



# Du corpus enregistré au corpus analysé : questions méthodologiques sur l'utilisation d'outils de requêtes informatisés

Isabel Colón de Carvajal

## ► To cite this version:

Isabel Colón de Carvajal. Du corpus enregistré au corpus analysé : questions méthodologiques sur l'utilisation d'outils de requêtes informatisés. Les cahiers de praxématique, Montpellier : Presses universitaires de la Méditerranée, 2006-, 2010, pp.313-326. <halshs-00630514v2>

**HAL Id: halshs-00630514**

**<https://halshs.archives-ouvertes.fr/halshs-00630514v2>**

Submitted on 3 Jul 2013

**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.

## Du corpus enregistré au corpus analysé : questions méthodologiques sur l'utilisation d'outils de requêtes informatisés

### Introduction

Je me suis intéressée dans mon travail de recherche (Colón de Carvajal, 2010) à la mobilisation des ressources technologiques (téléphone, webcam, ordinateurs, etc.) dans les interactions professionnelles, et en particulier dans des centres d'appels<sup>1</sup> ; je m'appuie sur les trois champs disciplinaires que sont l'Ethnométhodologie, l'Analyse Conversationnelle et les *Workplace Studies*. Afin d'analyser les pratiques des professionnels en situation naturelle, j'ai mis en place un dispositif d'enregistrement permettant d'observer à la fois l'activité des opérateurs dans leur environnement de travail, et celle des écrans sur lesquels agissent les professionnels. J'ai ainsi obtenu deux points de vue complémentaires de la même situation d'interaction. A partir de ces deux points de vue, j'ai développé une réflexion méthodologique propre au travail sur corpus, que je vais présenter dans cet article. Nous montrerons l'évolution entre le corpus enregistré, afin de répondre aux questions initiales de recherche, et le corpus analysé qui rend compte de problèmes pratiques émergents rendus pertinents par les participants dans leurs interactions.

Pour ce faire, je me fonde sur une analyse développée dans ma thèse. Lors de la transcription des données, j'ai remarqué que les participants rapportent ou reformulent à l'oral des informations qui sont écrites sur un écran, en utilisant des verbes introductifs du type « il dit que », fréquents dans le discours rapporté à l'oral. Je me suis alors interrogée sur le lien entre le cours d'action dans lequel sont engagés les participants et l'émergence de ces discours rapportés/reformulés, où la référence aux messages écrits peut potentiellement modifier le statut des écrans. Afin de mener à bien mon analyse, j'ai constitué une collection (Schegloff, 1987, 1988b, 1993) regroupant des cas où une thématization implicite ou explicite d'un écran est rendue manifeste par les participants. Dans une première étape, j'ai utilisé les outils de requêtes informatisés disponibles sur la plateforme CLAPI<sup>2</sup> développée au laboratoire ICAR<sup>3</sup>. J'ai en particulier recherché les différentes formes de verbes introductifs et leurs variations possibles en fonction du pronom personnel sujet. J'ai obtenu un certain nombre d'extraits permettant de développer une analyse approfondie.

Le travail que je présente ici permet de problématiser, d'une part, la recherche informatisée d'extraits à partir d'une forme donnée, d'autre part, le rapport entre les résultats obtenus avec ce type de recherche (approche quantitative) et mes attentes pour réaliser une analyse en termes d'action (approche qualitative). Plus particulièrement, je vais montrer que le chercheur ne peut pas se contenter d'une collection établie à partir d'une recherche informatisée seule, mais qu'il doit passer par d'autres étapes complémentaires pour affiner sa collection avant de produire l'analyse.

Dans un premier temps, je présenterai brièvement la situation d'interaction entre les participants du centre d'appels. Après une présentation de la plateforme CLAPI, j'expliquerai la manière dont j'ai constitué ma première collection à l'aide des outils de requêtes informatisés. Ensuite, je proposerai une réflexion sur les avantages et les désavantages du caractère informatisé de l'outil ; et j'expliquerai différents enrichissements possibles de la collection que j'ai dû réaliser pour rendre compte des éléments pertinents afin de produire une analyse linguistique et multimodale de l'activité des participants. J'appuierai ma réflexion sur trois extraits. Précisons enfin que l'objectif de mon article n'est pas de développer une analyse fine de ces exemples.

