

DOCUMENTS AUTHENTIQUES, ORAL, CORPUS

Alex Boulton

Equipe CRAPEL, ATILF–CNRS, Nancy–Université

Ce numéro spécial des *Mélanges CRAPEL* est issu d'un colloque international organisé en décembre 2007 par le CRAPEL et ICAR. Les trois thématiques principales qui apparaissent dans son titre – *documents authentiques, oral, corpus* – préoccupent la didactique depuis plusieurs décennies. Dans un premier temps, cette introduction survole leur place dans l'histoire des *Mélanges*, la revue du CRAPEL¹ ; ensuite, nous revenons au présent avec un résumé des articles contenus dans ce numéro spécial.

Une première recherche par mot-clé « document authentique » dans les *Mélanges* nous renvoie au début des années 1970, à commencer par *Documents non didactiques et formation en langues* de Duda *et al.* (1972). On remarque le choix du terme « non didactique » (même si l'expression « documents authentiques » figure également dans cet article), choisi peut-être en raison de sa définition négative : « tout document en langue étrangère dont la finalité n'est pas l'enseignement de cette langue » (p. 2). Cette idée de négation est soulignée par Abe *et al.* (1979 : 2) : « on peut définir le document authentique, de façon négative, comme *tout document produit à des fins autres que l'apprentissage d'une langue seconde*. En termes positifs, c'est un énoncé produit dans une situation réelle de communication... »

D'autres types de documents, même si on ne peut pas les appeler « authentiques », ne sont pas exclus de l'apprentissage (Abe *et al.*, 1979 : 3). Par exemple, on peut défendre le réenregistrement de documents oraux authentiques pour compléter les corpus, améliorer la qualité sonore, ou même pour éliminer les « bavures idiolectales » – du moins à des fins d'expression orale puisqu'il serait « sans intérêt, et probablement dangereux, de faire produire à tous ceux qui utiliseront ce cours, les hésitations, les répétitions, les erreurs d'encodage... » (Holec, 1974 : 22). Si leur utilisation n'est pas à rejeter définitivement, la préférence au CRAPEL a toujours été pour des documents authentiques tels quels, car eux

1. Tous les numéros des *Mélanges CRAPEL* (ainsi que de son précurseur les *Mélanges pédagogiques*) depuis 1970 sont disponibles gratuitement en ligne en format pdf : <http://revues.univ-nancy2.fr/melangesCrapel/>.

seuls rendent compte de « la langue telle qu'elle est réellement, complexe et multiforme » (Duda *et al.*, 1972 : 3) ; on voit ici un précurseur de la devise de Sinclair (2004) : « Trust the text ». L'objectif du CRAPEL dans ce sens était principalement d'écarter les « documents (écrits ou oraux) créés de toutes pièces en vue de l'enseignement d'une langue » (Abe *et al.*, 1979 : 4) ; si ce type de texte construit n'est pas non plus à rejeter de façon dogmatique, on préconise « infiniment plus de textes authentiques que de textes construits » (Holec, 1974 : 22).

C'est Holec (1990) qui en fait une synthèse générale dans les *Mélanges avec Des documents authentiques, pour quoi faire ?* puisque beaucoup d'articles antérieurs sont, comme il en était coutume, des comptes-rendus d'expériences sur le terrain. Par exemple, dans leur premier article, Duda *et al.* (1972) se préoccupent principalement de l'exploitation de textes littéraires ; dans un deuxième, *L'exploitation didactique de documents authentiques* publié l'année suivante, ils passent à l'oral avec des comptes-rendus de trois expériences d'exploitation de documents authentiques, dont deux sont basées sur des émissions radio et des interviews spontanées (Duda *et al.*, 1973). En effet, la compréhension orale constitue une préoccupation majeure dans les travaux du CRAPEL, car elle représente une aptitude à la fois essentielle pour beaucoup d'apprenants mais aussi sous-exploitée dans de nombreux manuels et approches. Holec (1970) a signalé cette importance avec un article intitulé *Compréhension orale en langue étrangère* dans le tout premier numéro des *Mélanges pédagogiques*.

