

**DES DOCUMENTS AUTHENTIQUES ORAUX AUX
CORPUS : UN DEFI POUR LA DIDACTIQUE DU FLE**

Jeanne-Marie Debaisieux

Equipe CRAPEL, ATILF–CNRS, Nancy–Université

Résumé

Après un rappel de la façon dont les documents authentiques sont intégrés comme corpus de découverte pour l'apprentissage/acquisition de l'expression orale dans la démarche du CRAPEL, l'article montrera comment l'évolution des technologies permet un saut quantitatif et qualitatif vers l'intégration des corpus selon les mêmes principes. Ce point sera développé au travers de la présentation d'un projet en cours au sein de l'équipe.

Abstract

Following a review of the ways authentic documents can be integrated into the acquisition/learning of spoken skills at the CRAPEL in the form of a discovery corpus, this article shows how the evolution of new technologies represents a quantitative and qualitative advance in the integration of corpora in precisely the same way. The point is developed and supported by a description of an on-going project at the CRAPEL.

Introduction

En didactique du français langue étrangère (FLE), l'utilisation des documents authentiques s'est peu à peu répandue mais ne concerne généralement que les activités de compréhension et n'intervient souvent que dans la phase de confrontation. Le recours à de tels documents pour l'apprentissage/enseignement de l'expression orale reste exceptionnel. Une des raisons les plus évoquées de ce choix est que ces documents seraient trop difficiles pour les apprenants. Dans ce domaine, le CRAPEL a depuis longtemps adopté un principe tout différent. Je montrerai dans une première partie selon quelles modalités les documents authentiques sont intégrés dans la méthodologie de l'expression orale. Je mettrai ensuite en évidence comment, grâce aux nouvelles technologies, peut s'opérer un saut à la fois quantitatif et qualitatif du document authentique isolé vers un corpus oral. Je terminerai en présentant un projet de recherche d'un centre de ressources en ligne intégrant ces nouvelles perspectives.

1. Les documents authentiques dans l'approche du CRAPEL

1.1. Leur fonction

L'approche méthodologique du CRAPEL se définit essentiellement par le fait qu'elle est une démarche d'apprentissage plus que d'enseignement. Elle se situe explicitement du côté de l'apprenant et travaille à l'amélioration des conditions de l'apprentissage, c'est-à-dire des activités menées par un individu de façon volontaire et consciente dans le but d'une maîtrise non réfléchie de cette langue. Elle peut être qualifiée de cognitiviste au sens où elle cherche à créer les conditions favorables au développement du processus interne d'appropriation de la langue, c'est-à-dire son acquisition.

Pour ce faire, les activités d'apprentissage sont organisées en trois phases que je rappellerai ici rapidement¹. Il s'agit tout d'abord, au travers d'activités dites de découverte, de permettre à l'apprenant d'observer les comportements langagiers afin de repérer et de sélectionner parmi les éléments de nature segmentale, supra-segmentale, stratégique ou interactionnel, ceux dont il vise l'acquisition. La deuxième phase d'activités dite « d'utilisation systématique » consiste pour l'apprenant à s'entraîner de façon répétée sur un point particulier afin de mettre en place des processus d'automatisation. La troisième phase, d'utilisation non systématique, consiste, au travers de simulations communicativement situées, à mettre en pratique les éléments sur lesquels il a travaillé.

1. Je renvoie à Gremmo (1997) pour une analyse détaillée de ces trois phases.

Il s'agit de favoriser ainsi la mise en place progressive d'un savoir faire qui consiste à produire des discours en langue étrangère, c'est-à-dire « à mettre en œuvre ses savoirs pour... dire des textes réalisant des intentions » (Holec, 1990 : 68) selon un procédé que Shiffrin et Schneider (1977) décrivent dans ces termes :

La transformation des savoirs déclaratifs en savoirs procéduraux implique le passage d'un traitement contrôlé qui requiert une grande attention et s'appuie sur la mémoire à court terme à un traitement automatique qui s'appuie sur des procédures automatisées stockées dans la mémoire à long terme.²

Dans ce cadre, la première fonction des documents authentiques, c'est-à-dire « tout document écrit, sonore, visuel, produit à d'autres fins que des fins didactiques » (Pieron, 1978 : 109), ou semi-authentiques³ est de donner à l'apprenant un accès aux multiples paramètres – linguistiques interactionnels, proxémiques (en cas de documents vidéo) – en jeu dans une interaction. Leur utilisation est donc un des fondements de la méthodologie en compréhension orale et en expression orale. Seule la mise au contact de véritables discours permet en effet à l'apprenant de traiter « par construction et vérification d'hypothèses, ce qu'il aura saisi des comportements langagiers auxquels il est exposé » (Holec, 1982 : 67) et par là même de progresser dans la connaissance de la langue qu'il vise à acquérir. Les documents authentiques fournissent l'exposition à la langue et son fonctionnement comme l'illustre la figure 1, extraite de Holec (1990 : 67).

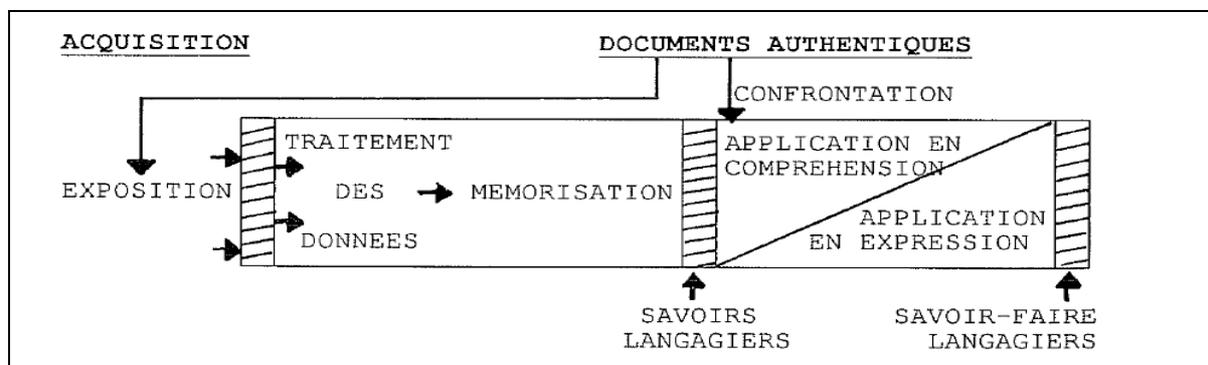


Figure 1. Fonction des documents authentiques.

Holec (1990 : 69) parle ainsi de la mise à disposition des documents authentiques comme « corpus de découverte ». A la même époque, certaines

2. « The transformation of declarative knowledge into procedural knowledge involves a transition from controlled processing, which requires a great deal of attention and use of short-term memory, to automatic processing, which operates on automatised procedures stored in long-term memory. »
3. Le terme semi-authentique ou réaliste recouvre deux types de documents oraux. Ce peut être des interactions réelles enregistrées en situation qui sont réenregistrées en laboratoire par nos soins, pour des raisons techniques touchant à la qualité sonore de l'enregistrement ou pour des raisons éthiques de droit des personnes enregistrées. Nous utilisons par ailleurs des enregistrements « improvisés par des anglophones auxquels ont été données des indications sur la situation » (Abe *et al.*, 1978 : 5).

approches anglo-saxonnes mettent l'accent sur la nécessité de cette phase de découverte. Ainsi Schmidt (1990) cité par Perez Paredes (2004 : 250) :

Les travaux d'inspiration cognitive sur l'apprentissage des langues ont mis en évidence que le fait de découvrir, c'est-à-dire de noter, d'observer et/ou de prêter une attention particulière à un fait linguistique spécifique constitue un pré-requis pour l'apprentissage.⁴

Dans la méthodologie pour l'expression orale conçue par le CRAPEL, ces documents sont mis à disposition des apprenants selon diverses modalités pédagogiques que je présenterai dans la section suivante.

