



# Exploitation, exploration et appropriation de l'archive. Étude sémiotique de la didactisation de la base de données CLAPI

Julien Thiburce, Biagio Ursi

## ► To cite this version:

Julien Thiburce, Biagio Ursi. Exploitation, exploration et appropriation de l'archive. Étude sémiotique de la didactisation de la base de données CLAPI. Signata - Annales des sémiotiques, 2021, Sémiotiques de l'archive, 12. hal-03104340

**HAL Id: hal-03104340**

**<https://hal.science/hal-03104340>**

Submitted on 16 Dec 2022

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License



**Signata**

Annales des sémiotiques / Annals of Semiotics

**12 | 2021**

**Sémiotiques de l'archive**

---

## Exploitation, exploration et appropriation de l'archive

Étude sémiotique de la didactisation de la base de données CLAPI

**Julien Thiburce et Biagio Ursi**

---



### Édition électronique

URL : <https://journals.openedition.org/signata/3040>

DOI : 10.4000/signata.3040

ISSN : 2565-7097

### Éditeur

Presses universitaires de Liège (PULg)

### Édition imprimée

ISBN : 9782875622693

ISSN : 2032-9806

Ce document vous est offert par Bibliothèque Diderot de Lyon - ENS



### Référence électronique

Julien Thiburce et Biagio Ursi, « Exploitation, exploration et appropriation de l'archive », *Signata* [En ligne], 12 | 2021, mis en ligne le 31 mai 2021, consulté le 16 décembre 2022. URL : <http://journals.openedition.org/signata/3040> ; DOI : <https://doi.org/10.4000/signata.3040>

---

Ce document a été généré automatiquement le 8 juin 2021.



Creative Commons - Attribution 4.0 International - CC BY 4.0

<https://creativecommons.org/licenses/by/4.0/>

---

# Exploitation, exploration et appropriation de l'archive

Étude sémiotique de la didactisation de la base de données CLAPI

Julien Thiburce et Biagio Ursi

---

## Introduction

- 1 Dans un dialogue entre étude sémiotique et recherche linguistique sur corpus, nous traitons le thème de l'archive à l'aune des pratiques mobilisées dans la recherche en sciences du langage pour enquêter sur des situations ordinaires et professionnelles, de leur enregistrement à leur implémentation et exploitation en tant que données en environnement numérique. Nous procédons à l'étude de la plateforme CLAPI (Corpus de Langues Parlées en Interaction), une ressource numérique qui constitue une base de données spécifique. En partant de notre propre expérience d'utilisateurs de cette base de données, nous retraçons notre parcours de l'interface et problématisons ensuite l'appropriation de cette archive d'enregistrements d'interactions langagières en situations naturelles.
- 2 À partir de cette appropriation, des pistes d'exploitation se profilent pour les usagers qui font de cette base de données un terrain de recherche. D'une part, en tant qu'objet à observer, l'archive devient un environnement de travail aux frontières mouvantes : le processus d'archivage, dans le temps, transforme l'épaisseur qualitative du corpus et la richesse quantitative des données traitées et mises à disposition. D'autre part, les finalités de l'utilisation de l'archive se transforment au fil des requêtes réalisées et des questions de recherche soulevées : les attentes initiales et préalables à la consultation de la base de données se trouvent petit à petit précisées et redéfinies. En tant qu'utilisateurs, notre pratique de la base de données peut se formuler dans les termes d'une double découverte : d'un côté, celle des potentialités de l'objet manipulé et de l'environnement de travail expérimenté ; de l'autre, celle de l'étendue et de la configuration de questions de recherche sollicitées par la consultation des données.

- 3 Nous questionnons en effet les potentialités offertes par la base de données à divers acteurs du monde de la recherche, s'inscrivant dans des projets variés et dans différents scénarios d'exploitation. Notre étude est traversée par une question de fond, celle de la nécessité et de la pertinence d'opérer une mise à jour des ressources sémiotiques ainsi qu'un ajustement des outils mis à disposition pour *exploiter, explorer* et *s'approprier* cette base de données – notions développées dans la section 4 de ce texte. Par ailleurs, un regard sur l'ergonomie de l'interface numérique et sur l'alimentation ou la réactualisation du contenu de l'archive sont indispensables.

## 1. Les bases de données du français parlé : des ressources pour les sciences du langage

- 4 La réflexion que nous proposons dans cet article est focalisée sur l'archive en tant qu'objet de recherche et ressource exploitable dans le cadre d'études en sciences du langage. Avant de nous pencher sur un cas spécifique, celui de la plateforme CLAPI et de la base CLAPI-FLE, nous présentons une vue panoramique sur les bases de données linguistiques en contexte francophone.
- 5 Depuis une vingtaine d'années, la réalisation et le développement des bases de données de l'oral en français ont impulsé un nouvel élan dans les sciences du langage. Cela a amené les chercheurs à proposer des bilans sur les ressources existantes (Bruxelles *et al.* 2009 ; Avanzi *et al.* 2016a) et des manuels pour la constitution de données (Baude *et al.* 2006). Les bases de données francophones, qui se sont affirmées en retard par rapport à d'autres langues européennes comme l'anglais ou le néerlandais (Bilger & Blanche-Benveniste 1999), sont désormais des repères pour les chercheurs travaillant dans tous les domaines de la linguistique et s'intéressant aux données naturelles de français parlé.
- 6 La nature des données présentées dans les bases de données francophones illustre la variété des traditions de recherche et aussi les objectifs scientifiques qui y sont attachés. Dans cette section, nous présentons trois bases de données à titre d'exemples ; d'autres projets portant sur la langue parlée seront également présentés.
- 7 La base ESLO (Enquêtes Sociolinguistiques à Orléans), développée par le laboratoire LLL d'Orléans, s'articule autour de deux « réservoirs » :
- ESLO1, « réalisé entre 1968 et 1971 [...] comprend 470 enregistrements d'une durée totale de 318 heures, ce qui représenterait [...] 4,5 millions de mots » ;
  - ESLO2, « en cours de réalisation depuis les années 2000, affiche un objectif de plus de six millions de mots pour 450 heures d'enregistrements » (Baude & Dugua 2016, § 32-33).
- 8 À partir d'un projet didactique, qui s'est concrétisé en une méthode d'enseignement du français langue seconde fondée sur des documents authentiques (Biggs & Dalwood 1976), l'objectif de ESLO1 était la documentation du français parlé en région orléanaise au prisme d'une variété de situations. Ceci est également l'objectif de ESLO2, plus riche en termes de contextes documentés. La présence des mêmes locuteurs dans les deux corpus représente un facteur primordial pour des études longitudinales et comparatives, dans une perspective sociolinguistique et, surtout, pour des recherches en diachronie – voir la contribution de Lotfi Abouda et Marie Skrovec (2018) pour un exemple de corpus ressource sur l'étude des formes grammaticales sur diachronie courte.

- 9 Le programme « Phonologie du Français Contemporain (PFC) : usages, variétés, structure » répond à une exigence au sein de la communauté des chercheurs : « la construction d'un important corpus de référence permettant de rendre compte de la diversité des usages oraux du français dans l'ensemble de la francophonie » (Racine *et al.* 2016, § 2). Les enquêtes menées sur le terrain s'articulent autour de quatre tâches s'inspirant des travaux sociolinguistiques de Labov : la lecture d'une liste de mots, la lecture d'un texte, un entretien semi-dirigé et un entretien libre entre deux ou trois locuteurs. À ce jour, le corpus PFC contient 16 enquêtes anonymisées (soit 164 locuteurs) et plus de 40 autres enquêtes sont en cours de traitement. Trois publics sont concernés<sup>1</sup> : les chercheurs travaillant dans les différents domaines de la linguistique (voir notamment les études phonétiques et phonologiques sur le *schwa* et la liaison) ; les enseignants/apprenants de français langue étrangère, première ou seconde ; le grand public intéressé par les accents du français ou par le patrimoine linguistique francophone en général. Les questions liées à la didactique et à l'apprentissage du français ont émergé très vite comme des axes de recherche pour la constitution de corpus ultérieurs. Ainsi, le projet InterPhonologie du Français Contemporain (IPFC), qui documente les productions d'apprenants de français langue étrangère, et le sous-projet PFC-Enseignement du Français (PFC-EF), destiné à l'exploitation pédagogique du corpus PFC, ont vu le jour dans les années 2000.
- 10 La base CLAPI (Corpus de Langue Parlée en Interaction) représente la ressource de référence pour les recherches de linguistique interactionnelle (mais des exploitations s'inscrivant dans d'autres approches sont également possibles) à partir d'une très grande variété de situations sociales : réunions de travail dans différents cadres, interactions de service, interactions en site commercial, visites privées, repas familiaux et amicaux, visites guidées, consultations médicales, appels téléphoniques privés et professionnels, situations de classe, conversations en ligne (voir groupe ICOR 2016). Les données présentes et requêtables dans cette base de données outillée sont dans une large mesure des données qui documentent des « *naturally occurring interactions* », des enregistrements qui « *should catch "natural interaction" as fully and faithfully as is practically possible* » (Ten Have 1999, p. 48).
- 11 Actuellement (octobre 2018), CLAPI comprend 66 « corpus » correspondant à 140 transcriptions de situations interactionnelles variées, alignées au signal audio(visuel), disponibles dans plusieurs formats. Parmi ces données, 46 heures sont téléchargeables sans condition d'accès et 63 heures (soit plus de 90 % de la base) sont interrogeables par des outils dédiés. Comme nous le verrons plus loin (§ 2.2), ces outils représentent de véritables ressources pour les études de linguistique de corpus, avec la possibilité d'un aperçu quantitatif et d'un éclairage qualitatif à la fois. L'étude des particules, comme par exemple « oh » (Groupe ICOR 2008a) ou, plus généralement, des marqueurs discursifs (Balthasar *et al.* 2007 ; Groupe ICOR 2008b, 2015), dans leurs caractérisations conversationnelles et séquentielles à travers l'étude de contextes différents, représente une mobilisation privilégiée de la base de données CLAPI. Un choix d'extraits représentatifs issus des corpus de CLAPI a donné lieu à la ressource CLAPI-FLE, qui est née à partir de l'exigence de didactiser des données conversationnelles, enregistrées en situation naturelle, comme nous le verrons plus loin (§ 2).
- 12 Ces trois archives<sup>2</sup> offrent une vision partielle mais suffisamment représentative des ressources actuellement mises à disposition de différentes communautés d'utilisateurs. D'autres pourraient être mentionnées, comme par exemple le « Corpus International

Écologique de la Langue Française » (CIEL-F), rassemblant des données interactionnelles constituées dans différents contextes et dans plusieurs pays francophones dans le monde (Mondada & Pfänder 2016), ou le corpus OFROM, visant à une cartographie des variétés de français parlées en Suisse romande (Avanzi *et al.* 2016b). Dans la recherche en linguistique, plusieurs projets financés par l'Agence Nationale de la Recherche ont aussi commencé à mener une réflexion à large échelle et à proposer une méthodologie d'annotation multi-niveau. Par exemple, le projet ORFÉO « Outils et Ressources sur le Français Écrit et Oral » (Benzitoun *et al.* 2016) a rassemblé quatre millions de mots de français oral et six millions de mots de français écrit à partir de corpus existants<sup>3</sup>. La mise à disposition de données annotées avec des formats unifiés et des transcriptions alignées au signal, s'est accompagnée à l'élaboration de protocoles d'annotation ou de manuels illustratifs, qui ont formalisés les choix opérés respectivement pour les annotations manuelles et pour les annotations semi-automatiques (Debaisieux et Benzitoun 2020).

