HAL
open science

L'apport du concordancier à l'analyse et à la remédiation des erreurs des apprenants dans les forums de discussion en ligne

Joseph Rézeau

To cite this version:

Joseph Rézeau. L'apport du concordancier à l'analyse et à la remédiation des erreurs des apprenants dans les forums de discussion en ligne. ALSIC - Apprentissage des Langues et Systèmes d'Information et de Communication, 2007, 10 (2). edutice-00142217

HAL Id: edutice-00142217

https://edutice.hal.science/edutice-00142217v1

Submitted on 17 Apr 2007

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

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

Apprentissage des Langues et Systèmes d'Information et de Communication

alsic.org ou alsic.u-strasbg.fr
Volume 10, 2007
article mis en ligne en mars 2007
Pratique et recherche

L'apport du concordancier à l'analyse et à la remédiation des erreurs des apprenants dans les forums de discussion en ligne

Joseph RÉZEAU
Université Rennes 2, France

Résumé : *La généralisation des logiciels de concordances et l'accès facile à des corpus numériques a permis le développement de la linguistique de corpus au cours de ces 25 dernières années. L'utilisation de ces logiciels pour l'analyse des productions des apprenants est plus récente (*[Granger98] *;* [ChambersOSullivan04]*). Avec l'augmentation des cours en ligne, dont les forums de discussions sont une composante quasi obligatoire, l'enseignant de langues a désormais accès à un corpus numérique de choix pour analyser les caractéristiques des productions semi-spontanées de ses étudiants. En nous appuyant sur des cas concrets de cours d'anglais en ligne à destination d'étudiants francophones niveau master, nous montrerons comment l'utilisation d'un concordancier peut contribuer à aider l'enseignant à analyser les erreurs typiques d'une population donnée et à y remédier.*

- 1. Introduction
- 2. La constitution du corpus
- 3. Observations sur les fréquences de vocabulaire
- 4. Observations d'erreurs syntaxiques typiques dues à la nativisation
- 5. Stratégies de remédiation
- 6. Conclusion : *"from data-driven learning to data-driven learning on the web"*
- Références

1. Introduction

Avec l'augmentation des cours en ligne, dont les forums de discussions sont une composante quasi obligatoire, l'enseignant de langues dispose désormais d'un accès à un corpus numérique de choix pour analyser les caractéristiques des productions semi-spontanées de ses étudiants. Nous parlons ici de productions semi-spontanées parce que, avec le courrier électronique et les forums de discussion en ligne, est apparu un nouveau genre d'expression, à mi-chemin entre l'oral et l'écrit, entre le spontané et le réfléchi. Ainsi, comme le montrent les travaux de Rachel Panckhurst sur le courrier électronique ([Panckhurst98], [Panckhurst99]), on trouve dans les écrits numériques à la fois des marques d'oralité (le style "informel" des courriers électroniques) et des marques d'écrit standard (le "ne" de la négation, par exemple). Sur le plan de la communication, les forums de discussion en ligne sont également à cheval sur la communication privée et la communication publique, comme l'écrit Mangenot, en se référant aux

forums de discussion en ligne.

> *L'écrit asynchrone tend à rapprocher ce mode de communication des situations classiques de production écrite (ou également du courrier électronique), mais la dimension publique des échanges, pour sa part, l'apparente plutôt à une situation de débat en cours présentiel.* [Mangenot02]

Dans cet article nous ne dirons rien de l'analyse de discours à laquelle on peut soumettre un tel corpus (cf. [Rézeau04]). Nous nous intéresserons uniquement à l'analyse d'un petit corpus de productions d'apprenants – relativement homogène – dans le but d'établir une typologie des erreurs les plus fréquentes, cette typologie pouvant déboucher sur des stratégies de remédiation.

Le fil conducteur de nos questions de recherche est l'ouvrage de Granger [Granger98]. Toutefois, notre corpus étant strictement limité aux productions de francophones étudiant l'anglais, nous ne pourrons établir – comme le fait cet ouvrage – de comparaisons avec des étudiants d'anglais de langue maternelle différente.

Notre propos étant avant tout pratique, nous appuierons chaque étape de cet article d'exemples concrets et d'explications des manipulations nécessaires. Nous montrerons tout d'abord comment se présentent les données brutes en ligne, et comment les récupérer sous une forme exploitable. Une fois les données formatées, le corpus sera exploité avec un logiciel de concordances qui nous fournira une liste des mots permettant de mettre en évidence la fréquence relative des mots (*types* et *tokens*), pour comparaison avec une liste de fréquence standard. Ces listes nous permettront d'étudier les cas de sur-emploi et de sous-emploi de certaines catégories de mots. Nous aurons également recours à la présentation *KWIC* (*keyword in context*) du concordancier pour mettre en évidence certains types d'erreurs syntaxiques ou d'emplois plus ou moins déviants du lexique. Enfin nous montrerons comment il est possible d'intervenir a posteriori dans les forums de discussion pour insérer des renvois à des règles ou conseils de remédiation.

2. La constitution du corpus

2.1. La population cible et la plateforme utilisée

Le corpus utilisé pour cette étude est le corpus *Cursus*, du nom donné à la plateforme d'EAD (enseignement à distance) de notre université. Il est constitué des messages de forums de plusieurs cours d'anglais en ligne, s'adressant à des étudiants d'un master "technologies de l'enseignement et de la formation" (1ère année en 2005 ; 1ère et 2e année en 2006) et d'un master "créateur multimédia" (2e année en 2006). Certains de ces étudiants n'ont pas fait d'anglais depuis des années (il s'agit par exemple d'enseignants du primaire ou du secondaire en reprise d'études). La philosophie générale de ce type de cours est celle de la langue de spécialité (LANSAD), qui s'appuie exclusivement sur des documents authentiques appartenant au domaine d'études considéré. L'un des objectifs "psycho-pédagogiques" du cours consiste à réconcilier les étudiants avec l'anglais, à leur montrer que, quel que soit leur niveau actuel, ils peuvent progresser. La plateforme d'enseignement en ligne utilisée est le logiciel libre *Moodle* [Moodle], mis à la disposition des enseignants à l'université de Rennes 2 depuis le printemps 2005.