1.2. Modalités de mise à disposition

1.2.1. Le Système d'Apprentissage Auto-Dirigé avec Soutien (SAAS)

Le dispositif ayant fait l'objet de nombreuses publications⁵, j'en rappellerai ici les grands principes. Le système est basé sur une démarche volontaire qui permet à un apprenant, à la fois de disposer d'un conseiller qu'il peut rencontrer lorsqu'il le désire pour des entretiens qui constituent autant de mises au point sur son apprentissage et d'avoir accès au matériel d'apprentissage. Les rencontres avec le conseiller sont à l'initiative de l'apprenant. Le conseiller y joue un rôle d'expert et d'accompagnateur de l'apprenant sur tous les aspects de l'apprentissage auxquels ce dernier se confronte : besoins, méthodologie, progression, représentations sur la langue, etc. Le principe pédagogique vise à permettre à un apprenant de développer simultanément ses capacités langagières et ses capacités d'apprentissage. Le conseiller est amené ainsi à suggérer des modalités d'activités à l'apprenant. Les extraits ci-dessous, tirés d'enregistrements de sessions de conseil effectuées au CRAPEL, illustrent, au travers des suggestions du conseiller⁶, comment nous envisageons l'utilisation des documents authentiques pour l'apprentissage de l'expression orale.

Suggestions du conseiller

- Vous **regardez systématiquement** dans les documents authentiques, **vous regardez** ce que disent les locuteurs natifs.
- Ce que vous pouvez faire c'est **regarder** dans les articles et **regarder** à chaque fois là où il y a du prétérit et là où il y a du présent perfect.
- Ce que je vous propose c'est de vous donner des documents sur cassettes où les gens parlent de leurs expériences professionnelles et cette fois-ci vous allez utiliser ces cassettes non pas pour travailler votre compréhension... mais pour **repérer des façons de dire...** des expressions, du vocabulaire... mais c'est pas mécanique, c'est pas une structure répétée en dehors de tout contexte, c'est

4. « Within cognitive perspective on language learning, it has been pointed out that noticing, that is noticing, observing or paying special attention to a particular language item, is generally a prerequisite for learning. »

5. Cf. notamment Cembalo et Holec (1973) ; Abe et Gremmo (1981, 1983) ; Ciekanski (2005).

6. Le principe du conseil est de suggérer et non d'imposer des activités à l'apprenant.

vraiment **pour vous faire écouter** tiens comment il fait celui là ah il fait ça toc **vous notez vous reprenez pour vous...** Ce seront des **expressions** que vous aurez **choisies** dont vous pensez que vous avez réellement besoin... (Ciekanski, 2005 : 466)

- [enregistrer une heure de conversation avec un anglophone] ça vous donnerait aussi un document de travail parce que ce que dira la personne anglophone... ce que cette personne là va dire vous êtes sûr que c'est de l'anglais euh du bon anglais de l'anglais authentique disons tel qu'il est vraiment parlé et à ce moment là vous pouvez après utiliser l'enregistrement de cette conversation... pour votre expression si **vous repérez** des expressions qui vous plaisent bien alors vous pouvez les travailler en les vous voyez en les répétant dans un premier temps et en vous entraînant vraiment à les redire. (Ciekanski, 2005 : 483)

Il s'agit bien (cf. les éléments en gras) d'inciter l'apprenant par des activités de repérage et de découverte à « récupérer » en quelque sorte les éléments qu'il jugerait utile de s'approprier pour son apprentissage. La démarche consiste à expliciter le rôle du travail d'observation et d'entraînement dans la démarche d'apprentissage et ce conformément à l'orientation du CRAPEL selon laquelle le développement de la capacité d'autonomie passe par l'enrichissement de la culture d'apprentissage⁷, c'est-à-dire par la prise de conscience par l'apprenant des tenants et aboutissants des activités d'apprentissage. La démarche n'est pas réservée à un dispositif d'apprentissage individuel et a été transposée dans les matériaux d'apprentissage conçus par l'équipe.

1.2.2. La mise à disposition dans les matériaux d'apprentissage

Les documents authentiques ou semi-authentiques sont un des fondements de l'activité éditoriale pédagogique du CRAPEL. Ils sont à la base des premiers modules « d'expression orale anglaise en apprentissage autodirigé » conçus entre 1974 et 1978. Plus récemment, la méthode *Vacances cubaines*, matériel d'enseignement/apprentissage de français destiné à la formation des professionnels du tourisme à Cuba⁸, présente un peu plus de quatre heures de documents audio semi-authentiques transcrits de façon intégrale. La méthode, conçue pour pouvoir « être utilisée directement par les apprenants, sans qu'il y ait nécessairement la médiation d'un enseignant » (Gremmo, 1997 : 65), présente une partie intitulée « Conseils pour apprendre » qui cible la formation des apprenants à la prise en charge progressive de leur apprentissage et dans laquelle sont fournies notamment des indications quant à l'utilisation des documents authentiques oraux pour l'expression orale. J'en présente ici un bref extrait :

De quelle façon les documents authentiques peuvent-ils vous aider à améliorer votre expression orale?

D'un côté, ils peuvent vous aider à **découvrir le fonctionnement de la langue** et à **vous entraîner à mettre en pratique certains aspects de la langue afin de les**

7. Je renvoie à Holec (1990) pour une présentation détaillée de ce concept.

8. CRAPEL et EAEHT (1998-1999) : trois volumes, trois cassettes audio. Voir Gremmo (1997) pour une présentation des options didactiques de ce matériel.

automatiser. Dans les documents authentiques, vous pouvez **observer** tous les aspects de la langue qui sont utiles pour mieux s'exprimer : les lexiques, les expressions adéquates, la grammaire, la prononciation et l'intonation. Les documents authentiques constituent **une sorte de base de données** de laquelle vous pourrez **extraire des informations utiles** sur la façon dont les français **s'expriment** dans des **situations particulières**.⁹

Il s'agit donc, comme le soulignent les éléments en gras, d'inciter les utilisateurs à s'appuyer sur les documents authentiques pour des activités de découverte et d'entraînement systématique. La démarche est explicitée : l'observation et l'entraînement permettent de mettre en place des automatismes concernant divers aspects du fonctionnement langagier. L'apprenant peut faire ses choix en fonction de ses propres critères et des situations communicativement ciblées dont il vise l'apprentissage.

1.3. Du côté des apprenants

La question que l'on est en droit de se poser est de savoir comment réagissent les apprenants à ces incitations à construire eux-mêmes leur matériel d'apprentissage. Les nombreuses sollicitations pour l'adaptation de *Vacances Cubaines* à d'autres pays¹⁰ permettent de penser que la méthode dans son ensemble est jugée pertinente pour la formation des professionnels du tourisme. Les entretiens de conseil nous donnent par ailleurs un accès direct aux réactions des apprenants. C'est ce qu'illustrent les extraits suivants.

- donc là j'ai essayé de repérer un peu les structures que du moins je connaissais mais bon qui pouvaient me ressortir (Ciekanski, 2005 : 413).
- ces cassettes [cassettes de conversation avec des natifs] elles sont intéressantes parce qu'il y a beaucoup d'expressions beaucoup de vocabulaire... et c'est vrai qu'à la relecture de la cassette je pouvais la noter donc la relecture de la cassette m'a permis de faire un travail de notes (Ciekanski, 2005 : 426).
- Ces documents, c'est vraiment une mine, parce que c'est des situations de la vie de tous les jours et on peut voir comment on doit dire les choses (Gremmo, 2000).

Les entretiens portent sur des problématiques très diverses, mais dans l'ensemble, les apprenants semblent saisir l'intérêt de la démarche¹¹. Il me paraît important de noter que cette pratique n'est pas réservée à des apprenants de niveau avancé ou spécialistes de l'apprentissage comme le montrent les extraits suivants dans lesquels une apprenante débutante ayant cessé sa scolarité après le certificat d'études, explique comment elle découvre certains principes du fonctionnement de la

9. Traduit de *Vacances cubaines : fonctions d'animation* – « Comment se servir de documents authentiques pour l'expression orale » Conseils pour apprendre (CRAPEL & EAEHT, 1998-1999 : 224). Les conseils sont rédigés en espagnol puisque la méthode s'adresse à des hispanophones.

