Corpus et perspectives pour l'enseignant : Compétences, formation, outils, besoins, activités, objectifs

Geoffrey Sockett

Édition électronique

URL : <http://journals.openedition.org/rdlc/1691>

DOI : 10.4000/rdlc.1691

ISSN : 1958-5772

Éditeur

ACEDLE

Référence électronique

Geoffrey Sockett, « Corpus et perspectives pour l'enseignant : Compétences, formation, outils, besoins, activités, objectifs », *Recherches en didactique des langues et des cultures* [En ligne], 11-1 | 2014, mis en ligne le 07 janvier 2014, consulté le 20 avril 2019. URL : <http://journals.openedition.org/rdlc/1691> ; DOI : 10.4000/rdlc.1691

Ce document a été généré automatiquement le 20 avril 2019.

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

Corpus et perspectives pour l'enseignant : Compétences, formation, outils, besoins, activités, objectifs

Geoffrey Sockett

Introduction

- 1 En abordant le domaine de l'apprentissage sur corpus (*data-driven learning* – DDL) du point de vue de l'enseignant, il est nécessaire de se poser un certain nombre de questions afin de mettre en perspective l'approche, la pratique de l'apprenant et celle de l'enseignant. Ce sont ces trois aspects que nous aborderons dans les paragraphes qui suivent. Cette démarche nous permettra notamment de poser la question de l'apprenant et des compétences qu'il pourra développer en mettant en œuvre de tels apprentissages. Il sera ainsi possible de questionner les pratiques de classe sous-entendues par l'apprentissage sur corpus afin de mieux préparer l'enseignant à appréhender une telle approche pédagogique.
- 2 Du point de vue du linguiste, certaines croyances par rapport à l'enseignement/apprentissage des langues en France aujourd'hui peuvent conduire à une recherche de méthodologies novatrices. Bon nombre de spécialistes constatent un faible niveau chez les étudiants en début de cursus universitaire en langues vivantes. D'autres considèrent que les aspects linguistiques à proprement parler sont insuffisamment approfondies chez des publics en langues qui se destinent à des métiers tels que l'enseignement et la traduction. Parmi les facettes de cette formation approfondie figurerait la linguistique de corpus qui sous-tend le DDL.
- 3 Du point de vue du didacticien, il est possible d'avancer d'autres hypothèses pour expliquer d'éventuelles faiblesses chez l'apprenant de langues vivantes en France aujourd'hui. Parmi les facteurs présentés par plusieurs études européennes récentes

(Commission européenne, 2012 ; Stevens & Shield, 2010) figurent la perception de la pertinence de la langue étudiée pour la vie réelle et l'exposition à la langue dans les médias tant traditionnels que nouveaux. Il est également important de prendre en compte le fait que les divers aspects de la langue ne sont pas absorbés par l'étudiant mais se construisent activement sur la base de son système langagier existant. Ainsi l'acquisition d'une langue est effectivement "*data-driven*" dans la mesure où l'on apprend notamment en remarquant des occurrences fréquentes dans notre exposition à des contenus. Parmi les phénomènes récents qui influent sur cette exposition figure l'apprentissage informel en ligne qui postule que l'apprenant consomme des quantités importantes de médias en langue cible et que cette consommation contribue de manière significative à sa construction de la langue.

- 4 Selon une enquête européenne sur les compétences linguistiques (Commission européenne, 2012), des pays tels que la France et l'Espagne se situent loin derrière certains pays du nord de l'Europe en termes de compétences dans la première langue vivante à la sortie du collège, avec en moyenne 86 % des collégiens français atteignant un niveau inférieur à B1 (moyenne de trois activités langagières) alors qu'ils ne sont que 18 % en Suède par exemple. L'enquête a donc cherché des explications par rapport à ce décalage du côté des apprenants ainsi que du côté de la formation des enseignants.