Ce petit retour aux débuts du CRAPEL montre que les documents authentiques et l'oral représentent des préoccupations de longue date, alors pourquoi donc le besoin d'un colloque international et d'un numéro spécial des *Mélanges* trente ans plus tard ? Plusieurs explications s'offrent à nous, à commencer par la constatation que les documents fabriqués sont toujours utilisés et que la compréhension orale reste une aptitude sous-entraînée – même si on peut espérer que cet état des choses se fait plus rare qu'il y a trente ans. Ensuite, lorsqu'on se replonge dans les premiers articles des *Mélanges* on peut être surpris par la fraîcheur des idées : rien que dans Duda *et al.* (1973) on retrouve non seulement les « documents authentiques » mais également l'« aptitude de compréhension orale » (p. 22), le contraste « écrit oralisé » vs. « spontané » (p. 16), ainsi que la notion de « sensibilisation » (p. 22). Ou encore Abe *et al.* (1979 : 3) qui devançant Widdowson (1998) de près de vingt ans lorsqu'ils se demandent : « peut-on extraire des énoncés de leur contexte, les détourner de leurs objectifs premiers dans le but d'en apprendre la langue, sans en détruire l'authenticité ? La réception, le décodage sont-ils forcément authentiques parce que le document l'est ? » Si ces concepts ne sont pas nouveaux ils sont toujours d'actualité ; il n'est donc guère surprenant que les articles dans ce numéro reviennent souvent aux mêmes questions qu'il y a trente ans. Mais en même temps le monde a changé et il y a du nouveau à dire ; nous laissons au lecteur le plaisir de le découvrir à travers les contributions à ce numéro spécial.

Tournons-nous maintenant vers le troisième mot-clé du titre du présent numéro : « corpus ». Encore une fois, le mot n'est pas nouveau dans la didactique des langues : Duda *et al.* (1973 : 4) parlent d'un « corpus d'expérimentation », et Holec (1990 : 65) d'un « *corpus for the discovery of language* ». Dans ce sens, le terme renvoie à un petit échantillon de documents à découvrir individuellement lors d'une écoute en temps réel. Cet usage du mot n'a pas disparu, comme en témoignent plusieurs articles dans ce numéro. Mais il a depuis les années 1970 acquis un nouveau sens lié à la linguistique de corpus, celui d'une collection électronique de millions de mots qui nécessite de nouvelles formes de consultation, en particulier à travers une lecture « verticale » des concordances. Les corpus oraux sont composés généralement de transcriptions écrites, ce qui peut paraître paradoxal pour l'apprentissage de l'oral ; mais les transcriptions peuvent être alignées avec le son, voire même avec la vidéo dans un corpus multimodal. Le tout rend possible une autre façon de découvrir la langue avec des recherches ciblées sur des échantillons immenses, soutenues par un va-et-vient entre écrit et audio. Ainsi les corpus n'ont pas comme seule utilité d'aider les concepteurs à décider quels éléments linguistiques enseigner, mais peuvent également servir en aval à l'apprenant directement dans ce que Tim Johns (par ex. 1988 ; Johns & King, 1991) a appelé le « *data-driven learning* », c'est-à-dire l'apprentissage de la langue à travers l'exploration des données linguistiques.

Par le passé, la France figurait parmi les pionniers de l'exploitation des grands corpus – même oraux – dans une perspective didactique. L'un des plus célèbres est bien entendu le *Français fondamental* (Gougenheim *et al.*, 1956/1964), qui a été fêté lors de son cinquantième anniversaire par un colloque international organisé à Lyon en 2005 par SIHFLES et ICAR ainsi que par un numéro spécial du *Français dans le monde* (Cortier & Bouchard, 2008). Le CREDIF, responsable du *Français fondamental*, explorait l'emploi de concordances déjà dans les années 1960, une idée développée par l'INALF avec ses « dictionnaires contextuels » de concordances sur microfiche pour différents domaines de spécialité (par ex. Descamps *et al.*, 1992). Le plus grand corpus « sérieux » (c'est-à-dire collecté à partir de sources soigneusement contrôlées et non pas à partir d'un recueil automatique sur Internet) du français qui existe aujourd'hui c'est *FranTexte* : plus de 200 millions de mots en collecte permanente par l'ATILF – Analyse et traitement informatique de la langue française, une unité mixte de recherche CNRS / Nancy-Université. Malgré ses nombreux atouts, ce corpus a des applications limitées en français langue étrangère (FLE) car il s'agit principalement de textes littéraires remontant jusqu'au 16^e siècle. Depuis quelques années, comme le constatent les deux premières contributrices à ce numéro, la France est à la traîne en matière de corpus, surtout de corpus oraux. Ce qui nous amène aux contributions de ce numéro spécial des *Mélanges CRAPEL*.

