Constitution et exploitation de corpus vidéo en linguistique interactionnelle : rendre disponibles les détails multimodaux de l'action située

The Production of Video Data in Interactional Linguistics : Making Available the Relevant Multimodal Details Organizing Situated Action

Lorenza Mondada

Édition électronique

URL : <http://journals.openedition.org/praxematique/1183>

DOI : 10.4000/praxematique.1183

ISSN : 2111-5044

Éditeur

Presses universitaires de la Méditerranée

Édition imprimée

Date de publication : 1 janvier 2010

Pagination : 327-350

ISBN : 978-2-36781-012-6

ISSN : 0765-4944

Référence électronique

Lorenza Mondada, « Constitution et exploitation de corpus vidéo en linguistique interactionnelle : rendre disponibles les détails multimodaux de l'action située », *Cahiers de praxématique* [En ligne], 54-55 | 2010, document 19, mis en ligne le 01 janvier 2013, consulté le 25 septembre 2020. URL : <http://journals.openedition.org/praxematique/1183> ; DOI : <https://doi.org/10.4000/praxematique.1183>

Lorenza Mondada

Laboratoire ICAR (U.M.R. 5191, université de Lyon) Institut universitaire de France

Constitution et exploitation de corpus vidéo en linguistique interactionnelle : rendre disponibles les détails multimodaux de l'action située

Introduction

Cet article a pour objectif de problématiser la question des observables pour l'analyse, en montrant qu'elle se pose au long de toutes les étapes du processus de constitution et d'exploitation des corpus. Nous en parlerons en termes de disponibilité (cf. Mondada, 2003) : en linguistique interactionnelle, le corpus est constitué de manière à rendre disponibles les détails rendus pertinents par les participants eux-mêmes dans l'organisation mutuellement intelligible de leur activité. Le but de l'analyse est de reconstituer les *méthodes*, *i.e.* les micro-pratiques systématiques par lesquelles les participants produisent leurs conduites de manière *accountable*, de façon à permettre une coordination de leur inter-action. Dans ce cadre, la constitution du corpus — depuis son enregistrement jusqu'à sa transcription — a pour but d'offrir une documentation de l'activité située qui prenne en compte les ressources mobilisées par les membres dans cette coordination et qui en permette la reconstitution fine, moment par moment, dans la transcription, l'annotation multimodale et l'analyse. Ces ressources et ces détails sont rendus localement pertinents et constituent ainsi la spécificité de l'activité : sa documentation relève d'une proto-analyse multiple (Macbeth, 1999 ; Mondada, 2006) qui ajuste les décisions techniques lors de l'enregistrement, mais aussi la transcription, de la manière la plus adéquate pour pouvoir les capturer. Cela fait du terrain, de l'enregistrement et de la transcription des moments proto-analytiques qui jouent un rôle important dans la compréhension de la dynamique située de

l'activité, indispensable pour en effectuer un enregistrement et une transcription appropriées.

Dans cet article, nous allons nous baser sur un exemple, tiré de l'enregistrement d'une visite guidée, pour développer ces exigences à la fois techniques et théoriques.

I. La constitution de corpus en analyse conversationnelle et linguistique interactionnelle

La linguistique interactionnelle inspirée de l'analyse conversationnelle se fonde sur l'observation des activités des participants dans des événements de la vie sociale ordinaire, dans des *naturally occurring interactions*, c'est-à-dire des interactions qui auraient eu lieu même en absence du chercheur et qui n'ont pas été provoquées par lui. Cette exigence dépend de deux principes fondamentaux de l'analyse, relatifs à l'importance du contexte et à l'importance de la temporalité.

Le premier renvoie au caractère à la fois indexical et systématique des activités langagières (*context-free* et *context-shaped*) : celles-ci s'organisent de manière localement située, en s'ajustant aux contingences qui affectent l'interaction, tout en contribuant réflexivement à (re)définir le contexte où elles ont lieu (Heritage, 1984). Transposer ces pratiques dans un autre contexte signifierait les altérer de manière radicale, puisqu'elles s'ajusteraient à d'autres contingences. Le second principe invite à travailler sur des *enregistrements* audio/vidéo d'interactions sociales afin de se fonder sur des données permettant de documenter l'émergence et le déploiement de ces pratiques *dans le temps* et donc *ni* sur des descriptions de celles-ci (dans des entretiens, dans des notes prises par le chercheur) *ni* sur des produits de celles-ci (par exemple dans des textes ou des objets issus de l'activité). En particulier, les enregistrements livrent les détails de l'organisation de ces pratiques — détails qui ne peuvent être ni imaginés ni reconstitués en leur absence, mais seulement constatés, identifiés dans un travail répété de visionnement et d'écoute, prolongé dans un travail minutieux de transcription (voir Auer, 1993 ; Bergmann, 1985 ; Mondada, 2008 a). Parce que la linguistique interactionnelle fonde la description du langage et des activités humaines sur l'interaction sociale, considérée comme organisée collectivement par les participants, d'une manière localement située, de façon incrémentale à travers son déploiement temporel et séquentiel, en mobilisant un large éventail de ressources