On observe une relation positive entre l'aptitude linguistique dans la langue testée et la perception que les élèves ont de la connaissance de cette langue par leurs parents. De même, si les élèves y sont exposés et l'utilisent à travers les médias traditionnels ou récents. Les différents indices ayant trait à la formation initiale et continue [des enseignants] n'ont que peu d'incidence sur l'aptitude linguistique [des apprenants]. (Commission européenne, 2012 : 13)

- 5 Pris ensemble, ces deux constats encouragent le chercheur à examiner davantage l'univers privé de l'apprenant plutôt que les pratiques de classe pour mieux comprendre les forces et faiblesses de l'apprenant moderne.
- 6 Lorsqu'on cherche à caractériser les défis de la mise en œuvre de l'apprentissage sur corpus ou DDL, un certain nombre de difficultés se présentent. L'approche reste minoritaire, s'avère chronophage, nécessite l'utilisation de logiciels qui manquent parfois de convivialité, et suscite souvent davantage d'enthousiasme chez certains enseignants que chez le public apprenant (Boulton, 2012).
- 7 Si l'apprentissage est effectivement une question de corpus, mais que les approches *data-driven* peinent à s'imposer, il serait utile de procéder à une analyse de l'apprentissage par corpus / DDL en tant qu'activité afin de déceler les origines de ce paradoxe.

Figure 1 – Triangle d'activité DDL

- 8 La Figure 1 reprend la description du DDL développée dans les recherches publiées (voir Boulton, 2010, pour une synthèse), en les resituant dans un triangle d'activité de troisième génération du type proposé par Engeström (1990). Si les trois aspects de l'activité évoqués en haut du triangle sont relativement transparents, s'adressant clairement à un public universitaire spécialiste de la langue, c'est dans le bas du schéma que les interrogations quant au DDL en tant qu'activité de classe se précisent. En effet, une activité de classe se définit largement en termes de règles (quels sont les processus de prise de décision autour de l'activité ?), de communauté (quel est l'impact de faire l'activité avec d'autres personnes ?) et la division du travail (qui fait quoi ?). Sur tous ces points, les réponses apportées par Boulton ne permettent pas de cerner la valeur ajoutée du DDL comme activité de classe par rapport à la même activité entreprise à titre individuel. Pour reprendre le terme utilisé par Johns (1997 : 101), si chaque apprenant est invité à devenir un "Sherlock Holmes" de l'apprentissage, rappelons toutefois que le détective en question ne travaillait pas souvent en équipe. Une activité autodirigée se prête donc mieux à un travail intellectuel qu'à un travail en cours. De même, la notion de communauté impliquée dans l'activité conjointe est faible car elle se limite à un ensemble d'étudiants qui n'ont pas entrepris l'activité dans un but précis et commun élaboré à partir des préférences et centres d'intérêt du groupe. La division du travail enfin implique que chaque apprenant résout une tâche choisie par l'enseignant ou par lui-même, sans qu'il y ait de valeur ajoutée dans la mise en œuvre de rôles complémentaires dans la salle de classe ; l'élément-clé d'une pédagogie DDL sera donc la mise en place d'une tâche globale à laquelle les différents participants contribuent. L'usage, enfin, de l'outil informatique en cours, qui peut dans certains cas favoriser la communication, risque dans celui du DDL de devenir un obstacle à l'interaction lorsque la tâche consiste à trouver une information dans des fichiers plutôt qu'à négocier le sens de l'information au travers d'outils de communication.

