Alors la question c'est...?
Questions pragmatiques et annotation pédagogique des corpus

Résumé

Dans le domaine de la didactique, les corpus oraux garantissent l'observation de conduites langagières dans des situations socialement et communicativement caractérisées et fournissent ainsi un outil d'exposition à la langue et à son fonctionnement indispensable à un apprentissage effectif. Nous développerons ici un exemple. Alors que les outils didactiques rattachent le plus souvent trois types de structures (inversion, morphème interrogatif et intonation) à la fonction «demande d'information», les corpus révèlent pour cette même fonction une grande diversité de formes, parmi lesquelles on peut relever des formes de reformulation et d'appel à inférences telles que:

1. c'est donc plutôt sportif que balade alors
2. donc toi tu es quand même plutôt corde alors

Bien que celles-ci ne relèvent pas formellement de la question, elles appellent chez l'interlocuteur une réaction en forme de réponse.

Ce type d'exemple montre comment le recours au corpus, loin de renforcer le mythe du locuteur natif idéal, permet à l'apprenant d'être confronté à la variation des productions langagières et ainsi de sélectionner celles qui lui paraissent les plus adéquates à sa personnalité et à la situation de communication. Ils soulignent également la nécessité d'introduire dans les corpus mis à disposition des apprenants, un marquage pragmatique des expressions. La communication fournira des pistes concrètes en ce sens.

1. Introduction

La description des langues rendue possible par l'étude des corpus a bouleversé bon nombre d'intuitions des locuteurs natifs. Le domaine le plus touché est certainement celui de la lexicographie, mais les corpus se révèlent indispensables pour tout travail descriptif de nos jours. En didactique des langues étrangères, les outils descriptifs (les dictionnaires monolingues et bilingues, les grammaires, les manuels d'usage…) ont subi une véritable révolution grâce aux informations tirées des corpus, mais les auteurs des outils pédagogiques (manuels de classe et autres supports) sont souvent plus lents à tirer profit des recherches en amont. Pour cette raison, depuis des années de nombreux chercheurs signalent un décalage important entre la langue présentée dans les manuels de cours et la langue «authentique» telle qu'ils la découvrent dans les corpus, surtout pour la langue parlée; très influents dans ce domaine sont les travaux de Carter (1998) et de Carter et McCarthy (1995).

Le problème de ce décalage serait d'autant plus sérieux en français, où les études de corpus sont moins avancées que dans bien d'autres langues (Debaisieux, 2000, 2005b). Particulièrement défaillante serait la description du langage parlé: jusqu'à ces derniers temps, «pour une bonne partie des linguistes, le français parlé n'apparaissait pas comme un objet d'étude légitime» (Blanche-Benveniste, 2000: 15). Une explication partielle réside sans doute dans les représentations répandues sur le langage de prestige, des idées bien ancrées sur ce qui est *correct*, *juste* ou *bon*, par opposition à ce qui est *incorrect*, *faux* ou *mauvais*. En effet, pourquoi décrire la réalité de l'usage langagier puisque le rôle de l'Académie Française est de «maintenir les qualités»? Elle «définit le bon usage» et «fixe l'usage de la langue» tout en critiquant d'autres formes.[1] Le Secrétaire perpétuel à l'époque, l'Immortel Maurice Druon (1999), nous met en garde de:

> [ramasser] dans le ruisseau tous les détritus tombés des poubelles langagières. Mais comme ces horreurs sont «dans un dictionnaire» chacun peut se croire encouragé à les utiliser… Voilà ce que l'on nous affirme être «le vrai français»… Cette volonté délibérée de «faire populaire», cette démagogie lexicale, ce gauchisme du vocabulaire sont, comme tous les gauchismes, un péché de l'esprit.

Chacun a droit à ses opinions. Toutefois, en tant qu'enseignants de langue, nous sommes en droit de nous poser des questions lorsque ces idées reçues se reflètent dans les manuels d'apprentissage de langue étrangère. L'apprenant doit-il apprendre une norme artificielle ou un langage qui se rapproche plus de la réalité?

Un décalage dans les manuels antérieurs à la révolution de corpus n'est guère surprenant: l'intuition et l'expérience étaient les moyens privilégiés pour établir le contenu des programmes de langues. Mais ce décalage persiste: pour Mauranen (2004a: 97), «descriptions emerging from… corpora have not found their way to teaching materials on a large scale yet» (voir aussi le nombre d'articles qui traitent ce point dans Hidalgo *et al*., 2007). On peut alors penser qu'il s'agit d'un choix délibéré, peut-être dans un but de simplification. Ce serait certainement compréhensible: même les matériels COBUILD sélectionnent les exemples les plus révélateurs dans ce but (Willis, 1991: vii). Pour que cet argument tienne, il faudrait démontrer que le langage présenté est effectivement plus «simple» que celui représenté dans les corpus.

Pour ce faire, nous nous focalisons ici sur un seul acte langagier. Il s'agit des questions, ou plus généralement des demandes d'informations, suivant l'intuition de l'un d'entre nous qui,

[1] Extraits du site officiel de l'Académie Française: «Le Rôle». http://www.academie-francaise.fr/role/index.html, page consultée février 2007.

en tant qu'apprenant du français langue étrangère, a le souvenir d'avoir travaillé longuement à l'école sur l'inversion comme forme interrogative alors que cette forme est plutôt rare en conversation quotidienne. Bien entendu, l'intuition peut être trompeuse, surtout en matière du langage parlé (Debaisieux, 2005a: 9). Qu'en est-il réellement du côté des manuels et du côté de la réalité révélée par une étude de corpus? Si décalage il y a, comment les corpus peuvent-ils nous aider à mieux décrire la réalité, à mieux décider quoi enseigner, à mieux apprendre ? Ce sont les questions que nous nous posons dans cet article, qui sert également à présenter un projet pour l'apprentissage du français langue étrangère en ligne destiné à des étudiants étrangers arrivant dans une université française.

2. Questions dans les manuels

Poser des questions est une fonction langagière fondamentale ; dès le début de son apprentissage, l'apprenant au niveau A1, selon le Cadre européen commun de référence pour les langues, doit pouvoir « poser à une personne des questions la concernant… et peut répondre au même type de questions… ».[2] Ces consignes sont interprétées grammaticalement ainsi par le Référentiel niveau A1.1 (Beacco, Ferrari, Lhote et Tagliante, 2005: 116-117):

> Au niveau A.1, **l'utilisateur/apprenant est capable d'utiliser**…
> 5.2.5 Structures interrogatives

GN V […] /Intonation/	*Tu viens?*
Est-ce que + GN + V […]	*Est-ce que tu viens?*
Interrog GN V	*Où tu vas?*
V + GN […] Interrog.	*Vous allez où?*

Il s'agit donc de comprendre mais aussi de produire ces formes grammaticales. Notons en passant qu'aucune mention n'est faite de l'inversion interrogative.

Du côté des manuels, nous avons choisi une méthode très utilisée en Grande Bretagne, la nouvelle édition d'*Encore Tricolore* qui se vante sur la quatrième de couverture de «a clear and comprehensive approach to grammar progression». En effet, chaque unité est organisée grammaticalement, accompagné d'un précis grammatical à la fin du livre. Le niveau 1 (Honnor et Mascie-Taylor, 2000) se destine aux élèves de 11 ans qui commencent à apprendre le français en Year 7, et le précis grammatical () résume les formes de demandes d'informations ainsi (p. 162):

Figure 1. Précis grammatical des questions dans *Encore Tricolore 1*.

[2] http://www.coe.int/T/DG4/Portfolio/?L=F&M=/main_pages/levelsf.html, page consultée janvier 2007.

7.2 Asking questions

There are several ways of asking a question in French.

- You can just raise your voice in a questioning way:
  Tu as des frères et des sœurs? Do you have brothers and sisters?
- You can add *Est-ce que* to the beginning of the sentence:
  Est-ce que tu as un animal? Do you have a pet?
- You can turn the verb around:
  Avez-vous des idées? Do you have any ideas?
  Jouez-vous au badminton? Do you play badminton?
- You can use *Qu'est-ce que (qu')* ...? meaning 'What ...?'.
  Qu'est-ce qu'il fait? What is he doing?
  Qu'est-ce que tu prends comme petit déjeuner? What do you have for breakfast?
  Qu'est-ce que tu aimes comme musique? What kind of music do you like?
- You can use a question word, e.g.
  Qui? Who?
  Où? Where?
  Quand? When?
  Pourquoi? Why?

Tous les exemples relèvent du domaine indiqué par Beacco *et al*. (2005 : 117), à savoir «poser à une personne des questions la concernant». Toutefois, aucune aide n'est fournie concernant la fréquence ou l'emploi des différentes formes, et toutes sont apparemment à maîtriser en production autant qu'en réception. On remarque surtout l'inversion parmi les formes proposées, une forme réputée difficile si on compare avec le précis grammatical du même manuel en deuxième année (Honnor et Mascie-Taylor, 2001: 158). La section *7.2 Asking questions* (Figure 2) est identique en tous points sauf en ce qui concerne l'inversion:

Figure 2. Inversion dans le précis grammatical des questions dans *Encore Tricolore 2*.