2.2. Les forums de discussion

Pour chaque cours, outre l'indispensable forum des nouvelles, le cours comporte un *Open Social Forum*, où l'expression est libre, et, pour chacun des thèmes étudiés, un forum spécifique. A l'intérieur d'un

forum, les étudiants peuvent commencer un fil de discussion ou répondre aux messages postés par leurs camarades. La capture d'écran de la Figure 1 montre la disposition des messages postés sur le forum de discussion de la "leçon" *Games in Education* : le sujet du tout premier message du fil de discussion apparaît sur fond orange ; les réponses successives à ce message sont affichées selon le format "réponses emboîtées" proposé par l'interface de la plateforme *Moodle*.

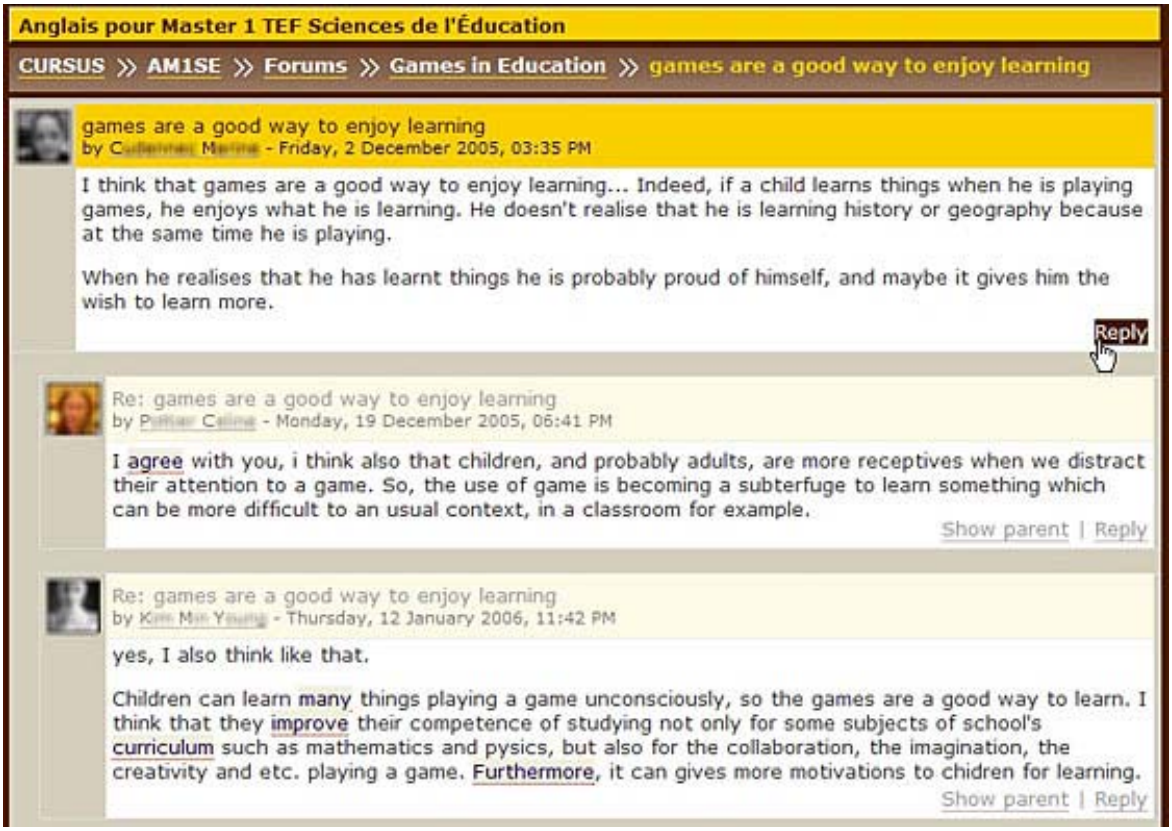


Figure 1 – Exemple de messages postés dans un fil de discussion sur un forum.

Caractéristiques du corpus *Cursus* :

- population totale d'étudiants ayant participé (en lisant ou en postant un message) aux forums de discussion : 106 ;
- nombre total de messages postés : 254 ;
- nombre total d'étudiants ayant posté au moins 1 message : 68 ;
- 21% des messages sont postés par seulement 7% des étudiants (les "hyperactifs") ;
- 45% des messages sont postés par 20% des étudiants ;
- 23% des étudiants ont posté un seul message.

Précisons enfin que la plateforme *Moodle* permet d'attribuer à chaque message posté sur un forum de discussions une note ou une appréciation. Lors de l'expérimentation rapportée ici, nous n'avons pas utilisé ce système de notation, mais les étudiants étaient avertis que leur participation aux divers forums ouverts sur le cours serait prise en compte dans l'évaluation de fin de semestre.

2.3. Le recueil des données

Les messages postés par les étudiants des 4 cours concernés ont été exportés au format texte, nettoyés