- 9 Ainsi la valeur ajoutée du DDL comme activité collective par rapport à l'activité individuelle n'émerge pas clairement d'une première analyse de l'activité. Ce manque d'attractivité du DDL en salle de classe s'explique par un ensemble de facteurs sous-jacents à cette démarche pédagogique. En effet, si l'un des buts du DDL est de simuler l'exposition à des structures trouvées dans un discours authentique en en augmentant les occurrences, cette augmentation se fait au prix de la suppression du contexte plus large de l'utilisation de la structure. Dépourvue de ce contexte, la structure sera perçue par l'apprenant comme une forme grammaticale plutôt que comme un élément ayant un sens dans un récit authentique. Ce risque est toutefois connu dans la recherche et des démarches pédagogiques orientées vers le discours font aussi partie du paysage DDL (par ex. Charles, 2007). L'apprentissage sur corpus sera donc compatible avec une pédagogie à visée actionnelle en tant que pré- ou post- tâche permettant une réflexion (prise de distance) sur la nature des structures mises en œuvre.
- 10 Si les arguments autour de l'attractivité du DDL ne se limitent pas à l'approche pédagogique, l'utilisation d'outils informatiques dans le cadre du cours pourrait représenter une motivation complémentaire pour l'apprenant. Là encore, la réticence des apprenants à se prendre au jeu nécessite réflexion. Si, il y a 20 ans, l'utilisation d'outils informatiques en classe pouvait représenter un facteur important de motivation des apprenants, aujourd'hui chaque apprenant a ses habitudes et préférences par rapport à l'utilisation de ces outils, tant dans la sphère privée qu'en cours. Ainsi il se peut que certains apprenants se positionnent davantage comme décideurs devant les tâches courantes impliquant l'outil informatique s'ils ont l'habitude dans d'autres contextes de choisir quel logiciel utiliser, quelle démarche de recherche d'informations adopter, etc. Proposer un cours dans lequel ces choix ne relèvent pas de l'apprenant ne favoriserait donc pas l'intersubjectivité et la participation aux activités.
- 11 Une approche DDL sera donc plus attractive à chaque fois que l'enseignant fait l'effort d'associer l'apprenant aux choix concernant le matériel, les logiciels et corpus choisis (voire produits), et de permettre à l'apprenant de retrouver un rôle auquel il est plus habitué, celui d'acteur devant l'outil informatique. L'ordinateur n'est souvent plus un objet qui appartient à l'Université et qui ne s'utilise que dans un contexte formel. L'apprenant se situe davantage comme propriétaire de l'outil et des logiciels et pratiques associés, et s'en sert autant dans le cadre privé des loisirs que dans le cadre des cours.
- 12 L'apprenant moderne a en effet un ensemble de pratiques langagières associées à l'outil informatique, que nous avons déjà décrit (Sokkett, 2014) comme "apprentissage informel en ligne", et qui relèvent principalement du domaine privé. L'exposition accélérée à des occurrences de structures dans la langue cible dans le cadre informel passe souvent par une surconsommation de séries télévisées et de films sur internet en langue cible (notamment en anglais), de musiques avec des paroles en langue cible et d'échanges dans le cadre de jeux en ligne multi-joueurs ou de réseautage social.
- 13 Arriver à la conclusion le web est un corpus n'est certes pas une nouveauté (Kilgarrieff & Grefenstette, 2003), mais encore faut-il déterminer en quoi une activité informelle en ligne peut ressembler à une activité DDL (Boulton, à paraître). Un corpus est souvent décrit comme un ensemble de textes lisible par une machine et représentatif d'un ensemble plus grand (McEnery & Wilson, 1996), alors que la recherche sur l'apprentissage informel met l'accent sur le fait que l'apprenant a déjà à sa disposition de grands ensembles de textes. Des études récentes (Sokkett, 2011 ; Kussyk & Sokkett, 2012) ont déjà permis de montrer que les ensembles de textes auxquels sont exposés des apprenants

dans un cadre informel présentent des agglomérats de langage (collocations plus ou moins fixes) à une fréquence similaire à celle de la vie réelle et que l'acquisition de ces agglomérats est possible sans attention explicite à la forme.

- 14 Dans ce cas, des activités informelles telles que l'écoute de centaines d'heures de chansons en anglais jumelée à la consultation de sites web proposant des transcriptions, ou encore le visionnement en ligne de centaines d'heures de séries télévisées en anglais avec sous-titrage en anglais et/ou en français, ressemblent à des activités DDL. Plusieurs études (Sockett, 2014 ; Kusk & Sockett, 2012) indiquent que ces activités concernent une majorité importante d'étudiants du secteur Lansad (langues pour spécialistes d'autres disciplines), et qu'elles ne sont donc pas limitées aux seuls spécialistes qui constituent souvent le public privilégié en DDL.