---

<sup>1</sup> Les résultats de cette recherche sont intégralement disponibles sur Internet : [http://theses.univ-lyon2.fr/documents/lyon2/2010/colon\\_de\\_carvajal\\_i](http://theses.univ-lyon2.fr/documents/lyon2/2010/colon_de_carvajal_i)

<sup>2</sup> <http://clapi.uni-lyon2.fr>

<sup>3</sup> <http://icar.univ-lyon2.fr>

## Situation d'interaction

Les données présentées ici sont issues du corpus LocBike présentant une situation d'interaction professionnelle en centre d'appels, dans laquelle des opérateurs et des clients interagissent par téléphone à distance. L'activité principale des opérateurs consiste à répondre aux demandes des clients, qui louent des vélos mis à disposition dans la ville. Les opérateurs gèrent les comptes des différents clients sur ordinateur, grâce à une plateforme logicielle qui regroupe les dossiers des usagers ainsi que l'ensemble des stations LocBike réparties dans la ville. De leur côté, les clients manipulent les stations LocBike. En sélectionnant une succession d'options sur un écran tactile, ils obtiennent l'autorisation d'emprunter un vélo pour une durée limitée.

De mon point de vue, l'intérêt analytique de ce corpus réside dans le fait que l'opérateur doit gérer l'échange avec le client à l'aide d'un ordinateur. L'objet 'ordinateur' a une fonction centrale dans cette situation d'interaction. L'activité des professionnels dans ce contexte est étroitement organisée autour de la manipulation d'objets technologiques : l'ordinateur (associé à la souris et au clavier), ainsi que le téléphone. Dans ce corpus, je me suis intéressée non seulement à la manipulation des technologies informatiques par les opérateurs, mais aussi à leur mobilisation par les appelants confrontés à des problèmes face aux terminaux de location disséminés dans l'espace public.

Ici, je me suis interrogée sur la manière dont l'objet 'écran' est convoqué, décrit, intégré, dans un « réseau socio-technique »<sup>4</sup> (Licoppe, 2008), flexible et redéfinissable dans la conversation entre les participants. L'objet 'écran' prend alors une place essentielle dans l'activité de service pour résoudre le problème d'un client. J'ai voulu montrer le lien entre le cours d'action dans lequel les participants sont engagés et l'émergence de discours rapportés ou reformulés, au cours duquel la référence aux messages écrits peut transformer les écrans et les systèmes informatiques en « actants », notamment lors du guidage de l'activité à distance en temps réel.

Avant de développer la manière dont j'ai constitué une collection d'extraits représentatifs pour mon étude, je propose une brève présentation de la plateforme CLAPI.

## Présentation de CLAPI

La plateforme CLAPI – Corpus de Langue Parlée en Interaction – a été développée en 2002 par le groupe ICOR (Interaction CORpus). Elle héberge un ensemble de corpus d'interactions enregistrés en situation réelle, dans des contextes variés (Groupe ICOR, 2008, 2009, 2010). Ces corpus ont été collectés depuis plus de vingt ans à partir de programmes de recherche individuels (mémoires, thèses) ou collectifs (projets d'équipe, réponses à des appels d'offres). Ils proviennent des membres du laboratoire, mais aussi d'autres équipes de recherche. Jusqu'à récemment, ces corpus étaient dispersés, sous la responsabilité de leur auteur, et ne bénéficiaient pas de la diffusion et de l'enrichissement qu'offre la plateforme CLAPI aujourd'hui, notamment du fait de l'utilisation d'outils de requête.



---

<sup>4</sup> La notion de « réseau socio-technique » a été développée par Latour (1992) en se basant sur le modèle de l'acteur-réseau où les personnes et les choses constituent des « actants », également mobilisables et mobilisés dans des réseaux sociotechniques hétérogènes (Licoppe, 2008 : 296). L'analyse du travail collectif est nécessaire pour produire les différents attachements et agencements sociotechniques capables d'agir à distance sur les situations, de les configurer et de les cadrer (Latour, 1994).

