

***WEB-BASED DDL: AN INSTANCE OF HOW AJAX TECHNOLOGIES CAN
IMPROVE LANGUAGE LEARNING WEB APPLICATIONS***

***DDL EN RÉSEAU: UN EXEMPLE D'UTILISATION DES TECHNOLOGIES AJAX
POUR AMÉLIORER LES OUTILS INTERNET D'APPRENTISSAGE DES LANGUES***

Simone Torsani

Université de Gênes – Doctorat en Langues et Nouvelles Technologies
Facoltà di Lingue e Letterature Straniere – Piazza Santa Sabina 2 -16146 Genova
s.torsani@gmail.com

Résumé: La contribution présentera un module pour la plateforme Moodle qui permet d'analyser des corpus textuels directement en ligne. La possibilité de réaliser cette activité en ligne offre différents avantages, aussi bien en ce qui concerne les aspects pratiques de l'utilisation des corpus pour l'enseignement des langues (disponibilité des appareils, coûts de réalisation), qu'en ce qui concerne l'utilisation des données dans un environnement d'apprentissage virtuel, qui permet d'intégrer ces données mêmes dans plusieurs activités linguistiques.

Riassunto: Il contributo presenterà un modulo per l'ambiente di apprendimento virtuale Moodle che permette di effettuare analisi di corpora linguistici. La possibilità di realizzare quest'analisi come attività in rete offre diversi vantaggi sia per quanto concerne gli aspetti pratici (disponibilità dei calcolatori e costi di realizzazione), sia per quello che concerne l'utilizzo dei dati che, in un ambiente di apprendimento virtuale, possono essere integrati in diverse attività.

Mots clés: corpus, concordance, DDL, enseignement des langues, FOAD

Parole chiave: corpora, concordanza, DDL, didattica delle lingue, didattica a distanza

1. INTRODUCTION. CORPUS POUR L'APPRENTISSAGE DES LANGUES

L'emploi des corpus dans l'apprentissage d'une langue seconde est une activité pratiquée désormais assez souvent par les enseignants et les recherches dans ce domaine ont mis en évidence ses caractéristiques, ses points forts et ses insuffisances (Tognini-Bonelli, 2001).

Qu'il se compose d'une « collection de textes » ou bien d'un simple ensemble de données, un corpus est toujours une « source d'évidences » (Stubbs, 2004), pouvant être utilisée avec profit dans bien des domaines de l'apprentissage des langues, de l'analyse des textes et de l'analyse stylistique. Même si nous nous bornons aux seules applications concernant directement l'apprentissage des langues¹, l'ampleur des applications possibles (lexique ou grammaire, à des niveaux différents et dans des buts variés) est remarquable et assez difficile à définir de façon satisfaisante : Porcelli (2004) montre des exemples d'analyse de corpus qui servent de support à l'apprentissage des langues de spécialité ; Chanier and Selva (2000) ont introduit l'utilisation des corpus pour la réalisation automatique d'exercices linguistiques dans leur système « Alexia ». Nombre de professeurs utilisent, d'après T. Johns², une approche inductive (« bottom-up » : Tognini-Bonelli, 2001) nommée *Data Driven Learning* (DDL) : c'est-à-dire une procédure empirique³ au moyen de laquelle les apprenants découvrent des « facts about the language they are learning for themselves » (Hunston, 2002). Une activité DDL pourrait être *grosso modo* décrite comme l'exploration d'un corpus à la découverte de faits linguistiques spécifiques, même sans avoir recours à des cadres théoriques préalables (Tognini-Bonelli, 2001) : Escoubas-Bienveniste (2002), par exemple, fait analyser un corpus de journaux français pour confirmer ou pour rejeter leurs conjectures à propos de l'emploi de « an » ou « année » en français.

Quoique les corpus existent depuis très longtemps, bien avant l'époque de l'emploi massif des ordinateurs⁴, ils n'ont été employés à des fins didactiques qu'au moment où les ordinateurs sont devenus des produits de masse⁵ : c'est justement à partir de ce moment que, la technologie étant essentielle – selon une opinion commune et acceptée (Stubbs: 107) – pour une analyse performante des corpus, l'exploitation de corpus commença à être couramment employée, et constitue à présent l'un des exemples les plus intuitifs de ALAO (Apprentissage des Langues Assisté par Ordinateur : anglais CALL). Chappelle (2001) reconnaît même dans la Linguistique des Corpus une discipline liée – ou même fondatrice – de la « Computer Application in Second Language Learning » (CASLA), un terme créé pour étendre le domaine du ALAO au *testing* et à la recherche sur l'apprentissage des langues lui-même.

