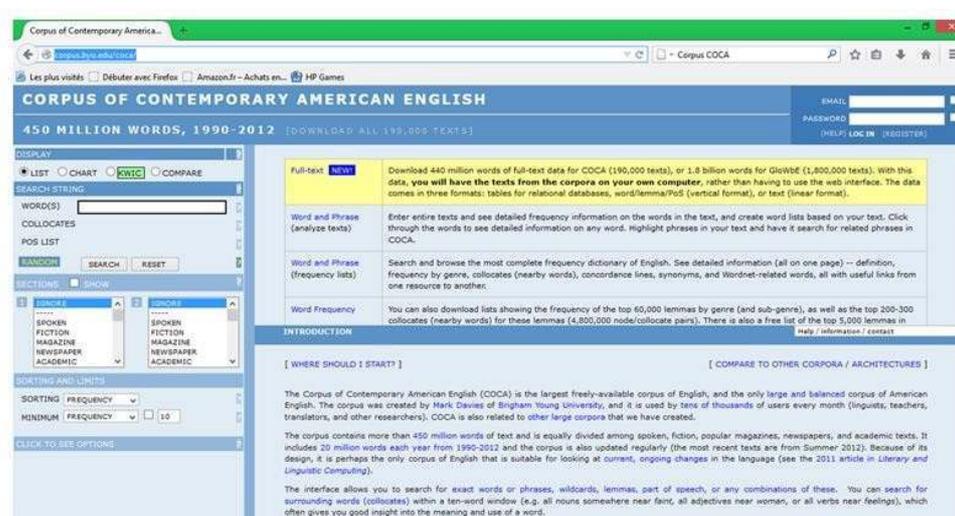
- 19 L'apprentissage sur corpus est sans doute un "*travail créateur*" (Holec, 1990, cité par Debaisieux, 2009 : 51). Les recherches de Debaisieux (2009) et de Schaeffer-Lacroix (2011) tendent à affirmer que les corpus sont porteurs d'une évolution des pratiques et des attitudes d'apprentissage. Il s'agit donc d'aider l'apprenant à se construire des repères langagiers mais aussi des repères d'apprentissage à partir des données consultées. On voit que dans le cas des corpus, l'"*autonomie physique*" (Mangenot, 1996) liée à un accès direct aux corpus doit être assortie d'une autonomie cognitive pour être efficace. Nous discuterons dans la section suivante des caractéristiques permettant aux corpus de répondre à cette autonomie cognitive.

## Corpus et autonomie

- 20 La démarche d'apprentissage qui sous-tend l'utilisation des corpus dans l'apprentissage des langues coïncide avec la culture langagière et la culture d'apprentissage préconisées dans le cadre de l'apprentissage autodirigé (Holec, 1990). On retrouve d'ailleurs l'expression "*corpus de découverte*" chez Holec (1990), et Debaisieux (2009) montre, à partir de l'analyse d'extraits d'entretiens de conseil (cet accompagnement particulier de l'apprentissage dans le cadre de l'apprentissage autodirigé), la façon dont les conseillers aident les apprenants à explorer et exploiter des données langagières issues de documents authentiques pour construire leur compréhension du fonctionnement de la langue. Comment expliquer alors l'absence de corpus dans les ressources sélectionnées par les centres de ressources en langues ?
- 21 Cela peut s'expliquer en partie par la nature de l'offre actuelle. D'une part, beaucoup de corpus institutionnels qui pourtant répondent aux formats standard (ex. la TEI - *Text Encoding Initiative*) et permettent ainsi une utilisation dans un grand nombre de contextes donnent l'impression d'avoir été créés pour répondre aux besoins des chercheurs plutôt qu'à ceux des apprenants. La grande quantité des données recueillies ne répond pas forcément aux besoins des apprenants : cet accès illimité aux données (plusieurs millions parfois) peut se transformer en excès et empêcher tout apprentissage. D'autre part, l'autre tendance repérée par Boulton (2008) est la création de corpus par les enseignants eux-mêmes pour un public ciblé, assortis de concordanciers tout aussi spécialisés ou d'outils issus de l'EAO (Enseignement Assisté par Ordinateur) également conçus de façon *ad hoc*. Ce travail de sélection et de constitution de corpus par les enseignants, même s'il est aujourd'hui facilité par les technologies disponibles comme le souligne Schaeffer-Lacroix (2011), est chronophage et demande une formation linguistique solide. En outre, la question de la mutualisation de ces corpus créés à des fins d'apprentissage ciblé demeure épineuse. Ainsi, la plupart de ces 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. De plus, ces corpus ne sont pas toujours pensés en vue d'analyses multiples, ce qui fait que les formats de conception choisis ne sont pas forcément pensés pour permettre l'utilisation de différents outils d'analyse.