CLAPI Corpus de Langues Parlées en Interaction Pour citer cette page : CLAPI, <http://clapi.univ-lyon2.fr>

Vous êtes connecté en tant qu'invis

**Bienvenue dans CLAPI banque de données et plateforme logicielle**

<p>Accueil</p> <p>Présentation des corpus</p> <p>Présentation des transcriptions</p> <p>Outils de requêtes</p> <p>Identification des formes</p> <p>Gestion de vos collections</p> <p>Changement de compte</p> <p>Contact</p>	<p><b>CLAPI</b> CLAPI est une banque de données outillée de Corpus de Langue Parlée en Interaction enregistrés en situation réelle, dans des contextes variés : Interactions professionnelles, institutionnelles ou privées, commerciales, didactiques, médicales, ...</p> <p><b>VOLUME</b> 45 corpus, 327 enregistrements (135 h), 514 transcriptions documentés par 75 descripteurs, 56 h balisées pour des traitements d'analyses et de requêtes</p> <p><b>PLATEFORME LOGICIELLE</b> -&gt; Un ensemble d'outils d'analyse automatique des données : fréquences, co-occurrences, répétitions, ... -&gt; Un outil de requêtes complexes pour mettre en évidence les conditions entre tokens et phénomènes interactionnels -&gt; Un concordancier aligné avec le signal par streaming audio/vidéo</p>	<p style="text-align: center;"><b>CONDITIONS D'ACCES</b></p> <p>==&gt; 35 h de données, soit plus de 70 % de la base en accès libre pour les analyses et les requêtes</p> <p>==&gt; 100 transcriptions documentant une grande diversité de situations interactionnelles</p> <p>16 h de données téléchargeables (transcriptions, enregistrements)</p> <p>Accès libre sur la totalité de la base aux éléments suivants : Descripteurs, Lexique, Fréquence des tokens, Synthèse des transcriptions</p>	<p><b>CORPUS D'INTERACTIONS</b> Nous vous invitons à consulter <a href="#">le site associé CORINTE</a>, développé au sein du groupe ICOR, dédié à la recherche multimodale sur CORPUS en INTERACTION</p> <p><b>CONVENTION ICOR</b> Vous pouvez télécharger <a href="#">la convention ICOR</a> avec son formatisme et des exemples d'utilisation</p> <p><b>INTEROPERABILITE</b>   Un export des descripteurs des corpus, des enregistrements, des transcriptions est disponible en Dublin Core-OLAC Un export des transcriptions (descripteurs et version textuelle de la transcription) est disponible en TEI</p>
--	---	---	---

Le groupe ICOR, Interaction CORPUS, est un collectif qui conçoit et développe la plateforme CLAPI et le site CORINTE. Actuellement, il est composé de Michel EERTI, Sylvie BRUXELLES, Carole ETIENNE, Emille JOURIN, Justine LASCAR, Lorenza MONDADA, Christian PLANTIN, Sandra TESTON, Véronique TRAVERSO, Daniel VALERO.

Les travaux du groupe ICOR ont bénéficié de la collaboration et de l'expertise de Kamel AOUCHE et Lukas BALTHASAR

Nous vous informons que les requêtes/analyses effectuées font l'objet d'une trace informatique à des fins d'optimisation

Image 1 : Page d'accueil de la plateforme CLAPI

La plateforme CLAPI s'appuie sur une médiathèque qui contient les données hébergées dans la banque de données mais également d'autres éléments documentant les corpus (documents papiers, etc.). Pour CLAPI, un corpus est composé de :

- Données primaires : les enregistrements,
- Données secondaires : les transcriptions et conventions de transcription ; et les notes de terrain du collecteur,
- Autres : les autorisations ; et les publications (ex : les rapports, les mémoires portant sur certains corpus CLAPI, originaux ou en version informatique).

La plateforme CLAPI comprend également une plateforme logicielle composée d'un ensemble d'outils d'analyse automatique comme par exemple l'analyse des fréquences d'un token<sup>5</sup>, de co-occurrences d'un token, de répétitions d'un token. Elle propose également un outil de requêtes complexes pour mettre en évidence les corrélations entre des tokens et des phénomènes interactionnels (chevauchements, pauses). Enfin, elle propose un concordancier aligné avec le signal par streaming audio/vidéo.

## Collection et utilisation d'outils de requêtes informatisés

Dans le cadre de mon travail de thèse, j'ai utilisé la plateforme CLAPI. Elle me permet d'établir une collection. Dans un espace de travail personnel<sup>6</sup> en ligne, j'ai chargé l'ensemble des transcriptions issues du corpus LocBike. Grâce aux outils de requêtes proposés par la plateforme CLAPI, j'ai cherché à repérer de manière automatique les énoncés produits par les clients ou par les opérateurs dans lesquels ils se focalisent sur l'objet 'écran'. Les requêtes outillées sont possibles en renseignant soit un phénomène précis, soit un token, à partir duquel s'effectue la recherche automatique. Mon objectif est d'analyser l'activité qui consiste à rapporter à l'oral une information écrite sur un écran. J'ai ainsi déterminé une liste de tokens pertinents et récurrents dans cette activité pouvant rendre compte d'une focalisation sur l'écran, et cela lorsqu'un participant le thématise ou le mobilise dans l'interaction. Cette liste a été établie de manière manuelle, et après la phase de transcription des données ; elle est composée des tokens suivants : « dit, met, demande, affiche, indique, créditer, recréditer ». Elle correspond principalement aux verbes introductifs qui précèdent le contenu informationnel affiché sur l'écran et qui est retransmis à l'oral.