- 13 Comme nous l'avons vu à l'aide de ESLO, PFC et CLAPI, les bases de données ne sont pas de simples ressources pour les communautés de chercheurs en linguistique ; elles questionnent aussi les usages que d'autres publics pourraient en faire dans une perspective appliquée. L'exploitation pour des finalités pédagogiques, en tant que dimension présente en amont du projet ou émergente au fil de sa réalisation, constitue une constante et nous amène à étudier un cas spécifique : celui de la plateforme CLAPI et la constitution de son volet enseignement.

## 2. (Re)configurer l'architecture du site : de CLAPI à CLAPI-FLE

### 2.1. Évaluer les besoins et mettre en œuvre un protocole

- 14 En proposant un « volet enseignement » de la base de données CLAPI, la constitution de CLAPI-FLE opère un passage d'un matériau conçu pour la recherche en sciences du langage et, plus précisément, en linguistique interactionnelle, à un matériau redéfini et ré-encadré pour une application didactique. CLAPI-FLE a en effet été créé pour répondre aux demandes et aux besoins des enseignants de FLE comme des apprenants. Cette conception d'une base de données appropriée à des usages particuliers dans le domaine de l'enseignement du français langue étrangère ne s'est déroulée ni hors sol, ni en abstraction. En concertation avec plusieurs centres d'étude de français à l'étranger, l'équipe LIS du laboratoire ICAR travaille depuis 2010 à la constitution d'un corpus propice à la formation d'enseignants et à l'enseignement du FLE, sur les différents niveaux de langues évalués et établis par le CECR (Cadre Européen Commun de Référence pour les langues) – du niveau A1 au niveau C2. En effet, afin d'évaluer la complexité des données de CLAPI, différentes expériences d'utilisation ont été réalisées à partir de la grille de notation élaborée par le CECR<sup>4</sup>. La catégorisation et la définition du degré de pertinence des données pour développer certaines compétences sociales et langagières chez les apprenants se sont donc effectuées à partir des pratiques de ces derniers et des formateurs, dans une démarche émergente (*bottom-up*) : c'est à partir de l'interprétation collaborative menée dans des groupes en classe de langue que différents niveaux de compréhension ont pu être repérés et annotés.

- 15 Par exemple, des séances de travail en groupe classe, avec des étudiants de niveau A2 à B1, ont consisté en l'utilisation d'enregistrements audio(visuels) de CLAPI choisis par l'enseignante pour illustrer différents phénomènes prosodiques (assimilations, élisions), pragmatiques (routines) et multidimensionnels (émotions) dans l'interaction (ICOR et Elisa Ravazzolo, Université de Trente, Italie). Toujours en situation de groupe en classe, avec des étudiants de niveau C1 futurs interprètes, des extraits portant sur le désaccord ont été utilisés à travers le déploiement de tests de compréhension et de reformulation (ICOR et Natacha Niemants, Universités de Bologne et de Macerata, Italie). Dans une autre configuration de travail, en entretiens individuels auprès de neuf étudiants de niveaux A2 à C1, d'autres extraits de CLAPI portant sur le désaccord ont été utilisés avec, cette fois-ci, des tests de compréhension, de reformulation, de détection de début et de fin de séquences et leur justification (Thomas *et al.* 2016). En vue de garantir l'exploitation expérimentale du matériau d'origine, une consigne a été appliquée de manière transversale à ces trois situations de travail :

L'enseignant(e) a donné des informations contextuelles et parfois lexicales, et constaté que les difficultés des élèves portaient davantage sur l'organisation de l'interaction, la co-construction de tours de parole par plusieurs locuteurs et la compréhension des tours brefs que sur le lexique. Les expériences ont également montré que le niveau de bruit (souvent considéré comme un frein à l'utilisation de données enregistrées dans des contextes sociaux naturels) n'est pas toujours problématique et constitue au contraire une aide à la compréhension.  
(Baldauf-Quilliatre *et al.* 2016, § 62)

- 16 Ce protocole permet aux concepteurs de l'interface CLAPI-FLE de déterminer les besoins des utilisateurs finaux et réduit d'éventuels écueils dus à des projections (inter)culturelles erronées. À partir de tels constats, les chercheurs qui travaillent à la construction d'un « volet enseignement » de l'archive CLAPI voient leurs questionnements se réduire et s'affiner à la fois. Le fait de savoir que les bruits ambiants captés à l'enregistrement ne sont pas des freins à la compréhension d'une situation d'interaction, nous pose questions quant aux retombées (i) sur le plan de la constitution des données (lors de l'enregistrement, devrait-on mobiliser des micros supplémentaires, focalisés non pas sur les prises de paroles des participants, mais sur leur environnement ?) et (ii) sur le plan de leur traitement en vue de leur implémentation dans une base de données (dans le traitement d'une piste audio, est-il pertinent et nécessaire de réduire ou d'augmenter certains contrastes, de manière à rendre (in)audibles certains événements sonores ?).
- 17 Aussi, au-delà de l'étude mise en œuvre pour évaluer les besoins des utilisateurs finaux, nous pouvons nous intéresser aux ressources sémiotiques mobilisées dans la conception d'interfaces numériques destinées à des publics experts et non experts en linguistique.

## 2.2. Regard sur l'interface

- 18 En portant notre regard sur l'interface de chacune des bases de données, nous sommes intéressés à comprendre les reconfigurations sémiotiques imposées par une repertinisation du corpus d'origine. Il est alors utile de mettre en regard deux « passages » comparables de CLAPI et CLAPI-FLE : leur page d'accueil et leur présentation du corpus. Comparables, ces pages le sont non seulement au niveau de l'architecture de l'interface, mais aussi au niveau de l'accessibilité des données. La page

d'accueil est un objet sémiotique pertinent à analyser du point de vue de l'adresse et de la destination d'un discours à différents publics. Si nous considérons ces sites en tant que production (hyper)textuelle, pour des questions méthodologiques relatives à notre approche herméneutique et philologique, il nous faut décrire la structure de la page CLAPI puis celle de CLAPI-FLE (§ 2.2.1) et expliciter le traitement et l'organisation des données (§ 2.2.2). Nous observons ensuite les variations instillées, les changements opérés et les modifications apportées dans la mise en valeur des données, à partir de l'émergence de différences entre l'un et l'autre des sites, tant sur le plan de l'expression que sur le plan du contenu (§ 2.3).

### 2.2.1. Focalisation sur la page d'accueil

- 19 La page d'accueil de CLAPI est un espace sur lequel le visiteur trouve une description sommaire du contenu du site. Nous allons décrire la structuration de cette page en trois modules<sup>5</sup>.

Figure 1a



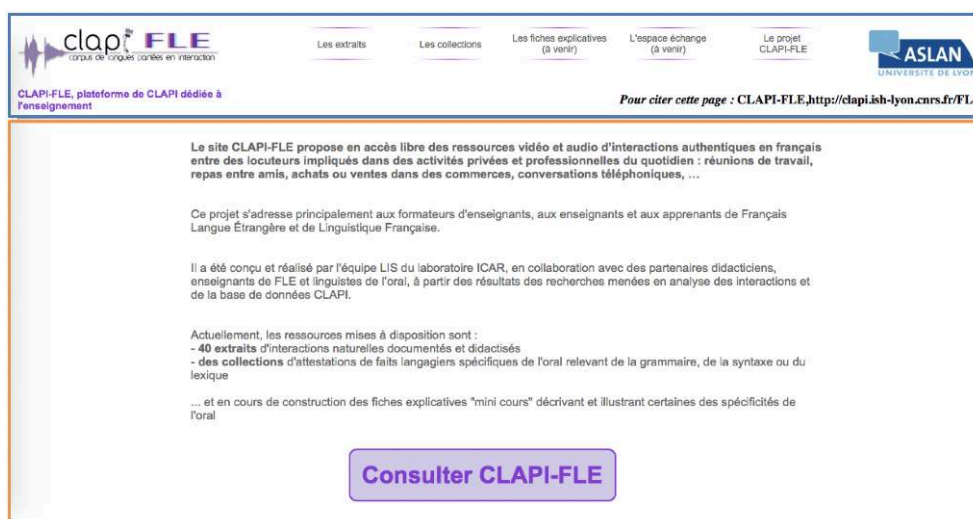
Site de la base CLAPI - page d'accueil<sup>6</sup>.

- 20 Le bandeau supérieur (encadré en bleu par nous-mêmes), sous forme d'onglets, propose un accès particulier au corpus constitué d'enregistrements audio et audiovisuel. L'onglet « Corpus » propose un accès à la totalité des enregistrements présents dans la banque de données. À travers l'onglet « Lexique », on peut accéder à différentes listes de mots réalisées à partir des transcriptions disponibles. L'onglet « Requêtes thématiques », quant à lui, favorise une approche multicritère du corpus, avec la possibilité de chercher les diverses répétitions présentes dans une même transcription. Aussi, comme nous pouvons le constater, un accès au site par identifiant et mot de passe est possible : selon que l'on est connecté en tant qu'utilisateur « invité » (sans code) ou « authentifié » (avec code à la suite d'une signature de convention de recherche), un nombre plus ou moins restreint de données est accessible.



- 21 Le module principal (encadré en orange) se trouve scindé en deux parties. Dans la partie de gauche, en guise de mot de « bienvenue », on trouve une brève description verbale du contenu du site, ainsi qu'une entrée directe aux données, renseignées en chiffres, dans l'encadré étiqueté « Accès libre à ». Cette partie pourrait être glosée : « voici le contenu du site auquel vous pouvez avoir librement accès en cliquant sur chacun des trois items ». Dans la partie de droite, on trouve un lien qui, lorsqu'on clique dessus, affiche la page d'accueil du site CLAPI-FLE dans un nouvel onglet : ce dernier peut ainsi être utilisé comme une extension sémiotique de CLAPI, comme un nouvel environnement de travail<sup>7</sup>. On trouve aussi trois autres ressources mises « à [...] disposition » des visiteurs du site CLAPI. Cette partie pourrait être glosée par : « des ressources sont mises à votre disposition pour saisir la méthode mobilisée pour recueillir, confectionner, intégrer, diffuser et analyser les enregistrements audio(visuels) (site Corinte) ; pour interpréter et produire des transcriptions selon les conventions ICOR ; pour garantir une interopérabilité des transcriptions aux formats d'encodage TEI et DUBLIN-CORE ». Enfin, en pied de page du site, un troisième module élabore une implémentation juridique des données – licence *Creative Commons*, référence à citer et copyright – et propose des liens qui contextualisent et encadrent cet environnement de travail – « Contact », « Qui sommes-nous », « Mentions légales ».
- 22 Dirigeons-nous maintenant vers la page d'accueil du site de CLAPI-FLE de manière à en décrire l'organisation.

Figure 1b



Site de la plateforme CLAPI-FLE - page d'accueil<sup>8</sup>.