Facteurs favorisant l'acquisition : la fréquence

- 15 Si l'acquisition d'une langue peut avoir lieu dans des contextes informels, loin de la salle de classe ou du pilotage par l'enseignant, cela suppose que la structure de la langue permette une telle acquisition. La démarche du DDL consiste à créer artificiellement une concentration de structures cibles pour en faciliter l'analyse et l'apprentissage. Toutefois la distribution statistique des structures dans une langue en milieu naturel comporte aussi une forte concentration de certaines structures, augmentant ainsi la visibilité de ces dernières du point de vue de l'apprenant.
- 16 Telle est la position de la grammaire des constructions (Goldberg, 1995 ; Bouveret & Legallois, 2012) qui soutient que toute langue présente une distribution conforme à la loi de Zipf-Mandelbrot selon laquelle le rang d'un mot ou d'une construction est inversement proportionnel à sa fréquence. Ainsi le dixième mot ou construction le plus fréquent dans une langue donnée est dix fois plus fréquent que le centième. Cela signifie concrètement que les structures fréquentes d'une langue sont très fréquentes. Sinclair (1991) indique par exemple qu'en anglais 14 % des occurrences de la structure *a * of* ne concernent que quatre mots (*lot, kind, number* et *couple*), et que 65 % des occurrences de *couple* se produisent dans *a * of*.
- 17 La grammaire des constructions considère aussi que les structures grammaticales s'apprennent en milieu naturel par l'apprentissage de l'exemple prédominant, ou prototype de chaque structure et ensuite seulement par une généralisation à d'autres exemples moins fréquents. Ainsi en anglais, Goldberg (1995) soutient que la structure ditransitive est apprise d'abord par les exemples *to give someone something* (*give me the ball, etc.*) et *to tell someone something* car ces deux exemples représentent plus de 60 % des occurrences de la structure dans les corpus étudiés.
- 18 Ainsi une perspective inspirée de l'étude des acquisitions en milieu informel et d'une grammaire des constructions considère qu'une grande partie du développement langagier est déjà *data-driven*, c'est-à-dire qu'elle est impulsée par des régularités dans les très nombreux documents auxquels l'apprenant/utilisateur de la langue est exposé. L'intérêt d'une démarche "*data-driven*" dans la salle de classe est donc de proposer une version systématisée de ces phénomènes en présentant une grande quantité de données comparables entre elles.

Changer nos pratiques pédagogiques

- 19 Cette perspective représente à la fois une bonne nouvelle et un défi pour l'enseignant d'anglais en particulier. La bonne nouvelle est certes celle de l'impact des activités informelles sur le développement langagier. Nos études suggèrent (Toffoli & Sockett, 2010 ; Sockett, 2014) que même les étudiants du secteur Lansad passent aujourd'hui un temps important à rechercher, écouter, visionner, lire et écrire en anglais dans leur temps libre en se servant d'un grand nombre d'outils numériques. Le cours d'anglais est donc potentiellement d'une grande pertinence pour l'apprenant/utilisateur de la langue aujourd'hui dans la mesure où il peut aider à développer les compétences mises en œuvre dans ces pratiques informelles.
- 20 Le défi que représente cet état de fait concerne la pertinence des cours d'anglais pour ce public. En effet, si l'apprenant d'aujourd'hui possède ses propres préférences et centres d'intérêt en termes de documents authentiques, son propre matériel *hardware* et *software* pour accéder à ces documents et ses propres démarches en matière de recherche, choix et consommation de ces ressources, le rôle de l'enseignant, jadis spécialiste du contenu dans une perspective transmissive, se voit quelque peu menacé.
- 21 Un apprenant qui passe une dizaine d'heures par semaine à visionner des séries, écouter des chansons et visiter des sites web en anglais est peut-être parfois plus exposé à la langue cible que son enseignant. L'époque où l'enseignant détenait un trésor de documents authentiques qui n'étaient accessibles à l'apprenant qu'en cours ou en bibliothèque sonore est donc révolue, et il serait important aujourd'hui de partager le rôle de spécialiste du contenu langagier avec l'apprenant (Toffoli & Sockett, à paraître).
- 22 Outre la connaissance d'innombrables ressources, souvent peu familières à l'enseignant, l'apprenant/utilisateur de la langue possède aussi une connaissance des outils numériques, techniques de téléchargement etc. qui peut dépasser celle de l'enseignant. C'est donc avec une certaine crainte de ne pas être à la hauteur que l'enseignant s'engage dans ce virage.