Un « corpus ne peut -évidemment- rien faire à lui seul » (Hunston 2001), mais ce sont plutôt les logiciels d'analyse qui permettent d'en tirer des résultats : normalement, cette analyse se fait grâce à un programme effectuant des concordances dans le format MCC (Mot-clé en contexte ; en anglais KWIC : *Key Word In Context*, Gavioli, 2005), aussi bien sur des corpus en ligne que sur des corpus préétablis résidant sur ordinateur (WordSmith, MonoConc).

¹ C'est-à-dire si l'on ne prend pas en considération les instructions sur l'utilisation professionnelle des corpus

² T. Johns a développé cette pratique à l'Université de Birmingham (Hunston, 2002, p.). Une bibliographie sur cet argument peut être trouvée dans sa page web (<http://www.eisu.bham.ac.uk/johnstf/biblio.htm>) et dans l'ouvrage de Hunston.

³ Sur la découverte des règles de grammaire (et l'évaluation d'hypothèses) dans des textes à l'intérieur des modèles didactiques, voir Balboni (1998: 89)

⁴ Sur l'histoire des concordances des corpus voir Tribble et Jones (1990) dans Stevens (1995); voir aussi Tognini-Bonelli "the use of a corpus of language as the basis for language description is as old as linguistic enquiry" (p. 50).

⁵ Les articles de Johns ont apparu dans les années quatre-vingts.

Je ne voudrais pas ici prendre position dans le débat autour du fondement scientifique du DDL⁶, ni sur le rôle qu'il pourrait avoir en classe, ni enfin sur la révolution supposée que les corpus introduiraient dans l'enseignement des langues⁷ ; je voudrais me borner à considérer si une activité, faite à distance, avec un corpus d'un serveur auquel tout le monde peut accéder, est susceptible de résoudre quelques-uns des problèmes pratiques que l'on rencontre normalement en classe et si, en vertu de recherches scientifiques ou d'épreuves empiriques, cette même activité pourrait avoir d'autres avantages.

Le logiciel présenté n'est pas révolutionnaire, ni d'un point de vue linguistique, ni technique : il est en fait assez simple et ne veut pas reproduire tous les résultats des logiciels traditionnels installés normalement sur l'ordinateur (*desktop application*). Il a néanmoins, même à première vue, des caractéristiques qui le différencient positivement par rapport aux logiciels installés. Par exemple, c'est un module adapté à Moodle⁸, un Environnement d'Apprentissage Virtuel « open source » parmi les plus diffusés : il permet donc aux usagers d'obtenir des concordances en ligne tout en restant à l'intérieur de la plateforme, exactement comme pour n'importe quel autre exercice ou activité de leur apprentissage. En outre, étant « open source », n'importe qui peut le modifier selon ses nécessités. Moodle en effet, comme beaucoup d'applications VLE, est une application réalisée avec PHP/MySQL⁹, un langage puissant, simple et ouvert. Mon logiciel se base, en outre, sur une approche AJAX.

Comme d'autres approches récentes pour la réalisation d'applications Web, AJAX, dont on parle beaucoup aujourd'hui, n'arrive pas véritablement à introduire de nouveaux éléments techniques par rapport au passé ; mais il permet de réaliser des applications tellement flexibles, riches et puissantes qu'elles peuvent imiter le comportement de logiciels installés sur les ordinateurs. C'est le cas, justement, de la concordance des corpus, une activité occupant normalement une quantité de mémoire et de temps trop importante pour être exécutée en ligne dans une unique page.

2 AVANTAGES DU DDL À DISTANCE PAR RAPPORT AU DDL COMME ACTIVITÉ EN CLASSE

Il y a beaucoup d'outils pour analyser des corpus en ligne, mais la plupart d'entre eux utilisent des corpus prédéfinis, comme le 'Brown corpus'¹⁰. Ils ne permettent donc pas aux utilisateurs d'effectuer des recherches sur leur textes, et quand cela est possible, la dimension des textes est très limitée. Le DDL en tant qu'activité à distance présente alors par rapport au DDL en classe des avantages (mais aussi des désavantages) qu'il est possible de récapituler en une série de points.

La classe ou le laboratoire, par exemple, ne peuvent pas toujours garantir un nombre d'ordinateurs suffisant pour que chaque étudiant puisse travailler (Hunston, 2002: 171). En classe, il est en outre nécessaire d'un côté que tous les ordinateurs aient parmi les programmes installés un logiciel pour l'analyse des corpus et de l'autre que les mêmes documents soient à la

⁶ Voir Hunston (pp. 192-197) pour des idées contrastantes sur l'usage des corpus pour l'apprentissage des langues.