- 23 Sur cette page, nous avons deux modules. Dans un premier module (encadré en bleu), on trouve le nom du site « CLAPI-FLE, plateforme de CLAPI dédiée à l'enseignement » et son logo (cliquable, alors que celui de CLAPI ne l'est pas). À ce jour (octobre 2018), différents onglets actifs (« Les extraits », « Les collections », « Le projet CLAPI-FLE ») ou non (« Les fiches explicatives », « L'espace d'échange ») permettent d'accéder à des pages spécifiques du site. On trouve également le logo du LabEx ASLAN, institution de la recherche ayant financé une partie du projet CLAPI-FLE. Sous ce premier module, un second (encadré en orange) décrit brièvement le contenu du site, énonce les acteurs auxquels ce site est adressé, présente les acteurs à l'initiative de sa conception et

explicite les intitulés des multiples onglets. En pied de ce module, on trouve un onglet violet de la couleur du logo qui invite le visiteur à « Consulter CLAPI-FLE ». Ici, il n'y a pas de module dédié aux informations légales. On peut alors se questionner sur la raison de cette absence : ces mentions sont-elles absentes de la page d'accueil parce qu'elles sont présentes en page d'accueil du site CLAPI ? Serait-ce également parce qu'elles sont mentionnées au sein de la page des « extraits » présents dans CLAPI-FLE (voir fig. 2b, encadré rouge) ?

- 24 Ces questions soulèvent le problème de l'emboîtement des sites (CLAPI > CLAPI-FLE) et de l'inclusion des pages (accueil > extraits), comme nous allons le voir en traitant de l'organisation des données pour chacun des deux sites.

## 2.2.2. Systématisation et présentation des données

- 25 Du point de vue de la constitution d'une archive d'enregistrements, le groupe ICOR a élaboré une typologie permettant d'identifier différents statuts du matériau constitué et traité dans une perspective de recherche scientifique. En parcourant chacun des deux sites, nous pouvons faire émerger trois dénominations pour trois statuts différents : *données* > *corpus* > *extraits*.
- 26 Le premier niveau d'inclusion, celui des « données », réfère aux enregistrements “bruts” et de première main, réalisés à partir de situations naturelles. Ces « données » peuvent être audio ou audio-visuelles. Archivées et implémentées sur le site de CLAPI ou CLAPI-FLE, ces « données » acquièrent le statut de « corpus » dès lors qu'elles sont organisées en un ensemble qui fait système. À ce deuxième niveau d'inclusion qu'est le « corpus », en effet, une systématisation est opérée par le biais du titrage des situations d'interactions documentées (« apéritif entre ami(e)s », « consultation juridique », « visite guidée », etc.) et par la mobilisation de descripteurs, de catégories d'indexation (« auteur », « année », « genres interactionnels », « source audio/vidéo », « locuteurs »). Enfin, à un troisième niveau, celui des « extraits », sont présentés des fragments d'interactions enregistrées catégorisées en tant que « corpus ».
- 27 Un phénomène tout à fait intéressant à observer et à enquêter sur le plan d'une (re)configuration sémiotique des données est celui du renversement de la présentation de cet emboîtement dans le processus même de didactisation du site d'origine.

## 2.3. Quelle mise en valeur des données pour quelle adresse ?

- 28 Dans le passage de CLAPI à CLAPI-FLE, on trouve une modification de l'indexation et de la présentation des données primaires sur les deux interfaces. Observons et décrivons la mise en valeur des données sur la plateforme CLAPI avant de nous diriger vers CLAPI-FLE.

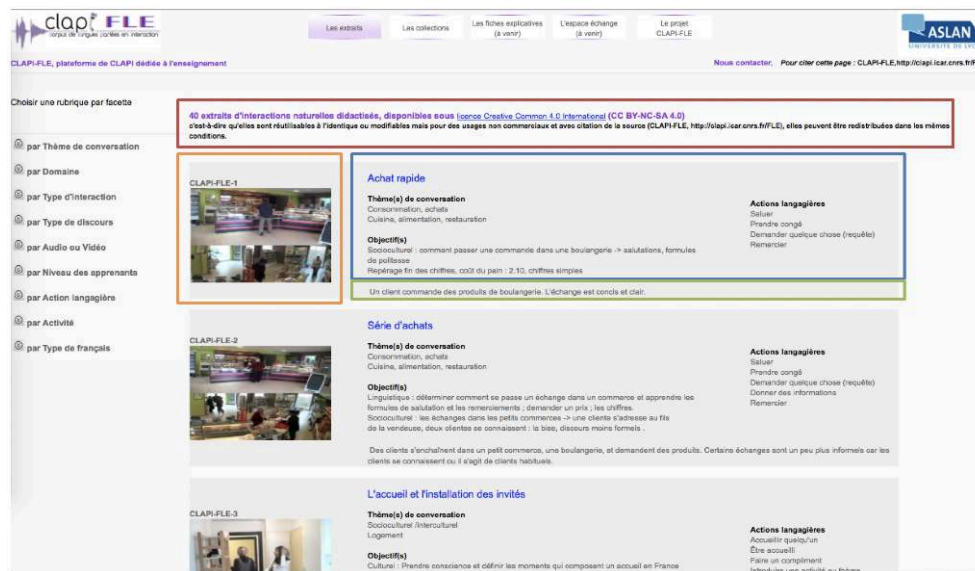
Figure 2a

Corpus	Auteur	Date de Collecte	audio vidéo	Nombre d'enregistrements	Durée	Extraits
<a href="#">Apéritif entre ami(e)s - chat</a>	G. Icor, S. Schwarz	2009	vidéo	1	31m55s	Jouer
<a href="#">Apéritif entre ami(e)s - glasgow</a>	G. Icor				21m23s	Jouer
<a href="#">Apéritif entre ami(e)s - pois</a>	G. Icor				37m32s	Jouer
<a href="#">Apéritif entre ami(e)s - rupture</a>	G. Icor				33m	Jouer
<a href="#">Bielefeld - situations de contact - cadre universitaire</a>	U. Kraft, E. Gülich, U. Dausendschön-Gay	1984-1998	audio	9	3h29m34s	Jouer

Site de la base CLAPI – page du corpus<sup>9</sup>.

- 29 Pour le site de CLAPI, la présentation des données est tabulaire. Comme nous pouvons le voir dans la fig. 2a, une première colonne est dédiée aux « corpus » sous forme de liens menant vers des pages internes au site. Chaque corpus est décrit par des entrées destinées aux auteurs, à la date de collecte, à la nature sémiotique du support, aux enregistrements que l'on y trouve et à la durée du corpus. Ce n'est que dans une dernière colonne, à l'extrémité droite du tableau, que l'utilisateur peut avoir un aperçu du corpus, par une image de l'enregistrement lorsqu'il s'agit d'un support vidéo et par un extrait d'une durée de quarante secondes tant pour les vidéos que pour les audios. Lorsque l'utilisateur du site passe ou arrête le curseur de la souris (ou du *trackpad*) sur une image, un agrandissement de celle-ci apparaît à l'écran (encadrée en orange en fig. 2a).
- 30 Ici, une primauté est donc accordée aux corpus, aux situations enquêtées. Nous pouvons envisager cette primauté au regard des finalités du site : rendre accessibles des enregistrements de pratiques langagières en situation naturelle à une communauté de chercheurs en sciences du langage, à partir d'une perspective interactionniste. De multiples situations d'interaction sont ainsi identifiées de manière à rendre compte de leur diversité. Cette primauté est modifiée sur le site CLAPI-FLE. Comme nous l'observons sur la fig. 2b ci-après, la plateforme CLAPI-FLE accorde une primauté aux extraits des situations, aussi bien dans la représentation visuelle de chaque corpus que dans leur dénomination.

Figure 2b

Site de la base CLAPI-FLE – page des extraits<sup>10</sup>.

- 31 La visualisation d'un extrait en image est première dans la représentation tabulaire (encadrée en orange) et sa description thématise le type d'interaction qu'elle exemplifie (encadrée en bleu). Chaque type est marqué par un titre en bleu et des focalisations sur trois caractéristiques structurales sont opérées par du texte en gras : le(s) thème(s) de conversation actualisés ; le(s) objectif(s) didactiques que permet de travailler cette ressource ; les actions langagières à l'œuvre dans l'échange entre les interactants. En guise de synopsis, on trouve un bref résumé du déroulement de l'échange (encadré en vert)<sup>11</sup>. Ici, le corpus est plus problématisé pour des finalités de pratiques d'enseignement et d'apprentissage que pour des finalités de recherches en linguistique.
- 32 Nous faisons l'hypothèse que, en orientant l'attention des visiteurs du site sur la scénographie d'une pratique socio-langagière archivée et documentée par le biais d'une capture d'écran, les concepteurs de la plateforme CLAPI-FLE non seulement mettent en valeur la dimension écologique et spontanée des interactions implémentées, mais aussi misent sur la puissance illustrative de l'image, qui l'emporte sur une description verbale. D'une certaine manière, la visualisation de l'interaction par l'image assure une prise phénoménologique de la scène représentée par le corpus en relation aux situations et aux espaces dans lesquels se trouvent les apprenants dans leurs pratiques ordinaires. Autrement dit, on met en avant les images pour que les visiteurs du site, qu'ils soient enseignants ou apprenants, puissent se figurer le milieu de l'interaction documenté en relation intime avec les environnements dont ils se font ou se feront une expérience au quotidien. On représente une boulangerie pour mettre en image les modalités d'interaction dans les achats de produits dans ce type de lieux : la *situation*, qui est une *occurrence* de pratiques ayant un ancrage spatial, temporel, actoriel et modal particulier, est alors érigée au rang de *type* de pratique *contextuelle*, et a valeur de *schématisation* (Fontanille 2011) à un niveau pragmatique<sup>12</sup>.
- 33 L'introduction d'une telle variation de mise en valeur des données, d'une base à l'autre, nous questionne sur les enjeux épistémologiques de la constitution d'une base de données, tant du point de vue de la spécification des outils mis à disposition de

professionnels de la recherche en sciences du langage (et plus largement en sciences humaines) que de la différenciation des usages qu'ils pourront en faire.

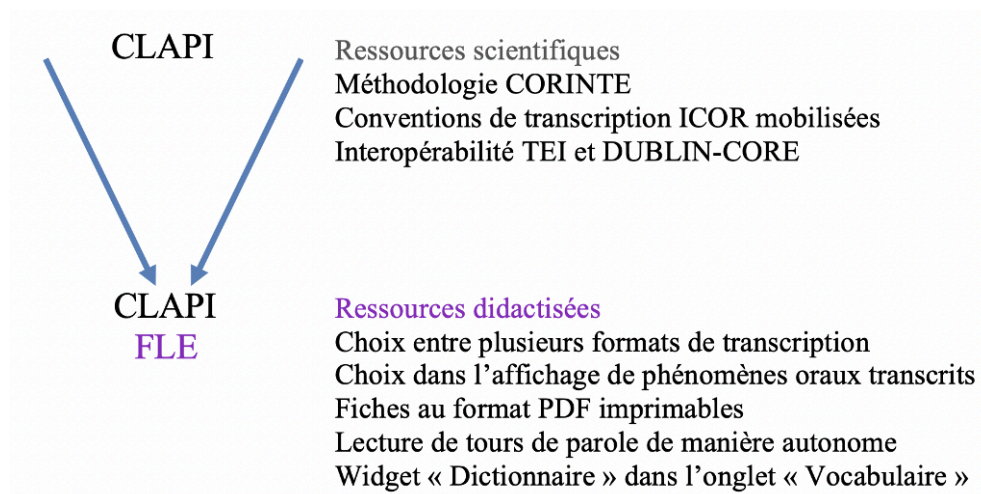
### 3. Spécification des outils, différenciation des usages. Vers une épistémologie de l'archive comme ressource professionnelle