vocales, verbales, visuelles et corporelles, qui sont publiquement déployées et interprétées *in situ*, les enregistrements vidéo constituent aujourd'hui la base de sa démarche.

Au lieu de présenter de manière générale les principes de l'approche du corpus en linguistique interactionnelle, nous allons expliciter, à partir d'un exemple, les exigences et contraintes exercées sur le terrain, l'enregistrement, la transformation des données premières, la transcription et son approfondissement en vue de l'analyse.

2. La visite guidée : prise en compte de la spécificité de l'activité et choix des technologies pour son enregistrement

Le cas empirique que nous allons développer ici concerne une activité sociale particulière, la visite guide. Plus particulièrement, il s'agit d'un enregistrement vidéo d'une visite guidée d'un site architectural ; cette visite est effectuée par un expert du site (Jean), qui s'adresse à un petit groupe de visiteurs, une personne connaissant bien le bâtiment pour l'avoir fréquenté pendant de nombreuses années (Sophie) et deux architectes (Yan et Élise) qui le découvrent pour la première fois mais qui en connaissent et apprécient l'auteur.

La visite guidée est une activité qui présente un certain nombre de spécificités : les participants sont mobiles, se déplaçant progressivement à travers les différents espaces du bâtiment ; la visite guidée implique non seulement de la parole en interaction mais aussi une attention individuelle et collective envers des détails de l'environnement ambiant, localement définis par les participants, qui peuvent s'y approcher, y pointer ou les identifier verbalement (voir Birkner et Stuckenbrok, 2010 ; De Stefani, 2010 ; Mondada, 2012 a, 2012 b ; Pitsch, 2013 pour des analyses de ce type d'activité).

Cette activité entraîne ainsi plusieurs défis pour l'enregistrement, comportant une situation de mobilité et la mobilisation de ressources multimodales variées. Afin de tenir compte de ces caractéristiques, l'enregistrement a été effectué par un dispositif multi-sources. Trois caméras ont été utilisées : une caméra steadycam, qui permet une stabilisation forte de l'image en déplacement, une caméra paluche, qui permet une très forte mobilité et ajustabilité, et des lunettes caméra, portées par un des participants (Yan), permettant de capturer les changements d'attention et l'orientation de la tête (figure 3A *infra*). Pour le son, deux micros cravates HF ont été placés sur deux participants (Yan

et Sophie) et ont été connectés avec la steadycam. En outre, de nombreuses autres caméras ont été disposées le long de tout le parcours, dans le but de documenter le tournage lui-même. Le dispositif a été testé au préalable dans des exercices de tournage dans le bâtiment, visant à familiariser les deux camérawomen avec l'écologie de l'activité, les difficultés du tournage à reculons, les obstacles constitués par les escaliers, les portes, les couloirs, etc. La pratique du terrain spécifique aux corpus vidéo se caractérise ainsi par le fait qu'elle se concentre surtout sur le repérage des activités pertinentes, de leur écologie, des difficultés techniques, sociales et éthiques que posera l'enregistrement vidéo.

Afin de montrer les choix à la fois analytiques et technologiques effectués par l'enregistrement, nous allons nous concentrer sur l'examen d'un extrait, que nous présenterons d'abord sous la forme d'une transcription verbale. Les conventions de transcription verbales et multimodales sont explicitées à la fin de cet article.