⁷ Voir Leech (1997), cité par Chapelle (2001: 38)

⁸ <http://moodle.org> est un environnement d'apprentissage très connu et utilisé, sa particularité est sa nature modulaire : beaucoup de développeurs ont produit des modules, intégrables dans l'environnement, pour de différentes activités didactiques.

⁹ PHP (<http://www.php.net>) est un langage pour la réalisation de pages HTML dynamiques, MySQL (<http://www.mysql.com>) est un système pour la gestion de bases de données relationnelles.

¹⁰ Sur le Collins CoBuild en ligne (<http://www.collins.co.uk/Corpus/CorpusSearch.aspx>), on peut chercher des corpus différents. Le "Online Concordancer" (v.4, http://www.lexutor.ca/concordancers/concord_e.html) utilise des corpus différents. Le "Spaceless Concorder" (<http://www.spaceless.com/concord/>) peut analyser les corpus des utilisateurs, mais la taille des documents est limitée à 50,000 mots.

disposition de chaque utilisateur, en même temps et pour tous les appareils : c'est seulement à ces conditions que l'activité d'analyse individuelle est possible dans le cadre d'une leçon présentielle. Il faut quand même souligner que toutes ces variables sont souvent difficiles à réaliser, dans les situations courantes de l'enseignement et des établissements scolaires. L'utilisation d'un système en ligne permet au contraire de résoudre ces problèmes: chaque étudiant utilise son propre ordinateur (ou il peut en utiliser un en autonomie, ce qui revient au même; en outre si le corpus défini par l'enseignant est dans un serveur commun, tout le monde peut analyser le même corpus en même temps. Une autre limite de l'analyse des corpus est qu'elle peut occuper trop de temps par rapport à la durée moyenne d'un cours (Hunston, 2002: 178): on peut résoudre même ce problème par une activité en ligne, à distance et indépendamment des contraintes de la structure du cours.

Un troisième élément de différence concerne l'autonomie de l'apprenant dans l'utilisation didactique du corpus. L'activité DDL en classe peut être dirigée ou assistée par un tuteur, mais elle peut aussi être conduite de façon autonome : si cette dernière option se révèle sans aucun doute plus motivante et plus adaptée pour des étudiants avancés, il paraît bien plus difficile qu'elle puisse être offerte à des débutants (Huston, 2002). En tout cas, le DDL autonome à distance, non dirigé directement, mais inséré dans des activités à la maison (comme la recherche d'une phraséologie pendant la traduction de textes de spécialité, etc.), se révèle très utile, si improbable que cela puisse paraître, pour atteindre l'objectif de stimuler puissamment toute une série de réflexions personnelles sur les faits de langue, ce qui représente, selon bien des spécialistes, l'un des points les plus importants du SLA (Balboni, 1999: 90).

Qui plus est, l'analyse des corpus à l'intérieur d'un environnement d'apprentissage offre des avantages non seulement pratiques mais aussi didactiques, puisqu'elle peut être intégrée dans la série d'activités collaboratives formant ce CSCL (Computer Supported Collaborative Learning : Apprentissage Collaboratif en Ligne), dont traite un vaste nombre de travaux, qui concernent du moins en partie la didactique des langues (Chapelle, 2001: 31-32). Pour les activités collaboratives, la pratique du "Tandem Network", c'est-à-dire l'échange de courriels électroniques entre deux étudiants qui parlent des langues différentes, est un exemple des plus banals dans notre domaine. (Porcelli-Dolci, 1999: 89).

Les informations obtenues grâce à l'analyse automatique d'un corpus peuvent en effet être utilisées avec succès dans cette perspective collaborative : elles sont contenues dans un corpus commun et sont disponibles pour tous à chaque moment, non seulement dans la classe, mais n'importe où quand les étudiants sont en train de travailler à leurs devoirs. S'ils travaillent ensemble, ils peuvent utiliser ces données pour vérifier des hypothèses sur la langue, ou pour en tirer des exemples et des matériaux sur lesquels discuter à propos de leur choix lexicaux, morphosyntaxiques ou stylistiques: ces données offrent, en bref, la possibilité de réaliser la négociation du savoir entre élèves, qui selon beaucoup est si importante pour l'apprentissage collaboratif (Littlewood, 2005: 518) et pour laquelle les développeurs d'environnements virtuels d'apprentissage ont réalisé beaucoup d'instruments comme les forums et les chats. Les observations de Porcelli et Dolci (1999) sur l'utilisation des courriels électroniques offrent, si elles sont étendues à ces contextes, les moyens et les cadres théoriques pour définir les limites de l'apprentissage en ligne des langues.