- 34 Lorsque nous parlons de *ressource professionnelle* dans le titre de cette section, nous voulons pointer deux dimensions impliquées dans la constitution d'une archive et dans la conception d'un dispositif de recherche en sciences du langage.
- 35 La *ressource professionnelle* renvoie, d'un côté, à la base de données développée par des professionnels scientifiques et, de l'autre, à l'interface conçue pour des professionnels de l'enseignement. Dans le premier cas, ce sont les pertinences des concepteurs qui sont principalement prises en compte et, dans le deuxième, ce sont celles des utilisateurs finaux. Si, dans certains cas, ces deux positions tenues vis-à-vis de l'interface se recoupent (être et concepteur et utilisateur), la réalisation d'une archive demande néanmoins d'envisager une multiplicité de scénarios d'usages et, ainsi, de se projeter d'autres utilisations que celles faites à partir de son propre champ disciplinaire ou de son propre objet de recherche. C'est pour cette raison même que le groupe ICOR, dont les membres travaillent en linguistique interactionnelle, a mis en place un partenariat avec des chercheurs et des étudiants de FLE : les concepteurs ont pu envisager les contraintes d'utilisation finales à partir des questions soulevées par les pratiques de ces professionnels.
- 36 Une épistémologie de l'archive du point de vue des pratiques professionnelles nécessite que l'on s'attarde sur les outils théoriques et méthodologiques tantôt spécifiques, tantôt transversaux aux disciplines des sciences du langage. Lorsque nous reprenons le document décrivant le projet CLAPI-FLE<sup>13</sup>, aussi synthétique qu'il soit, son ancrage théorique et pratique apparaît clairement. Sous le titre cadre théorique, nous pouvons lire que
- Le projet se situe au croisement de la linguistique interactionnelle, issue des travaux de Sacks, Schegloff et Jefferson (1974) sur l'organisation de la parole et des travaux de Garfinkel (1984) sur les pratiques sociales situées, et de l'approche interactionniste en acquisition des langues étrangères qui place l'interaction au centre du processus d'acquisition. Dans ce paradigme, l'apprenant n'est plus vu comme un individu intériorisant un système linguistique mais comme un acteur social qui développe des compétences langagières dans un contexte social donné, en lien avec d'autres acteurs sociaux. Nous avons également retenu la théorie de l'*input* (Krashen et Tracy 1983), qui explique le lien entre entendre et comprendre des productions de natifs, et l'acquisition de cette langue (*the input hypothesis or the comprehension hypothesis*).
- 37 Par là même, la centralité accordée aux actions des apprenants dans le processus d'acquisition et de développement de compétences interactionnelles, sociales et langagières pose la question des ressources implémentées et mises à disposition des utilisateurs.

### 3.1. Ressources à disposition et prise des utilisateurs

- 38 En vue de favoriser l'implémentation de données et de canaliser les pratiques du dispositif, des ressources sémiotiques sont mises « à [n]otre disposition » en tant qu'utilisateur, comme nous l'avons vu plus haut (§ 2.2.1). Ces ressources ont pour but (i) d'assurer une traduction linguistique et culturelle entre l'arrière-plan des chercheurs et celui des apprenants et (ii) de consolider le terrain de travail interprétatif et praxéologique des données.
- 39 Aussi, tous les chercheurs ne disposent pas des mêmes outils technologiques (ordinateurs avec différents systèmes d'exploitation tels que Windows, Mac OS, Linux, etc.) et ne pratiquent pas la recherche avec les mêmes méthodes (les transcriptions réalisées à partir de conventions établies entre certains acteurs ne sont ni partagées ni comprises par tous). Plus haut, lorsque nous avons problématisé l'emboîtement du matériau présent dans le site CLAPI à travers la suite *données > corpus > extraits*, nous avons questionné l'impossibilité d'accéder directement à CLAPI depuis la page CLAPI-FLE. Aussi bien au niveau des extraits que des ressources mises « à disposition », le passage se réalise sous forme d'entonnoir : les conventions de transcription et les autres ressources de CLAPI valent pour CLAPI-FLE, mais les ressources mises à disposition sur CLAPI-FLE ne valent pas pour CLAPI.

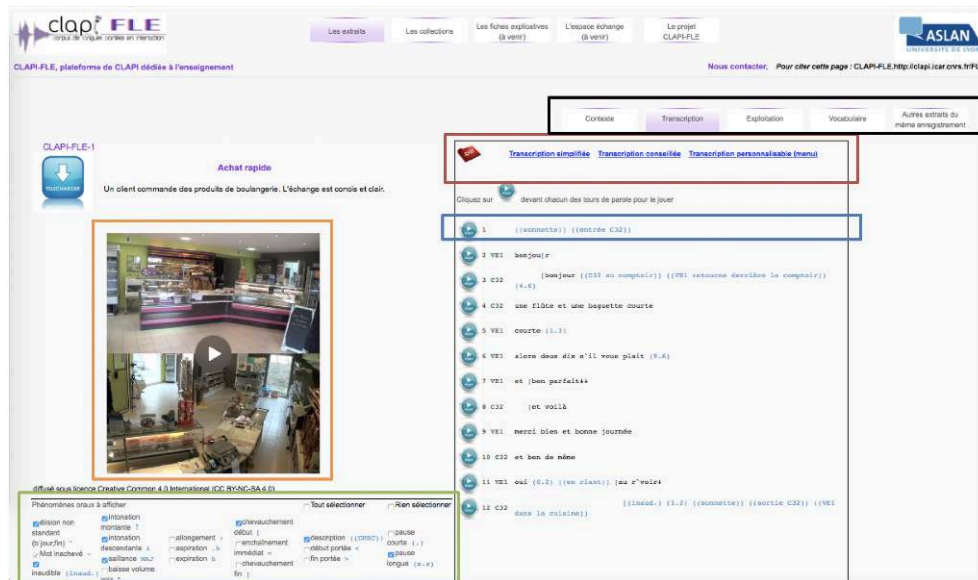
Figure 3



Schématisation de la structure des ressources de CLAPI à CLAPI-FLE : un entonnoir.



Figure 4



Extrait CLAPI-FLE-1, onglets « Les extraits » > « Transcription » - ressources didactisées.

- 40 Sur cette page comme sur d'autres, la configuration élaborée entre les données primaires (enregistrements) et les données secondaires (transcriptions, fiches descriptives, etc.) participe de la formation d'un matériau de travail quasiment prêt à l'emploi. Chacun des extraits se trouve accompagné de divers éléments accessibles via les onglets (encadrés en noir) : l'onglet « Contexte » permet d'avoir des informations sur les personnes qui ont réalisés les enregistrements, sur les participants de l'interaction et sur leurs échanges ; l'onglet « Exploitation » explicite le cadre de travail auquel est destiné cet extrait, l'exploitation que l'on peut en faire ; via l'onglet « Transcription », on a accès à une lecture des phénomènes oraux de l'extrait transcrits ; les mots transcrits et leur fréquence sont renseignés sous forme de liste par l'onglet « Vocabulaire » ; l'onglet « Autres extraits du même enregistrement » donne accès à d'autres échanges de la même interaction.
- 41 Sur la fig. 4 ci-dessus, c'est l'onglet « Transcription » qui est actif et mis en visibilité (surbrillance et violet dans l'encadré noir). Trois niveaux de transcription sont proposés aux formateurs d'enseignants, aux enseignants et aux étudiants de FLE, selon le niveau de granularité auquel ils se situent pour l'exploitation des échanges : la « transcription simplifiée » est une version orthographique des échanges verbaux qui se déploient dans l'interaction ; la « transcription conseillée » est celle réalisée à partir des conventions ICOR pour laquelle sont annotés les chevauchements, les pauses, les gestes corporels des participants ; la « transcription personnalisable (menu) » consiste à donner la possibilité aux visiteurs d'afficher ou non certains phénomènes oraux transcrits (« élision non standard », « mot inachevé », « intonation montante », début et fin de « chevauchement » de tours de parole, enchaînement immédiat de tours, etc.). Cette dernière fonctionnalité est l'une des transformations majeures du point de vue de la didactisation de l'interface CLAPI. D'une part, cette transformation concerne la possibilité de cocher et de décocher des options. D'autre part, cette transformation concerne le discours verbal mobilisé pour référer aux phénomènes à enquêter ; notamment, la classification des enregistrements en genres interactionnels sur CLAPI est décomposée sur CLAPI-FLE en plusieurs axes (« thèmes de conversation »,

« domaine », « type d'interaction », « type de discours »). En passant du champ de la recherche scientifique à celui des pratiques didactiques et éducatives, les transformations de la base de données vont donc bien au-delà d'une simple intervention esthétique sur l'interface.

- 42 En proposant différents degrés de granularité de transcription, on cherche en effet à donner une *prise* (Landowski 2009) à la fois sensible et épistémique aux utilisateurs de cette interface, en vue de faire entrer celle-ci dans un dispositif pédagogique composé d'une multiplicité de médiations (échanges oraux en salle de cours, apprentissage des pratiques socio-langagières par le visionnage de films, utilisation de manuels, récit d'expériences par l'autobiographie et l'autofiction, etc.). On voit ainsi la complexité qu'implique le passage d'une archive d'enregistrements, pour laquelle penser une typologie et un ordonnancement permettant l'indexation des corpus, au déploiement d'une plateforme d'exploitation de matériel didactique orientée vers des utilisateurs (*user-centered design*), pour laquelle gérer la tension entre des pratiques potentielles et des appropriations réalisées en acte.

### 3.2. Champs disciplinaires, exploitation et exploration de la base de données

- 43 Cette tension entre pratiques potentielles et pratiques réelles pose question vis-à-vis de la diversité des parcours de l'interface et de l'ajustement bilatéral des acteurs impliqués dans l'utilisation d'une base de données : de l'amont vers l'aval de son implémentation, c'est-à-dire des concepteurs vers les utilisateurs, et de l'aval vers l'amont, des utilisateurs vers les concepteurs. Ce mouvement de l'amont vers l'aval concerne non seulement la transformation du rôle des acteurs qui participent de l'implémentation et de l'utilisation de la base de données, mais aussi la transformation du champ disciplinaire dans lequel ils opèrent. En procédant à la spécification des fonctionnalités de CLAPI, le processus de didactisation implique, d'abord, une traversée disciplinaire, de la linguistique interactionnelle à la didactique des langues et des cultures ; ensuite, la formation d'un pont entre les membres individuels et collectifs du domaine de la recherche scientifique et ceux des pratiques éducatives.
- 44 Ainsi, toute pratique de l'archive par des professionnels de la recherche ou par des professionnels de l'éducation nous semble relever d'une mise en tension entre, d'un côté, une *exploitation* de la base de données en continuité avec les visées de celle-ci et, de l'autre, une *exploration* des potentialités du dispositif par des utilisateurs. L'*exploitation* a une aspectualité terminative en ce qu'elle renvoie à la "raison instrumentale" (Simondon [1958] 1989) d'une action menée à travers cette médiation technologique que constitue la « banque » de données. L'*exploitation* est guidée par la recherche d'une symétrie entre les raisons pour lesquelles cet outil est conçu et les modalités par lesquelles les utilisateurs s'en saisissent. L'*exploration* est marquée quant à elle d'une inchoativité, en ce que l'espace de la pratique des utilisateurs est en train de se définir. L'*exploration* met en jeu la recherche d'un terrain pertinent de la pratique en cours, sans pour autant former une trame complètement appropriée et applicable à des projets futurs. Entre ces deux dynamiques, deux questions méritent d'être posées dans le cadre de notre étude sémiotique de l'archive.
- 45 D'abord, dans la perspective d'un dialogue interdisciplinaire, nous nous posons la question de savoir dans quelle mesure des acteurs de disciplines autres que le FLE



sauraient ou pourraient se saisir d'une plateforme à finalité didactique pour mener un travail de recherche : par exemple, les acteurs de la recherche en sémiotique ne pourraient-ils pas trouver là un terrain d'étude foisonnant pour enquêter les processus de resémiotisation d'un objet culturel ? Ensuite, dans la transformation de l'archive de corpus en plateforme de travail, nous sommes intéressés par savoir jusqu'où les pratiques des utilisateurs peuvent s'approprier un dispositif implémenté pour une pratique particulière (recherche scientifique ou pratique d'enseignement) sans en dévoyer la pertinence : à partir de quand l'appropriation de la plateforme par les utilisateurs constitue-t-elle un « respect » ou bien un « viol » des propriétés de l'outil conçu et des données implémentées ?