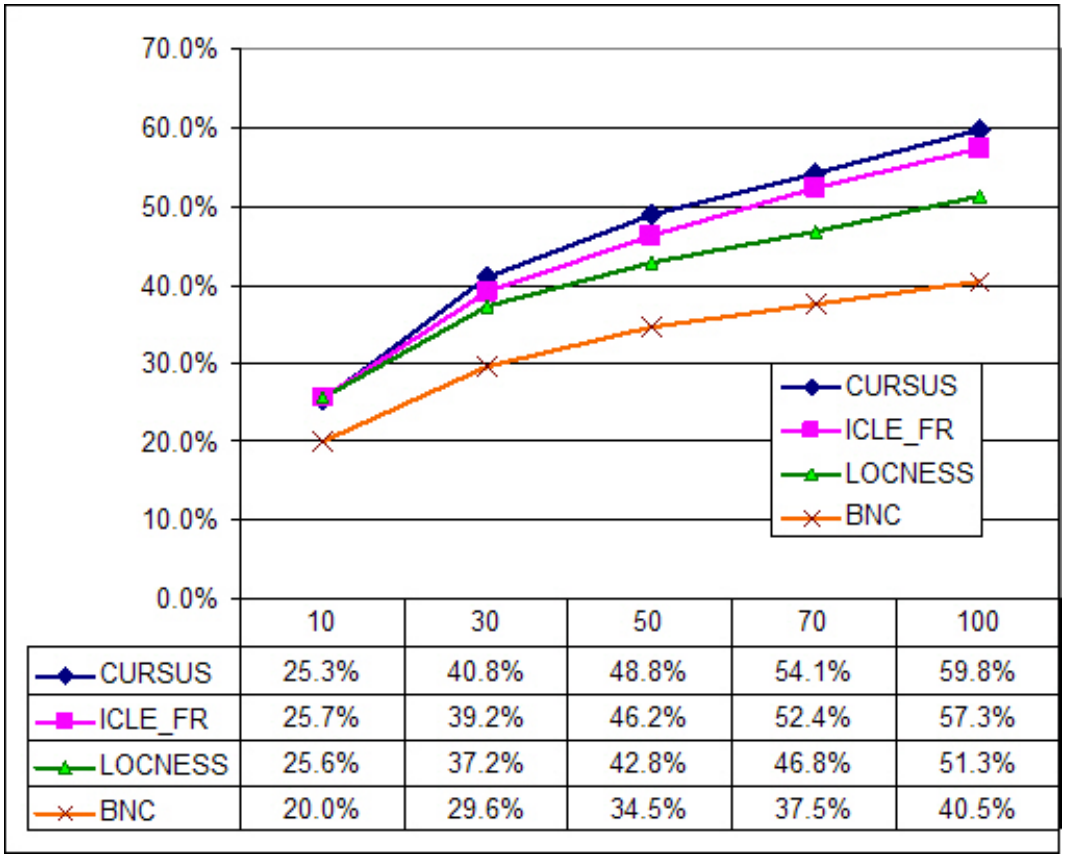
des éventuelles balises HTML, étiquetés (nom abrégé du cours et nom de l'étudiant) et enfin concaténés en un seul fichier au format texte pur. Ce travail est rendu fastidieux par l'absence de module d'exportation dans la plateforme *Moodle*, probablement parce que ce type d'exploitation des données est rare chez ses utilisateurs.

3. Observations sur les fréquences de vocabulaire

Un premier traitement possible avec les outils de traitement des données textuelles est l'établissement de listes de fréquence d'emploi du vocabulaire, que l'on peut comparer à des listes de fréquence d'emploi établies à partir d'autres corpus. Les études de Granger comparent des listes établies à partir de corpus relativement semblables en termes de genre (essais argumentatifs ou interviews) et de longueur. Les auteurs utilisent en particulier le corpus ICLE (*International Corpus of Learner English* ; environ 100 000 mots), composé de productions d'étudiants d'anglais d'une demi-douzaine de langues maternelles différentes, corpus qu'ils comparent avec par exemple le LOCNES (corpus d'essais rédigés par des étudiants anglophones). Faute d'avoir accès à ces corpus, nous nous sommes contenté de comparer la liste de fréquence du vocabulaire de notre petit corpus *Cursus* (27 000 mots) à la liste de fréquence établie à partir du gros *British National Corpus* [BNC] (BNC : 100 millions de mots). Nous sommes bien conscients que cette décision nous amène à comparer des corpus très dissemblables tant par leur volume que par leur contenu. Nous invitons donc le lecteur à garder à l'esprit cette discordance entre les objets comparés et à bien vouloir considérer que nous avons dû travailler avec les outils à notre disposition, comme étant *ad hoc* pour une réflexion davantage orientée vers la pratique que vers la recherche théorique.

3.1. Les 100 mots les plus fréquents

Si on s'intéresse aux 100 mots les plus fréquents d'un corpus d'apprenants d'anglais, quelle que soit son étendue, on peut penser avec Ringbom [Ringbom98] que, étant donné l'étendue limitée du vocabulaire dont disposent ces apprenants, les mots les plus fréquents seront sur-utilisés par rapport à un corpus de locuteurs natifs (*native speakers* NS). La taille du corpus a également un rôle à jouer. On voit sur la Figure 2 que les fréquences d'emploi du vocabulaire de notre corpus *Cursus* suivent de très près la courbe du corpus ICLE_FR. Nous remarquons tout comme [Ringbom98] que les 10 mots les plus fréquents représentent environ 25 % du vocabulaire total d'un corpus. C'est dans la zone des 30 à 100 premiers mots que les apprenants d'anglais ont tendance à sur-utiliser certains mots par rapport aux locuteurs natifs.



	10	30	50	70	100
CURSUS	25.3%	40.8%	48.8%	54.1%	59.8%
ICLE_FR	25.7%	39.2%	46.2%	52.4%	57.3%
LOCNESS	25.6%	37.2%	42.8%	46.8%	51.3%
BNC	20.0%	29.6%	34.5%	37.5%	40.5%

Figure 2 – Les 100 mots les plus fréquents dans 4 corpus comparés.

3.2. Méthodologie de recueil des données

Comme nous l'avons dit plus haut, n'ayant pas accès à des corpus de locuteurs natifs tels que le LOCNESS, nous avons décidé d'utiliser comme base de comparaison les listes de fréquence d'utilisation de mots anglais tirées du corpus BNC. Ces listes sont disponibles sur plusieurs sites Web, et nous avons utilisé le site compagnon de [LeechEtAl01]. Étant donné la nature hybride de notre corpus *Cursus* (à mi-chemin entre l'oral et l'écrit) il nous a paru intéressant d'utiliser également la liste de fréquence des mots du corpus d'anglais oral, disponible sur le même site Web (*BNC spoken*) [WordFreq]. Dans nos tableaux de comparaison de fréquence d'utilisation des mots ci-après, les indications chiffrées indiquent toujours ces fréquences en pourcentage par rapport aux divers corpus, mais le signe pourcent (%) sera omis, par simplification. Ainsi, par exemple, dans notre corpus *Cursus*, le mot *the* apparaît 1382 fois dans un corpus de 26880 mots, soit 5.14 (%) d'occurrences.