On peut donc faire un bref inventaire des contextes d'apprentissage linguistiques en ligne pour lesquels l'exploitation des corpus offre des avantages potentiels. Comme on l'a déjà signalé au auparavant, l'exploitation des corpus, en particulier l'analyse en ligne, devrait s'adapter surtout aux étudiants avancés et peut être utilisée dans des domaines spécifiques:

- recherche de phraséologie et de collocations, cette activité est particulièrement utile pour la formation à la traduction ou à la production (<http://www.edufle.net/Collocations-et-image-de-l>);

- recherche terminologique, où le corpus offre la possibilité d'analyser les usages et les domaines d'un terme;
- aide pour la traduction, pour laquelle les corpus offrent beaucoup de matériaux à recycler ou des exemples, comme on l'a déjà noté;
- analyse de textes à différents niveaux;

Enfin, le travail dans un Environnement d'Apprentissage peut être tracé (Mazzoni, 2004), c'est - à- dire enregistré et analysé automatiquement ou manuellement: ce qui présente des avantages aussi pour l'analyse des corpus. Il est donc possible de tracer les activités autour d'un corpus¹¹, et en particulier:

- contrôler si quelques phrases ou expressions ont été extraites d'un corpus, ce qui permet d'évaluer la validité d'un corpus spécifique ou bien des corpus en général;
- contrôler si les étudiants ont appris à exploiter un corpus;
- contrôler si les données extraites du corpus ont été utilisées pour la discussion commune;
- contrôler quels types de textes, spécialisés ou non, sont plus utiles que d'autres pour une activité déterminée, et pourquoi, et comment ;

Ce tracement peut faire partie d'une activité plus générale: on pourrait comparer les résultats de cette analyse avec les résultats du tracement d'autres activités, comme les forums, afin d'évaluer le processus global d'apprentissage.

3 L'APPROCHE AJAX ET LA CONCORDANCE DES CORPUS

L'acronyme AJAX, défini par Garret en 2005, est dérivé de Asynchronous Javascript And XML et consiste en une série de procédures et de stratégies qu'on peut utiliser pour rendre les applications en ligne le plus possible proches des applications installées sur les ordinateurs (Holzner, 2006: 9) ; en dépit de son nom, qui paraît n'indiquer que l'union de deux seules technologies, AJAX se réfère à des combinaisons variables de différentes technologies et d'objets utilisables pour le réseau, comme les langages côté serveur (ASP, PHP), les langages XML et la structure de la page : cette combinaison produit des interfaces riches et simples à utiliser. L'exemple qu'on va présenter utilise les technologies de la plateforme Moodle (PHP/MySQL), Javascript et les cadres pour obtenir, si non les mêmes prestations en termes de temps, au moins les mêmes résultats d'une application traditionnelle. Si une telle idée connaît aujourd'hui un remarquable succès, en réalité la combinaison de technologies diverses n'est pas nouvelle, et quand cette approche a été définie la première fois, on a justement noté : "les développeurs n'ont en fait pas attendu qu'on lui donne un nom pour «faire de l'Ajax»" (Catteau, Faugout, 2006: 17).

Comme tous les utilisateurs d'applications en ligne ont vérifié, ce ne sont pas les pages traditionnelles côté serveur qui sont interactives, mais l'application: chaque page permet, en général, de faire une seule opération et c'est justement l'interaction entre les différentes pages qui produit une opération complexe. Dans l'utilisation d'un forum web, par exemple, il y a une série de procédures standard que l'utilisateur doit suivre pour interagir avec le système, et chaque action correspond à une page différente (Asleson and Schutta, 2006: XIX). L'utilisateur en effet:

- ouvre la page du forum;
- ouvre une page dans laquelle il peut lire le message;
- ouvre une page dans laquelle il peut écrire un nouveau message et qui mène à
- une page dans laquelle les données sont envoyées et enregistrées

¹¹ Hunston (p. 170) affirme que le DDL est "hypotesised to improve general skills of using context to deduce meaning", this bebenefit having been not tested yet, an accurate activity tracking could provide data for further research.

- revient à la page principale

Une application AJAX, au contraire, pourrait, dans ce cas, faire communiquer le navigateur de façon invisible et continue avec le serveur, afin de lui faire élaborer et renvoyer les données sans changer de page. Ce n'est que l'une des stratégies que l'on pourrait utiliser pour obtenir le même résultat ; mais les avantages remarquables que cette stratégie offre sont à présent très appréciés par les développeurs en ligne, qui l'utilisent de plus en plus; voilà une raison supplémentaire de l'utilisation de cette technique dans mon logiciel.