- 46 Si nous n'avons pas de réponse immédiate à ces questions, nous pouvons au moins tenter de proposer quelques pistes de réflexion sur le phénomène de réappropriation des outils mis à disposition des utilisateurs, depuis nos propres expériences.

### 3.3. Réappropriation de la plateforme : *affordance* des outils et densité des corpus

- 47 Lorsque nous avons expérimenté la base de données CLAPI dans le cadre d'un travail de thèse de doctorat et d'ingénierie de corpus, nous ne nous étions pas forcément aperçus du caractère hautement spécifique de celle-ci. En immersion dans une pratique de la recherche développée au laboratoire ICAR à laquelle nous avons été formés, les termes mobilisés sur la plateforme de corpus de langues parlées en interaction ne nous paraissaient pas étrangers. La dimension familière de cet outil de travail et du discours qui la structure nous a néanmoins paru plus frappante lorsque nous sommes passés à son volet didactique. Comme l'espace discursif de CLAPI-FLE a été conçu pour des acteurs autres que ceux travaillant en linguistique interactionnelle, l'encadrement discursif des corpus exploités a été sensiblement transformé (voir notamment § 3.1). Ainsi, la question se pose à nous de savoir quelle utilisation peut-on se faire de la base de données CLAPI après s'être fait l'expérience de sa spécification en didactique : comment passe-t-on d'un site internet à l'autre, avec d'autres finalités et d'autres possibilités ? Et surtout, quels sont les effets rétroactifs et récursifs de notre parcours de CLAPI-FLE sur celui de CLAPI ?
- 48 Notre expérience de la base de données CLAPI-FLE nous a permis de dégager deux spécificités de cette plateforme : (a) la prééminence de la finalité pratique de son utilisation qui est enchevêtrée à (b) la systématité de la méthode déployée pour renseigner des situations interactionnelles et illustrer des cas. Lorsque nous sommes revenus à la base de données CLAPI, l'hétérogénéité des données primaires et secondaires perçue lors de nos premières utilisations est apparue comme le trait distinctif de cette plateforme par rapport à CLAPI-FLE. Loin d'être une multiplicité dispersée, cette hétérogénéité représente une base empirique rassemblant des corpus issus d'approches différentes. Si l'ensemble des données constituées à partir de diverses méthodes de travail pourrait laisser perplexes les utilisateurs de CLAPI, une approche transversale aux données est néanmoins rendue possible via l'outil « requêtes multicritères ». Par ce dernier, on peut envisager une recherche linguistique qui porte sur des situations discontinues en termes de genres interactionnels et d'acteurs impliqués. Autrement dit, cet outil permet de porter un regard panoramique sur des observables, en laissant la possibilité de se replonger à tout moment dans l'écologie

d'une situation en particulier. Les passerelles entre quantitatif et qualitatif permettent ainsi de consolider l'analyse : à la vision d'ensemble garantie par la dimension quantitative se télescope la dimension qualitative, en prenant en compte la pertinence du contexte et en appréciant l'épaisseur des occurrences situées, par des écoutes et des visionnages répétés.

- 49 De la même manière que des requêtes ont permis de faire émerger des extraits appropriés à l'apprentissage d'une langue, par exemple, au niveau A1 du CECR, est-ce qu'on pourrait envisager un parcours des données qui permettrait de faire émerger des phénomènes sémiotiques pas encore balisés, et que l'on ne pourrait pas repérer de manière automatique ? Notamment, la question se pose pour les phénomènes relevant de la multimodalité comme, par exemple, l'orientation des regards, les postures corporelles, la séquentialité des gestes, etc.
- 50 À partir de ce changement de perspective sur la plateforme outillée CLAPI, la recherche scientifique en tant que pratique *par* des professionnels et *pour* des professionnels mériterait d'être considérée d'un point de vue culturel en tant que pratique collective, ancrée historiquement, liée à la fois à la compréhension de phénomènes empiriques et à la transformation des représentations culturelles.

## 4. Vers une éthique de l'archive comme objet culturel

- 51 Dans cette section, nous revenons sur la structure de l'archive et nous la considérons comme manifestation spécifique d'un objet culturel. Nous nous penchons d'abord sur une réflexion éthique, que les chercheurs formulent et inscrivent dans la production de leurs objets scientifiques. Ensuite, nous considérons les chercheurs en tant qu'acteurs qui opèrent dans une socialité restreinte (communauté scientifique) et prennent part à des pratiques culturelles dans une socialité élargie (espace social). Ces pratiques présentent des spécificités, mais elles s'ouvrent à d'autres univers sémiotiques, elles tissent des liens avec d'autres acteurs sociaux, avec d'autres communautés, apportant ainsi une lumière nouvelle sur l'archive en tant qu'objet culturel. Le terme « culture » est ici appréhendé dans l'acception que peut en faire François Rastier (2011, pp. 2-6) en sémiotique : la pratique scientifique participe de la formation d'un domaine culturel aux côtés d'autres domaines, tels que les arts, les religions.

### 4.1. Du côté de la recherche : enjeux déontologiques et éthiques des objets scientifiques

- 52 Les travaux en linguistique ont été questionnés non seulement au prisme d'approches analytiques et de retombées sur les approches théoriques, mais aussi en termes d'enjeux méthodologiques et déontologiques, qui sont au cœur de la pratique de la recherche (Baude *et al.* 2006) : ces enjeux doivent être explicités en amont et questionner la production des corpus comme artefacts scientifiques. En principe, la constitution de grands corpus de l'oral ne présuppose pas les objets à investiguer : d'une part, dans un souci de préservation de l'écologie des pratiques, les situations enquêtées ne sont ni préfabriquées ni élicitées par le chercheur ; d'autre part, les phénomènes qui seront étudiés ne sont ni donnés *a priori* ni complètement définis à l'avance. En revanche, dans les faits, la possibilité d'exploiter les données doit se confronter au respect des acteurs sociaux et à la préservation de leur vie privée. À ce

sujet, l'exemple le plus évident qu'on peut rencontrer dans la base de données CLAPI est l'anonymisation des données personnelles. Pour tout enregistrement, le signal audio est bipé lors d'une mention de noms, adresses, numéros de téléphone et autres données personnelles – qu'il s'agisse de participants en coprésence physique ou convoqués en discours. Dans les transcriptions correspondantes, les (pré)noms et autres informations pouvant porter à l'identification des participants sont systématiquement remplacés par des pseudonymes et des données fictives.

- 53 La pratique de terrain est ainsi à appréhender dans son déroulement. Au moment de la prise de données, les chercheurs prévoient un moment d'échanges avec les participants, pour illustrer les enjeux scientifiques des enregistrements audiovisuels (un focus sur la multimodalité, notamment), ainsi que l'attention portée à la vie privée des personnes. Déjà dans la phase de constitution de corpus, nous retrouvons donc des aspects déontologiques<sup>14</sup> qui caractérisent les pratiques de la recherche, dans leur conception et leur déploiement sur le terrain : pour que le chercheur puisse obtenir le consentement des acteurs sociaux, il faut prévoir un document dédié et être préparé à répondre à d'éventuelles questions sur le dispositif d'enregistrement.
- 54 Pour ce qui concerne la phase d'archivage et la consultation des données sur une longue durée, d'autres enjeux se profilent pour les chercheurs et interrogent leur positionnement face à des questions éthiques. Par exemple, pour des finalités de patrimonialisation, afin de documenter les pratiques professionnelles en contexte médical, à quel niveau de granularité rendre publique des enregistrements de consultations gynécologiques ou de consultations de psychiatrie avec des réfugiés qui racontent leurs expériences traumatiques (projet ANR REMILAS<sup>15</sup>) ? Il n'en va pas de même de signaler la présence d'un corpus accessible sous convention de recherche que de le rendre librement téléchargeable et requêtable. La question s'est notamment posée pour deux projets de recherche au sein du laboratoire ICAR pour lesquels il a été décidé de ne pas intégrer les enregistrements audiovisuels produits à la base de données CLAPI. Au-delà de ces cas particuliers, le positionnement éthique des chercheurs est au cœur des possibilités de consultation et de patrimonialisation offertes par l'archivage. Dans le processus de publicisation de données vidéo dites « sensibles » (Paveau & Péréa 2012), à quel point les pratiques des chercheurs, dont les prérogatives sont encadrées juridiquement, sont-elles perçues et assumées, par les chercheurs eux-mêmes, comme (il)légitimes vis-à-vis de la relation tissée avec les participants ? À partir de quels critères déterminer l'intégration d'un corpus dans la base de données CLAPI et définir sa visibilité à la communauté scientifique (connexion en tant qu'« authentifié ») et, plus généralement, sur internet (en libre accès) ?
- 55 Dans le cadre d'une réflexion éthique, on peut s'interroger ainsi sur les obligations déontologiques des chercheurs, en tant que professionnels de la recherche scientifique, qui doivent se positionner envers leurs collaborateurs (Ingold 2013, pp. 327-331) dans la construction même de leur objet scientifique. En ce sens, les participants enregistrés ne sont pas considérés comme des simples acteurs ou figurants « sur scène », mais comme des personnes qui ont des droits, des attentes et peuvent également être sensibilisés et impliqués dans la démarche de recherche. La prise en compte de leur subjectivité et de leur individualité est primordiale pour la réalisation même des corpus. En particulier, la qualité des données enregistrées dépend « de la qualité de la relation avec les personnes ressources » (Baudé *et al.* 2006, p. 54), elle peut également ouvrir de nouvelles possibilités pour l'accès aux terrains d'enquête. De plus, pour ce qui concerne

la communication et la valorisation au-delà des frontières disciplinaires des sciences du langage, les vidéos issus de corpus « peuvent faciliter la création de nouveaux réseaux dans les domaines de la recherche et des institutions » (Piccoli & Ursi 2015, p. 11).

- 56 Cette dimension éthique permet d'envisager une appréciation en termes de valeurs qui se situe bien au-delà de la sphère des acteurs de la communauté scientifique, au sein de laquelle on produit les objets culturels. En suivant la proposition de François Rastier, on pourrait affirmer que « en tant qu'artefacts, les objets culturels sont indissociables des actions qui les produisent, les utilisent et leur donnent sens dans l'interprétation » (2011, p. 2). Et les liens entre les chercheurs et les acteurs sociaux sont une condition nécessaire pour que ces artefacts puissent être réalisés et puissent vivre comme objets culturels, représentations à la fois d'une sémiotisation scientifique et d'un univers symbolique social.