- You can turn the verb around:
  Allez-vous à la piscine? Are you going to the pool?
  Jouez-vous au badminton? Do you play badminton?
  Notice that if the verb ends in a vowel in the third person you have to add -t- when you turn it round:
  Joue-t-il au football? Does he play football?
  Lucie, a-t-elle ton adresse? Has Lucy got your address?
  In the perfect tense you only turn the auxiliary verb round:
  Avez-vous vu le film au cinéma Rex? Have you seen the film at the Rex cinema?
  As-tu écrit à Paul? Have you written to Paul?
  Jean et Pierre, sont-ils allés au match hier? Did Jean and Pierre go to the match yesterday?
  Sophie, a-t-elle téléphoné à Nicole? Did Sophie 'phone Nicole?

Tous les exemples relèvent encore de l'oral spontané, ce qui n'est pas surprenant puisque les formes interrogatives sont bien plus courantes en conversation qu'à l'écrit. Inspiré d'un grand

corpus en anglais, la grammaire de Biber, Johansson, Leech, Conrad et Finegan (1999: 211) trouve 23500 questions par million de mots en conversation, 7000 en fiction, et seulement 500 dans les autres registres principaux. Mais nos intuitions sont que ces exemples d'*Encore Tricolore* sonnent faux. Toutefois, rien ne garantit que nos intuitions soient plus fiables que celles des auteures de ces manuels. Pour les tester, nous aurons recours à un corpus.

3. Questions dans un corpus oral

Les corpus que nous analysons ici font partie d'un projet CRAPEL baptisé FLEURON: Français Langue Etrangère Universitaire – Ressources et Outils Numériques. Il s'agit d'un site internet destiné à aider les non francophones qui comptent venir étudier dans une université française (voir Boulton et Debaisieux, 2006). Notre préoccupation principale portant sur la compréhension et l'expression orales,[3] une partie du dispositif comprend un corpus oral enregistré dans des contextes universitaires. Pour l'instant, nous avons enregistré à peu près 75 heures dont la quasi-totalité est déjà transcrite (60 heures) et une trentaine déjà transcrites et alignées texte et son. Le corpus final comprendra une centaine d'heures d'enregistrements, ce qui équivaut à près de 700.000 mots selon les estimations générales de Biber *et al*. (1999: 27).

La taille du corpus permettra une variété suffisante pour répondre aux besoins divers d'utilisateurs différents. Il sera interrogeable en sous-corpus afin d'amoindrir le risque d'inonder l'apprenant avec trop de données (cf. Aston, 1997). Il s'agit donc pour nous d'un corpus «didactique» (Whistle, 1999), c'est-à-dire centré sur les besoins des apprenants plutôt que sur des critères de représentativité à des fins de recherches purement linguistiques (voir les travaux de Braun, par exemple 2007). Nous nous concentrons ici sur deux éléments comportant plus de 7.500 mots:

1. «Scolarité» (SCOL): 14 entretiens avec des étudiants (natifs ainsi que non natifs – locuteurs «experts») qui se présentent à l'accueil du service scolarité à Nancy 2 avec diverses questions (40 minutes).
2. «Escalade» (ESC): une étudiante qui interviewe un étudiant sur sa passion pour l'escalade (25 minutes).

Ces deux genres sont particulièrement riches en questions; en attendant un étiquetage pragmatique (voir Maynard et Leicher, 2007), nous avons parcouru ces deux extraits manuellement afin d'en retirer tous les exemples. Nous commençons avec l'inversion avant de passer aux autres formes proposées dans *Encore Tricolore 1* pour débutants FLE, à savoir

[3] Cf. Beeching (1997: 376): «I would argue that the first need of British students going to France is to *understand the spoken language* in non-academic contexts.»

est-ce que et *qu'est-ce que*, l'intonation, et l'emploi d'un mot-question. Ce survol restera bref; pour un traitement plus détaillé de la question en Français, voir Kerbrat-Orecchioni (1991).

3.1 Inversion

Sans compter *n'est-ce pas*, que nous pouvons considérer comme une forme figée et non pas comme une inversion propre, il n'y a aucun exemple d'inversion dans Escalade, et un seul dans Scolarité. Cette unique occurrence semble particulièrement polie, mais on peut difficilement généraliser à partir d'un seul exemple:

[SCOL] A: **puis-je** aussi vous demander d'ouvrir la porte
N: oui
A: pour que quelqu'un puisse rentrer[4]

Ceci semble confirmer notre intuition que l'inversion est rare au moins dans ces registres oraux, et indique un décalage entre la réalité et les manuels. Puisque, nous l'avons vu, cette forme est syntaxiquement complexe, son inclusion dans un manuel pour débutants est difficile à justifier.

Qu'en est-il des autres formes interrogatives décrites dans *Encore Tricolore*?

3.2 *Est-ce que* et *qu'est-ce que*

Dans son étude sur corpus en 1991, O'Connor di Vito trouve que 99% des questions en français se forment avec *est-ce que* ou intonation seule. Toutefois, *est-ce que* n'est pas très utilisé dans nos corpus: le seul exemple dans Escalade se situe au tout début de l'entretien et a tout l'air d'être une question préparée à l'avance:

[ESC] L1: **est-ce que** tu pourrais m'en dire plus

Est-ce que est certainement utile grâce à la simplicité syntaxique de son emploi, puisqu'il permet de construire un énoncé selon l'ordre SVO et ce, tant pour les questions fermées que pour les questions ouvertes. Néanmoins, il n'est pas très fréquent dans nos corpus et semble caractéristique d'une politesse marquée, comme le signale le fait qu'il est souvent accompagné de formes de politesse marquées: *s'il te plaît*, l'emploi du conditionnel (*pourrais*, *auriez*), une préparation (*je vais vous demander maintenant…*), justification (*parce que*), atténuation (*un peu*, *ou pas*), etc.:

[SCOL] A: [*se tournant vers une collègue*] Aurélie s'il te plait j'ai un problème parce que apo ne veut pas s'ouvrir **est-ce que** tu pourrais imprimer le

[4] Dans le corpus Scolarité, A = personnel administratif ; N = étudiant natif ; NN = étudiant non natif. Dans le corpus Escalade, L1 = intervieweur ; L2 = interviewé.

dossier du deux deux…
[SCOL] A: pas de souci simplement je vais vous demander maintenant **est-ce que** vous vous êtes renseignée sur les nouvelles formations parce que comme on est passé au LMD il y a de nouvelles offres de formation **est-ce que** vous avez vu un peu

Les questions ouvertes commençant par *qu'est-ce que* sont également peu fréquentes : on ne relève aucun exemple dans Escalade, est sur les dix occurrences du corpus Scolarité, neuf sont réalisées dans un énoncé qui semble être une forme quasi-figée dans ce contexte:

[SCOL] A: alors **qu'est-ce que** je peux faire pour vous

Le seul autre emploi de *qu'est-ce que* se trouve dans un énoncé d'une étudiante non native, une forme par ailleurs non canonique:

[SCOL] NN: je sais pas **qu'est-ce que** je dois faire pour mon titre de séjour

Avec si peu d'occurrences d'inversion et de *(qu')est-ce que*, nous devons nous retourner ailleurs pour trouver les formes de questions plus fréquentes.

3.3 Mot-question

Nous trouvons d'autres «mots-question» dans nos corpus. Les douze occurrences dans Escalade s'illustrent à travers les exemples suivants. Le mot-question peut se trouver au début ou en fin d'énoncé, ou bien être employé seul:

[ESC] L1: **pourquoi** tu me dis ça
[ESC] L1: c'est **où**
[ESC] L1: tu as commencé **où**
[ESC] L1: **où**

Quelques exemples de l'emploi d'un mot-question parmi les 25 occurrences dans le corpus Scolarité:

[SCOL] N: et euh **comment** ça se passe au niveau des cours
[SCOL] A: vous êtes de **quelle** nationalité
[SCOL] NN: je l'aurai **quand**

Autre avantage pour les apprenants anglophones c'est qu'il n'est pas inconnu en anglais, même s'il est plus représentatif d'autres registres, par exemple des quiz:

- Hans Christian Andersen, author of *The Princess and the Pea* and *The Ugly Duckling*, was **what nationality**?
- Homer, who wrote *The Iliad* and *The Odyssey* was **of what ethnicity**?
- Henrik Ibsen, who wrote *A Doll's House* and *Hedda Gabler*, had **which national**

origin?[5]

3.4 Intonation montante

Dans le cas des questions fermées, la marque de question consiste simplement à attribuer un schéma intonatif montant à l'énoncé. Il y en a 20 occurrences dans le corpus Escalade, 41 dans le corpus Scolarité, dont voici quelques exemples:

[ESC] L1: et ça t'a plu ↗
[ESC] L1: il y a pas un mec qui fait ça sur un sur les bâtiments ↗
[ESC] L1: il y en a beaucoup ↗
[SCOL] A: vous avez déjà passé les tests ↗
[SCOL] A: je peux peut-être prendre votre photo ↗
[SCOL] NN: demain ↗

Le recours à l'intonation seule est bien plus simple que d'autres formes interrogatives, même si les schémas intonatifs ne sont pas toujours aussi clairs. Ce procédé existe en anglais aussi, ce qui peut aider l'apprenant, mais il est plus marqué, souvent employé en écho et indicatif de surprise ou d'incrédulité. L'apprenant pourrait donc mal interpréter ce genre de question très fréquent s'il n'est pas sensibilisé aux différences, voire ne pas se rendre compte qu'il s'agisse d'une véritable question.