- 22 Toutefois, l'absence de corpus dans les ressources fléchées par les centres de ressources en langues peut également s'expliquer par le fait que les corpus ne répondent pas de façon satisfaisante aux caractéristiques nécessaires pour qu'une ressource soit utilisable sans la médiation d'un enseignant. Comme le rappellent Carette et Holec (1995 : 86), les ressources utilisables dans un apprentissage autodirigé répondent à des caractéristiques qui pallient l'absence d'un enseignant médiateur construisant toutes les passerelles requises entre ce qui doit être appris et celui qui apprend. Il s'agit ainsi de ressources prenant en compte la diversité des utilisateurs et des utilisations que ces derniers en font et qui n'est pas ou peu prévisible. Pour ce faire, les ressources doivent répondre à trois caractéristiques : l'adaptabilité, l'autosuffisance et l'accessibilité (Carette & Holec, 1995 : 87-89). Reprenant chacune de ces caractéristiques, chacun de ces critères, nous nous demanderons dans quelles mesures les corpus peuvent être utilisables en autonomie.
- 23 Pour Carette et Holec (1995), une ressource est *adaptable* lorsqu'elle est construite sans référence à un public donné et à ses caractéristiques propres. Il s'agit de ressources qui sont élaborées non en fonction d'objectifs d'apprentissage, de niveaux, de progressions préétablis, mais en tenant compte principalement d'objectifs d'acquisition (tel savoir-faire communicatif, tel savoir linguistique, etc.), et de principes méthodologiques généraux (l'authenticité de la langue, la couverture de l'ensemble des aspects langagiers à acquérir, etc.). Ce critère d'adaptabilité correspond à la plupart des corpus institutionnels, construits sans référence à un public d'utilisateurs particulier, puisque c'est l'objectif linguistique ou langagier qui a conditionné la sélection et la structuration des données. En ce sens, les corpus sont donc théoriquement utilisables par tout apprenant, en fonction de ses besoins langagiers. En outre, les corpus peuvent être un support pour diverses activités (support d'exploration ou d'observation du fonctionnement de la langue ; support à la production langagière ; support à l'autoévaluation des productions langagières).
- 24 Une ressource est *autosuffisante* lorsqu'elle est utilisable par l'apprenant sans qu'il soit nécessaire de faire appel à une aide extérieure (d'autres documents en complément ou la médiation d'un enseignant). Pour cela, la présentation de la ressource doit être entièrement compréhensible pour l'apprenant qui en fait usage. Les contenus doivent également être complets, c'est-à-dire qu'ils apportent tous les éléments nécessaires pour atteindre l'objectif d'acquisition visé (par exemple, l'exposition langagière doit couvrir toute la zone de fonctionnement communicatif ou linguistique en jeu ; les contenus doivent être en nombre suffisant, etc.). Ce critère d'autosuffisance attire l'attention à la fois sur la description du corpus (est-elle suffisamment claire et signifiante pour qu'un apprenant sache quoi faire avec les données proposées ?) et sur la mise à disposition et la compréhension des outils permettant d'effectuer les requêtes nécessaires à l'apprentissage visé. L'exemple de l'interface proposée par le COCA (*Corpus of Contemporary American English*)<sup>10</sup> montre l'importance accordée aujourd'hui à la médiation proposée pour soutenir l'utilisation du corpus (Figure 1).

Figure 1 – Interface de COCA.



- 25 L'interface propose différents éclairages, matérialisés par différents cadres indépendants les uns des autres, par lesquels l'utilisateur peut entrer dans le corpus, en fonction de son degré de familiarité avec l'outil. Dans la partie centrale, on trouve un tableau correspondant aux diverses modalités de consultation du corpus et un texte de présentation sur la nature du corpus et la façon dont il a été conçu (*Introduction*), adapté en fonction du statut de l'utilisateur (apprenant de langue ou chercheur) (*Where should I start ?*). Dans la colonne de gauche, l'interface propose un ensemble d'outils permettant différents types de requêtes à partir du corpus.
- 26 Les interfaces soutenant l'utilisabilité des corpus ont largement progressé ces dernières années et permettent ainsi plus facilement de répondre au critère d'autosuffisance des corpus pour une utilisation sans enseignant. Toutefois, un temps de familiarisation avec les différents outils demeure nécessaire, notamment pour comprendre la façon de formuler des requêtes appropriées et pour interpréter les résultats des requêtes obtenues.
- 27 Enfin, une ressource est *accessible* lorsqu'elle est disponible, lorsqu'elle permet un accès direct et rapide, et lorsqu'elle assure la transparence de son contenu. Comme nous venons de le voir, le travail particulier de conception de l'interface des corpus est essentiel pour assurer l'accessibilité des corpus. Si la plupart des corpus institutionnels offrent une interface riche pour l'utilisateur (même si le plus souvent l'utilisateur visé est d'abord un chercheur), il n'en est pas de même pour les corpus *ad hoc* créés par les enseignants et qui nécessitent la médiation de l'enseignant pour une utilisation pertinente par l'apprenant, en présentiel ou à distance. En outre, la question se pose différemment en fonction de la langue-cible étudiée puisque, s'il est aisé de disposer de corpus institutionnels pour l'anglais, il n'en est pas encore de même pour toutes les autres langues.
- 28 Le principal frein à l'utilisation de corpus en autonomie est lié à l'accessibilité cognitive des corpus. Même s'il est vrai que le développement de la familiarité des apprenants avec les technologies rend plus aisé le travail possible à partir de concordanciers, la difficulté principale demeure dans l'utilisation des données. Même si l'interface de COCA explique techniquement comment formuler une requête, elle n'aide pas l'apprenant à décider quoi

apprendre, ni comment interpréter les résultats obtenus, comme nous l'avons vu. La médiation permettant de rendre accessibles les données doit à la fois prendre en compte la grande quantité de données potentiellement consultées par l'apprenant et la complexité des requêtes nécessaires à l'apprenant pour son apprentissage (sur  $x$  critères de sélection).