## 4.2. Visualisations de corpus, accès et publicisation de données

- 57 La mise à disposition de données primaires et secondaires représente une étape primordiale dans la constitution d'une banque de données comme ressource telle que CLAPI. Les modalités de visualisation des corpus et de mise en valeur des données ont déjà fait l'objet de notre étude (§3.3). Ici, nous voulons attirer l'attention sur une dimension essentielle de l'appréhension de la base de données en tant que forme de vie (Fontanille 2015), dans le prolongement d'une réflexion sur les effets rétroactifs et récursifs de notre pratique de CLAPI-FLE sur le matériau d'origine. En particulier, rappelons-le, la base CLAPI rassemble des données recueillies par différents chercheurs et pour différentes finalités de recherche, les outils automatiques permettent d'effectuer des requêtes de manière transversale, sur des matériaux hétérogènes. Ainsi, par exemple, les transcriptions du corpus « Interactions parents-enfant - Léonard »<sup>16</sup> s'appuient sur des conventions qui présentent des traits communs à la convention ICOR<sup>17</sup>, mais aussi des spécificités (fig. 5), qui les distinguent de cette dernière (référence majoritaire pour les corpus de la base de données et actuellement employée comme convention pour la transcription de nouveaux corpus à intégrer).

Figure 5

Transcriptions- Les énoncés des adultes.

Ils sont transcrits de façon conventionnelle orthographiquement en essayant de respecter certaines caractéristiques de l'oral.

Les mots prononcés avec emphase sont notés en MAJUSCULES.

... = prise de parole coupée par l'intervention d'un autre participant.  
xxx = inaudible ou incompréhensible  
(?) = transcription incertaine

- Les énoncés de l'enfant.

Ils sont transcrits en phonétique (I.P.A. Times) jusqu'à Léonard 2;05 inclus, puis orthographiquement à partir de 2;06.

Dans les énoncés en phonétiques, certaines conventions ont été créées:

/ = pause  
// = précède une nouvelle prise de parole  
(soit celle d'un adulte, soit celle de l'enfant lui-même).  
? = schéma interrogatif  
! = schéma exclamatif  
xxx = inaudible ou incompréhensible  
: = allongement vocalique  
(?) = transcription incertaine

Dans les énoncés en orthographe, même conventions que pour les transcriptions des énoncés des adultes.

CLAPI - Corpus « Interactions parents-enfant - Léonard » - Conventions de transcription.

- 58 Pour ce corpus, les finalités de recherche d'un projet ANR et l'approche mobilisée représentent des paramètres qui influent sur la marge de manœuvre dont on dispose pour produire des données secondaires à partir des données enregistrées. En particulier, ces spécificités ont trait à une étude développementale qui se focalise sur l'apparition et le fonctionnement de marqueurs grammaticaux chez l'enfant de douze mois à trois ans, et comportent une granularité différente des modalités de transcription selon l'âge du sujet dont on étudie les productions praxiques et langagières. D'après les conventions illustrées en figure 5, avant deux ans et six mois, la réalisation phonétique des productions verbales de l'enfant s'avère un facteur primordial pour les champs d'étude de l'interaction et de l'acquisition langagières. Dans ce même corpus, les intonations ne sont pas notées, mais sont associées à des profils illocutoires (voir les termes « schéma interrogatif » et « schéma exclamatif »), transcrits à l'aide de signes empruntés au système de ponctuation. Par ailleurs, dans d'autres corpus de CLAPI, comme par exemple les corpus de sociologie urbaine (corpus « Lausanne – Genève » et « Paris Marais »), où les longs tours de parole des participants suivent des questions ou de courtes interventions de l'enquêteur, nous ne retrouverons pas ce type de détails. Et les pauses ne seront pas chronométrées.
- 59 Si l'on se penche sur une transcription issue de CLAPI-FLE, on pourra remarquer un degré de granularité différent et, de manière fort intéressante, négociable.

Figure 6

Extrait CLAPI-FLE-1 - Phénomènes oraux à afficher.

- 60 Ici, hormis la présence de flèches pour signaler les profils prosodiques (au lieu des notations « / » et « \ » de la convention ICOR), les phénomènes attestés dans les enregistrements peuvent être affichés dans la transcription. Mais, à partir d'un paramétrage par défaut (fig. 6), ces particularités peuvent être cochées ou décochées en fonction du niveau de compétence des apprenants-usagers. Ainsi, si l'on trouve des notations similaires à celles rencontrées pour le corpus « Interactions parents-enfant – Léonard », les intonations ne sont pas associées à des profils illocutoires particuliers, mais simplement à des contours perçus comme montants ou descendants, pouvant être interprétées comme réalisant des actions différentes (pas uniquement des interrogations ou des exclamations). La pertinence des détails conversationnels est marquée ici par les spécificités de la notation des productions verbales qui ne sont pas transcrites en phonétique (selon les conventions I.P.A.) : les particularités de l'oral sont signalées à l'aide de l'« orthographe adaptée » notée dans la fig. 6 par des phénomènes d'« élision non standard (b`jour, `fin) ». Dans CLAPI-FLE, donc, nous retrouvons une négociation des modalités de représentation du corpus, qui peut être plus ou moins fidèle à la dimension conversationnelle et, réciproquement, plus ou moins cohérente avec une représentation orthographique ou “graphocentrique” de l'oral. Dans ce cas, la resémantisation de ressources existantes met au centre l'utilisateur et pas les acteurs de la communauté scientifique (sauf en cas de syncrétisme de positions).
- 61 C'est l'utilisateur qui peut ainsi choisir les détails à afficher, qui peut s'appuyer sur ces données secondaires pour travailler sur les extraits proposés. La visée didactique de la plateforme CLAPI-FLE permet une visualisation qui répond aux compétences de la personne qui y travaille et l'explore. Un étudiant qui n'est pas familiarisé avec les conventions de transcription, contrairement aux chercheurs travaillant en syntaxe de l'oral ou en interaction, a un accès facilité et peut donc retrouver un environnement accessible, plus ergonomique que CLAPI, similaire à celui des textes qui sont à la base de sa scolarisation, des objets culturels omniprésents dans les contextes éducatifs et ordinaires.
- 62 Nous tenons à souligner ici que la reconfiguration de ressources existantes dans une interface numérique, comme le témoigne le passage de CLAPI à CLAPI-FLE, ainsi que la production de nouvelles données, avec des possibilités d'actualisation différentes et négociables, ouvre la voie à une évolution des pratiques. Les objets scientifiques se trouvent ainsi resémiotisés, et la valorisation qui en est faite englobe aussi des pratiques qui n'avaient pas été envisagées au début. La base CLAPI est donc une ressource qui ne se résout pas à un ensemble d'enregistrements déposés dans une médiathèque, mais qui vit en fonction des usages et des scénarios envisagés, modélisés et investis par les concepteurs et les utilisateurs, à partir d'exigences concrètes.

- 63 Cela nous amène enfin à retracer le parcours réalisé dans cette contribution et à proposer des perspectives de réflexion sur la notion d'*archive*, les objets qu'elle permet de saisir et les termes qui lui sont apparentés.

## Conclusion

- 64 L'une des questions que nous avons cherché à problématiser au fil de cette étude pourrait être formulée de la manière suivante : que veut dire le terme *archive* en environnement de travail numérique et en quoi se distingue-t-il d'expressions telles que *base de données* et *plateforme* ?
- 65 Dans la quatrième section, consacrée à une éthique de l'archive en tant que ressource professionnelle, on a cherché à montrer en quoi la transformation d'une banque de données en ressource *par* des chercheurs *pour* des pratiques éducatives questionne les propriétés des données primaires, leurs caractéristiques communes et leur hétérogénéité fondatrice. Du point de vue de la transformation du site dans la navigation que l'on en fait, une articulation s'opère entre l'*archive*, où se trouvent conservées des productions matérielles et discursives, la *base de données*, où des corpus se trouvent organisés en typologie et renseignés à l'aide de métadonnées, et la *plateforme*, où l'on tente non seulement de mettre en partage des ressources et des outils de travail, mais également de mettre en lien des acteurs de la recherche qui opèrent dans un même domaine. Nous avons également observé que, du point de vue des pratiques, la complexité de l'appropriation d'un environnement numérique de travail réside dans la dialectique entre une *exploitation* d'un outil mis à disposition, aux fonctions canalisées et déjà tramées, et une *exploration* par laquelle négocier l'expérience que l'on croit/veut/doit pouvoir faire de cet environnement, par rapport à d'autres médiations.
- 66 Dans la prolongation de ces questionnements sur la subjectivité des acteurs impliqués dans la constitution des enregistrements, on a cherché à mettre en saillance, dans la cinquième section, la dimension patrimoniale de l'archive de données scientifiques, celle-ci pouvant être appréhendée du point de vue d'une sémiotique des cultures. Une approche différentielle de deux sites sur lesquels se trouvent des enregistrements d'interactions ordinaires nous a notamment permis de questionner les choix qui incombent aux chercheurs quant à leur positionnement, en tant qu'acteurs d'un espace social travaillant dans tel domaine, s'inscrivant dans telle approche et dont les horizons de recherche ne sont pas tracés une fois pour toutes. Ces choix concernent aussi bien les modalités de constitution, de représentation et d'exploitation des données que la valorisation d'un projet de recherche, à quelque étape que ce soit.
- 67 À l'aune de l'ère numérique et de la société de l'enregistrement (Ferraris 2013) mentionnées par les directeurs de ce dossier thématique en introduction, les pratiques de la recherche devraient, elles aussi, être questionnées dans leur relation de contemporanéité avec les pratiques enquêtées : elles produisent une archive *du présent au présent*, qui ne peut faire abstraction de son ancrage socio-historique, comme en témoignent les données secondaires représentatives de différentes "perspectives scientifiques", chronologiquement repérables via les métadonnées.
- 68 La réappropriation d'un outil et la re-spécification de la pratique que l'on en fait questionnent ainsi le parcours sémiotique d'une médiation numérique et la



transformation de l'archive, entre ressource professionnelle et objet culturel. Dès lors, comment observer, à rebours de ses expériences, les tensions entre, d'un côté, l'usage programmé de l'outil inscrit dans la matérialité même du dispositif et, de l'autre côté, la réappropriation que l'on se fait d'une médiation ? Comment repérer les changements de focale opérés par notre regard sur les corpus et par la manière dont on s'en saisit pour faire émerger des questions de recherche ou pour y répondre ? Comment faire la différence entre l'inscription de la subjectivité collective (Gosselin 2010, p. 80) des chercheurs, qui passe par l'appropriation et la mise en commun de méthodes, et sa propre subjectivité individuelle, qui passe par la recherche, le travail et la mise en partage d'une voix personnelle ? Entre un outil/instrument fait pour un usage et un territoire/objet qui prenne en compte les stratifications et les *interpénétrations* actuelles (Basso Fossali 2012), l'appropriation de l'archive pointe le hiatus entre la recherche d'une adéquation aux attentes des concepteurs de la plateforme et l'exigence de creuser son propre sillon en tant qu'utilisateur, bon gré mal gré.