Deux articles issus des conférences plénières ouvrent ce numéro avec un survol des grandes questions de corpus oraux français. **Angela CHAMBERS** rappelle que seul « l'oral » peut servir à l'apprentissage de l'oral. S'interrogeant dans

un premier temps sur la nature « authentique » d'un corpus, elle contraste l'authenticité du texte à l'authenticité du contexte et de l'usage initialement prévue. Mais lorsque Widdowson (par ex. 1998) met ainsi en question la validité des corpus, Chambers remarque que cette logique amènerait à l'abandon de tout document authentique. Elle met l'accent plutôt sur l'authenticité de la tâche (une interrogation linguistique de corpus représentant une tâche authentique), et défend l'usage de « corpus pédagogiques » pour reprendre le terme de Braun (2005), plus faciles à connaître de par leur petite taille et dont l'exploitation ne se limite pas exclusivement aux concordances. Une discussion de trois corpus oraux français permet de situer ces arguments dans le contexte du FLE. **Jeanne-Marie DEBAISIEUX** continue dans la même lignée, constatant que le travail sur corpus prend du retard en FLE, en raison en partie du manque de corpus (oraux) disponibles. Toutefois, elle remarque que les outils et méthodologies de base se construisent sur des fondements plus anciens, notamment que l'exploitation d'un corpus repose sur l'apprentissage par la découverte à travers les documents authentiques. Elle retrace l'évolution de ces idées depuis les années 1980 avec un regard particulier sur la formation de l'apprenant, essentielle pour un usage autonome des « corpus de découverte ». Suit une analyse du potentiel des corpus oraux alignés texte-et-son, qui permet de constater que des lectures « horizontale » (traditionnelle) et « verticale » (concordances) sont complémentaires, chacune permettant un accès différent au discours. Ce potentiel sera exploité dans un projet CRAPEL en cours : FLEURON (Debaisieux & Boulton, 2007).

Les deux articles suivants se basent sur des corpus d'enregistrements de discours en salle de classe. Pour **Stéphanie O'RIORDAN**, l'objectif premier est la formation des futurs enseignants de langue. Ici elle se limite à une petite partie du corpus, à savoir des cours de français dispensés par des enseignants natifs et non natifs en Irlande. Cette approche permet d'analyser le recours à la langue maternelle à différents moments lors de l'enseignement de la grammaire. L'objectif n'est pas d'imposer aux futurs enseignants « la » méthode à adopter dans une culture qui prône une utilisation maximale de la langue cible, mais de présenter des extraits du corpus comme support de réflexion permettant une sensibilisation à diverses stratégies et possibilités. Ensuite, **Jean-Marc MANGIANTE** s'appuie sur les travaux de Braz (2007) pour l'analyse du discours des enseignants de disciplines non linguistiques (DNL). La constitution d'un corpus d'enregistrements réalisés en cours de géographie en Roumanie permet d'analyser l'alternance linguistique, et de réaliser une typologie de discours et de « rituels » en différentes situations de classe. Ce type de travail peut, selon Mangiante, aider l'harmonisation des rôles des enseignants de FLE (spécialistes de la langue) et de la discipline.