- 29 Le recours aux corpus en situation d'apprentissage en autonomie ne peut donc faire l'impasse d'une médiation spécifique pour conduire à un apprentissage effectif. De cette mise en regard des corpus avec les trois critères définissant la capacité d'une ressource à être utilisée dans un apprentissage autodirigé, il ressort que la question de la médiation des données rassemblées en corpus est fondamentale pour un travail en autonomie effectif. Si l'on regarde la façon dont les documents authentiques sont abordés en apprentissage autodirigé, on constate que la pédagogisation des documents authentiques passe par une proposition essentielle : la proposition d'activités dans lesquelles l'apprenant va apprendre à développer des techniques d'exploration de données en fonction de ses objectifs d'apprentissage et à choisir les corpus pertinents pour ses objectifs langagiers. Ces activités peuvent être préconstruites ou à construire par l'apprenant lui-même, à l'aide de documents de soutien aidant l'apprenant à savoir quoi apprendre et comment apprendre. Des documents permettant d'aider à apprendre en expliquant ce qu'est apprendre avec un corpus et comment le faire semblent ainsi nécessaires à une réelle pédagogisation des corpus.

## Quelle(s) autonomie(s) développer à partir des corpus ?

- 30 Après avoir discuté des potentialités et des limites de l'utilisation des corpus dans le cadre d'un apprentissage autodirigé, nous avons abordé le rôle fondamental joué par le guidage spécifique permettant une réelle utilisation des corpus en autonomie. L'accès direct aux données n'est pas une mise en autonomie des apprenants. La pédagogisation d'un corpus dépend des activités proposées ou suggérées par lesquelles l'apprenant va explorer et utiliser les données.
- 31 Nous nous intéresserons dans cette partie à l'utilisation des corpus au regard des différents types d'autonomie que ces derniers peuvent aider à développer (section 1). Nous verrons de quelle façon ces types d'autonomie sont ou ne sont pas pris en compte par les concepteurs des corpus, à partir de l'étude de deux corpus pédagogiques en libre accès (section 2).

## Facteurs de développement de l'autonomie

- 32 Rappelons avec Rivens Mompean et Eisenbeis (2009) qu'un simple accès direct aux ressources n'est pas suffisant pour conduire au développement de l'autonomie. Le "*présupposé implicite d'autonomie*" des apprenants (Linard, 2003) et qui est sous-jacent à l'utilisation des corpus en accès direct par les apprenants n'existe pas ; il est nécessaire de prendre en compte un éventuel déficit en autonomie afin de proposer un guidage approprié.
- 33 Germain et Netten (2004) redéfinissent différentes formes d'autonomie (autonomie générale, autonomie langagière et l'autonomie d'apprentissage). Nous discuterons des

corpus au regard de l'autonomie langagière et de l'autonomie d'apprentissage. Par *autonomie langagière*, Germain et Netten (2004 : n.p.) entendent "*la capacité de l'apprenant de prendre des initiatives langagières et d'utiliser avec spontanéité des énoncés nouveaux lors d'une situation authentique de communication en L2*". Selon ces auteurs, différentes marques indiquent l'autonomie langagière : la capacité de prendre des initiatives, des risques avec la langue, de produire des énoncés nouveaux, de soutenir une conversation, et d'utiliser de manière spontanée la langue cible. Trois critères semblent contribuer plus particulièrement au développement de l'autonomie langagière : la centration sur l'intérêt de l'apprenant, une approche de la langue axée sur l'interaction et la communication authentique, un nombre minimal d'heures intensives d'utilisation (les auteurs font référence à un volume de 250 heures d'utilisation sur cinq mois).

- 34 De quelle façon les corpus contribuent-ils à l'autonomie langagière des apprenants ? Les travaux portant sur l'utilisation des corpus pour l'apprentissage des langues (par ex. Chambers, 2009 ; Debaisieux, 2009 ; Frankenberg-Garcia, ce volume) soulignent principalement la façon dont l'utilisation des corpus peut engendrer une évolution de la culture langagière des apprenants, une évolution dans la compréhension du système de la langue et une plus grande facilité à produire des énoncés nouveaux. Ces conclusions vont dans le sens de Rastier (2004 : n.p.) :

L'accès aux corpus conduit ainsi à modifier le rapport entre théorie et pratique, tant en amont du processus de recherche, dans la formulation des hypothèses, qu'en aval, dans la recherche de contre-exemples ou de variations. C'est aussi le moyen de sortir des apories théoriques suscitées par la philosophie du langage : par exemple, l'analyse de corpus reste le seul moyen éprouvé pour relativiser la polysémie et contrôler l'ambiguïté ; ou encore, pour déterminer les valeurs des formes grammaticales (par exemple, le futur n'a pas les mêmes valeurs dans le discours juridique que dans le roman).

- 35 Par *autonomie d'apprentissage*, Germain et Netten (2004) font référence, à la suite des travaux d'Holec (1991 : 5), à "*la capacité de mener, activement et de manière indépendante un apprentissage de langue*". Dans la section précédente, nous avons montré que les corpus répondent de façon incomplète aux caractéristiques d'une ressource appropriée à un apprentissage autodirigé, tant que la notion d'autonomie est synonyme d'accès direct aux données des corpus. Selon l'étude menée par Rivens Mompean et Eisenbeis (2009 : 230), le processus d'autonomisation des apprenants serait soutenu par le fait que l'apprentissage est motivé par la recherche conduite par les apprenants. C'est l'utilisation de stratégies d'exploration (avec une dimension essai-erreur) qui contribuerait ainsi au développement de l'autonomie de l'apprenant (Rivens Mompean & Eisenbeis, 2009 : 235). Comme nous l'avons indiqué précédemment, l'utilisation des corpus invite la mise en œuvre de démarches relevant de la stochastique, caractéristique des pratiques d'autoformation. Toutefois, ces démarches d'exploration sur des corpus induisent en particulier des difficultés liées à la masse de ressources proposée par les corpus (notamment institutionnels) et au repérage des données à observer. Le principal levier d'autonomisation repéré par Rivens Mompean et Eisenbeis réside dans la réflexion que mène l'apprenant sur son apprentissage et sur les données langagières consultées, à travers différents types d'étayage (carnets de bord, entretiens, etc.). En outre, le dispositif proposé comme interface entre l'apprenant et les ressources n'est jamais neutre : il aide plus ou moins l'apprenant à s'orienter et à organiser son apprentissage, il est plus ou moins contraignant par rapport aux choix de l'apprenant, il est plus ou moins signifiant pour l'apprenant comme ressource permettant l'apprentissage.