Le choix des outils de requêtes disponibles sur la plateforme CLAPI permet un éventail d'options intéressantes en fonction de ce que je recherche à extraire des données. Parmi ces possibilités, j'ai retenu celle qui semblait être la plus pertinente et la plus efficace en terme de précision et de gain de

<sup>5</sup> Dans la plateforme CLAPI, un token est un mot dans une transcription, délimité par un espace (cf. glossaire sur le site CORINTE : [http://icar.univ-lyon2.fr/projets/corinte/bandeau\\_droit/glossaire.htm#to](http://icar.univ-lyon2.fr/projets/corinte/bandeau_droit/glossaire.htm#to)).

<sup>6</sup> L'espace personnel dans CLAPI correspond à un espace sécurisé, qui permet de transférer mes transcriptions sur la plateforme sans qu'elles ne soient visibles ou consultables par les visiteurs.

temps : la co-occurrence d'un token. Cet outil de requête a permis de rechercher automatiquement les co-occurrences gauches et droites d'un token donné, par fréquence décroissante, en ayant un accès direct au contexte (*i.e.* avec la transcription et le signal audio/vidéo aligné), notamment de la distribution des éléments sujets autour des verbes introductifs.

Après avoir sélectionné l'outil de requête, j'ai indiqué le token à rechercher. J'ai effectué la recherche sur l'ensemble des transcriptions de mon corpus LocBike disponible dans l'espace personnel de CLAPI. En indiquant le premier token listé « dit », puis en validant la recherche, l'outil a affiché une première information concernant le nombre d'attestations recensées du token « dit », à savoir 129 fois dans le corpus LocBike.

J'ai ensuite visualisé la liste des co-occurrences gauches et droites du token « dit » :

LISTE DES COOCCURENCES GAUCHES DU MOT <i>dit</i> PAR FREQUENCE DECROISSANTE					LISTE DES COOCCURENCES DROITES DU MOT <i>dit</i> PAR FREQUENCE DECROISSANTE				
DANS LA MEME PRODUCTION VERBALE(PV)					DANS LA MEME PRODUCTION VERBALE(PV)				
machine	vous	<i>dit</i>	==> 1	<a href="#">Voir</a>	<i>dit</i>	que	vous	==> 3	<a href="#">Voir</a>
on	vous	<i>dit</i>	==> 2	<a href="#">Voir</a>	<i>dit</i>	que	xxx	==> 1	<a href="#">Voir</a>
ça	vous	<i>dit</i>	==> 6	<a href="#">Voir</a>	<i>dit</i>	que	y	==> 1	<a href="#">Voir</a>
	<i>avait</i>	<i>dit</i>	==> 8	<a href="#">Voir</a>	<i>dit</i>	<i>de</i>		==> 8	<a href="#">Voir</a>
m'	<i>avait</i>	<i>dit</i>	==> 8	<a href="#">Voir</a>	<i>dit</i>	<i>de</i>		==> 2	<a href="#">Voir</a>
	<i>me</i>	<i>dit</i>	==> 8	<a href="#">Voir</a>	<i>dit</i>	de	d'	==> 1	<a href="#">Voir</a>
elle	me	<i>dit</i>	==> 2	<a href="#">Voir</a>	<i>dit</i>	de	de	==> 1	<a href="#">Voir</a>
i'	me	<i>dit</i>	==> 1	<a href="#">Voir</a>	<i>dit</i>	de	le	==> 1	<a href="#">Voir</a>
il	me	<i>dit</i>	==> 1	<a href="#">Voir</a>	<i>dit</i>	de	me	==> 1	<a href="#">Voir</a>
on	me	<i>dit</i>	==> 4	<a href="#">Voir</a>	<i>dit</i>	de	porter	==> 1	<a href="#">Voir</a>

Image 2 : Résultat partiel de la requête sur les co-occurrences du token « dit »

À ce stade de la recherche, la partie dite “outillée” de la requête est terminée sur ce premier token. J'ai donc ensuite procédé à une vérification manuelle des attestations recensées pour s'assurer qu'elles correspondaient bien à la problématique qui m'intéressait. Pour cela, j'ai affiché les passages de transcriptions correspondant aux attestations en cliquant sur le lien associé « Voir ». Pour chaque contexte d'énonciation, j'ai alors déterminé si telle ou telle attestation était un cas pertinent pour l'objet de mon analyse. L'exemple ci-après montre un cas où l'objet 'écran' est thématisé implicitement par l'opérateur lorsqu'il reformule une information provenant des messages affichés sur les écrans des stations de location de vélo.