10. Un volume de *Vacances en Amérique centrale* est paru en 2006 (CRAPEL & RECAFTUR) et un projet en cours concerne l'édition mexicaine de la méthode.

11. Les extraits sont tirés de la thèse de Ciekanski (2005) qui porte sur l'analyse de 31 entretiens de conseil sur une période allant de 15 jours à 6 mois pour un total de plus de 21h30 d'enregistrements.

langue anglaise. Le premier extrait porte sur la relation entre graphie et phonie, le second sur la façon dont elle a repéré, par analogie, le fonctionnement des phatiques. Les lettres C et A correspondent respectivement aux tours de parole de la conseillère et de l'apprenante¹².

- C366 oui donc petit à petit au fur et à mesure avec cette façon de travailler où vous avez à la fois le où vous avez à la fois la transcription et l'oral qui vous permet de comparer ça va vous permettre de faire les règles vous-même... vous allez vous-même tirer les règles de comment il faut faire pour ensuite pouvoir produire vous-même des mots que vous n'aurez jamais entendus ou jamais vus donc pouvoir produire des choses originales hein des énoncés à vous
- A 367 ah oui j'ai trouvé plein de trucs comme ça
- C 368 oui oui il faut être observatrice
- A 369 le i de *l'm* [prénom de A3] hein je dis bon c'est vrai que ça fait un grand *i* ben je dis au début je disais *i'm* (rires) alors et en entendant oh *l'm* bon ben je saurais le grand *i* c'est un ai
- C 370 et voilà (rires) comment vous avez trouvé vous-même la une règle (hum hum) de la langue oui oui
- A 371 voilà c'est des trucs (Ciekanski, 2005 : 626)
- A 151 ce qui m'a donné la puce à l'oreille c'est que quelques fois au début de la phrase elle le fait euh alors j'ai repassé comme il faut la cassette en réécoutant un peu plus et ben je dis euh qu'elle dit au début elle le redit aussi à la fin alors ça fait comme chez nous quand on sait pas quoi dire on fait euh ou ah (Ciekanski, 2005 : 635)

Ces extraits n'ont pas valeur de preuve et il serait intéressant de mener une étude sur ce qui peut soit faciliter soit gêner ce travail de repérage, mais il est clair, et c'est ce qui ressort de mon expérience personnelle de conseillère et d'enseignante, ainsi que celles des conseillers du SAAS, que les apprenants sont aptes à repérer et à interpréter efficacement les comportements langagiers qui leur sont donnés d'observer¹³. Or l'observation de documents authentiques constitue, tant pas la méthodologie active qu'elle implique que par l'objet à observer, une activité de réflexion sur la langue indispensable à l'apprentissage.

2. Les documents authentiques : une alternative aux documents pédagogiques

L'utilisation de documents authentiques dans les activités de découverte présente plusieurs avantages. Cela permet tout d'abord à l'apprenant de s'impliquer de façon active. Le travail de recherche et d'observation constitue en outre un savoir-faire méthodologique auquel il peut avoir recours tout le long de son apprentissage. C'est une démarche qui, en outre, respecte le caractère individuel des processus d'apprentissage : stratégies personnelles de prises d'indices, de construction et de vérification d'hypothèses. Enfin, le matériel langagier au sens le plus large proposé ainsi aux apprenants, constitue une alternative à l'indigence mainte fois décriée des

12. Les extraits sont reproduits selon les conventions de transcription de l'auteur.

13. Je renvoie à Debaisieux (2008) pour une brève description d'une expérience menée auprès d'apprenants espagnols.

matériaux pédagogiques qui, comme le signale Carton (1995 : 73), sont « élaborés dans une langue qui n'est ni vraiment de l'écrit, ni vraiment de l'oral et qui comportent des simplifications, voire des invraisemblances discursives. »

Ces derniers présentent en outre une partition radicale entre d'une part les présentations grammaticales dont Firth (1957) déplorait déjà le caractère artificiel¹⁴ et les dialogues censés représenter le fonctionnement de la langue en interaction. Cette séparation a des implications négatives sur l'apprentissage. Elle favorise notamment des représentations langagières – la langue sous forme d'un gros sac de mots et de listes de règles – qui aboutissent à des pratiques d'apprentissage peu efficaces, comme le dénonce explicitement Lightbown (1998), cité par Lyster (2004 : 324) :

Les leçons de grammaire décontextualisées... favorisent l'apprentissage et la catégorisation des formes plus que le lien entre ces formes et leur fonction communicative... Les faits langagiers appris de façon isolée dans des leçons de grammaire peuvent être disponibles en mémoire dans des contextes similaires (par exemple pendant un test de grammaire), mais sont difficilement restitués dans le contexte d'une interaction communicative.¹⁵

Il est donc important de revendiquer de nouveau le caractère indispensable des documents authentiques en tant qu'outils de découverte sur le fonctionnement de la langue, d'autant que les nouvelles technologies permettent aujourd'hui d'y avoir un accès à la fois plus aisé et plus systématique, au travers de la constitution de bases de données, comme je vais le montrer dans la section suivante.

3. Des documents authentiques aux corpus oraux

Le terme *corpus* semble porteur dans un certain nombre de domaines en Sciences humaines. En didactique des langues, la situation est tout autre et l'idée de recourir au corpus au sens moderne du terme, c'est-à-dire « une collection informatisée de textes authentiques (y compris de transcriptions de langue parlée) échantillonnée pour être représentative d'un genre particulier ou de variations langagières »¹⁶ (McEnery *et al.* 2006 : 5) pour l'apprentissage des langues, ne semble pas avoir trouvé d'écho en France. Les raisons de ce désintérêt sont multiples. La situation même des corpus oraux en France, dispersés, de petite taille et non accessibles¹⁷, fait que la majorité de la communauté des didacticiens les

14. « Many of the examples found in grammar books... are 'just nonsense' from a semantic point of view. »

15. « Decontextualised grammar lessons... emphasise the learning and categorising of form rather than relating these forms to their communicative functions... Language features learned in isolated grammar lessons may be remembered in similar contexts (e.g. during a grammar test) but hard to retrieve in the context of communicative interaction. »

16. « A collection of (1) *machine-readable* (2) *authentic* texts (including transcripts of spoken data) which is (3) *sampled* to be (4) *representative* of a particular language or language variety. »

17. Le projet RFP présenté par Detey *et al.* dans ce volume constitue une exception.

ignore. Les représentations sur la langue parlée comme lieu de « déviance » n'ont pas disparues des esprits alors même que de plus en plus de linguistes s'y intéressent. En outre il n'existe pratiquement aucun outil de consultation convivial, adapté à un usage d'apprentissage et surtout libre d'utilisation. En bref, tout ou presque est à construire. Or ces données apportent un éclairage nouveau sur certaines caractéristiques du discours oral que la didactique ne peut ignorer. Je traiterai tout d'abord dans la section qui suit un aspect qui me paraît important et qui concerne la présence de séquences préformées à l'oral. Je montrerai ensuite comment l'observation des corpus peut aboutir à interroger le sens le plus courant de certaines expressions. Enfin je montrerai comment le travail sur corpus permet d'aborder la dimension textuelle et la variation.

4. Ce qu'apporte l'observation des corpus oraux

4.1. L'importance des liens entre lexique et grammaire

Une des premières dimensions du fonctionnement langagier mis au jour grâce aux études sur corpus est l'importance des corrélations entre lexique et grammaire. Ce lien a été mis en évidence notamment par Sinclair (1991) qui montre que le langage n'est pas organisé selon un choix illimité de combinaisons entre grammaire et lexique – *the open-choice principle* – mais bien selon des corrélations fortes de choix de structures semi-préconstruites, souvent dictées par les genres de textes – *the idiom principle*.