- 69 La visée pratique de la plateforme se manifeste donc dans la mise à plat de la variété des situations enregistrées et des modalités de documentation, favorisant ainsi une recherche outillée sur une base commune à travers la médiation d'une interface. Néanmoins, la documentation élaborée sur des situations ordinaires peut être vue comme une mise en relief, par une *tridimensional understanding* (Pike 1967, p. 41), où les pratiques des participants, comme celles des chercheurs, peuvent être appréciées à l'aune d'une *stéréoscopie* selon une dialectique entre un regard externe à une communauté de pratiques (perspective *étique*) et un regard interne (perspective *émique*) (Basso Fossali 2015). Dans une perspective épistémologique, les pratiques de conception comme celles d'utilisation et de transformation de l'archive demandent ainsi à être questionnées au prisme des enjeux socio-symboliques relatifs au fait même de se revendiquer de tel ou tel champ, de tel ou tel paradigme disciplinaire. Les frontières disciplinaires qui délimitent le champ de pertinence d'une archive sont mises à l'épreuve par les regards multiples dont elle peut faire l'objet, de la mise à distance dogmatique jusqu'à l'implication dans son développement et son enrichissement, en passant par l'indifférence aprioristique et l'adoption curieuse.
- 70 Si une *distant reading* (Moretti 2013) permettrait une vue d'ensemble sur les données grâce à leur traitement quantitatif et automatisé, elle ne pourrait pas apporter à elle seule la spécificité d'un regard qualitatif indiciaire, d'une *deep reading* focalisée sur certains détails permettant de « remonter [...] à une réalité complexe qui n'est pas directement expérimentable » (Ginzburg 1989, p. 149). Les tensions entre les traces de subjectivité que les chercheurs font émerger dans leurs analyses et les traces de subjectivité qu'ils inscrivent dans leurs dispositifs de recherche débouchent sur une réflexion épistémologiquement fructueuse sur les statuts de l'archive, sur la conciliation entre de possibles vies multiples d'un même matériau documentaire. En interrogeant l'archive de manière réflexive, les chercheurs s'interrogent sur leurs propres pratiques de la recherche, sur les enchevêtrements entre une archive se présentant en tant qu'architecture déjà constituée et une archive reconfigurée par une fouille active et personnelle de données. En approchant l'archive, non pas en tant qu'objet figé et poussiéreux, mais en tant qu'espace toujours sujet à reconfiguration, la *distant reading* – qui permet de prendre de la distance, mais n'est pas la distance et peut être appréhendée comme une pratique située – appelle une *deep reading* – qui



permettrait un regard sur des transformations spécifiques de constitution, de consultation et de reconfiguration des archives, elles-mêmes sujettes à évolution.

---

## BIBLIOGRAPHIE

ABOUDA, Lotfi & SKROVEC, Marie (2018), « Pour une micro-diachronie de l'oral : le corpus ESLO-MD », in F. NEVEU, B. HARMEGNIES, L. HRIBA, S. PRÉVOST (éds), *6<sup>e</sup> Congrès Mondial de Linguistique Française - CMLF 2018*, Les Ulis, EDP Sciences. URL : <https://doi.org/10.1051/shsconf/20184611004>.

AVANZI, Mathieu, BÉGUELIN Marie-José, DIÉMOZ, Federica (dirs, 2016a), « Corpus de français parlé et français parlé des corpus », *Corpus* [En ligne], 15. URL : <http://journals.openedition.org/corpus/3060>.

AVANZI, Mathieu, BÉGUELIN Marie-José, DIÉMOZ, Federica (2016b), « De l'archive de la parole au corpus de référence : la base de données orales du français de Suisse romande (OFROM) », *Corpus* [En ligne], 15. URL : <https://journals.openedition.org/corpus/3060>.

BALDAUF-QUILLIATRE, Heike, COLÓN DE CARVAJAL, Isabel, ETIENNE, Carole, JOUIN-CHARDON, Émilie, TESTON-BONNARD, Sandra, & TRAVERSO, Véronique (2016), « CLAPI, une base de données multimodale pour la parole en interaction : apports et dilemmes », *Corpus* [En ligne], 15. URL : <http://journals.openedition.org/corpus/2991>.

BALTHASAR Lukas, BRUXELLES Sylvie, MONDADA Lorenza & TRAVERSO Véronique (2007), « Variations inter-actionnelles et changement catégoriel : L'exemple de 'attends' », in M. AUZANNEAU (dir.), *La Mise en œuvre des langages dans l'interaction*, Paris : L'Harmattan, pp. 299-319.

BASSO FOSSALI, Pierluigi (2008), « Éthique et sémiotique des destins croisés. La négociation de l'agir sensé entre formes de vie », *Protée. Revue internationale de théories et de pratiques sémiotiques*, Université du Québec, vol. 36, n° 1, pp. 59-69.

BASSO FOSSALI, Pierluigi (2012), « Possibilisation, disproportion, interpénétration : trois perspectives pour enquêter sur la productivité de la notion de forme de vie en sémiotique », *Nouveaux Actes Sémiotiques*, n° 115. DOI : 10.7202/019020ar.

BASSO FOSSALI, Pierluigi (2015), « Émancipation et disproportion : pour une problématisation de la notion de culture en sémiotique », in P. MARILLAUD et R. GAUTHIER (éds), *Actes du XXXV Colloque d'Albi - Langages et Signification, Juillet 2014, Albi, Université Jean-Jaurès, Toulouse, Culture et valeurs, CALS/CAMS*, pp. 65-81.

BAUDE, Olivier, BLANCHE-BENVENISTE, Claire, CALAS, Marie-France, CAPPEAU, Paul, CORDEREIX, Pascal et al. (2006), *Corpus oraux. Guide des bonnes pratiques*, Paris/Orléans : CNRS Éditions/Presses Universitaires d'Orléans. URL : <https://hal.archives-ouvertes.fr/hal-00357706/en/>.

BAUDE, Olivier & DUGUA, Céline (2016), « Les ESLO, du portrait sonore au paysage digital », *Corpus* [En ligne], vol. 15. URL : <https://journals.openedition.org/corpus/2924>.

BENZITOUN, Christophe, DEBAISIEUX, Jeanne-Marie & DEULOFEU, Henri-José (2016), « Le projet ORFÉO : un corpus d'étude pour le français contemporain », *Corpus* [En ligne], 15. URL : <https://journals.openedition.org/corpus/2936>.

- BIGGS, Patricia & DALWOOD, Mary (1976), *Les Orléanais ont la parole : Teaching Guide and Tapescript*, Londres, Longman (Livre du maître).
- BRUXELLES, Sylvie, MONDADA, Lorenza, SIMON, Anne-Catherine & TRAVERSO, Véronique (dirs, 2009), « Grands Corpus de français parlé : Bilan historique et perspectives de recherche », *Cahier de linguistique de Louvain*, vol. 33, n° 2, Louvain, Presses Universitaires de Louvain.
- DE CERTEAU, Michel ([1980] 1990), *L'Invention du quotidien*, tome 1 : Arts de faire, Paris, Gallimard.
- DEBAISIEUX, Jeanne-Marie et BENZITOUN, Christophe (dirs, 2020), « Orféo : un corpus et une plateforme pour l'étude du français contemporain », *Langages*, n° 219, pp. 9-24, Armand Colin. URL : <https://www.revues.armand-colin.com/lettres-langues/langages/langages-no-219-32020>.
- FONTANILLE, Jacques (2011), « L'analyse du cours d'action : des pratiques et des corps », *Semen* [En ligne], vol. 32. URL : <https://semen.revues.org/9396>.
- FONTANILLE, Jacques (2015), *Formes de vie*, Liège, Presses Universitaires de Liège, Sigilla. URL : <https://books.openedition.org/pulg/2207>.
- GINZBURG, Carlo (1989), *Mythes, emblèmes, traces. Morphologie et histoire*, Paris, Flammarion.
- GOSSELIN, Laurent (2010), *Les Modalités en français. La validation des représentations*, Amsterdam/New York, Rodopi.
- GRUPE ICOR (BERT, Michel, BRUXELLES, Sylvie, ETIENNE, Carole, MONDADA, Lorenza, TESTON-BONNARD, Sandra & TRAVERSO, Véronique) (2008a), « “Oh : :, oh là là, oh ben...” », les usages du marqueur « oh » en français parlé en interaction », in J. DURAND, B. HABERT et B. LAKS (éds), *1<sup>er</sup> Congrès Mondial de Linguistique Française – CMLF 2008*, Paris, Institut de Linguistique Française, pp. 685-701. URL : <https://doi.org/10.1051/cmlf08099>.
- GRUPE ICOR (BERT, Michel, BRUXELLES, Sylvie, ETIENNE, Carole, MONDADA, Lorenza & TRAVERSO, Véronique) (2008b), « Tool-assisted analysis of interactional corpora : « voilà » in the CLAPI database », *Journal of French Language Studies*, vol. 18, pp. 121-145.
- GRUPE ICOR (BRUXELLES, Sylvie, JOUIN-CHARDON, Émilie, TRAVERSO Véronique) & GUINAMARD, Isabelle (2015), « “Du coup” dans l'interaction orale en français : description de ses usages situés à partir d'une base de données multimédia, et considérations didactiques », in I. GUINAMARD, E. JOUIN-CHARDON, M. RISPAIL, V. TRAVERSO & TRINH D.T. (éds), *Langues parlées, interactions sociales : une variété d'usages pour l'apprentissage du français*, Paris, L'Harmattan, pp. 131-154.
- INGOLD, Tim (2013), *Marcher avec les dragons*, Bruxelles, Zones sensibles.
- LANDOWSKI, Eric (2009), « Avoir prise, donner prise », *Nouveaux Actes Sémiotiques* [en ligne], vol. 112. URL : <http://epublications.unilim.fr/revues/as/2852>.
- MONDADA, Lorenza & PFÄNDER, Stefan (2016), « Corpus international écologique de la langue française (CIEL-F) : un corpus pour la recherche comparée sur le français parlé », *Corpus* [En ligne], vol. 15. URL : <https://journals.openedition.org/corpus/2951>.
- MORETTI, Franco (2013), *Distant Reading*, London/New York, Verso.
- PAVEAU, Marie-Anne & PÉREA, François (dirs, 2012), « Corpus sensibles », *Cahiers de praxématique*, vol. 52. URL : <https://journals.openedition.org/praxematique/3334>.
- PICCOLI, Vanessa & URSI, Biagio (2015), « Le projet Petits Films : du retour aux participants à la valorisation des données », in I. COLÓN DE CARVAJAL & J.-P. MAITRE (éds), *ICODOC 2015 : Colloque Jeunes Chercheurs du Laboratoire ICAR*, Les Ulis, EDP Sciences. URL : <https://doi.org/10.1051/shsconf/20152001019>.

PIKE, Kenneth (1967), *Language in relation to a unified theory of the structure of human behavior*, The Hague, Mouton.

RACINE, Isabelle, DURAND, Jacques & ANDREASSEN, Helene N. (2016), « PFC, codages et représentations : la question du schwa », *Corpus* [En ligne], 15. URL : <https://journals.openedition.org/corpus/3014>.

RASTIER, François (2011), « Objets et performances sémiotiques. L'objectivation critique dans les sciences de la culture », *Texte !* vol. XVI, n° 1. URL: <http://www.revue-texto.net/index.php?id=2734>.

SIMONDON, Gilbert ([1958] 1989), *De l'existence des objets techniques*, Paris, Aubier.

TEN HAVE, Peter (1999), *Doing Conversation Analysis: A Practical Guide*, London, Sage.

THOMAS, Anita, GRANFELDT, Jonas, JOUIN-CHARDON, Émilie et ETIENNE, Carole (2016), « Conversations authentiques et CECR : compréhension globale d'interactions naturelles par des apprenants de FLE », *Cahiers de l'AFLS*, vol. 20, n° 1. URL : [https://afls.net/cahiers/20.1/2-CahiersAFLS20\(1\)-Thomas\\_al.pdf](https://afls.net/cahiers/20.1/2-CahiersAFLS20(1)-Thomas_al.pdf).

