De la *Capsule conversationnelle* à la classe

De la collecte aux usages

Olivier BAUDE(1), Catherine BRUMELOT(2)

(1) Université Paris Nanterre - UMR MoDyCo 7114 & CNRS-TGIR Huma-Num

(2) INALCO - Plidam

Résumé : La Capsule conversationnelle est un dispositif imaginé dans une collaboration art et sciences pour collecter un corpus de conversations utilisable en didactique des langues et des cultures. L'objectif ici est de présenter un projet né d'un besoin de documents authentiques dans le cadre d'un cours de FLE et, au-delà, de présenter des choix méthodologiques de constitution et d'exploitation d'un corpus de données langagières.

Mots-clés : apprentissage sur corpus, collecte, FLE, humanités numériques, usages.

***Abstract**: The interactive capsule is a framework thought in an art and science cooperation and intended to collect a corpus of conversations that can be used in language and culture teaching. The goal is first to present a project that originated from a need for authentic materials in classes where French is taught as a foreign language and more globally to present methodological choices to establish and use a corpus of language data.*

***Keywords**: data collection, data driven learning, digital humanities, French as a foreign language, uses.*

Introduction

Lors d'une micro projet mis en place en SDL (enseignement de la pragmatique et de l'interculturalité) et FLE (enseignement et didactique du FLE) et d'un projet de recherche (*ESLO : corpus de français parlé pour la recherche en linguistique et la didactique),* nous avons régulièrement été confrontés aux questions suivantes : comment stocker, traiter, diffuser, rendre visible, exposer, signaler, archiver … des corpus audiovisuels constitués de conversations quotidiennes ; comment mettre ces corpus numériques à disposition des chercheurs, linguistes, didacticiens mais également des enseignants ?

De ces questions est né un projet atypique et exploratoire qui prend la forme d'une *capsule conversationnelle*. Il a pour objectif la production et l'utilisation de corpus « écologiques » exploitables dans les classes.

Sur un plan tout à la fois théorique, technique et expérimental, le projet présenté ici traite de l'un des sujets les plus critiques dans l'évolution du monde multiculturel et multilingue contemporain, celui de la compréhension d'autres cultures et sociétés en collectant, archivant et analysant la parole vive des locuteurs de ces cultures et en utilisant ces données pour l'enseignement, la formation et la communication multilingue et interculturelle.

En répondant à des critères précis permettant l'archivage, la réutilisation et l'exploitation de collections de productions langagières, la linguistique de corpus et plus récemment le développement des humanités numériques apportent des solutions pour gérer des données.

Parallèlement aux grandes institutions de conservation et d'accès aux documents (BnF, INA), il existe ainsi différentes initiatives de plateformes qui partagent les mêmes caractéristiques et les mêmes objectifs (Medihal, Cocoon, Ortolang, Nakala-Huma-Num, …).

Toutefois, ces plateformes n'apportent qu'une partie intermédiaire des besoins. Elles doivent donc être complétées en amont par un travail de collecte, et en aval par des usages, notamment pédagogiques. Or ces trois éléments : collecte, corpus, usages sont liés par des contraintes théoriques et méthodologiques fortes qui créent une interdépendance qu'il convient d'expliciter et de maitriser.

Ainsi, le recours aux documents authentiques, ou tout au moins à des données langagières structurées, annotées et documentées est un enjeu repéré de l'enseignement du FLE/FLS et de la didactique des langues depuis plusieurs dizaines d'années. Mais, force est de constater qu'il n'existe pas actuellement de grandes bases de données permettant de répondre à ce besoin.

À l'autre bout de la chaine et faisant écho au projet de Ferdinand Brunot imaginé au début du XX^{e} siècle, quand ce dernier se lança dans le vaste projet des Archives de la parole et des collectes régionales, *La Capsule conversationnelle* est l'un des dispositifs de collecte sur l'ensemble du territoire français puis francophone. Elle facilite la captation de conversations ordinaires (discussions quotidiennes autour d'un café par exemple) tout en permettant la restitution dans un cadre artistique. Imaginé dans un dialogue entre artistes et linguistes, ce dispositif répond, sous la forme d'un espace physique convivial, aux exigences des plateformes de gestion des corpus et aux objectifs d'usages scientifiques et didactiques.