Pistes pour l'enseignant

- 23 L'accès massif à l'anglais (et d'autres langues) en ligne déplace le centre de gravité de l'autorité linguistique, qui ne sera plus celle de l'enseignant, vers celle des textes que rencontrent les apprenants dans leur ensemble. Valoriser les pratiques informelles d'apprenants en leur demandant de proposer des exemples d'utilisation de la langue tirés de leurs propres corpus de prédilection (séries télévisées ou films, par exemple) encourage une analyse plus approfondie de ces derniers tout en plaçant l'apprenant au centre du processus d'apprentissage.
- 24 Ainsi l'apport de l'enseignant dans le cours de langue est celui de guide, d'expert dans la recherche et l'évaluation d'exemples, de facilitateur d'échanges entre les apprenants eux-mêmes voire entre les apprenants et les corpus ou les documents. En effet, parmi les objectifs d'une pédagogie centrée sur l'apprenant figurent le développement des compétences d'évaluation et d'analyse de ressources, traitées souvent de manière assez passive par l'apprenant (Toffoli & Sockett, à paraître), et la gestion d'échanges autour de ces contenus proposés par les apprenants eux-mêmes.

- 25 Si cet article pose quatre questions dans son intitulé, il serait logique de conclure en y apportant des réponses à la lumière des perspectives présentées ici. La première est celle des compétences et de la formation des enseignants. Le récent rapport de la Commission européenne (2012) sur la question souligne que la formation initiale et continue des enseignants n'a que peu d'incidence sur l'aptitude linguistique et que ce sont des facteurs externes comme le domaine informel qui contribuent le plus au niveau de langue des apprenants dans un pays donné. Ainsi, la formation des enseignants de langues doit donner une place importante à ces facteurs et à la compréhension de l'univers de l'apprenant en général.
- 26 Si la formation ne permet pas de s'assurer de la pertinence d'un cours de langue aujourd'hui, il y a néanmoins un certain nombre de compétences pertinentes au vu des pratiques actuelles des apprenants. Pour ce faire, on pourrait les former à la pédagogie inversée et aux questions que soulève le déplacement de point de vue. En effet si plus de 75 % des étudiants du secteur Lansad par exemple visionnent régulièrement des émissions en version originale, il serait utile de développer un travail sur ce type de texte en dehors des heures de cours et de privilégier plutôt les échanges autour de ces pratiques informelles pendant le cours. L'une des difficultés constatées par des enseignants familiers des pratiques informelles est que le registre appris dans de tels contextes est perçu comme inapproprié. Ainsi, un autre type d'activité à privilégier en cours serait un travail contrastif sur les registres formels et informels dans les ensembles de textes concernés. La forte présence de fictions à substrat professionnel (FASP) dans les films et séries visionnés par les étudiants favoriserait sans doute ce genre de travail.
- 27 De manière générale, valoriser les pratiques informelles, chercher à les comprendre et en parler en cours permettrait d'atteindre une certaine intersubjectivité nécessaire à l'apprentissage, car cela implique de considérer l'apprenant comme utilisateur de la langue et donc comme un interlocuteur "valable" en la matière. La mise en œuvre d'outils comme les journaux de bord, les blogs ou la mise en place d'activités de réseautage social autour de ces questions pourrait également étayer ce processus.
- 28 Pour résumer les diverses activités pédagogiques possibles dans le cadre des pratiques informelles, on peut mentionner entre autres :
- proposer un travail sur des éléments de style des SMS ou statut Facebook dans la langue cible choisis par les étudiants ;
 - évoquer les situations de communication et les actes de parole qui contribuent à la réalisation d'une tâche collaborative dans le cadre des jeux en ligne multi-joueurs ;
 - proposer un atelier de production écrite d'aventures ou de scènes de fiction (fan fiction) en invitant les apprenants à écrire des scènes imaginaires pour leurs séries télévisées préférées ;
 - analyser les échanges professionnels et privés dans les fictions à substrat professionnels (romans, séries etc.) ;
 - travailler l'écriture collaborative d'articles à partir d'une analyse de publications scientifiques.
- 29 Si ces quelques pistes pratiques ne sont que des exemples de travaux sur corpus dans une perspective centrée sur l'apprenant, elles correspondent aussi aux objectifs proposés par Thorne et Reinhardt (2008) pour jeter des ponts entre enseignant et apprenant afin :
- de mieux comprendre les genres textuels conventionnels et en ligne ;
 - d'adapter sa production au contexte ;