3.5 Résumé

Nous avons analysé la fréquence des différents types de questions présentés dans *Encore Tricolore*, que nous avons résumé dans le Tableau 1:

Bien entendu, lorsqu'il s'agit de décider quelles formes enseigner, la fréquence n'est pas le seul critère. Mais il nous semble indéfendable de continuer à enseigner l'inversion auprès de débutants alors que cette forme est très rare à l'oral spontané, et que d'autres formes sont plus simples voire même transparentes pour les apprenants visés.

Tableau 1. Fréquence des formes dans nos deux extraits de corpus

	ESCALADE	SCOLARITÉ	TOTAL
inversion	0	1	1
est-ce que	1	4	5
qu'est-ce que	0	10	10
mot-question	12	24	36
intonation	20	41	51

[5] Extraits de http://www.funtrivia.com/trivia-quiz/People/Famous-Authors-What-Nationality-32302.html, page consultée février 2007.

4. Autres questions

Les formes que nous avons analysées jusque là sont celles présentées dans la plupart des manuels comme *Encore Tricolore*. Afin de voir s'il existe d'autres possibilités pour demander des informations, il nous faut déjà une définition opérationnelle de ce qu'est qu'une «question».

Selon le Nouveau Petit Robert (Rey-Debove et Rey, 2002), une question est une «demande qu'on adresse à quelqu'un en vue d'apprendre quelque chose de lui»; pour Grévisse (Goosse, 1994: 584), «par la phrase interrogative, on demande une information à l'interlocuteur». Ces définitions à usage généraliste sont très larges; on se tourne alors vers Kerbrat-Orecchioni (2001) pour une définition orientée vers les actes de parole:

> Nous considérerons comme question *tout énoncé qui se présente comme ayant pour finalité principale d'obtenir de son destinataire un apport d'informations*. (p. 86)
> La question est un acte initiatif qui généralement sollicite *fortement* une réaction verbale (ou à la rigueur un substitut gestuel). (p. 92)

Notre analyse se base alors sur tout acte linguistique qui obtient ce qui peut être qualifié d'une «réponse». Cette définition de la question amène à ne pas se limiter aux seuls énoncés habituellement répertoriés dans cette catégorie, mais s'appuie crucialement sur le contexte. Certains des phénomènes que nous repérons sont relativement simples, d'autres plus complexes ; dans la discussion qui suit, nous ne visons donc plus les seuls débutants.

4.1 Inachevées

Les formes de questions que nous avons examinées jusque là ne sont pas les seuls moyens d'obtenir une information. Un autre procédé courant est de commencer un énoncé et de laisser à son interlocuteur de compléter. Les réactions confirment qu'il s'agit bien d'une demande d'information, d'explicitation, souvent renforcée. Dans plusieurs cas on remarque l'emploi de *ou* dans le corpus Escalade et de *euh* ou une syllabe allongée dans le corpus Scolarité:

[ESC] L1: plutôt en salle **ou**…
L2: moi je s- euh me sers du support euh de la structure artificielle d'escalade
[ESC] L1: l'alpinisme c'est quand il y a des – de la glace **ou**…
L2: l'alpinisme c'est euh quoi qu'il arrive on va au sommet
[ESC] L1: tu fais cent vingt mètres en une fois **ou**…
L2: non cent vingt mètres on

[SCOL] N: donc là à l'UFR lettres je demande la fiche péda et puis **euh**…
A: l'emploi de temps de la troisième année de de culture et communication

[SCOL] N: et l'historique alors **euh**…
A: et l'historique on l'a ici
[SCOL] NN: déjà je fais ces papiers en en demandant **euh**…
A: vous essayez vous essayez d'av– de demander une dérogation d'inscription préalable

Il s'agit encore une fois d'un procédé relativement simple. On note qu'il dépasse la stricte opposition en question ouverte et question fermée. Il pourrait être utile d'attirer l'attention de l'apprenant sur la compréhension de ces formes, pour qu'il se rende compte qu'il s'agisse d'une question et que le locuteur attend une réponse.

4.2 Semi-assertions

Certains énoncés remplissent bien le rôle de «questions dénuées de tout marquage explicite» (Kerbrat-Orecchioni, 2001: 89), c'est-à-dire sans aucun signe apparent : lexical, syntaxique ou intonatif. Et pourtant l'interlocuteur les interprète souvent comme telles et répond de façon appropriée:

[SCOL] A: c'est juste là où il y a la queue
NN: **juste à côté ↘**
A: voilà

On peut alors parler de «semi-questions et de semi-assertions, et de toutes sortes de cas intermédiaires entre la question 'pure'… et la véritable assertion» (Kerbrat-Orecchioni, 2001: 90). Ces phénomènes sont peut-être «plutôt rares» dans certains contextes (Fontanay, 1991: 120), mais nous en avons relevé plusieurs cas dans nos corpus. Les quelques exemples suivants tirés du corpus Escalade montrent qu'il s'agit généralement d'une demande de confirmation: à la fin d'un échange, le locuteur pose ce qu'il a cru comprendre (souvent renforcé par *donc* ou *alors*) et attend une réaction – soit catégorique (*non* dans le premier exemple) soit mitigée (*tout dépend*).

[ESC] L1: ça t'intéresse pas ↘
L2: non sans plus
[ESC] L1: c'est plutôt sportif plus que ballade alors ↘
L2: tout dépend du : du niveau d'intensité des des voies choisies
[ESC] L1: donc toi tu es quand même plutôt corde alors ↘
L2: ben disons que

4.3 Tags, inserts, fragments, échos

Très souvent la question est formée à partir d'une assertion à laquelle s'accole un tag avec une intonation montante, une forme d' «insert» ou «response elicitor» dans la terminologie de

Biber *et al*. (1999: 93). Le tag peut également renforcer une question déjà signalée comme dans le dernier exemple ci-dessous:

[SCOL] N: donc le diplôme c'est pas avant l'année prochaine **c'est ça** ↗
A: exactement
[SCOL] A: vous avez reçu votre autorisation d'inscription **c'était ça hein** ↗
NN: oui
[SCOL] A: c'était l'année dernière **n'est-ce pas** ↗
N: oui
[SCOL] A: vous avez peut-être apporté la photo↗ **non** ↗
N: s–
A: elle est dedans

Le tag le plus fréquent dans ce corpus est *d'accord*; en général, ce tag suit une explication de la part de l'administratif, il y a souvent une pause avant, l'intonation est généralement montante:

[SCOL] A: il fallait juste ça **d'accord** ↗
[SCOL] A: il faut il faut euh que votre quand vous amenez les pièces que votre dossier soit complet **d'accord** ↗

Dans plusieurs autres cas, *d'accord* est d'abord posé par l'interlocuteur pour signaler la compréhension (intonation descendante) avant d'être repris par l'administratif (intonation montante):

[SCOL] A: mais faites vite car les inscriptions finissent le 20 novembre
NN: d'accord ↘
A: d'accord ↗

Même lorsque le schéma intonatif est montant, le statut de *d'accord* n'est pas clair: s'agit-il d'une véritable question?

> Si certains énoncés peuvent être clairement catégorisés comme réalisant un acte d'assertion, ou un acte de question, d'autres se situent quelque part sur ce continuum reliant l'assertion et la question, constituant ainsi une sorte d'acte-valise. (Kerbrat-Orecchioni, 2001: 91)

Dans une dizaine de cas, soit il n'y a pas de réaction audible, soit le locuteur continue l'explication sans attendre de réaction. Ceci est également le cas de *hein*, avec son intonation montante, qui atténue souvent une forme impérative:

[SCOL] A: je vais vous marquer les pièces qu'il vous faut **hein** ↗
[SCOL] A: vous sortez tous les tous les chèques et tout ça **hein** ↗ vous préparez tout
[SCOL] A: et bien remplir ça **hein** ↗ voilà et rapidement **hein** ↗

L'intonation montante ne sert donc pas forcément à marquer une véritable demande d'information; il peut également servir à gérer l'interaction. Pour Fontaney (1991: 140):

> Dans presque tous les exemples que j'ai trouvés dans les corpus, [«hein»] est une question *métadiscursive*. L'interrogatif «hein » ne pose pas de question sur la valeur de vérité de ce qui est asserté, mais sur la perception de cette assertion par

l'interlocuteur. L'intention précise dépend du contenu sémantique de la phrase et comprend des sens tels que « tu as compris ce que j'ai dit ?», «tu sais cela?», «tu te rends compte des conséquences?», «tu comprends mon point de vue ?», «tu es d'accord avec moi ?»…

Hein dans les exemples ci-dessus a la fonction de laisser à l'interlocuteur la possibilité d'interrompre si le message n'est pas compris. On peut signaler en passant d'autres cas similaires. Dans les exemples suivants, *oui* encourage l'interlocuteur à continuer; *pardon* demande répétition:

[SCOL] N: je suis venu au sujet d'une inscription au DULAPS
A: **oui ↗**
N: euh DULASP […] et puis là regardez j'ai amené…
[SCOL] NN: j'ai une copine qui l'a fait sur place
A: **pardon ↗**
NN: j'ai une copine qui l'a fait sur place

Enfin, ce genre de fonction n'est pas limité aux tags et inserts mais peut également s'exprimer à travers des fragments en écho. Ici la reprise avec intonation montante contribue à la co-construction de l'interaction ; dans le premier exemple elle sollicite une réponse, mais même lorsque ce n'est pas le cas comme dans le deuxième, la possibilité est ouverte et l'absence de réponse laisse supposer que l'interlocuteur donne un accord tacite:

[SCOL] NN: je viens m'inscrire en école doctorale
A: **en école doctorale ↗**
NN: oui
[SCOL] A: alors il faudra faire une une lettre au président
NN: **au président ↗**
A: pour lui demander…

4.4 Autour des questions

Jusque là nous nous sommes concentrés sur les questions mêmes, mais nous avons remarqué un certain nombre de phénomènes associés ; nous n'en retiendrons que trois ici. Le premier c'est que par moments une question ne se pose pas tout de suite, elle se prépare de façon explicite. Dans les prochains exemples il s'agit d'un signe de politesse, de souligner l'importance de la question à venir, de marquer un changement de sujet:

[SCOL] A: pas de souci **simplement je vais vous demander maintenant** est-ce que vous vous êtes renseignée sur les nouvelles formations parce que comme on est passé au LMD il y a de nouvelles offres de formation est-ce que vous avez vu un peu
[SCOL] N: **encore une question**
A: je vous écoute
N: pour avoir l'inscription de maîtrise sur le diplôme il faut faire une demande euh spéciale

Une autre forme de préparation a lieu lorsque le «vrai» sujet de la question est posé avant la question même, qui est par la suite pronominalisée ; on parle ici de topic ou de détaché à gauche:

[SCOL] NN: donc **le diplôme** c'est pas avant l'année prochaine prochaine c'est ça
[SCOL] NN: **les réponses** c'est toujours ça se fait toujours par courrier
[SCOL] N: **le dernier test** ça s'est passé quand

Dans le corpus Escalade, le détaché à gauche est souvent un pronom; s'il ne s'agit pas d'un topic au sens «sujet» de l'énoncé, mais la question porte néanmoins sur ce que le «toi» fait:

[ESC] L1: et **toi** ça va tu t'en sors bien
[ESC] L1: et **toi** tu tu fais quoi

Dans l'autre direction, on retrouve bon nombre d'antitopics (détachés à droite), où le sujet pronominal à l'intérieur de la question est explicité lexicalement après:

[ESC] L1: sinon c'est c'est c'est haut **comme falaise**
[ESC] L1: ça fait peur **cent vingt mètres**
[ESC] L1: ça fait longtemps que tu en fais **de l'escalade**

Comme c'était le cas pour le topic, nous retrouvons plusieurs cas où c'est un pronom qui est repris:

[ESC] L1: ça existe **ça**
[ESC] L1: c'est sûr **ça**
[ESC] L1: tu trouves ça intéressant **toi**
[ESC] L1: tu en as **toi**

Ce n'est peut-être pas une coïncidence que tous ces exemples proviennent du corpus Escalade qui relève d'un registre plus familier, un entretien entre deux étudiants qui se tutoient. Quant au sens et à la force pragmatique de ces détachés à droite, nous pouvons laisser au lecteur de comparer les deux occurrences suivantes, les deux avec une intonation montante:

[ESC] L1 : tu as des préférences
[ESC] L1: tu as des tu as une préférence toi

Si ces phénomènes sont particulièrement évidents en français, ils ne sont pas inconnus en anglais. Dans sa discussion de ce type de structures en anglais, Swan (2005: 504-505) tire de son corpus oral plusieurs exemples non limités aux questions. Le détachement à gauche serait plus fréquent que le détachement à droite, et une topicalisation lexicale bien plus fréquente que pronominale.

4.5 Résumé

Dans cette première partie nous avons vu que les formes traditionnellement proposées dans les manuels ne rendent pas compte de la réalité telle que nous la montrent les corpus. Il nous semble surtout indéfendable de continuer à présenter des formes complexes aux débutants si

elles sont rares (inversions) ; quant aux autres formes, des informations concernant l'emploi, la fréquence, le registre ne serait pas de trop. Par ailleurs, des formes fréquentes dans nos corpus ne sont pas représentées dans les manuels même pour les apprenants avancés (inachevées, semi-assertions, tags, échos) ; divers phénomènes associés aux questions (préparation, topic, antitopic) sont également négligés. Du point de vue de l'apprenant, certaines de ces formes sont syntaxiquement et cognitivement bien plus simples à produire que les formes traditionnelles, et bien plus naturelles à l'oral. L'existence souvent de formes équivalentes peut contribuer à les rendre plus transparentes, malgré les différences notamment en termes de fréquence et de registre. Pour cette raison, en compréhension, ces formes peuvent être mal interprétées, et il peut être utile d'attirer l'attention de l'apprenant là-dessus. Une première conclusion serait donc que le travail sur corpus peut informer ou inspirer les manuels.

Nous sommes partis à la recherche de décalages manuel / corpus, mais il se peut que les auteurs aient fait des choix délibérés en promouvant certaines formes et en en négligeant d'autres. La décision de privilégier l'inversion dans un manuel pour débutants ne se défend pas pour des raisons de simplification, mais il se peut que certaines formes se prêtent plus difficilement à un enseignement traditionnel (Boulton, 2006, sous presse). Dans la section suivante, nous examinerons les possibilités de contourner l'enseignement comme transmission du savoir grâce à l'exploitation directe des corpus par les apprenants : un processus qui met l'accent sur l'apprentissage et sur la découverte («noticing»). Il est question là d'un deuxième rôle pour les corpus, sujet de bon nombre de recherches dans les pays anglo-saxons: «data-driven learning» ou «discovery learning» (Johns, 1991), une approche qui s'accorde avec l'idée de «corpus de découverte» de Holec (1990). Une telle approche est-elle possible avec les corpus oraux (Mauranen, 2004a, 2004b; Simpson *et al*., 2000)?

5. Les corpus: outils d'apprentissage

Depuis sa création, le CRAPEL (Université Nancy 2) développe une méthodologie d'enseignement/apprentissage qui vise l'organisation des activités d'apprentissage proprement dites, c'est-à-dire des activités conscientes et organisées dont l'objectif est la mise en place de conditions favorisant l'acquisition, c'est-à-dire l'internalisation des savoirs et savoir-faire. En accord avec les théories de l'acquisition, ces activités sont organisées selon trois phases:

- **Une phase de découverte** qui permet à l'apprenant de sélectionner les traits linguistiques et discursifs des comportements langagiers dont il vise l'apprentissage;

c **Une phase d'entraînement** systématique par laquelle l'apprenant concentre son effort d'apprentissage sur les savoirs langagiers nécessaires, point par point au travers d'activités réitérées;
c **Une phase de mise en pratique** non systématique de type simulation contrôlée, qui amène l'apprenant à interagir de façon pertinente dans «des situations communicativement déterminées» (Gremmo, 1997: 68).

Cette démarche s'appuie sur l'utilisation de documents authentiques, c'est-à-dire non fabriqués à des fins didactiques, dans un objectif d'apprentissage. Mis à disposition des apprenants, les documents constitués d'extraits de textes écrits oraux fournissent, comme le signale Holec (1990: 26) «l'exposition à la langue et à son fonctionnement», puisqu'ils présentent, contrairement aux documents de type pédagogique, des comportements langagiers réels. L'apprentissage/acquisition se définissant comme une démarche empirique de mise en relation de formes et de significations en contexte, les corpus, définis dans ce cadre comme une collection organisée de documents authentiques, constituent, comme le signale Bernardini (2004: 17), «an ideal instrument to observe socially-established form/meaning pairings». Ils donnent la possibilité aux apprenants de se construire progressivement une interprétation des faits linguistico-pragmatiques qu'ils observent au travers d'une «activité structurante» (Besse et Porquier, 1984).

Cette méthodologie est pratiquée au sein du CRAPEL depuis de nombreuses années, en particulier au travers du Système d'Apprentissage Autodirigé avec Soutien. Dans ce cadre,[6] et notamment pour le travail en expression orale, l'apprenant est orienté par un conseiller vers des documents qui lui permettent d'observer «des réalisations linguistiques communicativement situées pour découvrir ce qu'il a à apprendre» (Gremmo, 1997). Impliqué dans un travail de recherche et d'observation, l'apprenant développe peu à peu un savoir-faire méthodologique de sélection et d'organisation de ce qu'il va apprendre en fonction de ses besoins, désirs ou motivations. Il apprend ainsi, avec l'aide du conseiller à «observer le fonctionnement langagier dans les documents authentiques» (Gremmo, 1997). La démarche relève donc de la construction de savoirs propres, respectant ainsi la diversité des stratégies du processus global d'appropriation qui consiste à «appréhender, traiter et à stocker les données langagières fournies […] et à les adapter à la connaissance antérieure par assimilation et accommodation» (Besse et Porquier, 1984: 244). Le plus souvent les apprenants adhèrent à cette façon de travailler. Ainsi un apprenant cité par Gremmo (1997): «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.»