Il permet en outre de préserver et rendre accessible des corpus oraux et d'en faciliter la réutilisation à des fins variées.

1 Des ressources ancrées dans la réalité

L'œuvre interactive appelée « capsule conversationnelle » a été conçue par l'artiste Guy Kayser et l'informaticien Gérard Parésys en collaboration avec des linguistes.

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799



Cette capsule modulable s'ouvre, se ferme au gré des besoins. Elle est équipée d'un dispositif numérique : tablette, micro, ordinateur afin de permettre le recueil de données langagières et donc la constitution d'un corpus dédié à différents usages. Le traitement numérique des données linguistiques ainsi que leur réutilisation au sein de plateformes technologiques en sont facilités.

La collecte s'organise actuellement *via* une application Web connectée à CoCoON[1] facilitant le recueil de consentement et le renseignement des métadonnées.

Grâce à l'accès à cette plateforme, outre le fait de pouvoir capter des interactions, la capsule permet l'écoute de documents sonores parmi un fonds de 50 langues.

Bien que la pertinence de la manipulation des corpus pour l'apprentissage des langues soit démontrée, l'emploi de telles données en classe reste singulier. De multiples raisons en sont à l'origine, parmi lesquelles le manque de formation des enseignants à la didactique de corpus et plus globalement à la sociolinguistique.

En effet, rares sont celles et ceux formé.e.s à l'utilisation de corpus en salle de classe. Le recours aux corpus est en effet généralement assimilé à un « défi » pour les enseignants. (Debaisieux, 2009)

Il semble pourtant nécessaire à bon nombre d'enseignants d'avoir recours à des supports audiovisuels et notamment à des supports oraux, plus authentiques que ceux proposés dans les manuels

[1] CoCoON pour « COllections de COrpus Oraux Numériques » est une plateforme technique qui accompagne les producteurs de ressources orales, à créer, structurer et archiver leurs corpus ; un corpus pouvant se composer d'enregistrements (en général audio) accompagnés éventuellement d'annotations de ces enregistrements. https://cocoon.huma-num.fr/exist/crdo/ [29 décembre 2019]

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

« Le décalage qui existe entre ce que montrent les données attestées, accessibles sous forme de corpus, et les présentations de phénomènes dans les manuels laisse parfois à désirer, et ne sont pas toujours prises en compte » (Tyne, 2012 : 5).

En effet, « les dialogues construits, inventés, créés pour les manuels de langue sont éloignés de la réalité des pratiques langagières en usage et ne proposent pas une exposition à la langue cible aux apprenants. Ils proposent souvent des interactions qui ne sont pas naturelles, une langue qui n'est pas utilisée ou qui n'existe que dans les manuels de langues » (André, 2018 : 73).

Il parait donc primordial de disposer de ressources audiovisuelles les plus « authentiques » possible, telles que des interactions entre locuteurs adultes, liées au contexte d'enseignement dans lequel se trouvent les étudiants de langue étrangère d'une part et liées à des situations auxquelles ces derniers sont confrontés dans leur vie de tous les jours d'autre part.

En tant que lieu « où l'on se sent en sécurité », cette capsule a le mérite de créer une situation favorable au recueil de conversations libres et/ou semi-dirigées.

Différents types d'interaction peuvent être par conséquent captés : des interactions naturelles authentiques telles que des conversations à bâtons rompus autour d'un café ou des interactions sollicitées tels que des entretiens sur différents sujets, des récits de vie, des anecdotes, etc. (André, 2016)

Le but est ici de collecter de la parole vive dans un ensemble de langues représentées et enseignées à l'Inalco : arabe, français, hongrois, malgache, tchèque, etc. Il est également d'élaborer *in fine* une méthodologie pour la constitution, l'archivage et l'exploitation de données linguistiques, notamment dans le cadre de l'enseignement (en présentiel et en distanciel) de langues dites rares ou à petits effectifs.

2 L'apport dans un contexte de faible didactisation

Ces corpus constitueraient ainsi un outil majeur prenant en considération les points de vue des étudiant.e.s en rapport avec leurs préoccupations et environnement immédiats : les études, les examens, les enseignants, leurs difficultés, ce qu'ils envisagent, leurs motivations, etc.