- de développer des compétences métacommunicatives et d'analyse critique ou réflexive ;
 - de créer des liens avec la vie réelle des étudiants ;
 - d'augmenter la participation des étudiants.
- 30 Ces objectifs impliquent la mise en œuvre d'outils qui seront parfois ceux des étudiants. Si l'enseignant est plus à l'aise avec un concordancier qu'il ne l'est avec le téléchargement de fichiers vidéos ou la transformation de vidéos *Youtube* en MP3, il serait intéressant d'instaurer un certain dialogue autour des pratiques afin de valoriser les choix des apprenants en la matière et de montrer l'intérêt complémentaire des outils plus 'scientifiques'.
- 31 Si, enfin, l'enseignant se pose la question de ses propres besoins par rapport au travail sur corpus, ces besoins peuvent se situer à trois niveaux. Premièrement, mieux connaître les pratiques d'apprenants en la matière favorise l'intersubjectivité et renforce ainsi l'efficacité du cours. Un apprenant qui sait que son rapport à la langue est compris par l'enseignant se sentira davantage concerné par un cours conçu pour favoriser le développement de ce rapport. Deuxièmement, la découverte de nouvelles pratiques et techniques participe au développement professionnel de l'enseignant, en lui offrant notamment des occasions de créer de nouvelles activités, de proposer des tâches revisitées, de mieux ancrer les apprentissages dans l'expérience de ses apprenants. Enfin, l'étude du domaine de l'informel représente pour l'enseignant une occasion de se mettre en recherche, de communiquer et de publier dans une perspective d'évolution professionnelle.

Conclusion

- 32 Le but de cette réflexion sur les corpus et l'apprentissage des langues n'est pas de suggérer que les approches DDL existantes n'aient pas leur place dans une pédagogie orientée vers des spécialistes de haut niveau. En effet, comme le suggère aussi Boulton (à paraître), le web n'est pas le *British National Corpus* et *Google* n'est pas un concordancier à l'instar de *WordSmith* ou d'*AntConc*. Il n'en demeure pas moins que l'apprentissage par corpus / *data-driven learning* sera plus pertinent pour un public élargi, lorsqu'il sera centré sur l'activité réelle de l'apprenant en langue cible et lorsqu'il contribuera à la réalisation de tâches de la vie réelle mettant en jeu la langue cible et dans lesquelles l'apprenant/utilisateur de la langue est déjà impliqué.

BIBLIOGRAPHIE

Boulton, A. (2010). "Learning outcomes from corpus consultation". In Moreno Jaén, M. Serrano Valverde, F. & Calzada Pérez, M. (dir.), *Exploring new paths in language pedagogy: lexis and corpus-based language teaching*. Londres: Equinox. pp. 129-144. Supplément électronique: "Empirical research in data-driven learning – A summary". Disponible en ligne. <http://arche.univ-lorraine.fr/course/view.php?id=616>