- 36 Nous proposons d'analyser deux corpus pédagogiques en libre accès (*Sacodeyl* et *Fleuron*) au regard de la prise en compte de l'autonomie langagière et de l'autonomie d'apprentissage par les concepteurs des corpus et de leur interface. Selon Debaisieux (2009 : 43) à propos des corpus : "il n'existe pratiquement aucun outil de consultation convivial, adapté à un usage d'apprentissage et surtout libre d'utilisation". Ces deux corpus pédagogiques répondent-ils à ces critères essentiels ?

## Analyse de deux corpus pédagogiques

### Sacodeyl

- 37 Le projet européen *Sacodeyl*<sup>11</sup> (2008) de l'université de Tübingen propose des corpus oraux à des fins pédagogiques sur le parler des adolescents de différents pays. *Sacodeyl* a pour thème la vie des jeunes européens. Ces jeunes, âgés de 13 à 18 ans, sont enregistrés sur vidéo dans leur langue maternelle (7 langues disponibles : allemand, anglais, espagnol, français, italien, lituanien et roumain). Les interviews se focalisent, d'une part, sur des sujets familiers (la famille, la vie quotidienne, les loisirs, les vacances, l'école, le travail, les projets d'avenir), et d'autre part, sur des sujets d'actualité comme l'interdiction de fumer dans les lieux publics, ou ce que l'Europe signifie pour eux.
- 38 Le dispositif qui sous-tend *Sacodeyl* vise à proposer des matériaux authentiques pour l'apprentissage des langues, suivant une approche qui associe l'étude de textes authentiques et des exercices et activités à base de données tirées d'un corpus (Braun, 2005). L'interface intègre une chaîne de traitement complète dont les outils peuvent être déployés du laboratoire à la salle de classe. Le dispositif de consultation et d'utilisation proposé permettrait une utilisation en autonomie :
- The overall objective is to help teachers and learners to proceed from decontextualised textual data to context-embedded discourse construction and thus to facilitate and promote authenticated language learning<sup>12</sup>. (Kohn, 2008 : 173)
- 39 Les corpus qui sont mis à disposition via une plateforme en libre accès sont préalablement annotés pour les thèmes, la grammaire, le lexique, les marqueurs de discours et les niveaux selon le CECR (Conseil de l'Europe, 2000). Des informations sont également données sur le contexte culturel et linguistique des corpus. Les exercices et activités visent à permettre à l'apprenant d'explorer et de repérer les régularités à partir des concordances obtenues et de construire des règles de fonctionnement.
- 40 L'intérêt de *Sacodeyl* pour un apprentissage en autonomie est de proposer non seulement un accès à des données authentiques et aux outils permettant l'exploration de concordances, mais surtout de proposer une diversité de médiations pédagogiques qui en fait un véritable corpus pédagogique, tel que complété dans le projet *Backbone*<sup>13</sup> :
- le choix des genres et des sujets pour un apprentissage de langues ;
  - le choix des annotations du corpus (thématiques, grammaticales, lexicales, caractéristiques textuelles, niveaux du CECRL) ;
  - les concordances, les activités proposées ; les propositions de consignes pour l'exploration et les activités communicatives ;
  - les fonctions de recherches ;
  - la possibilité d'utiliser les formes repérées dans des activités de communication collaboratives (Moodle) ;

- la transcription des documents audio, information sur le contexte de communication linguistique et culturel ;
  - un ensemble d'outils (aide à la transcription, l'annotation, l'organisation des données) permettant la création de son propre corpus (motivation, besoins personnels) ;
  - un projet de corpus ouvert ;
  - une interface permettant le catalogage / la curation des données (fichiers, description de la ressources et feuille ressources).
- 41 Ces différentes médiations vont dans le sens de Kettemann et Marko (proposer une présélection). Toutefois, le recours à ces présélections est ici laissé au choix de l'apprenant (dans un souci d'autonomie).