[6] Voir les articles de Bailly et Gremmo dans « Mélanges CRAPEL » n° 22 (1995) pour une description de ce dispositif d'apprentissage.

De façon paradoxale, l'exposition à des corpus authentiques permet en outre de démystifier le modèle du locuteur natif. Ces documents révèlent en effet l'importance des faits de variation dans les usages, que ce soit en fonction des situations de communication ou en fonction des locuteurs.[7]

Dans cette perspective, l'apport des nouvelles technologies dans la constitution de corpus paraît particulièrement intéressant pour les activités de découverte du fonctionnement de la langue, dans la mesure où le traitement informatique facilite la sélection et l'observation d'occurrences en contextes variés. Par l'accès informatique à des corpus, notamment au travers de concordanciers, les apprenants sont confrontés à « the repetition of similar linguistic patterns » (Tognini-Bonelli, 2001 : 41) et non à des faits linguistiques isolés, ce qui leur permet de se focaliser sur un point spécifique et de saisir les significations tant linguistiques que pragmatiques en contexte. Comme le souligne Bernardini (2004 : 21), «as a result of analysing large quantities of authentic data, [...] learners construct an interpretation for themselves, relying on lexico-syntactic, semantic and pragmatic observations».

Mais le passage à l'observation sur corpus informatisé ne représente pas un simple changement quantitatif. Il nécessite une réflexion sur les caractéristiques auxquelles doivent répondre les corpus pour constituer un outil d'apprentissage. C'est ce que présente la section suivante.

5.1 Caractéristiques des corpus

5.1.1 La représentativité d'usages diversifiés

Les corpus doivent tout d'abord, comme nous venons de le souligner, présenter suffisamment de diversité pour constituer un échantillon relativement représentatif permettant l'observation d'emplois déterminés (Habert, 2000). Il est impossible de viser à une représentation exhaustive des usages, mais il est fondamental de fournir des données mettant en jeu les différents paramètres de la variation, en particulier ceux touchant les « genres de textes », au sens de Bakhtine (1984) puisque ces genres déterminent l'organisation et l'interprétation des énoncés et permettent de mettre en évidence l'adaptation des comportements langagiers selon la situation et le statut et rôle des locuteurs. Il s'agit d'illustrer le fait que l'on ne dit pas n'importe quoi à n'importe qui dans n'importe quelle situation et de sensibiliser les apprenants à la notion d'énoncé approprié bien plus qu'à la notion d'énoncé correct. Sur ce point, il est important que les corpus soient constitués de productions de locuteurs natifs, mais

[7] Ainsi, «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» (Mparutsa *et al*., 1991: 130).

également de productions de « non-natifs compétents »[8]. Les productions de ces derniers permettent en effet l'observation de stratégies de contournement et de compensation dont la valeur heuristique est indéniable pour l'apprentissage / acquisition d'une compétence visant à l'efficacité de la communication et non au respect exclusif de la correction grammaticale.

5.1.2 Caractéristiques techniques

Afin que les apprenants puissent avoir accès à des informations précises et explicites sur les locuteurs et les situations de communication en jeu, qui leur permettront de mettre en relation les paramètres linguistiques et extralinguistiques dans l'interprétation des valeurs pragmatiques des faits observés, les corpus doivent être accompagnés de métadonnées explicites et aisément consultables. En outre, compte tenu de l'importance des phénomènes prosodiques et des techniques actuelles, il est possible et nécessaire aujourd'hui d'envisager la constitution de corpus alignés texte et son, permettant ainsi aux utilisateurs d'avoir un accès simultané aux faits segmentaux et suprasegmentaux. Nous verrons l'importance de ce point à propos de l'analyse des questions.

5.2. Modalités de consultation des corpus

5.2.1 Les type de requêtes

L'accès à de telles ressources, encore très inhabituel en France est l'objet d'un travail de recherche de l'équipe CRAPEL: nous en présentons ici les grandes lignes, dans le cadre du projet FLEURON. Le site, constitué de 8 modules traitant de façon spécifique les domaines suivants: expression orale, expression écrite, compréhension orale, compréhension écrite, sensibilisation interculturelle, conseils pour apprendre, conseils pratiques, permettra aux utilisateurs d'avoir accès à des documents authentiques extraits des diverses situations de communication auxquelles sont confrontés les étudiants dans leur vie académique au sens large: cours magistraux, travaux dirigés, interactions avec les enseignants, interactions dans les différents services de l'université (scolarité, vie étudiante, bibliothèque, résidence universitaire, interactions entre étudiants, situations de la vie quotidienne dans et hors de l'enceinte du campus). Une présentation modulaire, différenciant l'apprentissage des aptitudes langagières (lire, comprendre, parler, écrire), et n'imposant aucune progression

[8] Largement discutée dans les recherches anglo-saxonnes, cette position reste controversée, comme en témoigne la place qui lui est accordée dans les « Controversies in Applied Linguistics » de Seidlhofer (2003). Toutefois, elle reçoit de plus en plus de soutien comme modèle d'apprentissage (cf. le rôle de l'anglais comme langue véhiculaire de Seidlhofer [par ex. 2005] ou encore Jenkins [2007], ou les «successful users of English» de Prodromou [2003]). En particulier, c'est l'un des dix critères essentiels pour une grammaire parlée de McCarthy et Carter (2001), et l'incorporation de productions de locuteurs « experts » mais non natifs est prônée de plus en plus souvent dans la conception de corpus pour une exploitation didactique (par exemple, O'Keeffe *et al*., 2007; Aston, 2000; Mauranen, 2004a; Osborne, 2004; etc). Pour une analyse de ce concept spécifique au français, voir Adami *et al*. (2004).

linéaire permettra une navigation souple et motivée dans l'ensemble des outils mis à disposition. La consultation des corpus oraux est prévue selon deux modalités.

Il s'agit d'une part d'utiliser les corpus comme ressources pour des activités communicatives ciblées par un accès à des échantillons selon un système de catalogage établi à partir des métadonnées de chaque document. Le choix pourra être fait en fonction des besoins d'apprentissage ou de la situation de communication qu'illustre le document ou des types de locuteurs impliqués dans l'interaction. Pour la sélection du document et ses exploitations, l'apprenant aura accès à des fiches de conseils d'apprentissage qui présentent des suggestions d'activités d'apprentissage détaillées, incluant les aspects didactiques (objectifs visés, modalités d'évaluation, modalités de travail) et les aspects pratiques (matériel, manipulations), ainsi que des outils de sauvegarde de notes personnelles et de protocoles de journal d'apprentissage permettant de mémoriser le parcours d'apprentissage, les progressions, les difficultés rencontrées et d'établir des liens vers des documents ou des activités similaires.

D'autre part le corpus général ainsi constitué pourra faire l'objet d'activités de découverte et d'observation des fonctionnements langagiers mis en œuvre dans les différentes situations. Pour cela, les utilisateurs auront à disposition un concordancier autorisant des requêtes par chaîne de caractères sur l'ensemble ou un sous-ensemble du corpus. Les apprenants pourront ainsi constituer la liste des occurrences des items recherchés en contexte et avoir accès au texte intégral dans lequel se situe chaque occurrence ainsi qu'à l'enregistrement de la transcription.

La mise à disposition de telles ressources nécessite la mise en place d'outils pédagogiques permettant aux futurs utilisateurs d'apprendre à apprendre avec des corpus. De nombreux travaux ont souligné la difficulté des apprenants à s'approprier ce nouvel outil de travail. Le projet prévoit ainsi de constituer des mini-corpus d'entraînement pouvant faire appel à un marquage de formes spécifiques, afin de sensibiliser les apprenants à des fonctionnements langagiers qui risquent de passer inaperçus, d'autant qu'ils ne sont pas mentionnés dans les ouvrages de référence habituellement utilisés.

Les questions sans marques lexicales ou morphologiques spécifiques constituent un bon exemple de structures, pour lesquels une sensibilisation est nécessaire. Par exemple, pour le cas des énoncés questions, certaines structures, comme l'a noté Coveney (1995), apparaissent formellement identiques à des questions alors qu'elles n'en sont pas ; par exemple, les marqueurs de discours du type *n'est-ce pas?* ou *tu vois?* Par ailleurs, comme nous l'avons vu,

les énoncés questions peuvent n'être marqués que par des indices suprasegmentaux (intonation ouvrante ou implicative) et échappent donc à toute recherche par chaîne de caractères. Il est donc nécessaire d'envisager un marquage permettant aux apprenants de s'entraîner à repérer ces structures. Le projet FLEURON prévoit ainsi un marquage en actes de parole de quelques mini-corpus illustrant des formes non discrètes. Les difficultés de la démarche sont exposées dans la section suivante.