### Ext3-1a\_L26-16\_07'02 : Machine recréditée

```
OPE [euh ] donc j` vous débloque la carte sans aucune facturation\
CLI [ouais ]
(0.5)
OPE [dans ] deux minutes\
CLI [oui ]
(0.5)
OPE par contre/ si dans plusieurs jours la machine vous dit de recréditer
c'est que malgré tout [ça a été ] facturé alors que ça n` devrait
pas\
CLI [oui ]
(0.5)
CLI [oui ]
OPE [donc ] dans c` cas là\ il suffit de nous app`ler\ (.) et on
annul`ra la facture\
```

Au fur et à mesure de la vérification des attestations dans leur contexte, j'ai sélectionné une série d'extraits pour mon analyse dans un fichier de collection intégré dans CLAPI grâce à un code attribué à chacun d'eux. Ainsi, j'ai sauvegardé les extraits pertinents pour mon étude. Après avoir vérifié les attestations du token « dit », j'ai procédé à une nouvelle requête informatisée sur un nouveau token de la liste que j'avais établie. J'ai retenu au total quarante extraits, et j'ai analysé vingt-neuf d'entre eux quant à l'intégration de l'objet 'écran' comme « réseau socio technique ».

## Avantages et désavantages du caractère informatisé de l'outil

La plateforme CLAPI m'a fourni très rapidement une première sélection. Elle m'a également permis dans une seconde lecture, de donner la liste des contextes immédiats autour du verbe introductif. Cette recherche informatisée a été cependant étroitement dépendante d'une vérification manuelle dans les transcriptions d'origine. Le but était de ne pas exclure une forme grammaticale s'inscrivant dans une référence à une action réalisée sur l'objet 'écran'. En effet, j'ai remarqué que le temps des verbes pouvait être au présent et au passé. Il a donc été nécessaire de réaliser des requêtes automatiques sur le même verbe, en distinguant la forme au présent de celle au passé (ex : « dit » et « disait »)<sup>7</sup>.

En terme d'action, il est possible que l'objet 'écran' soit thématiqué dans le discours des participants sans que ceux-ci introduisent explicitement la référence à l'objet par une construction verbale de type « il dit que ». En l'absence de verbe introductif, l'outil de requête informatisé sur un token donné ne permet pas systématiquement de cibler un extrait où les actions réalisées sur l'écran sont rendues pertinentes pour les participants. Par exemple, dans le cas de lecture simultanée à l'écran, il est possible que le participant partage directement l'information affichée sur son écran. Ces extraits sont alors repérés par une lecture pas à pas des transcriptions du corpus et non par une requête informatisée. L'extrait ci-dessous présente un cas d'absence de verbe introductif, où l'opératrice et la cliente décrivent leurs représentations respectives des vélos mis en place sur une station. Elles collaborent à une activité de description en terme de « visualisation » de la station et l'opératrice s'appuie sur les informations qu'elle a sur son écran :

### Ext3-2\_L26-13\_42'12 : Tableau

```
1   CLI   #euh c'est c'est tous les matins tous les soirs normal`ment et
    ecr   #aff. page borne concernée-->
2       là euh: (.) $ce matin:\ (2.4) (d` façon) vous visualisez vous
    opeG   $clic icône état borne
3       visualisez pas donc j` peux pas vous m- donner mon numéro\
4       (0.7)
5   OPE   ben: #en fait si\ moi j'ai un tableau si vous voulez un genre
    ecr   ----#aff. liste vélo accroché à borne
6       d` tableau excel:/ $on va dire\ j'ai euh:: $ j` peux vous
    opeG   $descend ascenseur vertical$
7       dire que $par exemple euh:: `fin j'ai des j'ai- `fin j` peux
    opeG   $monte ascenseur vertical-->
8       pas vous dire\ (0.6) j'ai des vélos\ $ qui sont là bas\ (0.3)
    opeG   -----$
9       euh:: $j'ai: j'ai deux numéros d` vélos euh::: sûr\ $ (0.3)
    opeG   $descend ascenseur vertical $
10      i` $y en a d'autres où: j'ai pa:s `fin$ j'ai pas leurs
    opeG   $monte ascenseur vertical $
11      numéros d` vélo et: ça pourrait être les vôtres\ (0.3) mais i`
12      faudrait qu` je sache euh:\
13      (0.4)
14   CLI   c'est c'est à je croi:s (0.2) (le clé en èl) (0.4) j` sais pas
15      si vou::s voyez c'est\ i` en a beau[coup hein\]
16   OPE   [alors par ] contre c'est
17      vrai qu` j` la visualise pas euh: telle quelle quoi\ je je sais
18      pas comment elle est faite\
```