Cette technique, est dite "du cadre caché", car elle utilise une page, cachée à l'utilisateur mais présente, qui échange continuellement des données avec le serveur et permet des opérations qui seraient autrement impossibles. La concordance d'un corpus en offre un exemple très intéressant non seulement parce que la quantité de données qu'il faut y gérer est trop grande pour une seule page, mais aussi parce que les algorithmes qu'on utilise pour la réaliser peuvent consommer très rapidement la mémoire et le temps disponibles pour l'exécution d'un script¹². Il faut souligner qu'il existe déjà des outils en ligne pour l'analyse KWIC des corpus ; ils se basent toutefois sur d'autres technologies incompatibles avec lesquelles la majorité des environnements sont réalisés.

Les cadres étaient autrefois très populaires, et, tout comme le langage JavaScript, ils n'ont jamais disparu ; on les a pourtant critiqués quand on a commencé à prendre l'accessibilité comme standard pour la réalisation des pages web¹³. Ils consistent en une page qui contient au moins deux pages indépendantes à son intérieur: dans la technique du 'cadre caché' l'un des « frames » est la page vue par l'utilisateur et l'autre, la page cachée, est celle à travers laquelle l'application communique avec le serveur (cette page est liée à la page principale en utilisant JavaScript).

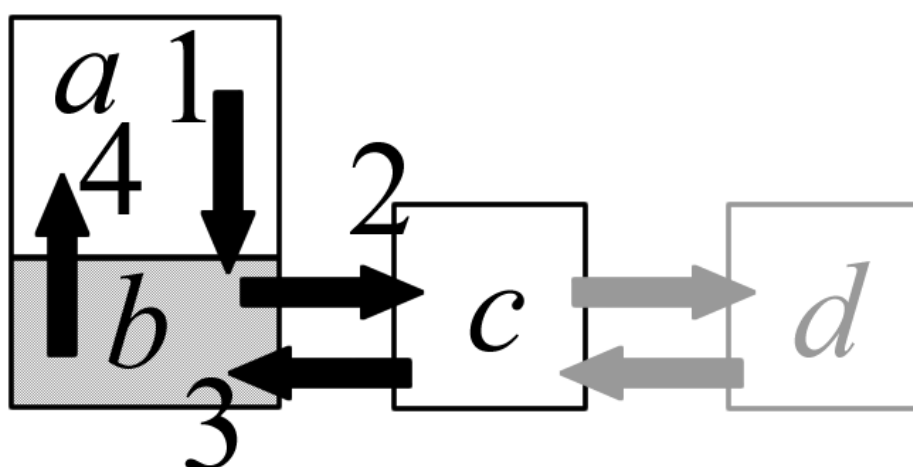


Illustration 1 Le fonctionnement du cadre caché

L'illustration montre le flux de travail du cadre caché. Les données du cadre visible (a) sont envoyées/récupérées par la page invisible (b) - qui les envoie au serveur (c) - au moyen de JavaScript. Le serveur exécute les opérations nécessaires (une activité supplémentaire, mais très commune, est la transmission/réception de données d'une base de données - d) et les renvoie au cadre caché, qui, à son tour, les restitue à la page visible, sans qu'il soit nécessaire de la mettre à jour afin de visualiser les nouvelles données. La transmission entre les deux pages est

¹² Selon la configuration typique du PHP les scripts n'ont que 30 secondes pour être exécutés.

¹³ Voir <http://www.w3.org/WAI/> pour des informations détaillées sur l'accessibilité. Javascript et les cadres sont toujours considérés des éléments (potentiellement) non accessibles.

déléguée à JavaScript qui peut travailler avec la structure de la page, l'arbre du DOM (Asleson and Schutta, 2006: 36-37); il n'est pas possible ici d'approfondir cet argument, mais il suffit de comprendre que les pages web ont une structure hiérarchique (l'arbre du DOM), JavaScript peut accéder sans problèmes aux éléments de cette structure. En outre, JavaScript peut charger automatiquement des pages dans le navigateur en manipulant l'adresse qui contient aussi des valeurs qui sont interprétées par la page, donc l'instruction

```
window.location.href="concord.php?offset="+offsetValue;
```

fait charger au navigateur une page avec une valeur (`offset`) qui est utilisée comme variable.

Cette technique permet à la page d'effectuer des opérations complexes sur des grandes quantités de données: un corpus textuel de 1.000.000 mots, par exemple, "pèse" sept mégabytes et la concordance KWIC a besoin d'algorithmes complexes pour extraire et ordonner les données. Cette technique permet d'obtenir les mêmes résultats d'un programme traditionnel comme Wordsmith, mais avec la limite du temps nécessaire pour l'exécution des opérations, un temps trop lointain de la vitesse des applications installées sur l'ordinateur, comme Wordsmith¹⁴.