5.2.2 Le repérage des unités par valeurs pragmatiques

Il est nécessaire en effet de penser les étiquettes « pragmatiques » en fonction de leur pertinence pour les besoins d'apprentissage. Comme le signale Kerbrat-Orecchionni (1991: 86): «L'acte de question recouvre de nombreux cas de figure […] qui s'opposent selon les raisons qui font que l'on sollicite telle ou telle information […] la nature particulière de l'information demandée […] la valeur taxémique de la question et ses effets sur la relation interpersonnelle». Variables auxquelles il faut rajouter «l'état relatif des savoirs des interlocuteurs» et «le type d'activité dans lequel s'inscrit l'échange» (Kerbrat-Orecchionni, 1991: 88). Compte tenu des multiples valeurs portées par des énoncés qui constituent un véritable «continuum reliant l'assertion et la question» (cf. *supra*), il ne paraît pas possible d'aboutir à une liste exhaustive d'étiquettes. Il revient donc aux concepteurs du corpus de sélectionner et d'étiqueter les comportements langagiers les plus représentatifs de la situation ou les plus susceptibles de faire l'objet d'observations, selon les corpus disponibles. Ce travail de recherche étant encore en chantier, nous nous contenterons de relever ici quelques exemples de cette diversité des valeurs pragmatiques attachées aux énoncés questions. On relève ainsi dans le corpus d'interactions à la scolarité:

c des demandes d'agir:
[SCOL] puis-je aussi vous demander d'ouvrir la porte
[SCOL] je vous laisse faire
[SCOL] je vous laisse vérifier

c des demandes de confirmation:
[ESC] donc toi tu es quand même plutôt corde alors
[SCOL] donc là j'aurais pas à fournir une enveloppe timbrée
[SCOL] donc vous avez pas en fait de de diplôme entre le bac et et ça

c des demandes de précisions:
[SCOL] et là je m'inscris euh
[SCOL] alors le chèque je ne savais pas à quel ordre il fallait le mettre
[SCOL] par contre euh culture euh connaissances des métiers de l'information c'est

c des formules d'accueil:
[SCOL] qu'est-ce que je peux faire pour vous

Une autre caractéristique de ces énoncés concerne la diversité des profils prosodiques relevés. Nous n'avons mené sur ce point aucune étude spécialisée, mais l'écoute des exemples nous a permis de constater intuitivement ce qui a été noté dans de nombreuses études[9] et résumé en ces termes par Kerbrat-Orecchioni (1991: 90): «On peut donc difficilement parler d'une intonation spécifique de la question». L'observation / écoute de corpus alignés permet ainsi de sensibiliser les apprenants à la diversité des usages authentiques, dont on peut raisonnablement penser qu'elle se retrouve dans leur propre langue.

5.3 Le découpage des unités

Un autre problème soulevé par les énoncés questions concerne leur aspect formel. Comme l'ont signalé de nombreux auteurs,[10] il est peu pertinent de chercher à reconstituer le découpage en phrases des productions écrites pour délimiter les unités de langue parlée. On constate, comme nous l'avons vu dans la première partie de cette étude, que les énoncés questions ne dérogent point à ce principe. On relève en effet des énoncés à structures non canoniques, présentant:

c des éléments détachés à gauche:
[ESC] et toi ça va
[SCOL] les réponses c'est toujours ça se fait toujours par courrier
[SCOL] le dernier test ça s'est passé quand

c des éléments détachés à droite:
[SCOL] vous l'avez sur vous le papier
[SCOL] vous l'avez déposé quand votre dossier

c des énoncés soulignés par des marques de renforcement que l'on peut assimiler aux «tags» de l'anglais:
[ESC] c'est plutôt sportif plus que ballade **alors**
[ESC] tu es obligé de continuer **alors**

r ainsi que de nombreux énoncés de type «inachevés»:
[ESC] tu fais cent vingt mètres en une fois **ou**
[SCOL] donc là à l'UFR Lettres je demande la fiche péda **et puis euh**

à propos desquels il est important de signaler, comme nous l'avions fait pour des énoncés similaires à valeur d'assertion,[11] qu'ils n'entraînent aucun incident de communication. On ne relève aucune tentative de «correction» ou de rectification de la part des locuteurs et les

[9] En particulier Grundsom (1973), Bolinger (1982), ou Stenström (1984).
[10] En particulier Blanche Benveniste *et al*. (1990) et Berrendonner (1991) pour le français; Leech (2000) pour l'anglais; Miller et Weinert (1998) pour l'allemand et le russe.
[11] Cf. Debaisieux (2000).

réactions des interlocuteurs ne signalent pas de problèmes d'incompréhension. Leurs réponses, comme on le voit dans les deux exemples suivants:

[SCOL] N: et je fais comment maintenant je peux faire euh une une autre type de demande ou euh
A: une lettre au président

[ESC] L1: donc l'escalade sans corde c'est toujours du solo il y a pas: --
L2: ben euh c'est euh on peut faire de l'escalade sans corde à deux mais euh - je vois pas l'intérêt

montrent au contraire qu'ils ont parfaitement compris ce qui constitue un appel à inférence et qu'ils y répondent pour une bonne continuation de l'interaction. En outre, une description pertinente des énoncés/questions implique la prise en compte d'un cotexte plus large, notamment des énoncés à valeur de «*pre*» (Schegloff, 1980) ou «*d'annonce*» (Roulet *et al*., 1987) comme dans:

[SCOL] N: **j'ai encore une question**
A: je vous écoute
N: pour avoir l'inscription de maîtrise sur le diplôme il faut faire une demande euh spéciale

qui sont à relier au problème de la sauvegarde des faces de l'interlocuteur : « C'est un fait généralement reconnu qu'une action impliquant une intrusion sur le territoire d'autrui est perçue comme moins menaçante pour sa face négative si elle est préalablement annoncée » (Roulet *et al*. 1987: 87).

La description doit de même prendre en compte les énoncés à valeur de justification, comme dans l'exemple suivant:

[SCOL] NN: je dois le refaire?/**parce qu'il y a des papiers je peux pas aller trouver**

qui permettent, selon Goffman (1987: 22) d'amoindrir: «ce que peut avoir d'offensant le fait de s'imposer à autrui avec une question.»

Il est clair qu'on ne peut restreindre les unités questions à des séquences de type phrases canoniques, si l'on souhaite sensibiliser les apprenants à des comportements langagiers appropriés. Ceci d'autant que les structures non canoniques représentent un pourcentage non négligeable des énoncés questions. En effet, sur la base des chiffres présentés dans le Tableau 1, on aboutit à la répartition suivante:

Tableau 2. Répartition en pourcentages des énoncés questions du corpus

inversion	0.67 %
est-ce que	3.35 %
mot-Q	**30.87 %**
intonation	**34.22 %**
inachevés + semi-assertions	**30.87 %**

Une approche descriptive des énoncés questions attestés amène à poser des regroupements de type macro-syntaxique au sens de Blanche-Benveniste *et al*. (1990) ou de Berrendonner (1991) dont la structure pourrait être représentée de la façon suivante:

Tableau 3. Organisation macro des énoncés questions

préparation	introduction	topic	noyau	antitopic	insistance	justification
	et	toi	tu fais quoi			
	donc	toi	tu es quand même plutôt corde		alors	
j'ai une question	pour avoir l'inscription		il faut faire une demande spécifique			
			je dois le refaire			parce qu'il y a des papiers …
			tu es allé un peu partout	toi		

Plusieurs arguments plaident pour la prise en compte de ce type de structures : d'une part, on peut difficilement prétendre enseigner la communication orale si l'on n'enseigne que la grammaire écrite. Par ailleurs, on relève ce type de structures macros dans d'autres langues. Comme le signale Leech (2000: 706): «More than one grammarian of speech has noted the capability of spoken language to build more complex structures, by adding c-units as «satellite slots» to the front or rear of a larger (usually clausal) unit».

Ainsi, Carter et McCarthy (1995: 151) représentent la structure «potentielle» des énoncés «conversationnels» en anglais:[12]

Tableau 4. Structural potential of the clause in conversation

pre-clause	clause	post-clause
TOPIC	CORE	TAIL

et commentent:

> Tails are therefore an important part of what may be called interpersonal grammar, that is to say speaker choices which signal the relationships between participants and position the speaker in terms of his/her stance or attitude. [...] the topic and the tail are optional, but when utilized, carry important interpersonal functions, as our and other's data show.

[12] Nous avons simplifié le tableau dont l'original présente également la liste des constituants et leurs formes.

En ce qui concerne les énoncés questions, on constate effectivement que les organisations de type topic ou de type antitopic:

[ESC] ça fait peur **cent vingt mètres**
[SCOL] **le dernier test** ça s'est passé quand

participent d'une stratégie de co-construction du discours. Ainsi dans l'échange suivant:

[ESC] L1: ouais **l'alpinisme** c'est quand il y a des - de la glace ou
L2: **l'alpinisme** c'est euh quoi qu'il arrive on va au sommet –

la structure de l'énoncé de L1 peut être interprétée comme la proposition d'un thème en tant que tel faisant ensuite l'objet d'un questionnement. C'est à partir de ce même objet de discours que L2 construit progressivement sa réponse en s'appuyant sur ce qui a été fourni par L1. Cette stratégie qui relève de la négociation du thème possède un double avantage: elle situe l'échange dans un mouvement positif de collaboration et peut servir à faciliter la mise en place des propos en procurant un temps de réflexion supplémentaire au locuteur. De même, les structures de type inachevées à contour implicatif, telle que:

[SCOL] NN: déjà je fais ces papiers en en demandant **euh**
A: l' admission préalable

semblent pouvoir être rapprochées intuitivement de ce que Fontaney (1991: 125) nomme une attitude de «croire savoir», qui permet à celui qui pose la question de ménager sa propre face (il a compris une partie des explications fournies) et celle de son interlocuteur puisque la question est amoindrie ; ceci nous semble particulièrement net dans les demandes d'informations auprès des services de scolarité.