Les choix opérés par les participants dans l'activité de « faire dire » (pour l'opératrice) ou « dire » (pour la cliente) le numéro de la bornette associé au numéro du vélo sont fortement liés au choix lexical des termes employés pour décrire la référence contextuelle sur laquelle s'appuient les participantes à distance, chacune essayant de rendre intelligible la représentation physique de leur environnement. Rétrospectivement, on comprend que l'emploi du verbe « visualiser » est

---

<sup>7</sup> En 2011, CLAPI va mettre à disposition un outil morphologique permettant de lister automatiquement les variantes conjuguées d'un verbe.

problématique dans le déroulement de l'interaction car les participantes ne s'alignent pas sur la même définition de ce verbe : l'opératrice utilise une représentation abstraite en liste sur l'écran : « tableau excel » (l. 6) ; la cliente utilise une disposition spatiale concrète des vélos sur la station : « c'est le clé en el » (l. 14).

Par ailleurs, la transcription de l'extrait obtenue grâce à l'outil est une transcription intermédiaire à partir de laquelle je fais des modifications. Sur l'extrait 3 (cf. Ex3-2), j'ai réorganisé les tours de parole pour rendre compte de l'explication continue de l'opératrice. Cette explication est ponctuée par des continueurs (l. 2,4,7) successifs et de pauses notées comme inter-tour dans la première version de la transcription. Voici la transcription modifiées, attribuées à la locutrice OPE :

### Ext3-1b\_L26-16\_07'02 : Machine recrédite

```

1   OPE   [euh ] donc j` vous débloque la carte sans aucune facturation\
2   CLI   [ouais]
3   OPE   (0.5) [dans] deux minutes\ (0.5) par contre/ si dans plusieurs&
4   CLI   [oui ]
5   OPE   &jours la machine vous dit de recréditer c'est que malgré tout
6         [ça a été] facturé alors que ça n` devrait pas\ (0.5) [donc]&
7   CLI   [oui ] [oui ]
8   OPE   &dans c` cas là\ il suffit de nous app`ler\ (.) et on annul`ra
9         la facture\

```

Cette seconde version de la transcription représente ainsi une première analyse de l'extrait que propose le chercheur. Comme le souligne Mondada (2002 : 70), « la transcription devient un parcours sans cesse recommencé » où elle rassemble les éléments pertinents pour le transcripteur. Elle devient alors « un véritable outil de travail, non pas en se substituant aux données enregistrées mais précisément dans un va-et-vient constant avec elles ».

### Enrichissement de la transcription pour l'analyse

Les transcriptions obtenues à partir des requêtes informatisées dans CLAPI sont des transcriptions de l'oral uniquement. En termes d'action, il était important pour moi de décrire certaines actions des opérateurs en train de se faire pendant l'interaction. Il a donc été essentiel d'enrichir mes transcriptions par des annotations multimodales. Le but était de rendre compte d'une part, des gestes, des regards, des mimiques faciales, des postures du corps, ainsi que des manipulations d'artefacts par les professionnels dans le cours d'actions ; et leurs conséquences sur l'activité de l'écran. Toutes ces ressources multimodales ont été décrites lorsqu'elles étaient « rendues pertinentes et intelligibles aux participants par un positionnement séquentiel et une temporalité finement synchrone au sein d'une organisation globale de l'action » (Mondada, 2008 : 129). Dans l'extrait 3-2, on peut lire un exemple d'annotations multimodales qui rend compte des actions pertinentes pour l'opératrice pendant la description des informations affichées sur son écran.