Ces conversations entre locuteurs natifs fourniraient un matériau pour la création de supports pédagogiques par et pour les enseignants et ce, dans les différentes langues enseignées/apprises.

Il est possible d'imaginer que ces corpus d'interactions d'une pluralité de locuteurs, aux profils et besoins variés, tout en permettant d'observer la langue produite, permettraient d'envisager une meilleure individualisation des parcours d'enseignement/apprentissage.

Ils pourraient ainsi donner lieu à un travail d'observation et de réflexion (phonologie, grammaire, variations) (Boulton & Tyne, 2014 : 9).

De plus, puisque les locuteurs de L2 acquièrent la dimension variationnelle de la langue principalement au contact de locuteurs natifs (Mougeon *et al.*, 2002), ces corpus

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

participeraient d'une image de la langue en usage aujourd'hui et permettraient ainsi d'observer et de réfléchir aux aspects variationnels.

« La nature du corpus doit permettre la prise en compte des phénomènes de variation de genre ou de registre et l'éventuelle identification et analyse des erreurs. » (Geneviève Bordet, 2014. ASP : 66)

Pour être exploitables, ces corpus se doivent d'être explorables et donc documentés et catégorisés.

Sont notamment et a minima enregistrées les métadonnées documentaires suivantes :

- Date, lieu, …
- Âge des locuteurs
- Profil des locuteurs (étudiants/enseignants/etc.)
- Thème de l'interaction
- Mots clés associés aux thèmes abordés

ainsi que les catégories / annotations :

- Niveau CECRL
- Thème
- Lexique
- Grammaire
- Phonologie
- Socioculturel
- Sociolinguistique.

Les données publiées étant ainsi *Facile à trouver, Accessible, Interopérable et Réutilisable (FAIR)*, ce dispositif répond donc parfaitement à la perspective préconisée aujourd'hui.

3 Un exemple de piste d'exploitation

Pour ce qui concerne le cas du français langue étrangères, Noda explique ainsi que (2011 : 56)

> « Les marqueurs discursifs, tels que *ben, bon, dis, écoute, hein, n'est-ce-pas, tiens* sont des éléments marginaux dans les programmes de français langue étrangère ; cependant leurs emplois ne peuvent manquer de poser

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

d'importants problèmes de compréhension et d'acquisition chez les non natifs. Or, il est nécessaire d'ancrer l'apprentissage de la langue dans une époque, dans un contexte social identifié. Cela permettra de comprendre et de faire comprendre les valeurs pragmatiques de ces différents marqueurs discursifs. »

En effet, de nombreux autres marqueurs discursifs, tels que *du coup, genre, grave, non, OK, quoi, trop, trop pas, sérieux,* etc. sont quant à eux de véritables freins à la compréhension pour les étudiants allophones.

Nous pouvons prendre deux exemples d'enregistrements à propos du marqueur discursif *genre*, l'un dit « sauvage » l'autre capté dans la capsule.

Tandis que l'enregistrement « sauvage » pose différents problèmes (questions juridiques, qualité sonore, etc.), celui réalisé dans la capsule conversationnelle offre la qualité sonore nécessaire pour une exploitation institutionnelle.

Genre « enregistrement sauvage » Inalco, janvier 2018	**Genre « capsule »** Orléans, 2018
« Vraiment des trucs, genre … » « J'pense qu'ça va être une exercice genre … » « Ok, c'est genre, OK » « Genre, en mode … « « Genre le médicament qui est … » « Genre, c'weekend, j'ai dû … » « Ah, oui, genre elle tombe, … » « C'est genre … » « Par contre, faut faire gaffe, genre … »	« Et euh et euh genre tu vois. En fait, savoir qu'il était en couple, genre j'étais un peu vicieuse. »

D'un point de vue didactique, outre des activités de repérage (observation) et de recherche des valeurs de *genre* dans ces deux conversations (analyse), il est intéressant de mener une séquence visant à développer la compréhension et l'utilisation de ce marqueur discursif afin d'ancrer l'apprentissage dans la langue d'aujourd'hui, c'est-à-dire à acquérir des compétences sociolangagières interactionnelles. De nombreuses ressources sont disponibles pour permettre de travailler la compréhension orale[2] ainsi que la compréhension écrite.[3]