	A	B	C	D	E	F	G	H	I	J	K
1		BNC total		BNC speech		CURSUS		COMPARAISON		COMPARAISON	
2	rang	mot	%	mot	%	mot	%	BNC	sur-/sous-emploi	BNC SPOKEN	sur-/sous-emploi
3	1	the	6.18	the	3.96	**the**	5.14	6.18	**-1.04**	3.96	**1.18**
4	2	of	2.94	I	2.94	**to-Inf**	2.93	1.63	**1.30**	1.49	**1.44**
5	3	and	2.68	you	2.60	**I**	2.65	0.89	**1.77**	2.94	-0.29
6	4	a	2.16	and	2.52	a	2.43	2.16	0.27	1.86	**0.57**
7	5	in	1.82	it	2.45	**of**	2.38	2.94	**-0.56**	1.46	**0.92**
8	6	to-Inf	1.63	a	1.86	**and**	2.26	2.68	**-0.42**	2.52	-0.26
9	7	it	1.09	's-Vb	1.77	in	2.12	1.82	0.30	1.16	**0.96**
10	8	is	1.00	to-Inf	1.49	**is**	1.73	1.00	**0.73**	1.02	**0.71**
11	9	to-Prep	0.93	of	1.46	**it**	1.58	1.09	**0.50**	2.45	**-0.87**
12	10	was	0.92	that	1.43	**that**	1.41	0.73	**0.68**	1.43	-0.02
13	11	I	0.89	-n't	1.22	for	1.09	0.84	0.25	0.62	**0.47**
14	12	for	0.84	in	1.16	with	0.95	0.66	0.29	0.44	**0.50**
15	13	that	0.73	we	1.04	**can**	0.92	0.24	**0.68**	0.36	**0.56**
16	14	you	0.70	is	1.02	**are**	0.92	0.47	**0.44**	0.47	**0.45**
17	15	he	0.68	do	0.96	**learning**	0.86	0.01	**0.85**	#N/A	#N/A
18	16	be	0.66	they	0.93	be	0.85	0.66	0.18	0.58	0.27
19	17	with	0.66	er	0.85	**we**	0.85	0.36	**0.49**	1.04	-0.20
20	18	on	0.65	was	0.81	but	0.83	0.46	0.38	0.64	0.20
21	19	by	0.51	yeah	0.79	have	0.82	0.47	0.34	0.75	0.07
22	20	at	0.48	have	0.75	you	0.80	0.70	0.11	2.60	**-1.79**
23	21	have	0.47	what	0.73	to-Prep	0.78	0.93	-0.15	0.70	0.09
24	22	are	0.47	he	0.73	this	0.74	0.46	0.27	0.56	0.17
25	23	not	0.46	that	0.72	not	0.69	0.46	0.23	0.47	0.22
26	24	this	0.46	to-Prep	0.70	**think**	0.68	0.09	**0.59**	0.40	0.28
27	25	's-Gen	0.46	but	0.64	**games**	0.66	0.01	**0.66**	#N/A	#N/A
28	26	but	0.46	for	0.62	on	0.64	0.65	-0.01	0.57	0.07
29	27	had	0.45	erm	0.60	's-Verb	0.61	0.35	0.26	#N/A	#N/A
30	28	they	0.43	be	0.58	-n't	0.60	0.33	0.26	1.22	**-0.63**

Tableau 1 – Comparaison des fréquences d'emploi dans les corpus BNC et *Cursus*.

3.3. Quelques sur- et sous-emplois caractéristiques

Nous suivrons les catégories définies par Granger et Rayson [GrangerRayson98] pour l'analyse des sur- et sous-emplois significatifs de vocabulaire par les apprenants d'anglais.

3.3.1. Les articles

Granger et Rayson ([GrangerRayson98] : 125) notent que les apprenants *francophones* ont tendance à sur-utiliser l'article indéfini *a* et à sous-utiliser l'article défini *the*. Nous faisons les mêmes constatations en comparant les listes de fréquence de notre corpus *Cursus* et du corpus BNC (Tableau 2).

	Cursus	BNC	différence	*BNC spoken*	différence
a	2.43	2.16	+0.27	1.86	+0.57
the	5.14	6.18	-1.04	3.96	+1.19
totaux	7.57	8.34	-0.77	5.82	+1.75

Tableau 2 – Emploi comparé des articles

Ces auteurs attribuent le sur-emploi de l'article indéfini *a* (et le sous-emploi proportionnel de l'article défini *the*) à un usage moins conforme aux normes de la langue écrite soutenue (qu'à celles de la langue familière ou orale) chez les apprenants. Notons d'ailleurs que le sous-emploi de l'article défini *the*

augmenterait encore si on déduisait les emplois fautifs de cet article précédant des indénombrables indéterminés. Le décompte de ces emplois, au nombre d'une trentaine dans notre corpus *Cursus*, augmenterait le sous-emploi de *the* de -1.04 à -1.14. En revanche, les auteurs cités notent un sur-emploi global des déterminants chez les apprenants par rapport aux locuteurs natifs, ce qui n'est pas corroboré par nos résultats. Le total des emplois des articles *a* et *the* dans *Cursus* est en effet inférieur (-0.77) à celui constaté dans le BNC. Mais la nature hybride de notre corpus nous incite à effectuer une comparaison avec le corpus *BNC speech*, dans lequel nous constatons une différence très sensible de l'emploi de ces déterminants (-2.58 par rapport au BNC total).

3.3.2. Les déterminants et pronoms indéfinis

Granger et Rayson [GrangerRayson98] font état chez les apprenants francophones d'un sur-emploi des déterminants indéfinis *a lot* et *lots*, que ces auteurs considèrent comme moins formels que le déterminant *many*. Notre corpus *Cursus* ne corrobore pas tout à fait ce sur-emploi, puisque *many* y est sur-employé à 0.09, et *a lot* à 0.08. Par ailleurs, ces mêmes apprenants francophones sur-utiliseraient les pronoms en *–body* au détriment de ceux en *–one*. Ces usages ne sont pas évidents dans notre corpus : *somebody* et *someone* sont utilisés à égalité, et l'on compte 3 occurrences de *nobody* et aucune de *no-one* (terme que les apprenants francophones ne semblent pas connaître). Notons enfin un emploi relativement élevé de *everybody*, dû à la fréquence d'utilisation de la salutation "*Hello / Hi everybody*" en tête de message (BNC : 0.01 ; *BNC spoken* : 0.03 ; *Cursus* : 0.08).

3.3.3. Coordination et subordination

Comme chez les auteurs cités, notre corpus révèle un sur-emploi des conjonctions de coordination *but* (+0.46) et *or* (+0.17) et un léger sous-emploi de *and* (-0.06). De même, la plupart des conjonctions de subordination sont sous-employées, sauf *because* et *if*, termes associés à la langue parlée, et qui sont sur-employés.