La fabrication de corpus par les étudiants fait l'objet des deux prochains articles. **Henry TYNE** décrit ce qui se passe lorsque les étudiants sont eux-mêmes responsables non seulement de la constitution mais aussi de la transcription et de l'analyse de leurs propres micro-corpus oraux. Toutes ces étapes sont perçues

comme des activités didactiques, même lorsque l'accent n'est pas sur l'apprentissage de la langue cible mais sur la variabilité de la langue pour un cours de sociolinguistique universitaire. Le dépouillement des réponses à un questionnaire détaillé permet une analyse quantitative des réactions des étudiants à ce processus ; une discussion de leurs commentaires permet un approfondissement de leurs perceptions. Il s'avère que l'ensemble des étudiants apprécie beaucoup le travail, même la transcription minutieuse, et estime avoir amélioré non seulement ses compétences sociolinguistiques mais aussi ses connaissances de la langue cible. **Jean-Marc HILGERT**² permet aux étudiants d'un Master FLE de découvrir les corpus tout en améliorant leurs compétences en TIC (technologies de l'information et de la communication) par la constitution de corpus oraux du français – conception, gestion, enregistrement, traitement des fichiers audio, transcription, création de glossaires et de pages web, etc. Cet article décrit les outils, étapes et consignes nécessaires pour mener à bien un tel projet, ainsi que l'évaluation d'une présentation finale par les étudiants, et montre comment ce type de « formation-production » est doublement motivant et utile puisque le corpus qui en résulte servira par la suite à des apprenants FLE. Son corpus est repris par **Emilia HILGERT**³, qui décrit l'exploitation en aval par les apprenants. Les techniques utilisées en salle informatique sont détaillées, de la transcription (« auto-dictée ») à l'écoute globale guidée. Un objectif majeur est la sensibilisation des étudiants à différents registres à partir du genre des « interviews » d'enseignants-chercheurs. Le corpus permet entre autres de comparer les particularités de l'oral et de l'écrit, puisque les transcriptions sont à nettoyer par les apprenants afin de les transformer en version écrite pour une mise en commun ultérieure.

Si la plupart des articles ici concernent le milieu universitaire, les deux suivants soulignent l'importance de prendre en compte l'usage réel de la langue à d'autres moments de l'apprentissage. **Emmanuelle GUERIN** s'intéresse ainsi à l'enseignement du français langue maternelle et surtout à la description grammaticale scolaire, un contexte qui nécessite de repenser la notion d'« authenticité » ainsi que l'opposition polaire « oral/écrit ». Une analyse de l'usage de *on* à l'oral mène à la conclusion d'une divergence entre les formes et usages enseignés (principalement appuyés par l'écrit littéraire) et ceux qui sont fréquemment attestés à l'oral (cf. Carter & McCarthy, 1995). Par ailleurs, il s'avère que l'oral est tout aussi complexe que l'écrit même littéraire, et il convient de l'analyser dans la situation de communication et non pas comme une forme « défailante » de l'écrit. **Hervé ADAMI** discute de l'emploi des documents authentiques chez les migrants, par définition dans un contexte d'immersion, autrement dit de langue seconde plutôt que de langue étrangère. Il remarque la nature « artificielle » des matériels même

-
2. Article accompagné de fichiers électroniques sur le CD annexe ainsi que sur le site des *Mélanges CRAPEL* : <http://revues.univ-nancy2.fr/melangesCrapel/>.
 3. Article accompagné de fichiers électroniques sur le CD annexe ainsi que sur le site des *Mélanges CRAPEL* : <http://revues.univ-nancy2.fr/melangesCrapel/>.

« réalistes » créés par les enseignants, ainsi que le manque de « naturel » lorsque des documents authentiques sont arrachés de leur contexte initialement prévu à des fins didactiques en salle de classe (cf. Widdowson, 1998). Il conclut qu'une simple dichotomie « authentique / non authentique » ne couvre pas la gamme des possibilités, et défend alors la notion d'un continuum d'authenticité.

Quels sont les facteurs responsables de la difficulté d'un document authentique oral ? **Alain KAMBER et Carine SKUPIEN**⁴ présentent quatorze extraits d'émissions radio à de futurs enseignants ainsi qu'à 52 apprenants de différents niveaux de compétence en langue. Les résultats d'un premier questionnaire montrent que les enseignants ont tendance à surestimer la difficulté par rapport aux apprenants. Par ailleurs, ces derniers désignent massivement le débit et le vocabulaire comme sources principales de difficultés plutôt que la nature spontanée de certains documents, avec leurs hésitations, interactions, bruits de fond, etc. Au contraire : l'écrit oralisé et les situations plus formelles sont perçues comme étant plus difficiles. Suite à une série de tests complémentaires, ces chercheurs concluent que le type de question et la nature du document interagissent pour déterminer la difficulté d'une activité de compréhension orale, et proposent une progression didactiquement motivée. **Nathalie SPANGHERO-GAILLARD et Pascal GAILLARD**⁵, quant à eux, ciblent un seul facteur souvent perçu par les enseignants comme source de difficulté : le rôle des bruits environnementaux dans la compréhension orale. Une série de trois courts dialogues a été présentée à des locuteurs natifs ainsi qu'à un groupe d'apprenants. Une première écoute avec bruits environnementaux permet une compréhension globale quasi-systématique, alors qu'aucun participant – natif ou non natif – ne réussit avec les dialogues seuls. Une deuxième écoute, une scène à la fois, donne des résultats médiocres, la conclusion étant que les bruits de fond ne sont pas une barrière mais une aide à la compréhension orale.