4 LE MODULE WORDTOOLS : UN EXEMPLE D'APPLICATION AJAX

On peut maintenant mieux approfondir le fonctionnement de l'application. Au niveau de l'utilisateur elle offre, comme nous l'avons dit, les mêmes résultats des programmes traditionnels, donc:

- Concordance des mots/séquences de texte : l'expérimentation effectuée afin de comparer les résultats du module avec le programme de référence (Wordsmith) a montré que le nombre de lignes extraites du même corpus est presque le même (Wordtools a extrait le 97% des lignes extraites par wordsmith). L'application offre la possibilité d'exécuter la recherche de mots simples ou de séquences de textes avec des caractères spéciaux (par exemple "livr*" permet de chercher livre, livraison etc.): cette deuxième modalité est beaucoup plus rapide que la première;
- Rangement des résultats : comme dans l'application de référence on peut très rapidement réordonner les lignes selon leurs collocations;
- Exportation dans des formats différents, comme XML, SQL ou texte simple;
- Avant-première des textes, pour voir les textes avant de choisir le plus convenable en cas de textes d'arguments différents, il sera possible de choisir les plus intéressants et les plus utiles pour une recherche déterminée;
- Création de listes de mots dans un document, pour voir les mots contenus dans un texte et leur fréquence;
- Visualisation du « plot », c'est-à-dire d'une image montrant en présentation graphique la diffusion d'un mot/séquence dans un texte;

Il est aussi intéressant d'observer le travail du programme au niveau technique, car il existe la possibilité de le modifier. Quand un document est chargé, il est divisé en une série de documents de dimensions plus petites (environ 100000 caractères), cette dimension étant analysable par l'exécution d'un seul script.

Quand on lance la recherche d'une concordance, c'est-à-dire quand les documents ont été choisis et on a spécifié le mot clé et les éventuelles collocations, le cadre caché analyse chaque petit document et envoie les résultats à une base de données. Cela se produit grâce au JavaScript qui, quand le processus d'analyse est terminé, permet à la page de se recharger chaque fois avec une valeur différente, correspondante au prochain document qu'il faut analyser.

¹⁴ <http://www.lexically.net/wordsmith/index.html>

208	be thine. nay, said jacob, i pray thee to take this gift which god hath sent me that i may find grace in thy sight, for meseemeth i see thy visage like the visage of
209	y and i am smooth? if my father take me to him and taste me and feel, i dread me that he shall think that i mock him, and shall give me his curse for the blessing. th
210	ereof such manner meat as thou knowest that i am wont to eat, and bring it to me that i may eat it, and that my soul may bless thee ere i die. which all these words r
211	especially for the youth of our land. it is indeed an encouraging sign. it means that the child of to-day is being introduced to the world's best in literature and sc
212	saac fulfilled these words and jacob gone out, when that esau came with his meat that he had gotten with hunting, entered in, and offered to his father saying: arise,
213	ther say to esau, thy brother: bring to me of thy venison, and make thereof meat that i may eat, and that i may bless thee tofore our lord ere i die. now my son, take
214	children of god, that is to say of seth, as religious, saw the daughters of men, that is to say of cain, and took them to their wives. this time was so much sin on th
215	eve not me, ne them that shall come after me, ne my kindred, but after the mercy that i have showed to thee, so do to me and to my land in which thou hast dwelled as
216	and into earth thou shalt go again. then adam, walling and sorrowing the misery that was to come of his posterity, named his wife eve, which is to say, mother of all
217	nd hills, for the water was grown higher fifteen cubits above all the mountains, that it should purge and wash the filth of the air. then was consumed all that was on
218	ly, that unneth [hardly] the country might suffice to their pasture, in so much that rumor and grudging began to sourde and arise between the herdmen of abram and th
219	of laban thine uncle. god almighty bless thee, and make thee grow and multiply, that thou mayst be increased into tourbes of people, and give to thee the blessings o
220	s meiny was also there. and their beasts began so sore to increase and multiply, that unneth [hardly] the country might suffice to their pasture, in so much that rum
221	the instruments, for they were found long after, but he was the finder of music, that is to say of consonants of accord, such as shepherds use in their delights and s
222	and what thou findest that is thine, take with thee. he, saying this, knew not that rachel had stolen her father's gods. then laban entered the tabernacle of jacob
223	ast eaten he may bless thee ere he die: to whom jacob answered: knowest thou not that my brother is rough and hairy and i am smooth? if my father take me to him and t
224	part. whereunto anon he answered: nay in no wise ye shall die, but god would not that ye should be like him in science, and knowing that when ye eat of this tree ye s
225	d in conclusion desired to have rebekah for his lord isaac; and if he would not, that he might depart and go into some other place, on the right side or on the left,
226	ral number. man was made to the image of god in his soul. here it is to be noted that he made not only the soul with the body, but he made both body and soul. as to t
227	ch is marvellous? and he blessed him in the same place. jacob called the name of that same place penuel, saying: i have seen our lord face to face, and my soul is mad
228	the east, and saw a pit in a field and three flocks of sheep lying by it, for of that pit were the beasts watered. and the mouth thereof was shut and closed with a gr
229	early of my life because of the daughters of heth, if jacob take to him a wife of that kindred, i will no longer live. isaac then called jacob and blessed him and comm
230	languages, in such wise that no man understood other. for tofore the building of that tower was but one manner speech in all the world, and there were made seventy-tw
231	pinion thirty sons and thirty daughters, and some fifty of that one and fifty of that other. we find no certainty of them in the bible. but all the days of adam livin