4 En guise de conclusion

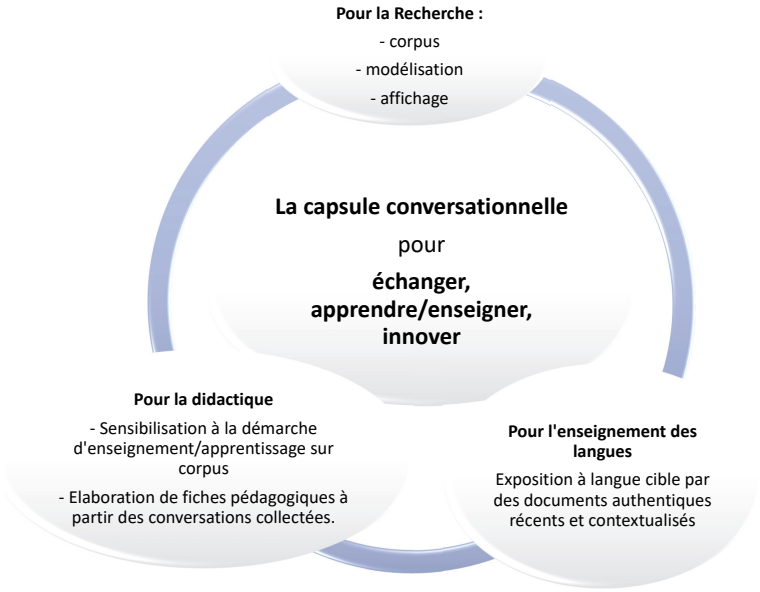
Il est reconnu que l'enseignement des langues est difficile à concevoir en l'absence de données orales provenant des échanges spontanés entre locuteurs. Or, ces données faisant régulièrement défaut, ce lieu d'échanges qu'est la capsule conversationnelle permettrait la collecte de parole vive entre locuteurs natifs, utile pour la recherche, pour l'enseignement des langues et plus généralement pour la didactique des langues. La

[2] https://langue-francaise.tv5monde.com/decouvrir/devenir-expert/lhumeur-de-linda/episode-4 (1.01 – 3.00)
https://www.francetvinfo.fr/replay-radio/modes-de-vie/les-tics-de-langage-ces-petits-mots-moteurs_1723295.html

[3] https://www.journaluniversitaire.com/tics-de-langage/
http://www.leparisien.fr/societe/ce-que-les-tics-de-langage-veulent-dire-20-04-2018-7675581.php

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

méthodologie utilisée serait également un moyen de sensibiliser les futurs enseignants à la démarche d'apprentissage sur corpus.



Ce projet intrinsèquement pluridisciplinaire puisqu'il implique la linguistique (linguistique de corpus, sociolinguistique, analyse du discours), la didactique des langues, les techniques et méthodes de collecte des données en SHS et, plus généralement, les humanités numériques (archivage et exploitation de corpus et méthodes numériques) a pour double objectif de :

- recueillir, archiver, décrire et exploiter pédagogiquement des données orales multilingues documentant la culture quotidienne dans des « grandes » langues vernaculaires et des langues dont le rayonnement géographique est plus restreint pour constituer un corpus de recherche en didactique des langues et des cultures,
- concevoir et expérimenter une plateforme à la fois scientifique et technique pour la collecte de la « parole vive », son archivage et ses exploitations pédagogiques.

De plus, plutôt que de penser en « grand corpus de référence » et en plateforme unique, l'intérêt serait d'avoir une agrégation de corpus qui répondrait plus facilement à la diversité des contenus et donc à la diversité des usages.

Bibliographie

Linguistique de corpus, archives numériques, Web sémantique

BAUDE, O., MARCHELLO-NIZIA, C., MONDADA, L., BLANCHE-BENVENISTE, C., CALAS, M.-F., CAPPEAU, P., CORDEREIX, P., LAMBERTERIE, I. D., GOURY, L., & JACOBSON, M. (2006). Corpus oraux : guide des bonnes pratiques. (O. Baude, éd.). CNRS Éditions et Presses universitaires d'Orléans.