De nombreux auteurs ont insisté sur l'importance de ces séquences préformées. Je reprends ici la définition de Wray qui me paraît suffisamment large pour recouvrir les différentes étiquettes relevées dans la littérature anglo-saxonne :

Une séquence continue ou discontinue de mots ou d'autres éléments qui est ou paraît préfabriquée : c'est-à-dire stockée et restituée de mémoire et non au travers d'une procédure de génération ou d'analyse à partir de la grammaire de la langue.¹⁸ (Wray, 2002 : 9, cité par Adolphs, 2008 : 51)

Ce lien étroit entre grammaire et lexique a longtemps été masqué par la séparation artificielle opérée dans les approches traditionnelles et dont la didactique s'est inspirée. En France, les études sur corpus, bien que limitées, sont arrivées aux mêmes conclusions. Ainsi Blanche-Benveniste (2000), étudiant les emplois de *dont* sur un corpus de 250.000 mots, aboutit aux résultats suivants :

- 63% des occurrences où *dont* est construit par un nom, présentent le substantif *façon* ou *manière* :

18. « A sequence, continuous or discontinuous, of words or other elements, which is or appears to be prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language's grammar. »

- la façon dont on réagit face aux problèmes [auto]¹⁹
- je suis très très content de la manière dont s- s'est passé le concert [grouper]
- 42% des occurrences où *dont* est construit par un verbe, présentent le verbe *parler* ou un synonyme :
 - le monstre dont dont je parlais tout à l'heure [job]

Ce type de régularité n'est visible qu'au travers d'outils tels que les concordanciers, qui permettent d'avoir accès à un grand nombre d'occurrences. On voit là, il me semble, l'intérêt de la lecture verticale, qui, loin de se réduire à une vision fragmentée de la langue, nous en laisse appréhender les régularités. Ainsi, dans un petit échantillon extrait d'une requête effectuée sur *qui* dans un corpus oral, on ne peut manquer de relever l'importance des structures en *il y en a qui* introduisant une sorte de « qualification par l'action » des personnes humaines (figure 2).

N°	Div	Contexte gauche	Occur	Contexte droit
1100	Cachemire	ait arnaquer globalement on savait pas trop où aller à	qui	faire confiance et là complètement par hasard c
32	escalade	euh un peu des placements tordus à chercher beaucoup à	qui	demande une grande lecture de voie - pas une vo
1422	foot1	oui Candela - avec un bon ballon qui était destiné à	qui	frappe de Thuram décidément il est inspiré hein
2954	phonét	ait je segmente pas du tout alors qu'il y en a il en a	qui	avait mais segmenté le truc depuis le début
897	cheval	L1 bon ça il y en a	qui	le font pas il y en a qui le font c'est mais ce
1616	guerre	ils allaient ils allaient chez les gens donc il y en a	qui	qui ramassaient de de l'argent et puis ils alla
898	cheval	L1 bon ça il y en a qui le font pas il y en a	qui	le font c'est mais ce qui est important si tu
150	jardinage	L2 ah oui oui oui enfin il y en a	qui	le font moi moi je m'en fous moi ah ah
1615	guerre	st-ce qui nous manquent encore il y en a qui il y en a	qui	manquaient de chauffage on avait pas on avait p
2306	retraites	ulpabilisent par rapport à leur situation et il y en a	qui	pensent qu'effectivement on est bien dans la fo
811	trombone	que c'est aussi un milieu assez difficile - il y en a	qui	dise entre guillemets qui dise que c'est un mil
1614	guerre	x trucs - qu'est-ce qui nous manquait encore il y en a	qui	il y en a qui manquaient de chauffage on avait
1604	guerre	L1 pour pour servir les gens il y en il y en a	qui	se sont fâché ils auraient voulu plusieurs abat
411	macdol	L2 peut-être il y en a	qui	la garde et euh la seule validité qui - tu voi
416	macdol	L2 ben euh tu sais euh on est pas parti il y en a	qui	sont pas partis en pause pas avant dix heures
1511	assoc	i compris un peu comment comment ça marchait il y en a	qui	sont /plus, un peu/ spécialisés dans la cuisin
3147	vichibure	e j'aurai re- risqué de le recopier alors si il y en a	qui	lise le livre ils pourront préciser hein + euh
2658	informatique	L2 parce qu'il y en a	qui	ont > des graveurs avec leur ordinateur pour de
151	jardinage	i peut-être tort mais j'y fais pas attention il y en a	qui	travaillent avec la lune moi je travaille jama
1601	guerre	s euh elle en donnait un petit bout à chacun il y en a	qui	disait rien qui le prenait et puis c'est tout n
593	Maroc	L3 oui mais il y en a	qui	se disent vrai musulman parce qu'ils pratiquent
152	jardinage	s la lune montante la lune descendante alors il y en a	qui	disent quand la lune est montante faut mettre c
1684	Itall	ns plus ou moins euh plus ou moins positives il y en a	qui	disent bon qu'est-ce qu'ils viennent faire ici
2966	phonét	euh en même temps je me je j- j'imagine que il y en a	qui	vont avoir euh tu vois ça va leur être passé au
169	jardinage	L2 il y en a qui le font avant l'hiver mais il y en a	qui	le font tout de suite au mois de février tu as
2958	phonét	nt bes- enfin + c'est c'est un repère même s'il y en a	qui	/ne, 0/ le font pas euh du tout le bah comme m

Figure 2. Interface du concordancier *Contextes* mis au point par Véronis²⁰, Université de Provence.

Ces séquences sont particulièrement fréquentes dans un oral non formaté par une tradition rhétorique et jouent un rôle dans l'apprentissage, que ce soit en langue étrangère ou en langue première, comme le souligne Mauranen (2004 : 99) :

Il semble que le fait de constituer les matériaux linguistiques en blocs en fonction d'unités récurrentes est une procédure spontanée qui participe de la compétence langagière. Cette procédure construit des représentations dans la mémoire à long terme d'unités séquentielles récurrentes préfabriquées et libère les procédures pour analyser de nouvelles significations apparaissant au fil du discours. En outre cela présente également des avantages au niveau socio-interactionnel puisque l'utilisation

19. Les noms entre crochets renvoient aux éléments du corpus en cours de constitution à Nancy.

20. <http://sites.univ-provence.fr/veronis/logiciels/Contextes/index-fr.html>, page consultée le 2/10/08.

de ces blocs aide le locuteur dans la gestion des situations discursives. En ce sens, il semble que les mêmes procédures spontanées soient en jeu dans l'acquisition d'une langue seconde.²¹

Un exemple très concret de l'intérêt pour l'apprentissage de tels groupements nous est fourni par l'étude de Carroll (1989, citée par Lyster 2004 : 330). L'auteur montre que les difficultés rencontrées par les anglophones dans l'apprentissage des déterminants en français sont liées au fait qu'ils considèrent déterminants et noms comme deux unités sémantiques et prosodiques distinctes, alors que les locuteurs natifs du français acquièrent et traitent déterminants et noms comme des « co-indexed chunks ».

4.2. Un accès aux sens les plus fréquents

Une autre dimension tient au fait que la signification des mots les plus fréquents n'est pas la signification proposée par les dictionnaires ou celles que nous dicte l'intuition. Sinclair (1991 : 112) prend ainsi l'exemple de *back* :

Le sens de *back* comme la partie postérieure du corps humain qui s'étend du cou à la région pelvienne (*Collins English Dictionary* (CED) 2^e édition, 1986, sens 1) n'est pas un sens très commun.²²

Pour le français, l'analyse des sens de la préposition *par* et de la locution *une fois* vont dans le même sens. Une analyse de l'emploi des 661 occurrences de *par* sur un corpus oral de 450.000 milles mots aboutit à la répartition suivante :

- 55% des emplois sont constitués par des groupes figés (*par exemple, par hasard*) et des prépositions composées (*par rapport, des emplois distributif*) ;
- 30% des emplois introduisent un « agent passif » ;
- 5% des emplois sont de type distributif.