Illustration 2 fenêtre de l'application

Comme le processus a besoin d'un peu de temps, il est aussi possible de choisir soit de visualiser les résultats pendant l'extraction, donc sans attendre la fin de l'analyse, soit de les visualiser au terme du processus ; cette deuxième option étant en tout cas plus rapide.

5 CONCORDANCE DE PAGES EN LIGNE

L'application offre aussi la possibilité d'effectuer la concordance de pages en ligne. Un volume très récent, *Corpus linguistics and the Web*, rassemble un nombre de contributions sur les derniers développements dans ce domaine particulier, dont le plus intéressant du point de vue technique est la réalisation du logiciel WebCorp pour l'analyse de la Toile comme corpus (pp. 47-69).

Les avantages d'Internet comme corpus par rapports aux corpus textuels traditionnels ont été exposés dans un grand nombre de travaux ; Fletcher (pp.25-47), par exemple, cite la fraîcheur, la nature exhaustive, la diversité de textes qu'on peut trouver en ligne, leur faible coût et leur représentativité.

L'internet offre l'accès à une quantité énorme de textes gratuits qui sont continuellement mis à jour et qui illustrent virtuellement toutes les variétés de langages, y compris ces langages émergents qu'il est difficile de retrouver dans les corpus textuels, dont le contenu vieillit très vite.

Dans la perspective des langues de spécialité, en particulier, il faut observer que les textes contenus dans les sites web des entreprises, dans les manuels d'instructions en ligne par exemple, représentent l'état le plus avancé et mis à jour du langage dans le domaine spécifique. En outre, la mondialisation de la production, ainsi que la localisation des produits, ont vigoureusement accéléré le développement de phénomènes linguistiques très intéressants, comme l'adoption ou la traduction de termes étrangers, qu'il serait très difficile de percevoir et de comprendre sans pouvoir accéder à des matériaux constamment actualisés. Cela n'est pas seulement vrai pour ce qui concerne les langues de spécialité: sur le web en effet on peut observer et analyser des modes de communication qui sont typiques du canal, comme les weblogs et les forums, et qui contiennent des variétés de langue qui seraient autrement inaccessibles.

D'un point de vue pédagogique il faut tenir en considération le fait que l'analyse de textes en ligne est comparable aux "raw corpus" dont parle Hunston (2002:171), c'est-à-dire à un matériel sur lequel l'enseignant n'a aucun contrôle, mais qui en revanche assure une grande motivation des apprenants, ces derniers partent vraiment à la découverte d'informations dans des corpus dont on ne sait souvent presque rien.

Du point de vue pratique, cette recherche est aisée : pour faire la concordance d'un site web il suffit d'insérer dans le module l'adresse de la page initiale. L'application analyse chaque page

de la façon qu'on a décrite plus haut, montre chaque fois les résultats sur l'écran et enregistre tous les liens à l'intérieur de la page pour continuer son analyse jusqu'à la fin du site.

Il faut aussi observer que cette fonction d'analyse de pages web n'est pas disponible sur tous les serveurs, pour des raisons de sécurité, et il faudrait s'informer avec l'administrateur du système pour savoir si cette opération est possible.

CONCLUSION

Comme je l'ai dit au début de cette communication, le travail qu'il est possible d'accomplir grâce au module Wordtools n'a apparemment rien de révolutionnaire. Le logiciel fait ce qu'on attend de lui, c'est-à-dire tout ce qu'on peut faire avec un corpus dans le cadre d'une formation linguistique spécialisée. Toutefois, il présente des caractéristiques intéressantes à souligner dans un colloque consacré au « changement ». Au moment où de nombreuses activités peuvent se déplacer en ligne, à partir des applications les plus communes pour arriver aux programmes les plus complexes jusqu'aux « disques durs » virtuels, cette petite application déplace dans une plateforme en ligne un travail très spécialisé, permettant donc de jouir d'un outil électronique à partir de tout ordinateur : elle se situe donc dans le mouvement général de migration sur la Toile, qui caractérise l'évolution de ces dernières années.