Plusieurs articles présentent des corpus, outils ou plateformes, à commencer par **Catherine CAWS** qui décrit le projet *FrancoToile* à la University of Victoria au Canada, un corpus de vidéos provenant de locuteurs francophones de par le monde ayant pour but une sensibilisation culturelle de la francophonie. L'apprenant peut travailler une vidéo et/ou la transcription dans son intégralité, ou bien effectuer des recherches par mot-clé. Une première expérimentation auprès d'une trentaine d'apprenants (majoritairement anglophones) de niveaux débutant à avancé se révèle très positive, autant dans les appréciations des apprenants que dans les résultats de tests. En conclusion, les TICs (notamment l'interface bilingue et l'exploitation d'hyperliens) semblent se prêter parfaitement à ce type d'outil multimédia, permettant une interaction dynamique qui serait impossible par l'intermédiaire d'autres types de

4. Article accompagné de fichiers électroniques sur le CD annexe ainsi que sur le site des *Mélanges CRAPEL* : <http://revues.univ-nancy2.fr/melangesCrapel/>.

5. Article accompagné de fichiers électroniques sur le CD annexe ainsi que sur le site des *Mélanges CRAPEL* : <http://revues.univ-nancy2.fr/melangesCrapel/>.

supports. **Sylvain DETEY, Chantal LYCHE, Atanas TCHOBANOV, Jacques DURAND et Bernard LAKS** enchaînent sur la notion de variabilité, domaine riche pour l'étude de corpus, analysée ici à travers le projet *Phonologie du français contemporain* (PFC). Ce corpus comporte quatre registres, allant d'une liste de mots lus jusqu'aux conversations informelles avec plus de 600 locuteurs à travers l'espace francophone dans le monde entier. Le PFC-EF vise l'enseignement du français, et une interface performante permet l'accès au corpus aligné texte-et-son et à ses ressources « prêtes à l'emploi » disponibles en ligne. **Lucia DRAGO** décrit une expérience d'apprentissage interculturel à partir de *GlossaNet*, un corpus en ligne de journaux francophones. Cet article expose en détails les étapes proposées à une vingtaine d'étudiants australiens du FLE pour découvrir les emplois contextualisés du terme-clé *identité nationale*. Une médiation par l'enseignante évite une longue formation aux outils informatiques mais permet néanmoins une analyse des concordances afin de cerner les colligations et collocations dominantes, et aussi une meilleure compréhension des connotations et autres aspects sémantiques. Une analyse de questionnaires et de blogs des étudiants indique que les activités sont bien accueillies et promeuvent une lecture de presse avertie. Enfin, **Joline BOULON** reprend les étapes d'apprentissage de Narcy (1990) pour l'utilisation de vidéos authentiques sur la plateforme SPIRAL utilisée à l'Université Lyon 1. Il s'agit d'une progression « bas-haut », en commençant par l'étape 0 qui sensibilise l'apprenant à différents types d'obstacles à la compréhension orale, notamment pour découper la chaîne sonore et préparer la prononciation des mots dans le document. L'étape 1 amène l'apprenant à une compréhension détaillée à travers des questions guidées sur des moments précis du document ; l'étape 2 vise une compréhension plus globale pour automatiser les processus ; l'étape 3 passe de la compréhension orale à la production, laissant plus de liberté à l'apprenant de recycler le langage à ses propres fins.