En complément de verbe, les valeurs de type *ainsi* comme dans *un bébé qui naît par césarienne* sont six fois plus nombreuses que les valeurs de type locatif qui sont attachées à une série de verbes récurrents, *passer par, partir par*. On relève aussi de nombreux exemples de structures du type [*commencer par* ou *finir par*

21. « It seems that chunking of linguistic material according to repeated sequential units is spontaneous and part of normal language processing. Chunking builds up long-term memory representations of repeated sequential units (such as prefabs), and frees processing capacity for analysing new meanings encountered in ongoing discourse (Ellis, 1996). Moreover, sociointeractional benefits also appear to follow from using prefabs in helping speakers manage discourse situations (Wray & Perkins 2000; Wray 2000). In these respects it seems that the same spontaneous processes take place in both first and second language acquisition. » (Mauranen, 2004 : 99)

22. « The meaning of 'back' as 'the posterior part of the human body, extending from the neck to the pelvis' (*Collins English Dictionary* [CED] 2nd edition, 1986, sense 1) is not a very common meaning. »

+infinitif] qui fonctionnent comme des marques d'aspect et dans lesquelles *par* a un sens très différent de borne et non de repère sur un trajet.

Les emplois de *une fois* sur un corpus oral divers de 300.000 mots se répartissent ainsi :

- 46% des occurrences concernent l'expression d'un accompli dans la présentation d'une succession d'actions²³ :
 - ***une fois*** *que les travaux sont commencés moi mon travail c'est de euh de veiller à ce que tout ce passe bien* [armée]
 - ***une fois*** *l'interview faite on fait la retranscription* [journaliste] ;
- 21% des occurrences présentent l'expression *encore une fois* ;
- pour les 33% restants, l'élément est le plus souvent en tête et a le sens de *un jour* et a pour fonction d'introduire une séquence narrative, comme dans l'exemple suivant :
 - *une fois on s'est levé un matin il avait sculpté des épées dans des fourreaux en plastique parce que voilà cadeau quoi ça je m'en rappellerai toujours c'était super* [Italie]

4.3. Un accès à la dimension textuelle

L'observation d'un corpus ne se réduit pas à une approche quantitative à contexte limité. Cette dernière ne permet pas de saisir ce qui constitue les différents paramètres en jeu dans les comportements langagiers attestés. Il est nécessaire d'ancrer le langagier dans un contexte plus large, incluant notamment les aspects séquentiels et prosodiques. C'est par exemple ce que permet le concordancier de Véronis (figure 3) : en un simple clic l'utilisateur passe de l'occurrence à son contexte textuel auquel il a accès par le texte et par le son.

La conception de tels outils adaptés à des requêtes d'apprenant ouvrent des possibilités dont on ignore encore toutes les implications²⁴. Néanmoins, on peut aisément imaginer une interface permettant aux utilisateurs d'avoir accès à des types de séquence dont l'inventaire reste à faire et dont le choix dépend étroitement du public et des objectifs visés. Il s'agit de permettre aux apprenants d'accéder concrètement à la partie la plus structurée de la dimension textuelle, c'est-à-dire à la séquence au sens d'Adam (1990) et qui constitue une possibilité de contourner l'inhérente hétérogénéité des types de texte dans les productions orales. Comme le souligne Rastier (2001 : n.p.) :

De grandes catégories de l'expression, comme la prose ou l'oral, conduisent à des regroupements oiseux (l'oral, de la brève de comptoir au réquisitoire, n'a évidemment pas plus d'unité que la prose.

23. Cappeau (2003) relève la fréquence de cette construction dans les recettes de cuisine.

24. Bien que la littérature de langue anglaise soit riche dans ce domaine ; cf. notamment Chambers (2005).

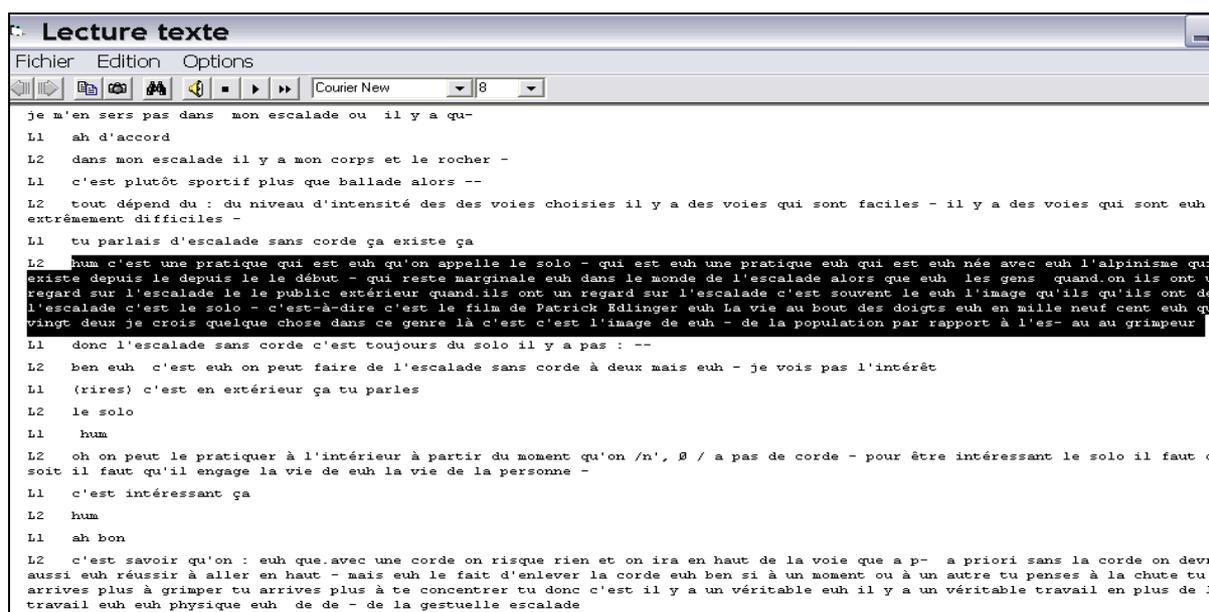


Figure 3. Interface du concordancier *Contextes* mis au point par Véronis, Université de Provence.

Les séquences sont de fait la manifestation « locale » des contraintes liées au genre dont l'importance dans les caractérisations des comportements langagiers a bien été soulignée par Bakhtine (1984 : 285) :

Apprendre à parler c'est apprendre à structurer des énoncés (parce que nous parlons par énoncés et non par propositions isolées et encore moins, bien entendu par mots isolés). Les genres du discours organisent notre parole de la même façon que l'organisent les formes grammaticales.

On peut donc penser des activités de découverte selon une double perspective : l'observation de micro-régularités au niveau des *chunks*, et l'observation de la façon dont ces régularités se distribuent selon les genres. On touche par là la dimension discursive que la didactique de la langue parlée n'a cessé de viser.²⁵

Bien que le travail d'analyse reste en grande partie à faire, notamment sur les productions orales en français, certaines tendances semblent se dégager. Ainsi dans l'extrait suivant qui présente le récit d'un incident, on relève un certain nombre de traits caractéristiques de ce que l'on peut nommer à la suite de Blanche-Benveniste (1995), le récit à l'oral.

euh donc une fois **ce que j'ai eu - c'est des jeunes qui** sont venus à ma caisse - alors - **un groupe de trois jeunes qui** sont venus - **je les ai servis** parce qu'il y avait **déjà** du monde en fait quand **il y a** du monde tout le monde dit euh et **moi** en fait quand **il y a** du monde c'est là que **je suis** calme - ... euh donc ce qui fait que **les jeunes ils** sont arrivés **ils** ont pris des grands menus - les grands menus Big Mac avec des grandes frites des grandes boissons et en plus de ça **ils** ont pris euh des glaces - donc trois glaces donc une glace pour chacun - alors donc **moi** toute speed euh **je** viens **je** dépose des trucs - et **ils** me disent non non **on veut** trois sacs différents - donc un sac pour chacun - bon moi **j'ai** fait confiance **je mettais** les trucs euh donc **j'ai** séparé - et au moment où **je suis** allée f- faire les glaces - alors **je**

25. Voir sur ce point l'article de Chambers dans ce volume.

faisais les glaces vite faites ta-ta-tac-tac et tout - **je reviens** avec les trois glaces **il y a plus** personne [MacDo1]

La première caractéristique touche l'organisation de l'information fournie qui suit strictement la chronologie des événements. La seconde est constituée par la fréquence élevée de constructions verbales simples : *ils ont pris des glaces, ils me disent, je viens, je dépose des trucs, je mettais les trucs, j'ai séparé, je faisais les glaces, je reviens*. La troisième concerne la nature des formes sujets. On relève qu'il n'y a aucun sujet de type lexical, mais des formes pronominales ou des formes à dispositifs : *c'est des jeunes qui, un groupe de jeunes qui, les jeunes ils, il y a du monde*.