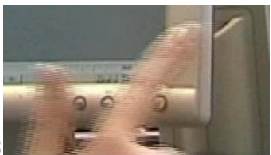
Dans l'extrait 3-3 (ci-dessous), je propose une autre manière d'annoter les actions multimodales en insérant des captures d'écran du geste réalisé par l'opératrice avec ses doigts. Ici, j'analyse ces ressources comme des mouvements co-verbaux tels que les caractérisent Cosnier (2000 : 333) dans son travail sur la voix, les gestes et le corps ; et Cosnier & Develotte (2011 : 32-33) dans leurs analyses des conversations en ligne.

### Ext3-3\_L24-15\_56'46 : Clé rib

```

1   OPE   c'est pas grave\ (0.3) trois [  doub]le vé point laville&
2   CLI   [(c'est)]
3   OPE   &point com\ en fait vous allez c'est la même adresse\ vous
4         effacez locbike dans l'adresse\ (0.8)
5   CLI   okay alors attendez donc\ (0.5) hop (1.6) donc trois double vé\
6         ensuite c'est euh\ laville/
7   OPE   oui tout attaché/ (..) en minuscule/ (0.4) et point com\
8         (2.1)
9   OPE   ensuite une fois qu` vous êtes sur c` si:te là vous allez
10  euh::: *vous allez voir sur euh y a:: y a un moteur d`

```

- img \*1 
- 11 **\*recherche/ sur la: \*droite/ de l'écran\ \*(0.3) tout en bas**
- img \*2  \*3  \*4 
- 12 **y a \*locbike\ (..) \*vous cliquez d`ssus/ (.) et**
- img \*5  \*6 
- 13 **\*ensuite vous \*allez dans les formules**
- img \*7  \*8 
- 14 **d'abonn`ment locbike/ (0.3) [et là-]**
- 15 CLI [alors ] attend-
- 16 (0.3)

L'opératrice indique à la cliente un « moteur de recherche » (l. 10-11) qui est situé « sur la droite/ de l'écran\ » (l. 11). La référence spatiale de la zone de moteur de recherche sur l'écran est exhibée de deux manières par l'opératrice : par la verbalisation « sur la droite/ de l'écran\ » ; et par un mouvement de sa main droite où elle dirige ses deux doigts en l'air vers la droite (cf. img 1 à 3, *mouvement co-verbal déictique*). Ensuite, elle oriente la cliente vers le bas de l'écran « tout en bas » (l. 11) pour lui indiquer la zone où est écrit « locbike » (l. 12). Elle baisse les doigts à partir de la courte pause de 0.3 secondes avant de signaler cet espace. Elle décrit alors l'action de cliquer sur le mot LocBike de deux manières différentes également : par la verbalisation « vous cliquez d`ssus/ » (l. 12) ; et par une rotation de son index du haut vers le bas (cf. img 5 et 6, *mouvement co-verbal illustratif*) qui commence sur le mot précédent « locbike ». Enfin, l'opératrice précise à la cliente d'aller « dans les formules d'abonn`ment locbike » (l. 13-14). Elle établit une co-référence de son instruction orale par un mouvement du doigt de la droite vers la gauche (cf. img 7 et 8, *mouvement co-verbal idéographique*).

L'usage de ressources multimodales vient ainsi renforcer, voire même dans le cas présent faciliter la description verbale faite par l'opératrice. En effet, à cet instant, la page du site internet, dont elle fait la description, n'est pas affichée sur son écran ; elle fait appel à sa mémoire et à ses connaissances professionnelles pour donner les instructions à la cliente. Elle incarne donc spatialement les actions et les espaces sur lesquels doit agir la cliente grâce à ses productions multimodales.

## Conclusion

L'utilisation des requêtes informatisées pour constituer une collection d'extraits est un outil performant pour rassembler des cas similaires à partir d'un objet d'analyse initial. Il a permis dans notre cas de rendre compte de la distribution des éléments sujets autour des verbes introductifs possibles qui précèdent un discours rapporté ou reformulé. Cela représente une étape intermédiaire intéressante dans la démarche d'analyse d'extraits telle que je la mène. En effet, elle a permis de repérer rapidement les différents emplois des verbes introductifs par les participants, et d'affiner ma réflexion par rapport à mon objet d'étude. Les résultats quantitatifs obtenus me permettent de développer une organisation cohérente à donner à l'analyse qualitative de mon objet d'étude. Ainsi, j'ai observé quatre activités dans lesquelles les participants convoquent le système informatique ou l'écran dans leurs interactions ; il s'agit de : i) formulations du problème, ii) guidages en direct, iii) partages d'informations écrites, et iv) recommandations prospectives. Le champ lexical des verbes



introdutifs est assez varié, mais les participants utilisent principalement trois verbes : « dire », « demander » et « mettre ».