### Fleuron

- 42 *Fleuron* (Français Langue Étrangère Universitaire : Ressources et Outils Numériques) est un corpus qui s'adresse principalement aux étudiants étrangers souhaitant se rendre dans une université française. Comme pour *Sacodeyl*, les données des corpus sont accessibles à partir d'un site Internet<sup>14</sup>. L'interface de *Fleuron* vise à développer chez l'apprenant sa capacité langagière et sa capacité d'apprentissage. Le recueil de données comprend des occurrences issues de la vie académique (cours, séminaires, conférences, examens oraux, soutenances, interactions entre enseignants et apprenants, entre apprenants et différents services universitaires, interaction entre étudiants en groupe de travail, etc.) ainsi que des occurrences de langue non académique (conversations, vie en dehors du campus, activité de loisirs, etc.). A cela s'ajoutent des témoignages d'étudiants étrangers et d'étudiants français sur la vie à l'université. Ces éléments permettent la prise en compte de la dimension interculturelle des interactions travaillées.
- 43 Les données de *Fleuron* sont organisées selon trois modalités : fournir des documents authentiques portant sur les situations spécifiées et pouvant donner lieu à des activités d'apprentissage concernant la compréhension et l'expression. Il s'agit de documents vidéo ou audio, accompagnés de transcriptions, pour lesquels on peut choisir de montrer ou non les sous-titres. Les documents sont accompagnés de fiches descriptives précises permettant leur sélection en fonction de critères variés, selon un système de catalogage informatif. Ce catalogage contribue à aider les apprenants à développer leurs critères de sélection. *Fleuron* propose également un concordancier qui permet de retrouver les contextes dans lesquels le mot recherché a été utilisé et le lien aux passages des documents sonores ou écrits où il apparaît. La présentation de *Fleuron* comprend des conseils pour apprendre et travailler avec le corpus, de façon à faciliter l'exploration et l'exploitation des données. Ces conseils relèvent de l'apprendre à apprendre tel que développé par le Crapel<sup>15</sup> et suggèrent des activités possibles à partir des corpus proposés, dans le but de travailler ou la compréhension ou la production.

### Limites de l'utilisation des corpus en situation d'autonomie ?

- 44 La description des deux corpus pédagogiques *Sacodeyl* et *Fleuron* souligne le souci accordé par les concepteurs des corpus à l'interface pédagogique proposée pour optimiser l'utilisation des corpus par les apprenants eux-mêmes. Les activités proposées permettent de rendre exploitables les corpus sans le recours à un enseignant. Toutefois, ces activités

de réflexion préparant une utilisation de la langue-cible ne sont pas toujours associées à des pistes d'autoévaluation, ce qui en réduit la portée.

- 45 L'une des pistes évoquées en sections 2 et 3 pour faciliter l'utilisation des corpus par les apprenants était de proposer à ces derniers des présélections, des requêtes précodées pour accompagner chez eux la construction de la signification. Toutefois, ces présélections, si elles facilitent l'accès aux données, peuvent également déformer la représentation de la langue que se font les apprenants. En effet, une sélection peut grossir le trait et mettre en avant des aspects de la langue de façon artificielle (que ce soit pour *Sacodeyl* sur "le parler jeune" ou pour *Fleuron*). Il reste donc essentiel de travailler avec les apprenants ce qu'est un corpus, à la fois comment on le construit, à partir de quels présupposés linguistiques ou sociolinguistiques, et ce qu'il permet de comprendre du fonctionnement de la langue. Pour Rastier (2004 : n.p.) :

Aucun corpus ne représente la langue... En revanche, il est adéquat ou non à une tâche en fonction de laquelle on peut déterminer les critères de sa représentativité et de son homogénéité.

- 46 Tout corpus dépend du point de vue qui a présidé à sa constitution et il est nécessaire pour les apprenants d'y être sensibilisés.

## Conclusion

- 47 Nous nous sommes intéressées dans ce travail aux applications pédagogiques des corpus en autonomie. Nous avons notamment discuté de l'accès direct au corpus qui est souvent la modalité proposée aux étudiants travaillant à partir de corpus à des fins d'apprentissage de langues. Nous avons interrogé la façon dont cet accès direct aux corpus proposé en classe est transférable en contexte d'apprentissage autodirigé ou en autonomie de façon plus large. Nous avons montré en quoi le recours aux corpus pouvait soutenir le développement d'une modalité d'apprentissage favorable à l'apprentissage en autonomie. Nous avons également discuté de la portée des corpus sur le développement de l'autonomie langagière. Enfin, nous avons souligné l'importance de concevoir l'interface d'accès aux corpus comme un dispositif d'apprentissage à part entière (en proposant notamment des activités et des présélections) pour que les corpus soient effectivement utilisables sans médiation externe.

- 48 L'autonomie des apprenants est souvent associée à l'autonomie des enseignants qui sont les premiers passeurs de ressources. Or, l'un des principaux freins à l'utilisation des corpus semble provenir des enseignants eux-mêmes. Pour McCarthy (2008), "*at present many teachers feel intimidated and unsure about corpora, and remain passive consumers of corpus-based materials rather than proactive in demanding them*"<sup>16</sup>. Pour Burton (2012), cela est renforcé par les représentations des éditeurs de ressources éducatives qui sont également peu favorables à ou intéressés par la diffusion de corpus pour l'apprentissage. Comment encourager le recours aux corpus dans ce contexte ? Selon Chambers (2009), la collaboration entre linguistiques et auteurs de manuels, ainsi que l'emploi des textes authentiques dans l'enseignement des langues, pourraient permettre aux enseignants de répondre à une description plus authentique et juste de la langue (Dewaele & Wourm, 2002). C'est le parti-pris qu'ont déjà choisi d'adopter de façon systématique certains manuels d'allemand pour le secondaire<sup>17</sup>. C'est sans doute cette exigence de nouvelles descriptions linguistiques qui permettra d'accueillir plus favorablement les corpus et de

préparer les apprenants à une utilisation de ces ressources dans la perspective d'un apprentissage tout au long de la vie.

---

## BIBLIOGRAPHIE

Albéro, B. (2000). *L'autoformation en contexte institutionnel. Du paradigme de l'instruction au paradigme de l'autonomie*. Paris : L'Harmattan.