Loin de constituer une difficulté, comme on l'entend si souvent dans la communauté franco-française des spécialistes du Français Langue Etrangère, la confrontation à des structures discursives dites « complexes », telles que celles que nous avons relevées, nous paraît être une possibilité pour les apprenants de mener les observations réflexives sur les faits linguistiques et la complexité des paramètres à prendre en compte pour leur interprétation, condition nécessaire à la construction d'un modèle propre. Ainsi que le relève McEnery *et al*. (2006: 197):

> it is precisely because they do not simply offer models to imitate, however, that corpus data seems valuable for learners. As Leech and Candlin (1986: xiv-xvi) observed well over a decade ago, data from corpora has to be interpreted subjectively. Their reality, from this point of view, is a characteristic not just of the data, but above all of the interpretative process. 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.

Ce type d'exemple relève des observables susceptibles d'enrichir une «pragmatique interlinguistique» (au sens de Riley, 2000) nécessaire dans l'apprentissage des langues, et

devrait être intégré dans le matériel fourni aux apprenants. Mais ceci implique une modification de certaines représentations en vigueur aujourd'hui. Il s'agit tout d'abord de reconsidérer la notion de compétence grammaticale, encore trop souvent analysée comme la simple «capacité de comprendre et d'exprimer du sens en produisant et en reconnaissant des phrases bien formées selon ces principe [la grammaire de la langue]»[13] au profit d'une notion plus large de comportement langagier descriptible en termes de savoir complexe de nature pragmatico-discursive. Pour ce faire, il est nécessaire d'intégrer dans les données soumises aux apprenants les productions effectives des locuteurs d'autant que, au-delà de leur apparente complexité, leur caractère « universel » peut jouer le rôle d'un facilitateur de transfert interlinguistique. Comme le souligne Leech (2000):

> ...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.

D'un point de vue didactique, l'intégration de telles structures dans des observables donnant lieu à des activités de découverte nous paraît ainsi faire le lien entre une approche strictement «form-focused» et une démarche de type communicative. En outre, la mise en place des activités d'observation des corpus qui donnent lieu à des stratégies de prises d'indices, de construction et vérification d'hypothèses permet de respecter le caractère éminemment individuel des processus d'apprentissage. Conférant à l'apprenant le rôle de chercheur (« student as researcher » pour reprendre la terminologie de la littérature anglo-saxonne), elles participent au développement de sa capacité à apprendre.

6. Pour ne pas conclure

Les possibilités d'une exploitation directe des corpus à des fins d'apprentissage sont relativement bien développées dans de nombreuses publications anglo-saxonnes. Ce qui manque le plus sont des recherches empiriques qui explorent l'efficacité d'une telle approche : Boulton (2007) recense une quarantaine d'articles qui vont dans ce sens, et trouve que la plupart s'intéressent aux attitudes ou aux pratiques des apprenants, ou à l'utilisation de corpus comme référence sans prendre en compte un quelconque apprentissage. En tout, seule une poignée d'études examinent les effets d'une exploration de corpus sur l'apprentissage

[13] Un cadre européen commun de référence pour les langues : Apprendre, enseigner, évaluer, chapitre 5 : Les compétences de l'utilisateur / apprenant - paragraphe 5.2.1.2. Compétence grammaticale. www.coe.int/t/dg4/linguistic/Source/Framework_FR.pdf, page consultée le 25 mars 2007.

même, notamment Cobb (1997a ; 1997b, 1999a, 1999b), Horst et Cobb (2001), Allan (2006), Koosha et Jafarpour (2006), et Sun et Wang (2003).

En France, l'utilisation de corpus est quasiment absente des pratiques d'enseignement/ apprentissage. Plusieurs facteurs concourent à cette situation: l'absence de corpus accessibles, notamment de corpus oraux ; la prégnance des représentations conservatrices sur la langue; le manque de descriptions de la langue parlée; et la réticence ou le scepticisme envers les démarches autonomisantes. A cela s'ajoute le manque de dialogue entre la communauté des didacticiens et celle des linguistes. En ce qui nous concerne, sans tomber dans un optimisme béat, nous pensons qu'il y aurait intérêt à dépasser ces résistances et à intégrer dans l'enseignement/apprentissage un outil qui présente suffisamment de possibilités pour ne plus être négligé. Notre position s'appuie sur les expériences accumulées lors de sessions d'apprentissage autodirigé avec soutien qui nous ont montré que les apprenants ont la capacité d'observer des corpus authentiques pour une analyse des faits linguistiques en contexte.

Références

Adami, H., André, V., Bailly, S., Carton, F., Castillo, D., Poncet, F. et Riley P. (2004) L'étranger compétent: Un nouvel objectif pour la didactique des langues étrangères. In: D. Groux et H. Holec (éds.) *Une identité plurielle: Mélanges offerts à Louis Porcher*. Paris: L'Harmattan, 535-550.

Allan, R. (2006) Data-driven learning and vocabulary: Investigating the use of concordances with advanced learners of English. *Centre for language and communication studies, occasional paper*, 66. Dublin: Trinity College Dublin.

Aston, G. (1997) Small and large corpora in language learning. In: B. Lewandowska-Tomaszczyket J. Melia, J. (éds.) *Practical applications in language corpora*. Lodz: Lodz University Press, 51-62.

Aston, G. (2000) Corpora and language teaching. In: L. Burnard et T. McEnery (éds.) *Rethinking language pedagogy from a corpus perspective*. Frankfurt : Peter Lang, 7-17.

Bailly, S. (1995) La formation de conseiller. *Mélanges CRAPEL*, 22, 63-83.

Bakhtine, M. (1984) [1952] Les genres du discours. In: *Esthétique de la création verbale*. Paris: Gallimard, 265-308.

Beacco, J.-C., Ferrari, M. de, Lhote, G. et Tagliante, C. (2005) *Niveau A1.1 pour le français. (Publics adultes peu francophones, scolarisés, peu ou non scolarisés). Référentiel et certification (DILF) pour les premiers acquis en français.* Paris: Didier.

Beeching, K. (1997) French for specific purposes: The case for spoken corpora. *Applied linguistics*, 18/3, 374-394.

Bernardini, S. (2004) Corpora in the classroom: An overview and some reflections on future developments. In: J. Sinclair (éd.) *How to use corpora in language teaching*. Amsterdam : John Benjamins, 15-36.

Berrendonner, A. (1991) Pour une macro-syntaxe. In D. Willems (éd.) *Données orales et théories linguistiques*. Paris-Louvain: Duculot, 25-31.

Besse, H. et Porquier, R. (1984). *Grammaire et didactique des langues*. Paris: Credif, Hathier-Didier.

Biber, D., Johansson, S., Leech, G., Conrad, S. et Finegan, E. (1999) *Longman grammar of spoken and written English*. London: Pearson.

Blanche-Benveniste, C. (2000) Corpus de français parlé. In: M.Bilger (éd.) *Corpus: Méthodologie et applications linguistiques*. Paris: Honoré Champion, 15-25.

Blanche-Benveniste, C., *et al.* (1990) *Le français parlé. Etudes grammaticales*. Paris: CNRS.

Boulton, A. (2006) Tricky to teach, easier to learn: Empirical evidence for data-driven learning. *American association of applied corpus linguistics (AAACL)*. Northern Arizona University, USA. 20-22 octobre.

Boulton, A. (2007). But where's the proof? The need for empirical evidence for data-driven learning. *Proceedings of BAAL 40: Technology, ideology and practice in applied linguistics*. University of Edinburgh.

Boulton, A. (sous presse) Looking (for) empirical evidence of DDL at lower levels. *Practical applications of language and computers (PALC 2007)*. In Lewandowska-Tomaszczyk, B. (Ed.) *PALC 2007*. (*Lodz studies in language*). Frankfurt: Peter Lang.

Boulton, A. et Debaisieux, J.-M. (2006) The wise whys and hows of spoken corpora in language learning and teaching. *3rd Inter-varietal applied corpus studies (IVACS) group international conference : Language at the interface*. University of Nottingham, UK. 23-24 juin.

Braun, S. (2007) Designing and exploiting small multimedia corpora for autonomous learning and teaching. In: E. Hidalgo, L. Quereda et J. Santana (éds.) *Corpora in the foreign language classroom*. Amsterdam: Rodopi, 31-46.

Carter, R. (1998) Orders of reality: CANCODE, communication, and culture. *ELT journal*, 52/1: 43-56.

Carter, R. et McCarthy, M. (1995) Grammar and the spoken language. *Applied linguistics*, 16/2: 141-158.