3. Un extrait : montrer un détail architectural aux co-participants

L'extrait sur lequel nous allons nous pencher est reproduit ci-dessous. Il commence alors que le groupe est en train de descendre une rampe reliant le rez-de-chaussée et le premier étage. Yan attire l'attention des autres participants sur un détail de l'architecture de la toiture vitrée qui surplombe la rampe :

Transcription 1 (Cep1/56.27- Lunc52.12)

1		(1.2) # (2.0)
	fig	#fig.1
2	YAN	c'est vraiment c'que : j- moi j'adore/ (.) °c'est/° (0.5) ces
3		plans là où on est : dans une cour : be/ avec eh (0.6) les deux
4		autres courbes qui repartent/ celle-ci qui vient/ les
5		fenêtres/
6	SOP?	hm
7	YAN	c'est une x- c'est des détails que j'adore hein\
8		(0.6)
9	JEA	ouais\
10		(0.6)

Ce fragment est caractérisé par un tour complexe de Yan, qui pointe vers un détail architectural, attire l'attention des co-participants sur

lui et le décrit. Malgré l'apparence monologique de cette description, nous verrons qu'elle est le produit d'une intense coordination entre les co-participants et donc un achèvement interactionnel. La description de cet extrait nous amènera à thématiser aussi bien la manière dont il a été filmé que la manière dont il est transcrit aux fins de l'analyse de la pratique qu'il documente : l'introduction d'un référent et sa description pour le groupe.

4. Reconstituer la tâche du montreur : l'accomplissement d'un focus d'attention commun

Si on reprend ce fragment en s'appuyant sur l'enregistrement vidéo, on remarque qu'il est produit dans un contexte spécifique, ayant des implications importantes pour la coordination et l'ajustement entre les participants. La référence est en effet introduite par Yan pendant qu'il descend la rampe, en suivant avec retard le reste du groupe qui le précède et qui l'attend après avoir passé le virage en épingle à cheveux (figure 1).

Figure 1. — Position des participants au tout début de l'extrait transcrit.

La position de Yan au début de cet extrait est problématique pour l'analyse : il est hors du cadre de la steadycam et aussi de la caméra paluche. Les deux caméras précèdent en effet le groupe et le virage rend impossible une prise de vue qui intègre Yan. Par contre, une des caméras qui a suivi le tournage permet de vérifier que Yan reste un bon moment en regardant vers le haut, alors même que l'équipe de tournage est déjà arrivée à la fin de la rampe (figure 2) :

Figure 2

La position de Yan constitue aussi un problème pour lui : dans un contexte de mobilité (il s'est arrêté dans la trajectoire de sa marche, tout comme les autres devant lui), de focus disparates de l'attention (chaque participant est en train de regarder autre chose) et de non compacité du groupe des participants (figure 1), il fait face à une tâche interactionnelle complexe. Cette tâche se fonde de manière cruciale sur l'établissement d'un espace interactionnel (Mondada, 2005 b, 2009) fonctionnel pour l'activité, *i.e.* sur le réarrangement des postures corporelles des participants de manière à permettre une coordination mutuelle et une attention conjointe. La production pas à pas de la description référentielle de Yan répond à ces exigences interactionnelles, d'une manière qui intègre la mobilisation des constructions grammaticales dans un ensemble varié mais cohérent de ressources multimodales.

Reprenons le début de l'extrait en nous focalisant cette fois sur les regards et les gestes de Yan :

Transcription 1.1

- 1 (1.3) # (1.7)*(0.1) # (0.2)
 yan reg au-dessus*reg le groupe----->
 fig #fig.3AB #fig.4AB
 2 YAN c'est vraiment c'que : ** j- moi j'adore/* (.) °c'est/° (0.5)*
 -->*reg v haut-----*
 +.....*pointe----->

Figure 3 a. — Vue des lunettes caméra.

Figure 3 b. — Vue synchrone sur les participants.

Figure 4 a. — Vue des lunettes caméra.

Figure 4 b. — Vue synchrone sur les participants.

La construction complexe de Yan est très étroitement coordonnée d'une part avec sa propre marche, son regard et ses gestes, et d'autre part avec son *monitoring* de la marche des co-participants. Cette coordination est reconstituée grâce aux lunettes caméra, qui permettent de documenter les mouvements de la tête et donc d'inférer les changements de la direction des regards. En revanche, elles ne permettent pas de reconstituer le regard lui-même, puisqu'elles se situent manifestement un peu au-dessus du focus du regard (comme le montre la figure 2 b : les cadrages de la lunette caméra sont abondants dans la partie supérieure du cadre mais insuffisants pour la partie inférieure) (voir sur les usages des lunettes caméra Zouinar, Relieu, Salembier, 2004).