***Cursus* comparé à :**	***after***	***although***	***before***	***until***	***when***	***while***	***because***	***if***
BNC	-0.05	-0.03	+0.04	+0.01	+0.12	-0.02	+0.37	+0.21
BNC spoken	-0.01	+0.01	+0.05	+0.03	+0.06	+0.03	+0.37	0

Tableau 3 – Emploi comparé des conjonctions de coordination et de subordination.

3.3.4. Divers

Le pronom personnel de la première personne du singulier, *I*, est le mot qui montre le plus grand écart d'emploi avec la liste du corpus du BNC, avec un sur-emploi de +1.77. Ceci n'a rien de surprenant, le genre du forum de discussion amenant les étudiants à affirmer à mainte reprise leur opinion personnelle. On compte dans le corpus *Cursus* pas moins de 130 occurrences de "*I think*", 30 de "*I agree*" (ou *"*I'm agree*"), etc. En revanche, ce même pronom montre un sous-emploi de 0.29 par rapport au corpus *BNC spoken*, ce qui indique une petite distanciation par rapport à la langue parlée. Bien évidemment, malgré leur nature hybride, il manque tout de même aux "conversations" des forums un interlocuteur que l'on peut directement interpeller, ce qui explique le sous-emploi du pronom de 2e personne *you* dans le corpus *Cursus* par rapport au *BNC spoken* (-1.79).

On trouve un grand nombre d'adverbes en *–ly* dans notre corpus, la plupart étant des décalques

relativement transparents d'adverbes français. Mais on ne trouve effectivement aucune occurrence des adverbes mentionnés par [GrangerRayson98] comme étant sous-employés : *amplifiers (greatly, truly, widely, readily, highly), disjuncts (importantly, traditionally, effectively* [1 occurrence]*), modal adverbs (possibly, supposedly), time adverbs in –ly (newly, currently, previously, ultimately)*. En revanche, les adverbes utilisés dans la langue parlée et mentionnés par [GrangerRayson98] sont presque tous présents : *in fact, of course, anyway, indeed, absolutely, really, certainly, now*.

3.4. Implications didactiques

Pour [GrangerRayson98], l'analyse des fréquences de vocabulaire des corpus d'apprenants d'anglais met en évidence que leurs écrits présentent toutes les caractéristiques d'une langue orale et pratiquement aucune caractéristique de la langue académique écrite. Étant donné la nature bivalente des forums de discussion, il est encore moins surprenant que les formes de la langue orale prédominent dans notre corpus. Cependant, si notre but est de promouvoir chez nos étudiants une langue argumentative se rapprochant de la langue académique écrite, il nous faudra développer des stratégies visant à enrichir leur vocabulaire. [GrangerRayson98] recommandent "*greater exposure to good quality expository or argumentative writing, as found [...] in the editorials of quality newspapers*". On peut également penser que les textes soumis à la lecture et à l'étude de nos étudiants peuvent servir de modèle. Si nous prenons l'exemple des adverbes pas ou peu employés dans notre corpus *Cursus*, on constate que certains d'entre eux sont effectivement employés dans le corpus de ces textes, comme : *greatly (1) ; highly (2) ; importantly (2) ; previously (3) ; readily (1) ; ultimately (1) ; widely (2)*. Comme on le verra plus loin, il est possible d'utiliser la technique des renvois automatiques depuis les messages des forums vers des glossaires lexicaux ou grammaticaux pour attirer l'attention des étudiants sur les règles d'emploi de certains mots.

4. Observations d'erreurs syntaxiques typiques dues à la nativisation

Pour Demaizière & Narcy-Combes,

> *la nativisation [...] fait que l'apprenant perçoit et analyse toute nouvelle donnée langagière en L2 selon des critères déjà en place, qui lui sont personnels [...]. Un apprentissage de L2 ne peut donc être efficace que s'il permet de passer par un travail cognitif que l'on peut qualifier de dénativisation, qui permettra d'effacer les représentations erronées et de les remplacer par celles qui correspondent bien au système de la L2.* [DemaizièreNarcyCombes05]

Dans une optique pratique de remédiation, nous avons recherché les erreurs les plus fréquentes et qui nous semblent les plus graves et les plus "fossilisées" chez les sujets de notre étude. L'inventaire qui suit n'a pas nécessairement d'ordre logique autre que *ad hoc*.

Les noms indénombrables employés au pluriel : *knowledges (12) ; informations (12)*

Contexte gauche	Occurrence	Contexte droit
example, or instead you have or know how to get	informations	on those i like we could exchange from this ob
b way"; I would like to know if you can give me	informations	about methods to improve my English. I have be
e reception of the presidency. if you want more	informations	: www.letudiant.fr Good visit
game or more likely kill the gameplay. Too much	informations	and less interactivity with the story can des
f we had critical glance behind the abundance of	informations	. All of resources are not forcing true!
k there are particular presentations formats of	informations	like pop-up, dynamic illustrations... Students
story. The second drawback is the importance of	informations	given to players, sometimes the game is much
e can admit the Web constitute a vast resouce of	informations	I think that Education must stay a human's job
hith our personnal interrest and maybe find some	informations	in English to keep the aim and motivation to w
think is not normal. Please, can you give some	informations	. Thanks. AG.
rom the classroom and with the pupils at all the	informations	that they need. I think that in a pedagogical
ny things are new, you can't easily organize the	informations	and select the points to work. You have done t
Contexte gauche	**Occurrence**	**Contexte droit**
ote about Piaget's theories is their influence on	knowledges	about children's thinking. All the work of this
prevailing idea about the way human beings build	knowledges	and skills : logic thinking is an old intellect
ortant for children's learning. Most of the time,	knowledges	and skills are presented them with games and it
in Master EAD, we are all engaged in a way where	knowledges	and skills value is obvious. For the pleasure.
r many possibilities of games in order to develop	knowledges	and skills. Children are the active builders of
For the pleasure. And because of all the "things"	knowledges	bring to us in understanding didactics. In the
h the part of text who speak about all differents	Knowledges	computer game can develop. And I think introduc
ion, handlings and experiments... to be caught at	knowledges	game...
work in network is to make easier the training of	knowledges	taught by using technologies in network to allo
lved in this discussion : games favor creativity,	knowledges	, skills, acquisitions on a personnal way and ry
es are necessary to assure a good construction of	knowledges	.
namic vision that he gives at the construction of	knowledges	. And the profusion of new technologies allows t
erience and learning of others to build their own	knowledges	. But i think guides are necessary to assure a g

Figure 3 – Concordances de type KWIC pour *informations* et *knowledges*.