Les mêmes tendances semblent se retrouver dans les extraits de conversation, comme le suivant :

- L1 non mais euh renseigne toi sur les bourses étudiantes par contre fais les
L2 mais non **j'ai pas** le droit moi je suis d'ici
L1 oui mais euh **tu les fait** au mois de au mois de avril
L2 avril
L1 **tu vas** sur le site du CROUS **tu** t'inscris <
L2 ah oui pour l'année prochaine >
L1 pour l'année prochaine
L2 oui
L1 **tu fais** une simulation < et **tu**
L2 ah oui > non mais là là **j'aurai** droit l'année prochaine c'est sûr
L1 les bourses étudiantes
L2 ah oui oui c'est sûr que **j'aurai droit** de toute façon
L1 **c'est pas** fait pareil **c'est** pour ça
L2 oui oui **c'est** sûr [entre amis]

L'analyse systématique de ce type de phénomène reste bien sûr à faire. Néanmoins, je formulerai deux remarques à propos de ces grandes tendances : elles correspondent à certains faits relevés par Klein (1989 : 111) et qu'il nomme « les principes pragmatiques qui semblent à l'œuvre dans les tentatives précoces des apprenants pour construire des énoncés ». Elles recoupent d'autre part l'hypothèse de Du Bois (2003 : 33) dite de la « preferred argument structure » selon laquelle « dans les discours spontanés, certaines configurations d'arguments sont systématiquement préférées à d'autres alternatives grammaticalement possibles »²⁶.

Concrètement, les productions orales spontanées montrent la tendance suivante : les constructions verbales n'y présentent qu'un seul argument lexicalisé et ce dernier occupe le plus souvent la fonction d'objet. Cette organisation grammaticale est pour l'auteur le résultat de contraintes pragmatiques touchant la double tendance à ne présenter qu'une seule information nouvelle par bloc verbal et à ne pas la présenter en premier. La séquence suivante, que l'on peut qualifier de séquence explicative, présente une structure différente :

ben **il faut quand même se renseigner** un petit peu **c'est pareil quand que** tu achètes un **que tu achètes un chien ou que tu achètes un poisson -faut se**

26. « In spontaneous discourse, certain configurations of arguments are systematically preferred over other grammatically possible alternatives. »

renseigner parce que si tu achètes un un : je sais pas moi moi un - gros chien et que tu vis euh **dans euh vingt mètres carrés** il va devenir complètement dingue **c'est exactement pour des poissons c'est exactement pareil pour des poissons tu vas pas mettre : c'est pareil un chien et un chat dans vingt mètres carrés - c'est exactement pareil pour des poissons - tu vas pas mettre un Piranha** avec euh je sais pas par exemple des Néons ça va faire **un très mauvais très mauvais mélange donc il faut euh il faut se renseigner** euh il faut regarder euh c'est comme tout quoi [aquariophile]

La séquence se caractérise par un nombre important de structures que l'on pourrait appeler complexes : *quand tu... que tu... faut se... / parce que si tu... tu vas*. Mais on relève par ailleurs de nombreuses répétitions avec mise en parallèle de structures syntaxiques de type topic-comment et contraste d'éléments lexicaux qui compensent cette apparente complexité. La présentation en grille (figure 4)²⁷ permet de mieux saisir l'organisation de ce passage.

		il faut quand même se renseigner un petit peu	
		c'est pareil quand	
que tu achètes un			
que tu achètes un	chien		
ou que tu achètes un	poisson -	faut se renseigner	parce que si tu
achètes un gros chien			
et que tu vis euh dans vingt mètres carrés		il va devenir complètement dingue	
		c'est exactement	
	pour des poissons	c'est exactement pareil	
	pour des poissons	tu vas pas mettre	
		c'est pareil	
		un chien et un chat dans vingt mètres carrés	
	tu vas pas mettre un Piranha	avec des Néons	
		ça va faire un très mauvais	
		très mauvais mélange	
donc		il faut euh	
		il faut se renseigner	

Figure 4. Extrait de la séquence explicative.

La visualisation qu'apporte l'utilisation de la dimension verticale met en évidence les répétitions de structures syntaxiques similaires et le grand nombre de répétitions lexicales. Ce qui me paraît extrêmement intéressant pour le didacticien, c'est le fait qu'une telle structuration semble se retrouver dans de nombreuses langues. Comme le note Leech (2000 : 714-715) :

27. La présentation consiste à présenter verticalement les éléments qui occupent la même fonction syntaxique, mais permet également d'entasser les énoncés qui présentent le même type de structures. Voir sur ce sujet, les études du Gars, notamment Bilger (1982) et Blanche-Benveniste (1997).

Le fait que des traits caractéristiques de la grammaire orale se retrouvent dans des langues parlées différentes est un avantage pour les apprenants. Cela suggère que les locuteurs non natifs d'une langue peuvent adapter des stratégies universelles de la langue parlée à la grammaire de la langue cible.²⁸

On peut en conclure que l'observation de telles régularités est d'un intérêt certain pour l'apprenant. Du moins on peut affirmer qu'il y trouvera des fonctionnements langagiers plus proches de sa propre pratique en langue maternelle.

4.4. Une confrontation à la variation

La troisième dimension apportée par l'accès au corpus concerne la variation. On pourrait estimer que le recours au corpus renforce le modèle du locuteur natif par rapport auquel il est de bon ton aujourd'hui de prendre du recul. En fait, c'est l'inverse qui semble se produire. D'une part la diversité des productions va à l'encontre d'un modèle unique et permet aux apprenants de constater à quel point les productions effectives sont diversifiées par rapport au modèle singulièrement appauvri des méthodes de langue ou des présentations grammaticales traditionnelles. Cette diversité joue à deux niveaux : elle interdit une stricte identification et permet au contraire la construction de modèles propres. Comme le souligne Bernardini (2004 : 24) :

Au travers de l'analyse de données authentiques massives, les apprenants construisent leur propre interprétation en s'appuyant sur leurs observations au niveau lexical, syntaxique, sémantique et pragmatique.²⁹

Elle transforme aussi le statut de la production « native ». Alors que la pauvreté des modèles proposés par les documents pédagogiques poussent à l'imitation, la richesse et la diversité des corpus impliquent recherche et sélection. Ce travail d'observation sur des données hétérogènes, qui par certains aspects sont proches des productions de l'apprenant, comme par exemple les traces de la mise en mots pointée par l'apprenante (A151, section 1.3 *supra*), contribue plutôt à une démythification du locuteur natif. C'est le point de vue également de Mparutsa *et al.* (1991 : 130) cités par Bernardini (2004) :

L'utilisation d'un concordancier remet en question le rôle du texte choisi dans la procédure d'apprentissage. Le texte ne constitue plus une inviolable autorité, mais un objet que les étudiants peuvent interroger et explorer.³⁰

28. « It is also a comfort to the learner that features characteristic of spoken grammar tend to be found in the spoken variety of different languages. This suggests that NNS learners of a language may be able to adapt universal strategies for handling speech for use with the target-language grammar. »

29. « As a result of analysing large quantities of authentic data..., learners construct an interpretation for themselves, relying on lexico-syntactic, semantic and pragmatic observations. »