C'est pourquoi elle peut se révéler très utile pour des cours à distance *via* Internet et contribuer ainsi à l'essor d'un monde FOAD connaissant aujourd'hui un développement spectaculaire, dont on n'arrive pas encore à bien entrevoir tous les effets. En particulier, elle peut offrir un outil précieux pour l'apprentissage des langues, tout comme fournir des éléments essentiels à la réflexion lexicale et stylistique, à l'activité de traduction et, dans tous ces cas, à des typologies différenciées de travail, individuel ou collaboratif. Ce sont là autant de caractéristiques qui s'accordent parfaitement aux mutations profondes de notre panorama de formation.

Finalement, cette application a été conçue dans une perspective « AJAX », qui est novatrice dans ses approches de programmation ; en outre, elle a été réalisée pour une plateforme « ouverte », diffusée et fréquemment modifiée par les développeurs : ce n'est qu'un des éléments significatifs du « changement » qui marque non seulement le panorama de la formation, mais encore le débat technique, la réflexion politique et, surtout, les comportements individuels. Ce qui pourra aussi marquer, en dernière conclusion, le sort même d'une application susceptible d'être modifiée, perfectionnée, ou délaissée, par la communauté (nombreuse) des développeurs de Moodle.

REFERENCES

- ASLESON R., SCHUTTA N.T., *Foundations of Ajax*, New York, Apress, 2006
- BALBONI P.E., *Tecniche didattiche per l'educazione linguistica*, Torino, UTET, 1998
- CATTEAU B., FAUGOUT N., *Ajax, le guide complet*, Paris, Micro Applications, 2006
- CHANIER T., SELVA T., *Général automatique d'activités Lexicales dans le Système Alexia*, in Sciences Techniques Educatives, vol 7,2 Paris, Hermès, 2000, pp. 385-412
- CHAPELLE C.A., *Computer Applications in Second Language Acquisition*, Cambridge, Cambridge University Press, 2001
- ESCOUBAS-BIENVENISTE *An où année? Un exemple d'analyse contextuelle guidée par corpus comme aide à la sélection lexicale por les apprenants italo-phones*, in Hédiard M. (ed.) *Le nuove tecnologie nella ricerca linguistica e glottodidattica*, Cassino, 2002,
- GARRETT, J. J., *Ajax: A New Approach to Web Applications*, 2005 [en ligne]. Disponible sur : <<http://www.adaptivepath.com/publications/essays/archives/000385.php>>, consulté le 09.03.2007
- GAVIOLI, L., *Exploring Corpora for ESP Learning*, Amsterdam, John Benjamins Publishing Company, 2005
- HOLZNER S., *Ajax for Dummies*, Hoboken, NJ, Wiley, 2006
- HUNDT M., NESSELHAUF N., BIEWER C., *Corpus Linguistics and the Web*, Amsterdam, New York, Rodopi, 2007
- HUNSTON S., *Corpora in Applied Linguistics*, Cambridge, Cambridge University Press, 2002
- LITTLEWOOD W., *Second Language Learning* in Davies A., Elder C. (eds.), *The Handbook of Applied Linguistics*, Malden, Blackwell, 2004
- MAZZONI E., *Strumenti per un approccio quantitativo allo studio delle interazioni. Il software Net Miner e i Log File*, dans Form@re, 2004 [en ligne]. Disponible sur : <http://www.formare.erickson.it/archivio/maggio_04/5mazzoni.html> (consulté le 10.03.2007)
- PORCELLI G., *Comunicare in lingua straniera. Il lessico*, Torino, Utet, 2004
- PORCELLI G. AND DOLCI R., *Multimedialità e insegnamenti linguistici*, Torino, Utet, 1999
- STUBBS M., *Language Corpora*, in Davies A., Elder C. (eds.), *The Handbook of Applied Linguistics*, Malden, Blackwell, 2004
- TOGNINI-BONELLI E., *Corpus Linguistics at Work*, Amsterdam/Philadelphia, John Benjamins, 2001
- VANCE STEVENS, *Concordancing with Language Learners: Why? When? What?* in CAELL Journal, vol 6 #2, 1995 [en ligne]. Disponible sur :: <http://www.eisu.bham.ac.uk/johnstf/stevens.htm> (consulté le 10.03.2007)
- WICHMANN A, FLIEGELSTONE S, MCENERY T., KNOWLES G., *Teaching and Language Corpora*, New York, Longman, 1997