La Figure 3, ci-dessus, montre un exemple typique de concordance de type *KWIC (Key Word In Context*) dans lequel les occurrences du mot recherché figurent dans la colonne centrale. La concordance pour *informations* est triée en ordre alphabétique sur le contexte de gauche, tandis que celle pour *knowledges* est triée sur le contexte de droite.

Emploi erroné de l'article défini :

- pour les mots abstraits indéterminés : *the creativity (2) ; the education (3) ; the e-learning (10) ; the knowledge (5) ; the learning (8) ; the life (2) ; the society (4) ; the teaching (2) ;*

- avec le génitif : *the Piaget's theories.*

Erreurs sur la morphologie de l'article défini : *the Web can be an other way to learn ; the teacher can present her in an other way ; we can consider e-learning in an other way.*

Emploi du pronom relatif *who* pour les choses : 14 occurrences (sur 60 occurrences totales de *who*), par exemple : *Thank you Suzanne for this adress, **who** can also be very interesting to show to my pupils when I will go back to school.*

Construction de verbes : *I am agree* ou *I'm agree* : 7 (sur 48 occurrences du verbe *agree*) ; *I don't / I can't agree* sont utilisés, mais le verbe *disagree* semble ignoré des étudiants.

Les quantifieurs :

- emploi de *much* + nom pluriel : *But why there are so much resistances... ?*

- emploi de *too much* + adjectif/participe passé : *Bandura reproach to these bad models of being too much mediatized.*

5. Stratégies de remédiation

Si, comme nous l'avons vu plus haut, les discussions sur les forums s'apparentent à une situation de débat en cours présentiel, le rôle de correcteur de l'enseignant devra cependant prendre des formes différentes. La nature asynchrone de ces discussions entraîne une correction nécessairement décalée dans le temps, contrairement à la situation de cours en présentiel où l'enseignant peut réagir *hic et nunc*. En outre, la nature écrite du support implique une correction également écrite et donc permanente, qui attacherait de façon psychologiquement difficilement supportable une erreur à un apprenant, si l'enseignant postait un message mettant ladite erreur en évidence. On pourrait certes envisager un message privé envoyé uniquement à l'apprenant fautif. Ce serait se priver alors de l'impact potentiel d'une telle correction sur l'ensemble du groupe. La littérature du domaine s'est depuis longtemps posé cette question et semble avoir abouti au consensus suivant : ne signaler les erreurs en public que lorsqu'elles sont pléthoriques chez de nombreux apprenants.

5.1. Détournement de l'outil "Glossaire" dans *Moodle*

Une fonctionnalité de la plateforme *Moodle*, légèrement détournée de son objectif initial, nous apporte une solution qui évite les inconvénients mentionnés. *Moodle* permet de créer des glossaires, composés de façon classique d'une entrée et d'une "définition" (voir Figure 5). A partir du moment où un mot figure dans l'un des glossaires d'un cours, si ce mot figure dans un texte quelconque du cours, y compris dans les messages des forums, il y sera dynamiquement muni d'un hyperlien de renvoi à son entrée dans le glossaire correspondant. Ce système permet d'attirer l'attention de l'étudiant sur une erreur de syntaxe ou de lexique selon deux modalités différentes.

5.1.1. Lien vers une erreur prévue

Lorsqu'un étudiant poste un message dans un fil de discussion, il parcourt évidemment les messages déjà postés par ses camarades ou l'enseignant. Ces messages peuvent comporter des hyperliens vers des entrées déjà rédigées par l'enseignant dans le glossaire "grammatical" (voir Figure 4). Si l'étudiant envisage de reprendre dans son message certains mots ou expressions déjà utilisés par ses camarades, il peut consulter ces hyperliens. Ensuite, lorsque l'étudiant a rédigé son propre message et l'a envoyé, il dispose de 30 minutes pour le modifier. Son message s'affiche et, s'il comporte des mots déjà entrés dans le glossaire, ces mots s'affichent avec leur hyperlien. Il peut donc à ce moment-là consulter le glossaire et – le cas échéant – rectifier des erreurs. D'après nos observations, cette possibilité de revenir sur la rédaction de son propre message est malheureusement rarement utilisée par nos étudiants.

5.1.2. Création d'une nouvelle entrée

Lorsque l'enseignant prend connaissance des messages postés dans les forums d'un cours, s'il constate une erreur non répertoriée, il peut rédiger dans le glossaire *General Grammar* une entrée comportant une règle, des exemples et surtout un mot-clé formulé de telle façon que le lien depuis l'erreur de l'élève sera automatiquement et dynamiquement construit par *Moodle* lors du prochain affichage du fil de discussion.

5.1.3. Exemple

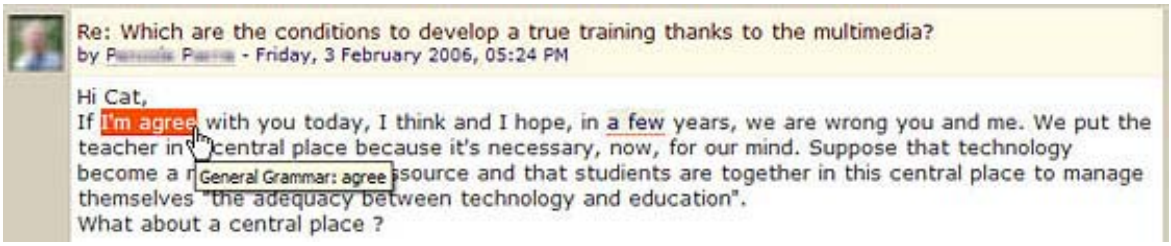


Figure 4 – Exemple de message d'un forum.

Lorsque Pierre relit son message, il voit que les mots ou expressions "*I'm agree*" et "*a few*" sont munis d'un hyperlien. Un clic sur "*I'm agree*" ouvre la fenêtre popup suivante :

Figure 5 – Exemple d'article du glossaire *General Grammar*.