La combinaison des deux prises de vue permet de reconstituer la manière dont, en descendant la rampe, Yan contemple pendant un long moment la verrière au-dessus de sa tête (figure 3 ab). C'est l'architecture telle qu'elle lui apparaît au travers de ces vitres qu'il décrira ensuite. Il est toujours silencieux alors qu'il a encore le regard levé au-dessus de sa tête : 0.3 secondes avant de commencer son tour, il baisse la tête et regarde le groupe qui le précède (durant « c'est vraiment c'que : » 2). Le début de son tour est ainsi occupé à une vérification et sécurisation de l'attention des participants sur lui-même et sur ce qu'il est en train de faire (figure 4 ab). Dans ce contexte, les ressources utilisées — deux débuts de constructions bipartites successives — sont caractérisées par un fort potentiel projectif : elles ouvrent une, voire

plusieurs, constructions, projetant une suite, sans que le référent visé soit encore spécifié.

5. Reconstituer les activités des interlocuteurs

Ces constructions syntaxiques sont coordonnées avec ce que Yan découvre en regardant vers ses interlocuteurs : pendant qu'il produit la première partie de la première construction (« c'est vraiment ») Jean et Sophie, sont en train de le regarder, alors qu'Élise ne le regarde pas encore.

Ce type d'observation montre la nécessité, pour l'analyse interactionnelle comme pour l'analyse linguistique en général, de tenir compte de la temporalité fine d'une multiplicité de conduites multimodales : la parole, les gestes, les regards, les postures du corps — chacune dotée de sa propre temporalité émergente, qui est coordonnée tout en ne coïncidant pas totalement avec les autres. On comprend ainsi qu'à l'exigence de disposer d'enregistrements rendant disponibles ces détails s'associe l'exigence d'une transcription qui en documente la synchronisation fine (Mondada, 2008 a). La transcription alignée de ce fragment avec le logiciel ELAN (<http://tla.mpi.nl/tools/tla-tools/elan/>) permet de représenter avec précision ces différentes temporalités des orientations des co-participants — difficiles à noter à la main à cause de la multiplicité des niveaux auxquels elles se déploient :

Transcription 1.2. (« c'est vraiment c'que j- moi j'adore/ »)

Dans la transcription 1.2, la colonne verticale surlignée correspond au temps pendant lequel est énoncé « c'est vraiment ». Durant la pause qui précède et la durée de cette construction, les regards des trois co-participants se portent progressivement sur lui : d'abord Jean, puis

Sophie, et enfin Élise. « C’est vraiment » est un segment qui projette une suite, qui est même fortement projectif, mais qui est informativement vide. Dans ce sens, il constitue une ressource verbale propice au contexte d’attente dans lequel se trouve le locuteur.

Yan entame sa construction pseudo-clivée « c’que j- moi j’adore/ » au moment où Élise, la dernière à poser son regard sur lui, commence à le regarder. La pseudo-clivée représente une nouvelle construction fortement projective, qui prépare ainsi la suite, l’acte référentiel proprement dit. Durant la première partie de la pseudo-clivée, Yan dirige les regards des participants de sa personne vers le point au-dessus de lui qu’il va décrire : il lève le regard vers le haut et commence à pointer verticalement. Ces deux mouvements sont tous deux des instructions à regarder dans une certaine direction (Hindmarsh et Heath, 2000). Ainsi, le début de la pseudo-clivée effectue un pas en avant dans la préparation de la mention du référent, qui ne vient qu’ensuite (Moncada, 2005 b) : ce n’est qu’une fois qu’il a assuré les regards de ses interlocuteurs sur lui que Yan les redirige vers le futur référent.

Les co-participants réorientent en effet lentement leur regard vers le haut :

Transcription 1.3. (« ces plans là où on est : dans une courbe/ avec eh »)

Figure 5 a. — Vue des lunettes caméra.

Figure 5 b. — Vue synchrone sur les participants.

À la fin de la pause qui précède le syntagme démonstratif introduisant le référent (« ces plans là ») Yan baisse progressivement son geste et son regard (figure 5) et surveille à nouveau ses interlocuteurs : durant la production du NP, il voit Sophie et Élise regardant vers le haut, alors que Jean ne le regarde toujours pas (figure 6). Ce n'est qu'après que le SN ait été produit que tout le monde regarde vers le référent (figure 7).