Du point de vue de l'apprentissage, seule dimension sur laquelle le didacticien peut avoir prise, c'est justement la confrontation à cette diversité qui permet la création d'hypothèses, de construction de règles de la part de l'apprenant et favorise donc la capacité à « créer des discours » (Holec : 1990). Comme le rappellent McEnery *et al.* (2006 : 197) :

C'est précisément parce qu'ils n'offrent pas un modèle simple à imiter que les données de corpus présentent un intérêt pour les apprenants. Les données doivent être interprétées de façon subjective. Pour les apprenants, la réalité des données de corpus réside essentiellement dans la possibilité qu'elles leur offrent d'être interprétées et de créer leurs propres modèles.³¹

Les possibilités de constitution de données alignées texte et son et l'accès libre à des outils de requête permettent aujourd'hui d'envisager cette mise à disposition. Mais il ne s'agit pas d'un simple changement quantitatif. L'objectif est de faire coïncider dans une même ressource un ensemble de documents authentiques, supports d'activités communicatives et outils facilitant un travail de recherche dans cet ensemble. On peut espérer remédier ainsi à l'aspect schizophrénique du traitement de la production orale en didactique des langues qui consiste dans le meilleur des cas, lorsqu'on utilise des documents authentiques, à juxtaposer des productions langagières attestées à des précis grammaticaux qui s'inspirent strictement de la grammaire traditionnelle donc du modèle de l'écrit, comme le dénonçait Halliday (1985 : 97) : « Ce que l'on a nommé 'la grammaire traditionnelle' et qui a été repris par le système éducatif était une théorie de la langue écrite. »³²

C'est dans cet esprit que le CRAPEL s'est lancé dans un projet de constitution d'une base de ressources en français destinée à des étudiants étrangers désirant améliorer leur français universitaire avant ou pendant un séjour dans une université française. Je rappelle ici les grandes lignes de ce projet³³ dont l'objectif est de faciliter l'intégration de étudiants dans leur université d'accueil.

5. Le projet FLEURON (*Français langue étrangère universitaire : ressources et outils numériques*)

Le projet est parti du constat qu'il n'existe pratiquement aucune ressource répondant aux demandes du public particulièrement hétérogène constitué par les

30. « The experience of using the concordancer... challenges the role of a set text in the learning process. The text shifts from being an inviolable authority to something which students can question, explore. »

31. « It is precisely because they do not simply offer models to imitate, however, that corpus data seems valuable for learners... Data from corpora has to be interpreted subjectively... For learners, the reality of corpus data would seem principally to lie in the extent to which they can interpret them to create models of their own. »

32. « The so-called 'traditional grammar' that came into the 'grammar school' was a theory of written language. »

33. Le projet a démarré en 2005 et a fait l'objet d'une présentation dans Debaisieux (2008). Par la suite, le projet a dû être abandonné pour des raisons internes. Il reprend depuis quelques mois.

étudiants étrangers souhaitant se rendre dans une université française. Nous avons donc opté pour la constitution d'un site jouant l'interface entre les apprenants et différents types de données. A l'instar de tous les dispositifs de formation conçus par le CRAPEL, les ressources ne tendent pas à l'exhaustivité, mais sont représentatives des situations de communication les plus fréquentes auxquelles les étudiants sont susceptibles d'être confrontés, sachant que cette représentativité est toujours problématique³⁴. Le dispositif est ouvert vers d'autres ressources et vise à développer chez les utilisateurs leur capacité à travailler simultanément leur capacité langagière et leur capacité d'apprentissage, c'est-à-dire à utiliser toutes les ressources possibles de leur environnement comme moyens d'apprentissage. Les ressources se répartissent donc entre ressources en langue et ressources pour apprendre. Pour les premières, le choix a été fait d'une représentation du français académique au sens large. Sont donc prévus : des données orales de la vie académique (cours, séminaires, conférences, examens oraux, soutenances, interactions entre enseignants et étudiants, entre étudiants et différents services universitaires, interactions entre étudiants en groupe de travail, etc.), et des données de langue « non académique » (conversations, vie en dehors du campus, activités de loisirs, etc.). A cela se rajoutent des témoignages d'étudiants étrangers et d'étudiants français sur la vie à l'université. Ces données ont pour objectif de prendre en compte de façon explicite la dimension interculturelle, mais aussi d'offrir des échantillons de non-natifs compétents dont les stratégies interactionnelles constituent un objet d'observation intéressant pour d'autres apprenants.

Les ressources en langue seront organisées selon trois modalités : il s'agit d'une part de fournir des documents authentiques portant sur les situations spécifiées et pouvant donner lieu à des activités d'apprentissage concernant la compréhension et l'expression. Les documents étant accompagnés de fiches descriptives précises, ils pourront être sélectionnés en fonction de critères divers selon un système de catalogage informatique. Par ailleurs ces documents viendront enrichir des données tout venant d'environ 400.000 mots regroupant des conversations courantes, des entretiens, des récits constitués en archive orale et qui pourront faire l'objet de requêtes au travers d'un concordancier. A cela devrait se rajouter une série de modules regroupant de façon systématique certains types de séquences illustrant des macro-actes dont la liste et les caractéristiques sont encore à l'étude.

Il s'agit, on le voit de concilier, dans une tentative de cohérence maximale, approche communicative et approche *form-focused* et de faire en sorte que ce travail d'observation des formes prenne appui sur des comportements langagiers effectifs et non sur les ersatz habituellement convoqués dans les précis grammaticaux.

34. Il me paraît intéressant de noter que cette question était déjà abordée dans les premiers travaux du CRAPEL. C'est ainsi que Pieron (1978 : 107) signale, à propos des documents authentiques oraux qu'il a rassemblés en vue de la formation de travailleurs étrangers : « on peut bien sûr interroger la représentativité du corpus ainsi constitué. »

Les ressources pour apprendre recouvrent ce que nous appelons traditionnellement les « documents d'aide à l'apprentissage³⁵ » et qui concernent en particulier les interrogations méthodologiques telles que : « Quoi et comment apprendre ? », « Que faire avec un document ? », « Comment évaluer ses progrès ? » etc.

Une deuxième série de documents d'aide à l'apprentissage concerne l'utilisation du corpus tout venant : comment formuler une requête ? Comment observer une requête ? Quels sont les éléments qui paraissent pertinents pour le sens, la construction ? etc.

Pour ne pas conclure

Les difficultés à mener un tel projet ne doivent pas être minimisées, nous en avons d'ailleurs fait l'expérience directe. La première réside dans l'ignorance de la problématique qui semble caractériser la didactique du FLE en France. De nombreuses réticences se font sentir, liées sans doute au fait que les didacticiens « échaudés » semblent convaincus qu'ils se trouvent là devant un nouvel artefact de linguiste ou, ce qui revient au même, devant un retour d'une certaine linguistique appliquée. Il serait bon de pouvoir dans ce domaine renouer un dialogue entre didacticiens et linguistes de la parole. D'autres difficultés sont liées à l'absence de volonté institutionnelle de développer en France les recherches sur corpus oraux, ou du moins de les soutenir logiquement. Or le travail est long, coûteux, non reconnu.

Je ne reviendrai pas ici sur la nécessaire transformation des représentations et de la langue parlée et du processus d'apprentissage qu'implique le recours à une telle démarche. J'espère néanmoins avoir su montrer comment l'accès direct aux corpus peut concourir à l'autonomisation des apprenants. Je reprend à mon compte cette affirmation d'Aston (2001 : 41) : « L'aspect le plus attractif des corpus pour l'apprentissage des langues réside peut-être dans les possibilités qu'ils offrent pour un apprentissage autonome. »³⁶

On pourrait ainsi « rêver » plus loin et penser que la fréquentation de tels outils est susceptible d'inciter les apprenants à construire leur propre corpus³⁷, tout comme les traducteurs constituent leur propre dictionnaire et à l'image de ce que contait l'écrivain grecque de langue française Alexakis lors d'une conférence à Thessalonique en septembre 2007 :

Quand je suis arrivé en France, je me promenais dans la rue en enregistrant à la sauvette ce que les gens disaient, et de retour en Grèce après 20 ans d'absence, j'ai dû faire la même chose.