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799

DI VITO S., L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE, Linx [En ligne], 68-69 2013, mis en ligne le 19 novembre 2015, consulté le 15 avril 2019. URL : http://journals.openedition.org/linx/1519 ; DOI : 10.4000/linx.1519

NODA, H. (2011). *« Emplois des marqueurs discursifs et intersubjectivité : autour de hein »,* Cahiers de praxématique [En ligne], 56

OUVRARD, L. (à paraître 2019), Ressources multimédias pour une didactique renouvelée de la langue et de la littérature malgaches, Paris : Éditions des Archives Contemporaines, 187 pages. ISBN : 9782813002853.

STOCKINGER, P.(éd.), 2011, *Nouveaux usages des archives audiovisuelles numériques*, Lavoisier-Hermès.

STOCKINGER, P., 2012, *Audiovisual Archives. Digital Text and Discourse Analysis*, ISTE-JohnWiley.

WEBER, C. (2013). *Pour une didactique de l'oralité : enseigner le français tel qu'il est parlé.* Didier/Paris.

Pédagogie numérique et enseignement des langues

ANDRE, V. (2018). Nouvelles actions didactiques : faire de la soicilinguistique de corpus pour enseingner et apprendre à interargir en français langue étrangère. Action Didactique, 1, 71-88. http://univ-bejaia.dz/pdf/ad1/Andre.pdf

ANDRE, V., BOULTON, A., CANUT, E., TYNE, H. (2013) (éds.). « Corpus et apprentissage du français ». Linx, 68-69.

BOULTON, A., & CHAMBERS, A., DEBAISIEUX, J-M., *et al.* (2009). Des documents authentiques oraux aux corpus : questions d'apprentissage en didactique des langues. Mélanges CRAPEL, n° 31.

DEBAISIEUX, J.-M. (2009). « Les corpus oraux : situation, exploitation linguistique, bilan et perspectives », Mélanges Crapel, 31, p. 36-56.

FORLOT, G., OUVRARD, L. (dir.), 2020, *Variation linguistique et enseignement des langues. Le cas des langues moins enseignées*, Paris : Presses de l'INALCO.

OUVRARD, L. (2015). "Conversations en ligne : description, perspectives", in Lemay, D., Potolia, A. (dir.), Enseignement / Apprentissage des langues et pratiques numériques émergentes, Paris ; Éditions des archives contemporaines.

TYNE, H. (2012). Variation dans l'enseignement-apprentissage d'une langue 2. Le Français Aujourd'hui, Armand Colin / Dunod ; Association française des professeurs de français ; Association française des enseignants de français (AFEF), 176, p. 103-112.

https://hal.archives-ouvertes.fr/hal-00702751/document (consulté le 29 décembre 2019)

WEBER, C. (2013), *Pour une didactique de l'oralité. Enseigner le français tel qu'il est parlé* http://www.editionsdidier.com/article/pour-une-didactique-de-l-oralite/, Paris, Didier, 334p.

WACHS S. et WEBER C. (dir.) (2016). Rapport effectué pour l'observatoire de la langue

française dans le cadre de l'appel à projet de l'observatoire des pratiques linguistiques

(DGLFLF) *Numérique et textualité : observation, description et analyse des pratiques contemporaines*

Sitographie

https://langue-francaise.tv5monde.com/decouvrir/devenir-expert/lhumeur-de-linda/episode-4 (1.01 – 3.00) [13/07/2020]

https://www.journaluniversitaire.com/tics-de-langage/ [13/07/2020]

http://www.leparisien.fr/societe/ce-que-les-tics-de-langage-veulent-dire-20-04-2018-7675581.php [13/07/2020]

https://www.francetvinfo.fr/replay-radio/modes-de-vie/les-tics-de-langage-ces-petits-mots-moteurs/1723295.html [13/07/2020]

Published in : Louise Ouvrard et Odile Farge (dir.), *Corpus audiovisuels*, Editions des archives contemporaines, ISBN : 9782813003799, doi : 10.17184/eac.9782813003799