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

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. pp. 225-234.

Bernardini, S. (2002). "Exploring new directions for discovery learning". In Kettemann, B. & G. Marko (dir.). *Teaching and learning by doing corpus analysis*. Amsterdam : Rodopi. pp. 165-182.

Boulton, A. (2008). "Esprit de corpus : promouvoir l'exploitation de corpus en apprentissage des langues". *Texte et Corpus*, vol. 3. pp. 37-46. Disponible en ligne. [http://web.univ-ubs.fr/corpus/jlc5/ACTES/ACTES\\_JLC07\\_boulton.pdf](http://web.univ-ubs.fr/corpus/jlc5/ACTES/ACTES_JLC07_boulton.pdf)

Boulton, A. (dir.) (2009). Des documents authentiques oraux aux corpus : questions d'apprentissage en didactique des langues". *Mélanges CRAPEL*, vol. 31. pp. 5-14.

Boulton, A. 2011. "Bringing corpora to the masses: free and easy tools for interdisciplinary language studies". In Kübler, N. (dir.) *Corpora, language, teaching, and resources: from theory to practice*. Berne : Peter Lang. pp. 69-96.

Boulton, A. & Tyne, H. (2014). *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*. Paris: Didier.

Braun, S. (2005). "From pedagogically relevant corpora to authentic language learning content". *ReCALL*, vol. 17, n° 1. pp. 47-64.

Braun, S. & Chambers, A. (2006). "Elektronische Ressourcen für den Fremdsprachenunterricht". In Jung, U. (dir.). *Praktische Handreichung für Fremdsprachenlehrer*. Frankfurt : Peter Lang. pp. 330-337. Disponible en ligne. <http://epubs.surrey.ac.uk/cgi/viewcontent.cgi?article=1006&context=translation>

Burton, G. (2012). "Corpora and course books: destined to be strangers forever? " *Corpora*, n° 7 (1), pp. 91-108.

Carette, E. & Holec, H. (1995). "Quels matériels pour les centres de ressources ?" *Mélanges CRAPEL*, vol. 22. pp. 85-94.

Chambers, A. (2007). "Popularising corpus consultation by language learners and teachers". In Hidalgo, E., Quereda, L. & Santana, J. (dir.) *Corpora in the foreign language classroom*. Amsterdam : Rodopi. pp. 3-16.

- Chambers, A. (2009). "Les corpus oraux en français langue étrangère : authenticité et pédagogie". *Mélanges CRAPEL*, vol. 31. pp. 15-34.
- Chanier, T. & Ciekanski, M. (2010). "Utilité du partage des corpus pour l'analyse des interactions en ligne en situation d'apprentissage : un exemple d'approche méthodologique autour d'une base de corpus d'apprentissage". *Alsic*, vol. 13. Disponible en ligne. <http://alsic.revues.org/1666>
- Chini, D. (2004). *Entre didactique et psycholinguistique, l'activité en classe de langue : quel rôle pour la problématique énonciative ?* Note de synthèse en vue de l'obtention de l'HDR. Université de Pau et des Pays de l'Adour. Disponible en ligne. <http://acedle.org/IMG/pdf/DChiniHDR.pdf>
- Cobb, T. (1997). "Is there any measurable learning from hands on concordancing?". *System*, vol. 25, n° 3. pp. 301-315.
- Conseil de l'Europe (2000). *Cadre européen commun de référence pour les langues vivantes : Apprendre, enseigner, évaluer*. Strasbourg : Conseil de l'Europe. Disponible en ligne. [http://www.coe.int/t/dg4/linguistic/Source/Framework\\_fr.pdf](http://www.coe.int/t/dg4/linguistic/Source/Framework_fr.pdf)
- Debaisieux, J.-M. (2009). "Des documents authentiques oraux aux corpus : un défi pour la didactique du FLE". *Mélanges CRAPEL*, vol. 31. pp. 35-56.
- Dewaele, J.-M. & Wourm, N. (2002). "L'acquisition de la compétence sociopragmatique en langue étrangère". *Revue Française de Linguistique Appliquée*, vol. 7, n° 2. pp. 129-143.
- Dubois, M., Kamber, A. & Skupien Dekens, C. (2010). "L'exploitation didactique des documents authentiques audio et vidéo dans l'enseignement des langues étrangères". *Bulletin VALS-ASLA*, vol. 92. pp. 1-7.
- Duda, R. & Tyne, H. (2010). "Authenticity and autonomy in language learning". *Bulletin Suisse de Linguistique Appliquée*, vol. 92. pp. 87-106.
- Ellis, N. (2001). "Constructions, chunking, and connectionism: the emergence of second language structure". In Doughty, C. & Long, M. (dir.). *Handbook of second language acquisition*. Oxford: Blackwell. pp. 63-103.
- Germain, C. & Netten, J. (2004). "Facteurs de développement de l'autonomie langagière en FLE/FLS". *Alsic*, vol. 7. Disponible en ligne. <http://alsic.u-strasbourg.fr>
- Grosbois, M. (2012). *Didactique des langues et technologies*. Paris : Presses de l'Université Paris-Sorbonne.
- Habert, B., Nazarenko, A. & Salem, A. (1997). *Les linguistiques de corpus*. Paris: Armand Colin.
- Halliday, M.A.K. (1989). "Part A". In Halliday, M.A.K. & Hasan, R. (dir.). *Language, context, and text: aspects of language in a social-semiotic perspective*. Oxford : Oxford University Press. pp. 55-79.
- Holec, H. (1990). "Qu'est-ce qu'apprendre à apprendre ?". *Mélanges Pédagogiques*, vol. 1990. pp. 75-87. Disponible en ligne. <http://www.atilf.fr/spip.php?article3658>
- Holec, H. (1991). "Autonomie de l'apprenant : de l'enseignement à l'apprentissage". *Éducation Permanente*, vol. 107. pp. 1-5.
- Johns, T. (1991). "Should you be persuaded: two examples of data-driven learning". In Johns, T. & King, P. (dir.). *Classroom concordancing. English Language Research Journal*, vol. 4. pp. 1-16.
- Kettemann, B. & Marko, G. (2011). "Data-driving critical discourse analysis: learning about language and ideology by autonomously exploring a corpus (of creationist literature)". In Kübler, N. (dir.). *Corpora, language, teaching, and resources: from theory to practice*. Berne: Peter Lang. pp. 20-48.