35. Voir les *Mélanges Pédagogiques* n°22 pour une description de ce type de matériel.

36. « Perhaps the greatest attraction of corpora for language pedagogy is their potential for autonomous learning. »

37. Voir sur ce point l'article de Tyne dans ce volume.

BIBLIOGRAPHIE

- ABE, Danielle & Marie-José GREMMO, 1981. « Apprentissage auto-dirigé : quand les chiffres parlent. » *Mélanges pédagogiques*, p. 4-18.
- ABE, Danielle & Marie-José GREMMO, 1983. « Enseignement/apprentissage : vers une redéfinition du rôle de l'enseignant. » *Mélanges pédagogiques*, p. 103-117.
- ABE, Danielle, Carolyn HENNER-STANCHINA & Odile REGENT, 1978. « Apprentissage de l'expression orale en autonomie : implications de l'approche fonctionnelle. » *Mélanges pédagogiques*, p.1-17.
- ADAM, Jean-Michel, 1990. *Eléments de linguistique textuelle*. Liège : Mardaga.
- ADOLPHS, Svenja, 2008. *Investigating Pragmatic Functions in Spoken Discourse. Studies in Corpus Linguistics*, 30. Amsterdam : John Benjamins.
- ALEXAKIS, Vassilis, 2007. Conférence d'ouverture du 6^e Congrès Panhellénique des professeurs de français. Thessalonique, 22-24 septembre.
- ASTON, Guy, 2001. *Learning with Corpora*. Houston : Athelstan.
- BAKHTINE, M. 1984. *Esthétique de la création verbale*. Paris : Gallimard.
- BERNARDINI, Silvia, 2004. « Corpora in the classroom : an overview and some reflections on future developments. » In J. Sinclair (ed.), *How to Use Corpora in Language Teaching*. Amsterdam : John Benjamins, p. 15-36.
- BILGER, Mireille, 1982. « Contribution à l'analyse en grilles. » *Recherches sur le français parlé*, 4. Université de Provence, p. 195-215.
- BLANCHE-BENVENISTE, Claire, 1995. « De la rareté de certains phénomènes syntaxiques en français parlé. » *Journal of French Language Studies*, 5, p. 17-29.
- BLANCHE-BENVENISTE, Claire, 1997. « De l'utilité du corpus linguistique. » *Corpus : de leur constitution à leur exploitation. Revue française de linguistique appliquée*, 1/2, p. 25-42.
- BLANCHE-BENVENISTE, Claire, 2000. « Convergences de matériel grammatical permettant d'établir des typologies textuelles. » In M. Bilger (éd.), *Linguistique sur corpus : Etudes et réflexions. Cahiers de l'université de Perpignan*, 31. Presses Universitaires de Perpignan, p. 103-116.
- CAPPEAU, Paul, 2003. « Vous mélangez tout ? De l'ordre dans les recettes. » In B. Combettes, C. Schnedecker & A. Theissen (éds.), *Ordre et distinction dans la langue et le discours*. Paris : Champion, p. 87-100.
- CARROLL, S. 1989. « Input and SLA: adults' sensitivity to different sorts of cues to French gender. » *Language Learning*, 49, p. 37- 52.

CARTON, Francis, 1995. « L'apprentissage différencié des quatre aptitudes. » *Verbum*, 18/1. Nancy : Presses Universitaires de Nancy, p. 63-73.

CEMBALO, Sam Michel & Henri HOLEC, 1973. « Les langues aux adultes : pour une pédagogie de l'autonomie. » *Mélanges pédagogiques*, p.1-10.

CHAMBERS, Angela, 2005. « Integrating corpus consultation in language studies. » *Language Learning & Technology*, 9/2, p. 111-125.

CIEKANSKI, Maud, 2005. *L'accompagnement à l'autoformation en langue étrangère : contribution à l'analyse des pratiques professionnelles*. Université Nancy 2 : Thèse de doctorat.

CRAPEL & EAEHT, 1998-1999. *Vacances cubaines. Tome 1 : fonctions d'animation ; Tome 2 : relations avec la clientèle ; Tome 3 : fonctions de service*. La Habana : La Editorial Academia.

CRAPEL & RECAFTUR, 2006. *Vacances en Amérique centrale. Tome 1 : service au client et vente de prestations ; Tome 2 : gestion de groupes et animation*. San José, Costa Rica : Ediciones Perro Azul.

DEBAISIEUX, Jeanne-Marie, 2008. « Corpus oraux et didactique des langues : un rendez-vous à ne pas manquer. » *Le français dans le monde : recherches et applications. Quel oral enseigner, cinquante ans après le Français fondamental ?* Paris : CLE international FIPF, p. 102-114.

DU BOIS, John, Lorraine KUMPF & William ASHBY (eds.) 2003. *Preferred Argument Structure*. Amsterdam : John Benjamins.

FIRTH, J.R. 1957. « A synopsis of linguistic theory, 1930-1955. » *Studies in Linguistic Analysis*. Philological Society, p. 1-32.

GREMMO, Marie-José, 1997. « Matériaux d'enseignement et apprentissage de langue : les options didactiques des 'modules de français pour les professionnels du tourisme'. » *Mélanges pédagogiques*, p. 63-79.

GREMMO, Marie-José, 2000. « Autodirection et innovation : raisons d'être d'un réseau européen. » *Mélanges pédagogiques*, p. 13-27.

HALLIDAY, M.A.K., 1985. *Spoken and Written Language*. Oxford: Oxford University Press.

HOLEC, Henri, 1982. « L'approche communicative, cru 1982. » *Mélanges pédagogiques*, p. 65-79.

HOLEC, Henri 1990. « Des documents authentiques, pour quoi faire ? » *Mélanges pédagogiques*, p. 65-74.

KLEIN, Wolfgang, 1989. *L'acquisition de langue étrangère*. Paris : Colin.

- LEECH, Geoffrey, 2000. « Grammars of spoken English : new outcomes of corpus-oriented research. » *Language Learning*, 50/4, p. 675-724.
- LIGHTBOWN, Patsy, 1998. « The importance of timing in focus on form. » In C. Doughty & J. Williams (eds.), *Focus on Form in Classroom Second Language Acquisition*. Cambridge : Cambridge University Press, p. 177-196.
- LYSTER, R. 2004. « Research on form-focused instruction in immersion classrooms: implication for theory and practice. » *Journal of French Language Studies*, 14/3, p. 321-341.
- MAURANEN, Anna, 2004. « Spoken corpus for an ordinary learner. » In J. Sinclair (ed.), *How to Use Corpora in Language Teaching*. Amsterdam : John Benjamins, p. 89-105.
- McENERY, Tony, Richard XIAO & Yoko TONO, 2006. *Corpus-Based Language Studies*. London : Routledge.
- MPARUTSA, C., A. LOVE & A. MORISSON, 1991. « Bringing concord to the ESP classroom. » In T. Johns & P. King (eds.), *Classroom Concordancing. ELR Journal*, 4, p. 115-134.
- PEREZ-PAREDES, Pascual, 2004. « Learner oral corpora and network-based language teaching. » In J. Sinclair (ed.), *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, p. 249-268.
- PIERON, C. 1978. « Français fonctionnel et travailleurs étrangers : expériences avec des débutants. » *Mélanges pédagogiques*, p. 105-141.
- RASTIER, François, 2001. « Eléments de théorie des genres. » *Texto !* http://www.revue-texto.net/1996-2007/Inedits/Rastier/Rastier_Elements.html, page consultée le 25/11/00.
- SCHMIDT, R. 1990. « The role of consciousness in second language learning. » *Applied Linguistics*, 11, p. 17-46.
- SHIFFRIN, R. & W. SCHNEIDER, 1977. « Controlled and automatic human information processing: Perceptual learning, automatic, attending and a general theory. » *Psychological Review*, 84, p. 127-190.
- SINCLAIR, John, 1991. *Corpus, Concordance, Collocation*. Oxford : Oxford University Press.
- WRAY, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge : Cambridge University Press.