Figure 6.

Figure 7.

La disponibilité de la vue (que l'on peut inférer) de Yan et d'une vue globale sur les participants permet de montrer la coordination extrêmement précise entre la constitution d'un focus d'attention commune sur le référent et l'introduction de ce référent. Dans ce cadre, le choix des ressources syntaxiques est ajusté à la temporalité des mouvements attentionnels collectifs, et fonctionne à la fois comme un attracteur d'attention et comme un retardateur de la progressivité de l'activité tant que l'attention des co-participants n'est pas focalisée sur l'objet vers lequel pointe le locuteur.

6. La stabilisation de l'attention et de la référence

On comprend donc que le locuteur ne se limite pas à nommer le référent vers lequel il a attiré l'attention de ses interlocuteurs : la manière

La transcription sur ELAN permet de noter les regards des interlocuteurs et leur temporalité — Élise étant celle qui le regarde d'abord, suivie de Sophie puis de Jean. Alors que Sophie regarde plus longuement vers le haut, Élise et Jean glissent immédiatement leur regard sur Yan.

Enfin, les co-participants modifient la direction de leur regard — et avec elle la pertinence de l'action à venir : ils se mettent à regarder vers l'avant (voir les formes hexagonales), projetant ainsi la complétude de la description et la poursuite de la marche :

Transcription 1.6. (« les deux autres courbes qui repartent/ celles-ci... »)

Après avoir suivi le pointage de Yan vers le référent et regardé ses gesticulations, les co-participants projettent ainsi la clôture de sa description en regardant vers l'avant et en se remettant à marcher. Yan s'ajuste à cette modification du regard, à la fois par la structuration de son tour et par la position de son corps : il effectue un ajout, en miroir, de la même construction qu'au début (7), alors qu'il se remet à marcher vers ses interlocuteurs :

Transcription 1.7.

6	SOP ?	hm
7	YAN	c'est une x- c'est des détails* que j'adore hein\ -->*reg en face de lui en marchant-->
8		(0.6)
9	JEA	ouais\ (0.6)
10		

Alors que les participants ont déjà anticipé la fin de la description, notamment en se remettant à marcher (Jean) ou en regardant devant eux dans le sens de leur marche (Élise — alors que Sophie le fait

au début du tour conclusif, 7), et qu'ils n'ont pas dit un mot en réponse, Yan organise une fin observable de sa description, marquée par la reprise de « c'est... que j'adore » (7), constituant ainsi *in fine* une construction pivot (« c'est que j'adore c'est **NP + liste** c'est que j'adore »). À cet endroit, il obtient un « ouais\ » de Jean (7).

Cet extrait se caractérise par une orientation très précise envers les actions des autres participants et par la constitution d'un focus d'attention commun, qui est la condition de possibilité pour que le référent et son élaboration descriptive puissent avoir lieu. Durant le développement descriptif, cette attention continue à être monitorée, dans une structuration du tour qui produit plusieurs pauses et expansions, offrant des opportunités aux participants de suivre mais aussi de répondre. En absence d'une telle réponse, une clôture est organisée de manière reconnaissable.

On observe ainsi que le choix de la pseudo-clivée et de la construction pivot est finement ajusté à cette écologie attentionnelle, d'une manière qui relève de la temporalité de l'émergence du tour. D'une part, la pseudo-clivée permet d'ouvrir un fort potentiel projectif, qui, tout en n'introduisant pas immédiatement le référent à regarder, mobilise l'attention des co-participants ; son déroulement pas à pas permet à la fois de projeter et de ralentir le référent, de manière à ce que le locuteur ajuste sa syntaxe *in progress* à l'attention de ses co-participants et à l'établissement d'un espace interactionnel adéquat à ce qu'il va faire. Cette propriété « ouvrante » de la pseudo-clivée (Hopper, 2004 ; Günthner et Hopper, 2010) est ici utilisée de manière située, incorporée (*embodied*) et interactionnellement ajustée au caractère temporellement émergent du tour et à la conduite des co-participants. D'autre part, la forme en miroir de la construction pivot est produite ici de manière émergente, bricolée pas à pas par des ajouts successifs et par une sorte de rétro-construction recyclant le début ouvrant, face à l'absence de réponse des interlocuteurs et la nécessité de clôturer la *Gestalt* au moment où les participants détournent leur attention du locuteur. On voit ainsi que le choix et l'usage de ces constructions grammaticales est finement ajusté à la temporalité de l'organisation interactionnelle des activités des participants, et intervient parmi d'autres ressources multimodales.