5.1.4. Remarque

On voit dans l'exemple de la Figure 4 que l'expression "*in a few years*" est munie d'un hyperlien bien que son emploi ne soit pas une erreur. De même, un emploi correct du verbe *agree* sera muni d'un hyperlien renvoyant à la règle d'emploi. Il serait possible mais extrêmement fastidieux de recenser et de créer des hyperliens uniquement dans les cas d'emploi erroné de formes lexicales ou syntaxiques. Par exemple, il faudrait faire pointer vers l'entrée *agree* du glossaire grammatical seulement les formes comme "*I'm agree*", "*I'm not agree*", "*Cathy is not agree (with me)*", etc. Il nous a paru plus simple d'afficher l'hyperlien dans tous les cas, ce qui permet aux étudiants de consulter le lien aussi bien dans un but de remédiation que de consolidation. Ainsi le lien n'est pas automatiquement synonyme d'erreur, mais une invitation à en savoir plus sur un point donné.

5.2. Cette stratégie est-elle efficace ?

La plateforme *Moodle* dispose d'un système de suivi et d'enregistrement (*log*) d'un bon nombre des actions effectuées par l'utilisateur, et en particulier de la consultation des hyperliens menant vers des articles des glossaires de cours. Moyennant un travail non négligeable de récupération de données, il est ainsi possible pour l'enseignant de savoir quelles entrées sont consultées, par qui, et à quel moment précis (*timestamp*). Malheureusement aucune donnée n'indique la nature du document *à partir duquel* s'est fait le clic sur l'hyperlien. Pour le savoir, il faut consulter l'historique de chaque étudiant. Une analyse sommaire des historiques nous amène à constater que les hyperliens présents dans les messages des forums sont relativement peu consultés. Les textes qui servent de support au cours comportent un nombre assez important d'hyperliens renvoyant à un glossaire de type lexical ou de type encyclopédique, liens qui sont souvent consultés de manière systématique par les étudiants au cours de leur lecture. En revanche, lorsque les étudiants prennent connaissance des messages postés par leurs camarades dans les forums (ou lorsqu'ils relisent leurs propres messages antérieurs – s'ils les relisent), la motivation

communicationnelle (prendre connaissance du contenu du message) semble l'emporter sur la motivation linguistique (s'interroger sur la forme du message).

5.2.1. Un exemple (Pierre P.)

Le 27 janvier 2006 à 22:38, Pierre P. a posté ce message :

`Hi, I think it's time for me to begin,only for yhe second time, a discussion on the net. Last term I have not use these possibilities. I did not understand that it was important, for me, to get points...and to exchange with you...of course ! So I go to try to write to you what I think about this last text.` **`I'm agree`**`, with the autor, when he writes that computers and contents are linked. etc.`

Le lendemain, 28 janvier à 22:04, Pierre P. a reçu une copie de son message par courriel. En effet, la plateforme *Moodle* permet à l'enseignant ou aux étudiants de choisir d'être avertis par courriel dès qu'un nouveau message a été posté dans un forum de discussion. Cet étudiant a alors sans doute été intrigué par l'hyperlien automatiquement placé dans son texte et visible dans le message reçu dans son logiciel de messagerie. Il a cliqué sur cet hyperlien et a été dirigé vers le site de la plateforme du cours d'anglais, directement à l'entrée *agree* du glossaire grammatical, entrée qu'il a pu lire à ce moment-là.

Le 3 février 2006 à 17:24, Pierre P. a posté dans un autre fil de discussion le message suivant :

`Hi Cat, If` **`I'm agree`** `with you today, I think and I hope, in a few years, we are wrong you and me. We put the teacher in a central place because it's necessary, now, for our mind. etc.`

Apparemment la simple lecture de la règle du fonctionnement de *agree* consultée une semaine auparavant n'a pas eu l'effet de dénativisation de l'erreur souhaité.

Le 7 mars 2006 à 18:56, Pierre P. a posté le message suivant :

`Hi, I'm come back. I have read what you have written and` **`I agree`** `with you but.....I think i'ts not "a lot of things wich gets ease to learn" but probably the way of learning.`

Cette fois-ci la forme correcte est employée. On peut s'interroger sur les raisons du délai nécessaire à l'utilisation correcte de la forme. Il faudrait multiplier les observations des sujets ainsi que les observations dans le temps pour en apprendre un peu plus sur les mécanismes de dénativisation à l'œuvre.

6. Conclusion : *"from data-driven learning to data-driven learning on the web"*

Tim Johns [Johns88a] a le premier lancé l'idée d'utiliser des exemples de concordances authentiques pour favoriser la réflexion des apprenants (*data-driven learning*). Dans le cadre d'un apprentissage en ligne, nous disposons maintenant d'outils de concordances en ligne qui pourraient être mis à profit pour un tel type de réflexion. Gaskell et Cobb [GaskellCobb04] de l'Université du Québec à Montréal poursuivent tout naturellement l'idée de Tim Johns en proposant sur leur site *The Compleat Lexical Tutor* [LexTutor] des activités de "*data-driven language learning on the web*". Bien que nous utilisions depuis 20 ans des concordanciers afin d'extraire de corpus de langue authentique des matériaux pour fabriquer des exercices de langue, nous n'avions encore jamais envisagé d'utiliser un concordancier en

ligne pour favoriser la réflexion des étudiants. Nous allons y songer, en nous appuyant sur l'argument avancé par Gaskell & Cobb :

> *We argue that an example-based acquisition theory translates into an instructional theory only if we can do the following things: (a) vastly increase the number of examples that L2 learners are exposed to in a given unit of time, (b) organize these examples so their patterns are highlighted, (c) get learners to attend to the examples, and (d) give systematic feedback on the success of interpreting the examples.* [GaskellCobb04]

Lors d'une prochaine expérimentation, toujours sur la plateforme *Moodle* de notre université, et avec le même public d'étudiants, nous envisageons d'intégrer des renvois systématiques à des extraits de concordances dans notre glossaire grammatical afin d'en étudier l'impact sur la correction des erreurs des apprenants. La difficulté prévisible est que la plateforme *Moodle* ne permet pas actuellement de garder la trace des consultations de ces extraits de concordances.

Références

Les liens externes étaient valides à la date de publication.