On termine ce numéro spécial des *Mélanges CRAPEL* avec un résumé de la table ronde qui avait servi à ouvrir le débat lors du colloque. Partant des transcriptions de leurs interventions et des questions posées par le public (ce qui constitue par lui-même un corpus oral avec d'éventuelles applications didactiques), les contributeurs réussissent à regrouper plusieurs fils conducteurs du débat : **Richard DUDA, Emmanuelle CARETTE, Francis CARTON, Chantal PARPETTE et Jean-Marc MANGIANTE.**

Nous remercions vivement tous les participants au colloque et tous les contributeurs à ce numéro des *Mélanges CRAPEL*, ainsi que les relecteurs pour leur travail indispensable : Virginie André, Sophie Bailly, Alex Boulton, Emmanuelle Carette, Francis Carton, Anne Chateau, Jeanne-Marie Debaisieux, Richard Duda, Jean-Marc Mangiante, Claude Normand, Myriam Pereiro, Chantal Parpette, Henry Tyne.

BIBLIOGRAPHIE

(Les articles des *Mélanges* sont disponibles en ligne : <http://revues.univ-nancy2.fr/melangesCrapel/>.)

- ABE, Danielle, Francis CARTON, Sam Michel CEMBALO & Odile REGENT, 1979. « Didactique et authentique : du document à la pédagogie. » *Mélanges pédagogiques*, p. 1-14.
- BRAUN, Sabine, 2005 « From pedagogically relevant corpora to authentic language learning contents. » *ReCALL*, 17/1, p. 47-64.
- BRAZ, Adelino, 2007. *Pour une réflexion sur l'enseignement bilingue*. Santander : Consejería de Educación de Cantabria.
- CARTER, Ronald & Michael McCARTHY, 1995. « Grammar and the spoken language. » *Applied Linguistics*, 16, p. 141-158.
- CORTIER, Claude & Robert BOUCHARD (eds.) 2008. *Le français dans le monde. Quel oral enseigner, cinquante ans après le Français fondamental ?* Paris : CLE international FIPF.
- DEBAISIEUX, Jeanne-Marie & Alex BOULTON, 2007. « Alors la question c'est... ? Questions pragmatiques et annotation pédagogique des corpus. » *Cahiers de l'AFLS*, 13/2, p. 31-59. <http://www.afsl.net/cahiers/cover13.2.php>, page consultée le 16/10/08.
- DESCAMPS, J-L, M-A. MOCHET, T. LEWIN, B. LAMIZET & D. COSTE, 1992. *Sémantique et Concordances (suivi du Dictionnaire contextuel de français politique)*. Paris : INALF / Klincksieck.
- DUDA, Richard, Edith ESCH & Jean-Pierre LAURENS, 1972. Documents non didactiques et formation en langues. *Mélanges Pédagogiques*, p. 1-47.
- DUDA, Richard, Jean-Pierre LAURENS & Sylviane REMY, 1973. « L'exploitation didactique de documents authentiques. » *Mélanges pédagogiques*, p. 1-24.
- GOUGENHEIM, G., R. MICHEA, P. RIVENC & A.SAUVAGEOT, 1956/1964. *L'élaboration du français fondamental (1^e degré)*. Paris : Didier.
- HOLEC, Henri, 1970. « Compréhension orale en langue étrangère. » *Mélanges pédagogiques*, p. 1-16.
- HOLEC, Henri, 1974. « Cours initial d'anglais oral : une approche fonctionnelle. » *Mélanges pédagogiques*, p. 13-28.
- HOLEC, Henri, 1990. « Des documents authentiques, pour quoi faire ? » *Mélanges pédagogiques* ; p. 65-74.
- JOHNS, Tim, 1988. « Implications et applications des logiciels de concordance dans la salle de classe. » *Les langues modernes*, 82/5, p. 29-45.

JOHNS, Tim & Philip KING (eds.) 1991. *Classroom Concordancing. English Language Research Journal*, 4.

NARCY, Jean-Paul, 1990. *Apprendre une langue étrangère, didactique des langues : le cas de l'anglais*. Paris : Les Editions d'Organisation.

SINCLAIR, John McHardy, 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

WIDDOWSON, Henry, 1998. « Context, community and authentic language. » *TESOL Quarterly*, 32/4, p. 705-716.