Les conditions de possibilité d'une telle description reposent de manière cruciale sur la disponibilité de données vidéo délivrant les

détails multimodaux de l'action, de manière continue et coordonnée. Cela suppose une prise de vue qui soit focalisée sur le groupe dans son ensemble, tout en documentant chaque mouvement en particulier. Cela est souvent rendu difficile par le fait que ces détails appartiennent à des échelles de visibilité différentes, pouvant concerner une partie de la tête (comme pour le regard, le sourire) ou un positionnement des mains (pour les gestes) mais aussi le corps des participants tout entier (comme la marche). Ces échelles visuelles très différentes ne peuvent être documentées que par l'adoption de multiples caméras — à synchroniser entre elles. En outre, ces multiples dimensions posent également des problèmes pour la transcription : elles rendent nécessaire la multiplication des niveaux d'annotation, en même temps qu'elles exigent de les considérer en tant qu'ensemble coordonné pour pouvoir rendre compte de leur fonctionnement interactionnel. Cela est rendu possible par l'adoption d'un logiciel d'alignement permettant de multiplier autant de lignes d'annotation que nécessaire — chose qu'il n'est pratiquement pas possible de faire manuellement.

7. Un deuxième extrait : constituer des collections

En analyse conversationnelle, il existe deux manières possibles d'organiser l'analyse : la première consiste à proposer une analyse de cas singulier (*single case analysis*), portant sur une donnée qui est parcourue dans sa complexité ; la seconde consiste à effectuer une analyse de collection, qui se focalise sur un phénomène particulier et qui vise à en montrer la récurrence et la systématicité. Voici comment Schegloff caractérise la première : « *the resources of past work on a range of phenomena and organizational domains in talk-in-interaction are brought to bear on the analytic explication of a single fragment of talk* » (1987 : 101) alors que la seconde consiste à utiliser « *a set of fragments, then, to explicate a single phenomenon or a single domain of phenomena* » (1987 : 101). Alors que la première façon de faire ne prétend pas approfondir l'analyse d'un phénomène, mais plutôt exploiter les connaissances analytiques existantes pour comprendre un cas singulier dans sa complexité, voire découvrir des pertinences non encore décrites, la seconde permet d'envisager l'étude d'un objet inédit dans sa systématicité et généralisabilité. Alors que la première se concentre sur un fragment, la seconde collationne plusieurs fragments, qui documentent la même

action, dans le même environnement séquentiel, faisant intervenir des ressources similaires : ce qui caractérise la collection est un faisceau de propriétés praxéologiques, séquentielles et formelles (Schegloff, 1996 ; voir Mondada, 2005 a, 2008 b). La démarche consistant à construire des collections est devenue une exigence classique en analyse conversationnelle ; elle a toutefois été massivement menée sur des données audio, alors qu'elle s'est révélée beaucoup plus complexe et exigeante dès qu'il s'agit de tenir compte des détails multimodaux (voir Mondada, 2008 b, à paraître sur les enjeux des collections multimodales). Ces enjeux sont bien illustrés par le cas étudié dans cet article : face à la complexité et à la multiplicité des détails mobilisés par les participants dans l'extrait 1, on peut se demander s'il est possible de trouver d'autres occurrences de ces phénomènes complexes. Nous allons ainsi nous pencher sur un deuxième extrait, enregistré durant la même visite, qui se situe lui aussi après la clôture de l'activité précédente, comporte lui aussi l'introduction d'un nouveau référent, là encore dans une écologie attentionnelle fragmentée. L'identification d'une seconde occurrence représente le premier pas vers l'établissement d'une collection permettant de montrer le caractère méthodique et systématique de la pratique étudiée : afin de vérifier celui-ci dans le cas des procédés d'introduction du référent identifiés dans le premier exemple nous nous penchons sur un second extrait.