L'analyse qualitative associée à la recherche quantitative sur les données m'a permis de montrer que le sujet du verbe introductif renvoie : i) soit à l'écran, au système impliqué dans un réseau socio-technique. Il est alors défini par des pronoms « il », « elle » (la borne, la station), « on », « ça », ou encore nommé « la machine » ; ii) soit à une personne désignée par « vous » ou « je ». Lorsqu'il est question de l'écran, les participants produisent des discours rapportés ou reformulés des messages affichés, ce qui confère au système informatique un rôle d'actant dans l'activité de résolution d'un problème. Lorsque les verbes renvoient à une personne, les participants sont alors engagés dans des séquences d'instructions ou de descriptions en temps réel. Dans ce cas, il s'agit de verbes d'actions (ex : « vous allez sur » ; « je clique sur », « j'accepte », « je recommence », « vous validez ») et la référence au système est ainsi marquée de manière implicite à travers l'action instruite en train de se faire ou à faire par un participant.

J'ai enfin montré lors des explications sur les avantages et les désavantages du caractère informatisé de l'outil que, pour obtenir une analyse linguistique et multimodale, un traitement supplémentaire des résultats quantitatifs était essentiel. Nous souscrivons ainsi pleinement à l'affirmation suivante : « l'analyse des corpus (...) continue à se fonder sur une analyse qualitative élaborée à toutes les phases du processus de recherche et qui permet de relancer et d'affiner la recherche quantitative » (Groupe ICOR, 2009 : 266).

## Bibliographie

- Colón de Carvajal, I. (2010). *La mobilisation des artefacts technologiques dans l'interaction : analyse linguistique et multimodale des pratiques professionnelles en centre d'appels* (thèse de doctorat, Université Lumière Lyon 2).
- Cosnier, J. (2000). V comme la Voix, les Gestes, le Corps. Dans B. Cerquiglini (dir.), *Tu parles ! Le français dans tous ses états* (p. 325-338). Paris, France : Flammarion.
- Cosnier, J. et Develotte, C. (2011). Le face à face en ligne, approche éthologique. Dans B. Colombat et C. Van Den Avenne (dir.), *Décrire la conversation en ligne : le face à face distanciel* (p. 27-50). Lyon, France : ENS Editions.
- Groupe ICOR. (2008). Tool-assisted analysis of interactional corpora: "voilà" in the CLAPI database. *Journal of French Language Studies*, 18, 121-145.
- Groupe ICOR. (2009). Exploitation de la plateforme CLAPI : le cas de voilà dans les chevauchements. *Cahiers de linguistique*, (33)2, 243-268.
- Groupe ICOR. (2010). Grands corpus et linguistique outillée pour l'étude du français en interaction (plateforme CLAPI et corpus CIEL). *Pratiques*, 147/148, 17-34.
- Latour, B. (1992). *Aramis ou l'amour des techniques*. Paris, France : La Découverte.
- Latour, B. (1994). Une sociologie sans objet ? Remarques sur l'interobjectivité. *Sociologie du travail*, 36(4), 587-607.
- Licoppe, C. (2008). Dans le carré de l'activité : perspectives internationales sur le travail et l'activité. *Sociologie du travail*, 50(3), 287-302.
- Mondada, L. (2002). Pratiques de transcription et effets de catégorisation. Dans B. Bonu (dir.), *Transcrire l'interaction* (numéro spécial *Cahiers de praxématique*), 39, 45-75.
- Mondada, L. (2008). Documenter l'articulation des ressources multimodales dans le temps : la transcription d'enregistrements vidéos d'interactions. Dans M. Bilger (dir.), *Données orales, les enjeux de la transcription* (p. 127-155). Perpignan, France : Presses Universitaires de Perpignan.
- Schegloff, E. A. (1987). Analysing single episodes of interaction: An exercise in conversation analysis. *Social Psychology Quarterly*, 50(2), 101-114.
- Schegloff, E. A. (1988). On an actual virtual servo-mechanism for guessing bad news: A single-case conjecture. *Social Problems*, 35(4), 442-457.
- Schegloff, E. A. (1993). Reflections on quantification in the study of conversation. *Research on Language and Social Interaction*, 26, 99-128.