## NOTES

1. Ces trois types de public sont affichés dans la page d'accueil du site PFC : <http://www.projet-pfc.net>.
2. Ici, nous utilisons le terme *archive* dans son acception générique, comme réservoir de corpus que l'on peut consulter à partir d'une matrice numérique. Une problématisation épistémologique et, donc, terminologique de l'archive sera proposée en fin d'article. Ces archives se trouvent centralisées sur ORTOLANG, une infrastructure en réseau offrant un réservoir de données (corpus, lexiques, dictionnaires, etc.) et d'outils sur la langue et son traitement (<https://www.ortolang.fr/>).
3. Le projet ORFÉO (ANR 12-CORP-0005), dirigé par Jeanne-Marie Debaisieux et mené de 2013 à 2017, s'inscrit dans une réflexion sur l'annotation multi-niveau de données de français oral qui a été initiée par le projet « Corpus prosodique de référence du français parlé » (Rhapsodie, ANR 07-CORP-030-01), dirigé par Anne Lacheret et mené de 2007 à 2012. Plus récemment, le projet franco-allemand « Segmentation de Corpus Oraux » (SegCor, ANR-15-FRAL-0004), coordonné par Véronique Traverso (ICAR, ENS de Lyon) et Thomas Schmidt (IDS, Mannheim), vise au développement d'une méthode de segmentation utilisable pour l'analyse de la parole-en-interaction sur différents niveaux et pour différentes communautés de chercheurs, en s'appuyant sur des ressources existantes (des corpus issus des bases CLAPI et ESLO pour le français, et de la base de données FOLK pour l'allemand), avec la prise en compte des aspects contrastifs entre les deux langues.
4. Pour accéder au référentiel du CECR : <https://www.coe.int/fr/web/common-european-framework-reference-languages>.
5. Si l'on se dirige vers la page du site du corpus orléanais ESLO (<http://eslo.huma-num.fr>), on voit qu'une autre configuration a été choisie, notamment par des entrées sur un bandeau vertical et un visuel en en-tête du site.
6. Pour accéder à cette page, voir le site : [http://clapi.icar.cnrs.fr/V3\\_Accueil.php?interface\\_langue=FR](http://clapi.icar.cnrs.fr/V3_Accueil.php?interface_langue=FR).
7. L'ouverture d'un nouvel onglet questionne notre utilisation de l'interface, la délimitation d'un périmètre du site où naviguer et la formation d'un territoire particulier de la recherche. La frontière entre ces deux environnements de travail aurait pu être amoindrie, par le changement

de page sur un même onglet, ou elle aurait pu être accentuée, par l'ouverture d'une nouvelle fenêtre.

8. Pour accéder à cette page, voir le site : <http://clapi.icar.cnrs.fr/FLE/accueil.php>. En observant le chemin de destination, on observe que l'architecture du site est différente de la structure construite par l'interface : si CLAPI-FLE apparaît en tant que page interne au site CLAPI (<http://clapi.icar.cnrs.fr/FLE/>), l'expérience que nous nous faisons de l'interface, en tant qu'utilisateur, est celle de se diriger vers un autre site.

9. Pour accéder à cette page, voir le site : [http://clapi.icar.cnrs.fr/V3\\_Corpus\\_Liste.php?interface\\_langue=FR](http://clapi.icar.cnrs.fr/V3_Corpus_Liste.php?interface_langue=FR).

10. Pour accéder à cette page, voir le site : [http://clapi.icar.cnrs.fr/FLE/liste\\_extraits.php](http://clapi.icar.cnrs.fr/FLE/liste_extraits.php).

11. Si ces synopsis sont pour la plupart d'ordre descriptif seulement, il y a des cas où une évaluation qualitative de l'échange est élaborée. Une telle différence peut s'apprécier notamment entre le synopsis de l'extrait intitulé CLAPI-FLE-2 (« Des clients s'enchaînent dans un petit commerce, une boulangerie, et demandent des produits. Certains échanges sont un peu plus informels car les clients se connaissent ou il s'agit de clients habituels ») et l'extrait CLAPI-FLE-1 (« Un client commande des produits de boulangerie. L'échange est concis et clair »).

12. On entrevoit une perspective heuristique intéressante pour tester la finesse et l'efficacité du dispositif : celle d'enquêter sur les *stratégies* d'usage et *tactiques* rusées (de Certeau [1980] 1990) mises en œuvre par les apprenants dans le passage du travail sur une interaction archivée et documentée à leurs propres pratiques langagières *in vivo* que l'on pourrait enregistrer. La mise en regard des enregistrements sur lesquels ils ont travaillé (sur CLAPI-FLE) et des enregistrements de leurs propres pratiques pourrait notamment permettre de travailler sur les contraintes propres à chaque situation et sur les réalisations inventives des apprenants.

13. Voir : [https://aslan.universite-lyon.fr/medias/fichier/ficheprojet-clapi-fle\\_1534513249286-pdf](https://aslan.universite-lyon.fr/medias/fichier/ficheprojet-clapi-fle_1534513249286-pdf).

14. En continuité avec la proposition de Pierluigi Basso Fossali (2008) entre *déontologie* et *éthique*, la première est conçue ici comme l'ensemble des obligations dont une catégorie professionnelle ou un groupe qui opère dans le domaine de la recherche se dote pour définir sa pratique par rapport aux usagers ou aux acteurs sociaux concernés. Dans le cadre de l'enregistrement de données, un document permet aux acteurs de donner leur consentement éclairé, selon des modalités établies par le laboratoire ICAR et validées par les institutions de la recherche scientifique (CNRS). Trois types de documents sont prévus à ces effets : un pour les participants majeurs, un autre pour les mineurs (consentement parental) et un dernier pour les institutions publiques ou privées impliquées dans les interactions enregistrées.

15. Le projet « Réfugiés, migrants et leurs langues face aux services de santé » (REMILAS), mené de 2016 à 2019 et coordonné par Véronique Traverso (ICAR, ENS de Lyon), a été financé par l'ANR dans le cadre du programme Flash Asile. Une description des axes de recherche et des activités proposées par les partenaires du projet (le laboratoire ICAR de Lyon et l'observatoire Orspere-Samdarra du Centre hospitalier Le Vinatier de Bron) est consultable à l'adresse suivante : <http://www.icar.cnrs.fr/sites/projet-remilas/>.

16. Les enregistrements de ce corpus, avec d'autres enregistrements, ont servis de base empirique pour le projet ANR « Léonard – Acquisition du langage et grammaticalisation » (<http://anr-leonard.ens-lsh.fr>), dirigé par Aliyah Morgenstern.

17. La convention ICOR est consultable à l'adresse suivante : [http://icar.cnrs.fr/projets/corinte/documents/2013\\_Conv\\_ICOR\\_250313.pdf](http://icar.cnrs.fr/projets/corinte/documents/2013_Conv_ICOR_250313.pdf).

---

## RÉSUMÉS

Dans un dialogue entre étude sémiotique et recherche linguistique sur corpus, cet article traite la notion d'archive à l'aune de pratiques mobilisées en sciences du langage, de la constitution d'enregistrements audio(visuels) d'interactions sociales en situation naturelle jusqu'à l'implémentation et à l'exploitation des données constituées dans une plateforme numérique. Nous proposons une analyse contrastive entre l'organisation de la plateforme CLAPI (Corpus de LANGues Parlées en Interaction), une base de données de divers enregistrements documentant plusieurs types d'interaction, et l'organisation de CLAPI-FLE, son « volet enseignement ». Nous nous penchons sur les passages d'un matériau conçu pour la recherche en sciences du langage (plus précisément en linguistique interactionnelle) à un matériau redéfini et ré-encadré pour une application didactique. En pointant les nuances qui émergent entre les notions d'*exploitation*, d'*exploration* et d'*appropriation* de l'archive, cette étude de cas propose un regard épistémologique sur les enjeux éthiques et socio-symboliques relatifs au respect des participants aux enquêtes, à leur implication sur le terrain et au positionnement de chercheurs dans des communautés différentes.

At the crossroads of semiotics and corpus-based linguistics, this article deals with the notion of *archive* in the light of practices used in language sciences, from the constitution of audio(visual) recordings of social interactions in natural settings to the implementation and exploitation of the data in a digital platform. We propose a contrastive analysis between the organization of the CLAPI platform (*Corpus de LANGues Parlées en Interaction*), a database of various recordings documenting several types of interaction, and the organization of CLAPI-FLE, a teaching-oriented database stemming from CLAPI. We focus our attention on the transitions from material designed for research in language sciences (more specifically in interactional linguistics) to material redefined and re-framed for didactic purposes. By pointing out the nuances that emerge between the notions of *exploitation*, *exploration*, and *appropriation* of the archive, this case study offers an epistemological look at the ethical and socio-symbolic issues related to the respect towards participants, to their involvement in the fieldwork, and to the positioning of researchers within diverse communities.

## INDEX

**Keywords :** archive, FLE French as a Foreign Language, ethics, corpus linguistics, semiotics

**Mots-clés :** archive, FLE, éthique, linguistique de corpus, sémiotique

## AUTEURS

### JULIEN THIBURCE

En couplant les outils de la sémiotique et ceux de l'analyse conversationnelle, ses travaux portent sur les dynamiques de co-construction du sens et sur les formes de négociation d'expériences, de savoirs et de valeurs à travers des pratiques sociales et langagières en interaction. Dans une approche sur corpus, il mobilise notamment des méthodes visuelles (photos et vidéos *in situ*) pour enquêter sur les transformations des relations matérielles, affectives et symboliques entre des agents sociaux et leurs environnements.

Actuellement chercheur postdoctoral CNRS, Julien Thiburce codirige le projet PrisM (Prisons et

Musées), une étude des représentations sur les prisons dans les musées en partenariat avec le Musée International de la Croix-Rouge et du Croissant Rouge (Genève), le Musée des Confluences (Lyon) et le Mémorial national de la prison Montluc (Lyon).

Travaux : <https://cv.archives-ouvertes.fr/julien-thiburce>

Courriel : [julien.thiburce\[at\]ens-lyon.fr](mailto:julien.thiburce[at]ens-lyon.fr)

### **BIAGIO URSI**

Linguiste interactionniste, il a travaillé sur plusieurs projets d'annotation de corpus oraux, dans une perspective linguistique multidimensionnelle (ORFÉO, SegCor) et appliquée, pour l'enseignement (CLAPI-FLE) et l'apprentissage (FLEURON) du français. Ses travaux portent sur l'analyse de données audiovisuelles enregistrées en situation naturelle et documentant différents contextes, en français et en italien. En particulier, il s'intéresse à la co-construction du sens et de l'expérience sensorielle (notamment le toucher), telles qu'elles sont réalisées dans la temporalité de la parole en interaction, par l'instauration de cadres de participation multiples et changeants. Les ressources corporelles, matérielles et technologiques mobilisées par les participants représentent ses éléments d'analyse privilégiés, au même titre que les ressources linguistiques. Il est actuellement attaché temporaire d'enseignement et de recherche à l'Université d'Aix-Marseille.

Travaux : <https://cv.archives-ouvertes.fr/biagio-ursi>

Courriel : [biagio.ursi\[at\]univ-amu.fr](mailto:biagio.ursi[at]univ-amu.fr)