- Kohn, K. (2008). "Telos Language Partner: 'do it yourself' authoring for content-based language learning". In Gimeno, A. (dir.). *Computer assisted language learning: authoring tools for web-based courseware development*. Valencia : Universidad Politécnic de Valencia. pp. 157-174.
- Landure, C. & Boulton, A. (2010). "Corpus et autocorrection pour l'apprentissage des langues". *ASp*, vol. 57. pp. 11-30. Disponible en ligne. <http://asp.revues.org/931>
- Linard, M. (2003). "Autoformation, éthique et technologies : enjeux et paradoxes de l'autonomie". In Albéro, B. (dir.). *Autoformation et enseignement supérieur*. Paris : Hermès/Lavoisier. pp. 241-263.
- Lüdeling, A. & Walter, M. (2009). "Korpuslinguistik für Deutsch als Fremdsprache-Sprachvermittlung und Spracherwerbsforschung". Version élargie de Lüdeling/Walter (2010) *Korpuslinguistik. HSK 35 : Deutsch als Fremdsprache*. Berlin : Mouton de Gruyter.
- Mangenot, F. (1996). "Informatique et autonomie dans l'apprentissage des langues". In Mariani, L. (dir.). *L'autonomia nell'apprendimento linguistico. Actes du congrès "Lingua e Nuova Didattica (LEND) de Montecatini"*. Florence: Nuova Italia Editrice, Quaderni del LEND.
- McCarthy, M. (2008). "Accessing and interpreting corpus information in the teacher education context". *Language Teaching*, vol. 41, n° 4. pp. 563-74.
- McEnergy, T., Xiao, R. & Tono, Y. (2006). *Corpus-based language studies*. Londres : Routledge.
- O'Keeffe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- Ormrod, J.E. (1995). *Educational psychology: principles and applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Randall, M. (2007). *Memory, psychology and second language learning*. Amsterdam/Philadelphia : John Benjamins.
- Rastier, F. (2004). "Enjeux épistémologiques de la linguistique de corpus". *Texte !*, vol. juin. Disponible en ligne. [http://www.revue-texto.net/Inedits/Rastier/Rastier\\_Enjeux.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html)
- Rivens Mompean, A. & Eisenbeis, M. (2009). "Autoformation en langues : quel guidage pour l'autonomisation ?" *Les Cahiers de l'Acedle*, vol. 6, n° 1. p. 221-244.
- Rouet, J.-F., Coutelet, B. & Dinet, J. (2004). "La recherche d'informations dans les documents complexes : processus cognitifs, apprentissage et développement". *Conférence invitées à la "Journée d'étude sur le traitement cognitif des systèmes d'informations complexes (JETCSIC)"*. Disponible en ligne. <http://tecfa.unige.ch/tecfa/jetcsic2004/papers/jfrjetcsic2004.pdf>
- Schaeffer-Lacroix, E. (2011). "Création de corpus réfléchis pour l'enseignement de la production écrite : l'exemple de l'allemand, langue étrangère". In Estève, I., Faure, A., Guitton, A., Mout, T., Souque, A. & Toutati, A. (dir.). *Autour des langues et du langage, vol. 2 - perspective pluridisciplinaire*. Grenoble : Presses universitaires de Grenoble. pp. 173-180.
- Sinclair, J. (1996). *Preliminary recommendations on corpus typology*. Technical report EAGLES (Expert Advisory Group on Language Engineering Standards).
- Tremblay, N. (2003). *L'autoformation pour apprendre autrement*. Montréal : Les Presses Universitaires de Montréal.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam : John Benjamins.
- Tyne, H. (2009). "Corpus oraux par et pour l'apprenant". *Mélanges CRAPEL*, vol. 31. pp. 91-112.

Uzar, R. (2003). A toolbox for translation quality assessment. In Lewandowska-Tomaszczyk, B. (dir.). *Practical applications in language and computers: PALC 2003*. Frankfurt : Peter Lang, pp. 153-162.