- Boulton, A. (2012). "Beyond concordancing: multiple affordances of corpora in university language degrees". In Macaire, D. & Boulton, A. (dir.), *Languages, cultures and virtual communities. Procedia : Social and Behavioral Sciences*, vol. 34. pp. 33-38. Disponible en ligne. <http://dx.doi.org/10.1016/j.sbspro.2012.02.008>
- Boulton, A. (À paraître). "Applying data-driven learning to the web." In Leńko-Szymańska, A. & Boulton, A. (dir.), *Multiple affordances of language corpora for data-driven learning*. Amsterdam: John Benjamins.
- Bouveret, M. & Legallois, D. (2012). *Constructions in French*. Amsterdam: John Benjamins.
- Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, vol. 6, n° 4. pp. 289-302.
- Commission européenne. (2012). *Première enquête européenne sur les compétences linguistiques*. Disponible en ligne. <http://u.jimdo.com/www31/o/sfa78b0f074d77f8d/download/m82ab46441470568b/1349800450/Premi%C3%A8re+Enqu%C3%Aate+Europ%C3%A9enne+sur+les+Comp%C3%A9tences+Linguistiques.pdf>
- Engeström, Y. (1990). *Learning, working and imagining: twelve studies in activity theory*. Helsinki: Orienta Konsultit Oy.
- Goldberg, A. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Johns, T. (1990). "From printout to handout: grammar and vocabulary teaching in the context of data-driven learning". *CALL Austria*, vol. 10. pp. 14-34.
- Johns, T. (1997). "Contexts: the background, development and trialling of a concordance-based CALL program". In Wichmann, A., Fligelstone, S., McEnery, T. & Knowles, G. (dir.), *Teaching and language corpora*. Harlow: Addison Wesley Longman pp. 100-115.
- Kilgariff, A. & Grefenstette, G. (dir.). 2003. *Web as corpus. Computational Linguistics*, vol. 29, n° 3.
- Kusyk, M. & Sockett, G. (2012). "From informal resource usage to incidental language acquisition: language uptake from online television viewing in English". *ASp*, vol. 62. pp. 45-65.
- McEnery, T. & Wilson, A. (1996). *Corpus linguistics*. Edinburg: Edinburgh University Press.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Socket, G. (2011). "From the cultural hegemony of English to online informal learning: cluster frequency as an indicator of relevance in authentic documents". *ASp*, vol. 60. pp. 5-20.
- Socket, G. (2014). *The online informal learning of English*. Basingstoke: Palgrave Macmillan.
- Stevens, A. & Shield, L. (2010). *Étude sur l'impact des technologies de l'information et de la communication (TIC) et des nouveaux médias sur l'apprentissage des langues. Commission Européenne*. Disponible en ligne. http://eacea.ec.europa.eu/llp/studies/documents/study_impact_ict_new_media_language_learning/final_report_fr.pdf
- Thorne, S. L. & Reinhardt, J. (2008). "Bridging activities, new media literacies, and advanced foreign language proficiency". *CALICO Journal*, vol. 25, n° 3. pp. 558-572.
- Toffoli, D. & Socket, G. (2010). "How non-specialist students of English practice informal learning using web 2.0 tools". *ASp*, vol. 58. pp. 125-144.

Toffoli, D. & Sockett, G. (À paraître). "University teachers' perceptions of online informal learning of English (OILE)". *Computer Assisted Language Learning*. DOI : 10.1080/09588221.2013.776970.

RÉSUMÉS

L'apprentissage sur corpus ("data-driven learning") est une pédagogie importante pour les apprenants avancés et spécialistes en langues vivantes. Son attractivité pour un public plus large mérite réflexion, dans la mesure où les activités informelles d'apprenants de langues mettent aujourd'hui ceux-ci en contact régulier avec une large gamme de ressources en ligne en langue cible. L'article se propose donc d'explorer des pistes pédagogiques qui permettraient de tirer profit des données linguistiques qui se cachent derrière de telles activités de loisir.

Data driven learning is a key pedagogical approach for advanced, specialist learners of modern languages. Its applicability to a wider audience needs careful thought, particularly since many foreign language learners today are involved in a range of informal activities which place them in contact with a large quantity of online resources in the target language. This article considers pedagogical solutions which may allow the linguistic data hidden in these activities to be put to good use in the classroom.

INDEX

Mots-clés : apprentissage sur corpus, data-driven learning, apprentissage informel, grammaire des constructions, pédagogie inversée

Keywords : data driven learning, informal learning, construction grammar, flipped classes

AUTEUR

GEOFFREY SOCKETT

Geoffrey Sockett est professeur des universités en didactique des langues et membre de l'équipe EDA de l'Université Paris Descartes. Il est spécialiste de l'apprentissage informel de l'anglais en ligne.

Courriel : geoffrey.sockett@parisdescartes.fr

Toile : <http://eda.recherche.parisdescartes.fr>

Adresse : Université Paris Descartes, Faculté des sciences humaines et sociales-Sorbonne, 45 rue des Saints-Pères – Bâtiment Jacob, 75270 Paris cedex 06