Les participants se trouvent dans une salle de cours du bâtiment visité ; alors que Jean, Sophie et Élise s'arrêtent au pied de la porte, Yan traverse la pièce, s'approche d'une fenêtre et contemple une partie extérieure du même bâtiment. Après un moment de silence, il utilise une construction pseudo-clivée similaire à la précédente. À nouveau, nous prenons connaissance de la transcription verbale avant de nous pencher sur les détails incarnés de sa production :

Transcription 2. (Cep2/15.27--Clun1.11.30)

1		(3.#7)
	fig	#fig.8
2	YAN	pis moi c'que j'adore/ c'est que: à chaque: endroit où on est
3		à l'intérieur/ on a une conscien:ce du propre bâtiment/
4		(0.9)
5		et c'est euh c'est un jeu qui : °>qui qui<° qu'il affectionne
6		beaucoup j'ai l'impression/ parce que:
7		(0.8)
8	YAN	alors [bon on est] toujours un peu tous sensibles &
9	ELI	[ah c'est vrai xxxxx]

- 10 YAN & nous les archis à c::::e genre d'détail/ mais c'est °vrai que:°
 11 (0.2)
 12 JEA de voir le bâtiment de l'int[érieur

Yan introduit par la pseudo-clivée « moi c'que j'adore/ c'est que: ... » une construction avec que-phrase (Müller, 2008) (*vs* avec SN dans le premier extrait) qui décrit ce qu'il voit. Alors que ligne 2 son tour pourrait être complet, après une longue pause, il continue en commençant par « et », qui permet de faire le lien avec ce qui précède, et avec une clivée (5-6) qui s'ouvre sur une subordonnée en « parce que » laissée inachevée. Après une nouvelle pause pendant laquelle personne ne répond (7), Yan relativise ce qu'il vient de dire, en le rapportant à sa catégorie professionnelle (8, 10). C'est alors qu'en chevauchement avec le début de ce tour, Élise répond avec un change-of-state token (Heritage, 1984) et, à la fin du tour, à nouveau inachevé (10), Jean répond en proposant une complétion collaborative (12).

Contrairement à ce qui se passait dans la première occurrence, celle-ci est donc reçue avec plusieurs prolongements collaboratifs des interlocuteurs — sans doute sollicités par le caractère inachevé des énoncés de Yan. Mais de façon similaire à ce qui se passait dans le premier fragment, le problème pratique initial que résout Yan avec la pseudo-clivée est le même. Les autres participants sont restés au fond de la salle, de sorte qu'ils ne sont pas en mesure d'apercevoir ce qu'il dit voir par la fenêtre. Yan répond à ce problème pratique en produisant un tour qui prend appui sur une construction ouvrante et projetante comme la pseudo-clivée et dont la temporalité va s'ajuster finement au déplacement de ses interlocuteurs.

8. Constitution et stabilisation de l'attention et de l'espace interactionnel commun

L'examen de la vidéo montre que, comme dans l'extrait précédent, Yan se livre d'abord à une contemplation visuelle silencieuse et individuelle près de la fenêtre (figure 8), dont on peut reconstituer l'objet grâce aux lunettes caméra.

Figure 8. — Position des participants au tout début de l'extrait 2.

Avant de commencer son tour, Yan se tourne vers ses interlocuteurs et découvre qu'ils sont loin de lui. À nouveau nous reprenons la transcription sous forme alignée avec ELAN :

Transcription 2.1.

Lorsque Yan commence la pseudo-clivée, ses interlocuteurs ne bougent pas, et Sophie se détourne même de lui, en regardant en arrière, vers Jean. Ce n'est que durant le segment suivant que les participants commencent à se mettre en mouvement vers Yan :

Transcription 2.2.

On comprend alors que l'ajout d'une insertion (« à chaque endroit où on est à l'intérieur » 2-3) entre la première et la deuxième partie de la pseudo-clivée permet à Yan d'ajuster son tour à la temporalité lente du déplacement de ses interlocuteurs. Durant cette insertion, ils commencent en effet à se rapprocher, de manière plus précoce (Sophie) ou plus tardive (Élise). Ce n'est que lorsque la dernière personne s'est mise en mouvement (Élise), que Yan entame la deuxième partie de sa pseudo-clivée, où il dit ce qu'il voit par la fenêtre. Son placement séquentiel de ce segment montre la pertinence de la position et du regard de ses interlocuteurs pour l'organisation émergente de sa syntaxe.

En outre, à ce moment-là il change de posture, comme il l'avait fait dans l'extrait précédent : il se tourne à nouveau vers la fenêtre et fait un pointage avec le menton vers ce qui se trouve au-delà d'elle. Même si les autres participants ne peuvent pas voir ce qu'il voit, ils sont en train de s'approcher de lui pour partager son focus visuel.