Cobb, T. (1997a) *From concord to lexicon: Development and test of a corpus-based lexical tutor.* Montréal : Concordia University, Thèse de doctorat. http://www.er.uqam.ca/nobel/r21270/webthesis/Thesis0.html, page consultée mars 2006.

Cobb, T. (1997b) Is there any measurable learning from hands-on concordancing? *System,* 25/3: 301-315.

Cobb, T. (1999a) Breadth and depth of lexical acquisition with hands-on concordancing. *CALL*, 12/4: 345-360.

Cobb, T. (1999b) Giving learners something to do with concordance output. *IMELT 99*. Hong Kong Polytechnic University, novembre. http://elc.polyu.edu.hk/conference/papers/Cobb%20-%20Giving%20learners %20something%20to%20do%20with%20concordance%20output.htm, page consultée octobre 2007.

Coveney, A. (1995) The use of the QU-final interrogative structure in spoken French. *Journal of French Language Studies*, 5/2: 143-171.

Debaisieux, J.-M. (2000) Réalités du discours oral et didactique du français langue étrangère. *Une didactique des langues pour demain (Le français dans le monde: Recherches et applications)*. Paris: CLE, 143-156.

Debaisieux, J.-M. (2005a) Les corpus oraux : Situation, exploitation linguistique, bilan et perspectives. *De la linguistique de corpus à la relation "partie/tout". Scolia* (Strasbourg : Publications de l'Université Marc Bloch), 19: 9-40.

Debaisieux, J.-M. (2005b) Corpus oraux et didactique des langues: Un rendez-vous à ne pas manquer. *Français fondamental, corpus oraux: 50 ans de travaux et d'enjeux*. Université Lyon 2, 8-10 décembre.

Druon, M. (1999) Discours sur l'état de la langue. Séance publique annuelle de l'Académie Française. http://www.academie-francaise.fr/immortels/discours_SPA/druon_99.html, page consultée janvier 2007.

Fontaney, L. (1991) A la lumière de l'intonation. In: C. Kerbrat-Orecchioni (éds.) *La question*. Lyon: Presses Universitaires de Lyon, 113-161.

Goffman, E. (1987) *Façons de parler*. Paris: Editions de Minuit.

Goosse, A. (éd.) (1994) *Grévisse: Le bon usage (13ème édition)*. Paris: Duculot-De Boeck.

Gremmo, M.-J. (1995) Conseiller n'est pas enseigner: Le rôle du conseiller dans l'entretien de conseil. *Mélanges CRAPEL*, 22, 33-61.

Gremmo, M.-J. (1997) Matériaux d'enseignement et apprentissage des langues. *Mélanges Pédagogiques*, CRAPEL, Université Nancy 2, 63-81.

Habert, B. (2000) Des corpus représentatifs: De quoi, pour quoi, comment? In: M. Bilger (éd.) *Linguistique sur corpus: Etudes et réflexions. (Cahiers de l'université de Perpignan, 31)*. Perpignan: Presses Universitaires de Perpignan, 11-58.

Hidalgo, E., Quereda, L. et Santana, J. (éds.) (2007) *Corpora in the foreign language classroom*. Amsterdam: Rodopi.

Holec, H. (1990) Des documents authentiques, pour quoi faire? *Mélanges CRAPEL*, 65-74.

Honnor, S. et Mascie-Taylor, H. (2000) *Encore tricolore 1 (nouvelle édition)*. Cheltenham: Nelson-Thornes.

Honnor, S. et Mascie-Taylor, H. (2001) *Encore tricolore 2 (nouvelle édition)*. Cheltenham: Nelson-Thornes.

Horst, M. et Cobb, T. (2001) Growing academic vocabulary with a collaborative on-line database. In: B. Morrison, D. Gardner, K. Keobke et M. Spratt (éds.) *ELT perspectives on IT and multimedia : Selected papers from the ITMELT conference 2001*. Hong Kong: Polytechnic University, 189-225. http://elc.polyu.edu.hk/conference/papers2001/cobb.htm, page consultée octobre 2007.

Jenkins, J. (2007) *English as a Lingua Franca: Attitude and Identity*. Oxford: Oxford University Press.

Johns, T. (1991) Should you be persuaded: Two examples of data-driven learning. In T. Johns et P. King, (éds.) *Classroom concordancing (English language research journal)*, 4: 1-16.

Kerbrat-Orecchioni, C. (éd.) (1991) *La question.* Lyon: Presses Universitaires de Lyon.

Kerbrat-Orecchioni, C. (2001) *Les actes de langage dans le discours: Théorie et fonctionnement*. Paris: Nathan Université.

Koosha, M. et Jafarpour, A. (2006) Data-driven learning and teaching collocation of prepositions : The case of Iranian EFL adult learners. *Asian EFL journal quarterly*, 8/4, p. 192-209. http://www.asian-efl-journal.com/December_2006_EBook.pdf, page consultée août 2007.

Leech, G. (2000) Grammars of spoken English: New outcomes of corpus-oriented research. *Language learning*, 50/4: 675-724.

Mauranen, A. (2004a) Spoken corpus for an ordinary learner. In: J. Sinclair (éd.) *How to use corpora in language teaching.* Amsterdam: John Benjamins, 89-105.

Mauranen, A. (2004b) Speech corpora in the classroom. In: G. Aston, S. Bernardini et D. Stewart (éds.) *Corpora and language learners*. Amsterdam: John Benjamins, 195-211.

Maynard, C. et Leicher, S. (2007) Pragmatic annotation of an academic spoken corpus for pedagogical purposes. In: E. Fitzpatrick (éd.) *Corpus linguistics beyond the word: Corpus research from phrase to discourse*. Amsterdam: Rodopi, 107-115.

McCarthy, M. et Carter, R. (2001) Ten criteria for a spoken grammar. In: E. Hinkel et S. Fotos (éds.) *New perspectives on grammar teaching in second language classrooms*. Mahwah, NJ: Lawrence Erlbaum Associates, 51-75.

McEnery, T., Xiao, R. et Tono, Y. (2006) *Corpus-based language studies*. London: Routledge.

Miller, J. et Weinert, R. (1998) *Spontaneous spoken language: Syntax and discourse.* Oxford : Clarendon Press.

Mparutsa, C., Love, A. et Morrisson, A. (1991) Bringing concord to the ESP classroom. In: T. Johnset P. King (éds.) *Classroom concordancing (English language research journal)*, 4: 115-134.

O'Connor di Vito, N. (1991) Incorporating native-speaker norms in second language materials. *Applied linguistics*, 12/4: 383-396.

O'Keeffe, A., McCarthy, M. et Carter, R. (2007) *From corpus to classroom : Language use and language teaching*. Cambridge: Cambridge University Press.

Osborne, J. (2004) Top-down and bottom-up approaches to corpora in language teaching. In: U. Connor et T. Upton (éds.) *Applied corpus linguistics : A multidimensional perspective*. Amsterdam: Rodopi, 251-265.

Prodromou, L. (2003) In search of the successful user of English : How a corpus of non-native speaker language could impact on EFL teaching. *Modern English teacher*, 12/2: 5-14.

Rey-Debove, J. et Rey, A. (2002) *Nouveau petit Robert*. Paris: Dictionnaires Le Robert.

Riley, P. (2000) «Je vous ai compris», aspects ethnologiques de la compréhension. *Une didactique des langues pour demain (Le français dans le monde: Recherches et applications)*. Paris: CLE, 79-95.
Römer, U. (2004) Comparing real and ideal language learner input: The use of an EFL textbook corpus in linguistics and language teaching. In: G. Aston, S. Bernardini et D. Stewart (éds.) *Corpora and language learners*. Amsterdam: John Benjamins, 151-168.
Roulet, E., Auchlin, A., Moeschler, J., Rubattel, C. et Schelling, M. (1987) *L'articulation du discours en français contemporain (2ème édition)*. Berne: Peter Lang.
Schegloff, E. (1980) Preliminaries to preliminaries: "Can I ask you a question?" *Sociological inquiry*, 50: 104-152.
Seidlhofer, B. (éds.) (2003) *Controversies in applied linguistics: Corpus linguistics and language teaching*. Oxford: Oxford University Press.
Seidlhofer, B. (2005) English as a lingua franca. *ELT journal*, 59: 339-341.
Simpson, R., Lucka, B. et Ovens, J. (2000) Methodological challenges of planning a spoken corpus with pedagogical outcomes. In: L. Burnard et T. McEnery (éds.) *Rethinking language pedagogy from a corpus perspective*. Frankfurt: Peter Lang, 43-49.
Sun, Y-C. et Wang, L-Y. (2003) Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *CALL*, 16/1: 83-94.
Swan, M. (2005) *Practical English usage (3rd edition)*. Oxford: Oxford University Press.
Tognini-Bonelli, E. (2001) *Corpus linguistics at work.* Amsterdam: John Benjamins.
Whistle, J. (1999) Concordancing with students using an 'off-the-web' corpus. *ReCall*, 11: 74-80.
Willis, D. (1991) *Collins COBUILD student's grammar*. London: HarperCollins.

Jeanne-Marie Debaisieux et Alex Boulton

CRAPEL– ATILF/CNRS, Université de Nancy 2