#### Sites Internet

COCA : *Corpus Of Contemporary American English*. <http://corpus.byu.edu/coca/>

*Fleuron: Français Langue Etrangère Ressources et Outils Numériques*. <https://apps.atilf.fr/fleuron2/>

*Sacodeyl: System-Aided Compilation and Open Distribution of European Youth Language*. <http://www.um.es/sacodeyl/>

## NOTES

1. L'aspect le plus attractif des corpus pour l'apprentissage des langues réside peut-être dans leur potentiel pour l'apprentissage en autonomie.
2. L'enseignant par conséquent agit comme un guide ou un conseiller et non plus comme un médiateur du savoir ou comme un évaluateur de la performance de l'étudiant.
3. Si la consultation de corpus doit être une activité commune aux apprenants de toute la gamme des études de langues (l'apprentissage général des langues, les études littéraires, l'apprentissage des langues sur objectifs spécifiques, la traduction, etc.), il semble nécessaire de penser son évolution dans un contexte plus large. C'est peut-être en dehors de la salle de classe qu'aura lieu la prochaine étape importante dans ce domaine de recherche.
4. *Mélanges CRAPEL, ALSIC, LLTJ, SISAL, System*.
5. Relevé effectué à partir des sites Internet proposés par quelques centres de ressources en langues universitaires français : le Centre de langues de l'Université Lille 3, le Centre de Langues de l'Université Lyon 2, l'interface Netvibes des Centres de Ressources en Langues de l'Université de Strasbourg.
6. A l'exception de la mention au corpus *Fleuron* (Français Langue Etrangère Ressources et Outils Numériques) sur le blogue FLE de *Coalea* (Conseil et Apprentissage des Langues Etrangères En Autonomie) ; <http://coalea.blogspot.fr/>.
7. Pour nous, au sens d'Halliday (1989).
8. Une approche de l'enseignement dans laquelle les étudiants interagissent avec leur environnement – en explorant et en manipulant des objets, en étant aux prises avec des questions et des controverses, ou en expérimentant.
9. Nous pensons que les explorations des étudiants bénéficieraient utilement d'un peu plus d'intervention et de feedback de la part des enseignants et d'un peu moins d'autonomie. L'accent pourrait être mis sur la différence entre les phénomènes uniques, singuliers, et les structures régulières et ordinaires.
10. COCA : <http://corpus.byu.edu/coca/>.
11. *Sacodeyl: System-Aided Compilation and Open Distribution of European Youth Language*. <http://www.um.es/sacodeyl/>
12. L'objectif global est d'aider les enseignants et les apprenants à aller de données textuelles décontextualisées vers la construction d'un discours contextualisé et ainsi de faciliter et de promouvoir un apprentissage rendu lui-même plus authentique de la langue.
13. *Backbone* project : <http://projects.ael.uni-tuebingen.de/backbone/moodle>.
14. *Fleuron* : <https://apps.atilf.fr/fleuron2/>
15. *Crapel* : Centre de Recherches et d'Applications Pédagogiques en langues, équipe didactique des langues et sociolinguistique du laboratoire ATILF, UMR 7118

16. A l'heure actuelle, beaucoup d'enseignants se sentent intimidés et incertains à propos des corpus et restent des consommateurs passifs de matériels basés sur corpus plutôt que d'être proactifs en les exigeant.

17. *Kiosk Seconde* (Hatier), Collection Macaire, avec un travail de réflexion sur la conceptualisation et l'approche méta de la grammaire par corpus, menée par René Métrich (Atilf) et les auteurs.

---

## RÉSUMÉS

Les corpus, en tant que ressources d'apprentissage, connaissent un intérêt grandissant en didactique des langues étrangères. Il subsiste néanmoins un écart entre le consensus théorique qui se dégage des recherches menées sur l'utilisation effective des corpus et leur intégration dans les pratiques en contexte d'apprentissage. Nous nous intéresserons plus particulièrement à l'utilisation directe des corpus par les apprenants, utilisation souvent qualifiée d' "utilisation en autonomie" par les auteurs du domaine. Nous chercherons, d'une part, à cerner la nature des corpus en tant que ressource d'apprentissage : les corpus répondent-ils aux caractéristiques d'une ressource adaptée à l'apprentissage en autonomie ? Nous discuterons, d'autre part, des freins et des leviers liés à l'utilisation des corpus dans les apprentissages en autonomie, en questionnant plus particulièrement la nature de la médiation à proposer autour de l'accès aux corpus et à leur exploitation.

Corpora as learning resources are of growing interest in language teaching and learning. Nevertheless, there is still a gap between the theoretical consensus that emerges from research on corpora for learning purposes and the integration of corpora into teaching and learning practices. This article focuses on the direct use of corpora by learners themselves, which is often described as 'self-access'. We seek first to better understand the nature of the corpora as a learning resource: are corpora a suitable independent learning resource? We then discuss the advantages and obstacles related to the use of corpora in autonomous learning, questioning in particular the type of the mediation offered to support access to corpora and their exploitation.

## INDEX

**Mots-clés** : autonomie, autonomisation, corpus, apprentissage "orienté par les données", médiation

**Keywords** : autonomy, self-access, corpora, data-driven learning, guidance

## AUTEUR

### MAUD CIEKANSKI

Maud Ciekanski est maître de conférences au laboratoire ATILF (UMR 7118), Equipe Didactique des Langues et Sociolinguistique (Crapel) à l'université de Lorraine.

Ses domaines de recherche concernent la didactique des langues, l'analyse du discours et les sciences de l'éducation. Ses travaux portent sur les pratiques langagières et formatives des conseillers en apprentissage autodirigé des langues, les environnements d'apprentissage en ligne multimodaux ainsi que les pratiques langagières qui en découlent, et les apprentissages de langues informels.

Courriel : [maud.ciekanski@univ-lorraine.fr](mailto:maud.ciekanski@univ-lorraine.fr)

Adresse : ATILF CNRS, 44 avenue de la Libération, BP 30687, 54063 Nancy cedex.