Bibliographie

[BongaertsEtAl88]
Bongaerts, T. *et al*. (dir.) (1988). *Computer Applications in Language Learning.* Dordrecht : Foris Publications.

[ChambersOSullivan04]
Chambers, A. & O'Sullivan, I. (2004). "Corpus consultation and advanced learners' writing skills in French". *ReCALL*, vol. 16, n° 1. pp. 158-172.

[DemaizièreNarcyCombes05]
Demaizière, F. & Narcy-Combes, J.-P. (2005). "Méthodologie de la recherche didactique : nativisation, tâches et TIC". *Alsic (Apprentissage des langues et Systèmes d'Information et de Communication)*, vol. 8. pp. 45-64. http://alsic.u-strasbg.fr/v08/demaiziere/alsic_v08_14-rec8.htm

[GaskellCobb04]
Gaskell, D. & Cobb, T. (2004). "Can learners use concordance feedback for writing errors?". *System*, vol. 32, n° 3. pp. 301-319. http://www.lextutor.ca/cv/pdf/concordance_feedback.pdf

[Granger98]
Granger, S. (dir.) (1998). *Learner English on Computer*. London & New York : Addison Wesley Longman.

[GrangerRayson98]
Granger, S. and Rayson, P. (1998). "Automatic Profiling of Learner Texts". In Granger, S. (dir.). pp. 119-131.

[Johns88a]
Johns, T. (1988a). "Whence and Whither Classroom Concordancing". In Bongaerts *et al.* pp. 9-34.

[Johns88b]
Johns, T. (1988b). "Implications et applications des logiciels de concordance dans la salle de classe". *Les Langues Modernes*, n° 5. pp. 29-45 (traduction française de Johns, 1988a).

[LeechEtAl01]
Leech, G., Rayson, P. & Wilson, A. (2001). *Word Frequencies in Written and Spoken English: based on the British National Corpus*. London : Longman.

[Mangenot02]
Mangenot, F. (2002). "Communication écrite entre étudiants par forum Internet : un nouveau genre d'écrit universitaire ?". *Enjeux*, n° 54. pp. 166-182. Namur : CEDOCEF. http://w3.u-grenoble3.fr/espace_pedagogique/mangenot-bruxelles.rtf

[Panckhurst98]
Panckhurst, R. (1998). "Analyse linguistique du courrier électronique". Actes du colloque *Les relations entre individus médiatisées par les réseaux informatiques, GRESICO*, Vannes. Paris : L'Harmattan. http://recherche.univ-montp3.fr/rachel/spip/IMG/pdf/gresico.pdf

[Panckhurst99]
Panckhurst, R. (1999). "Analyse linguistique assistée par ordinateur du courriel". In Anis, J. (dir.) *Internet, communication et langue française.* Paris : Hermès. http://recherche.univ-montp3.fr/rachel/spip/IMG/pdf/Anis.pdf

[Rézeau04]
Rézeau, J. (2004). "L'apport original du forum de discussion de WebCT dans un cours d'anglais appliqué à l'archéologie". *Les cahiers de l'Acedle*, n° 1. http://acedle.u-strasbg.fr/article.php3?id_article=399

[Ringbom98]
Ringbom, H. (1998). "Vocabulary frequencies in advanced learner English : A cross-linguistic approach". In Granger, S. (dir.). pp. 41-52.

Logiciels utilisés

Moodle. Plateforme d'enseignement à distance. http://moodle.org

Scott, M. *Oxford WordSmith Tools*. http://www.oup.co.uk/episbn/0-19-459400-9

Véronis, J. *Contextes, Concordancier pour les corpus écrits et oraux*. http://www.up.univ-mrs.fr/veronis/logiciels/Contextes/index-fr.html

Sites Internet

[BNC]
British National Corpus. http://www.natcorp.ox.ac.uk/

[WordFreq]
Companion Website for: Word Frequencies in Written and Spoken English: based on the British National Corpus. http://www.comp.lancs.ac.uk/ucrel/bncfreq/

- List 1.2: Rank frequency list for the whole corpus (not lemmatized): http://www.comp.lancs.ac.uk/ucrel/bncfreq/lists/1_2_all_freq.txt
- List 2.2: Rank frequency order: spoken English (not lemmatized): http://www.comp.lancs.ac.uk/ucrel/bncfreq/lists/2_2_spokenvwritten.txt

[LexTutor]
The Compleat Lexical Tutor. http://www.lextutor.ca/

Bibliographie complémentaire

Lamy, M.-N. & Klarskov, H. (2006). "Using concordance programs in the Modern Foreign Languages classroom". In Davies, G. (dir). *ICT4 LT: Information and Communications Technology for Language Teachers*. http://www.ict4lt.org/en/en_mod2-4.htm

McEnery, T. & Wilson, A. (2006). "Corpus linguistics". In Davies, G. (dir). *ICT4 LT: Information and Communications Technology for Language Teachers*. http://www.ict4lt.org/en/en_mod3-4.htm

Rézeau, J. (2001). "Concordances in the classroom: The evidence of the data". In Chambers, A. & Davies, G. *ICT and Language Learning: A European Perspective*. Lisse : Swets & Zeitlinger. pp. 147-166.

Tribble, C. & Jones, G. (1997). *Concordances In The Classroom : a resource book for teachers*. Houston : Athelstan.

À propos de l'auteur

Joseph RÉZEAU est maître de conférences en anglais de spécialité à l'université Rennes 2 où il est membre de l'équipe de recherche DILEM-LIDILE. Il s'intéresse à l'utilisation de l'ordinateur pour l'enseignement des langues depuis le début des années 1980. Il s'intéresse à la conception de dispositifs d'apprentissage multimédias et interactifs, ainsi qu'à l'analyse de la trace du parcours des apprenants utilisateurs de tels dispositifs. Il a développé pour la plateforme d'enseignement en ligne *Moodle* plusieurs utilitaires, dont un nouveau type de question faisant appel aux expressions régulières pour une analyse de réponse poussée et un feedback plus approprié.
Courriel : joseph@rezeau.org
Toile : http://rezeau.org
Adresse : Département d'anglais - Université Rennes 2 - F-35043 Rennes Cedex

Date de réception de l'article : 24 août 2006 ; date d'acceptation : 8 janvier 2007

© Apprentissage des Langues et Systèmes d'Information et de Communication, mars 2007