La pseudo-clivée complétée, Yan se tait, et personne ne dit rien. En fait, les participants sont encore engagés dans leur marche vers la fenêtre (figure 9) et ne sont donc pas encore en mesure de répondre à son énoncé. À la fin des 0.9 secondes de pause (4), Yan se tourne à nouveau vers eux, puis ajoute une clivée (5-6), durant laquelle le groupe continue à avancer. La clivée permet ainsi de « meubler » l'attente de la reconjonction des participants, leur permettant d'accéder finalement au référent et le cas échéant de produire une réponse.

La pause suivante de 0.8 secondes (7) est importante pour l'organisation du regard collectif. En effet, les co-participants sont arrivés à proximité de Yan et commencent à regarder par la fenêtre :

Transcription 2.3.

La partie grisée de la transcription 2.3 correspond à la durée de la pause de 0.8 secondes : Jean et Sophie viennent d'arriver à la fenêtre, Élise s'en approche un peu plus tard. À la fin de la pause, une première réaction de Sophie est audible (« ah c'est vrai ») et Jean sourit à Yan. Ces réponses restent toutefois minimales et le commentaire sur les architectes ajouté par Yan permet d'offrir une position séquentielle renouvelée où une réponse plus substantielle est possible. Celle-ci est non seulement attendue, mais aussi sollicitée, par son tour qui est laissé incomplet (10).

Cet agencement de ressources offre à Jean l'occasion de redire celle qui était la deuxième partie de la pseudo-clivée, par un tour (« de voir le bâtiment de l'intérieur » 12) qui commence par « de » : bien que du point de vue strictement syntaxique cela ne s'appuie sur aucune construction précédente, le fait de commencer par « de » permet d'exhiber un lien syntaxique avec le tour précédent et de signaler que la contribution de Jean y participe — reconnaissant ainsi Yan comme premier locuteur, et exhibant la participation du second locuteur, Jean, à ce qui devient un énoncé collectif.

La description référentielle de Yan est ainsi caractérisée par une série de dispositifs qui retardent et projettent à la fois ce qui est à venir : le retardement permet qu'un espace interactionnel adéquat à son action se constitue progressivement, grâce à la reconjonction des corps de tous les participants ; la projection permet de maintenir les attentes d'une action à venir dont l'attention conjointe et l'arrangement d'un espace interactionnel commun constituent la préparation. À nouveau, la construction pseudo-clivée permet de réaliser cette préparation de la projection.

9. L'imbrication de ressources multimodales dans la temporalité émergente de l'action : vers une systématisation

Les extraits 1 et 2 partagent une série de caractéristiques séquentielles qui fondent la collection :

- Yan commence par contempler l'objet de manière individuelle et silencieuse ;
- Il se tourne vers ses interlocuteurs avant d'initier son tour ;

- Au début de son tour, il constate que les participants ne le regardent pas ou sont dans une position qui rend impossible le partage de son focus visuel ;
- Il entame une pseudo-clivée : durant la première partie, les participants se mettent en mouvement et adoptent une nouvelle position dans l'espace, qui va leur donner accès au référent visé ;
- La seconde partie de la construction, décrivant le référent, est introduite et développée (y compris par des moyens gestuels) alors que les participants commencent à partager un focus d'attention commun ;
- L'expansion de la seconde partie de la pseudo-clivée offre l'occasion de stabiliser ce focus ;
- La clôture de la description s'oriente vers une réponse — attendue et projetée — des participants.

Cette synthèse démontre d'une part l'intérêt de complexifier l'analyse grammaticale premièrement par l'intégration de la corporéité et la prise en compte de toutes les ressources multimodales ; deuxièmement par la prise en compte de la temporalité finement coordonnée du déploiement de ces ressources. Les arrangements syntaxiques et plus généralement séquentiels de ressources sont ainsi des dispositifs qui constituent progressivement, pas à pas, de manière sensible aux contingences de l'action et de l'interaction, une *Gestalt multimodale* émergente.

D'autre part, cette synthèse montre l'importance de la disponibilité des détails multimodaux dans les données : sans enregistrement vidéo, il aurait été difficile de démontrer le lien entre syntaxe et coordination des mouvements des interlocuteurs ; sans un enregistrement vidéo qui couvre plusieurs niveaux de détail du corps (du regard et mouvements de la tête à la marche du corps tout entier) il n'aurait pas été possible de montrer que la coordination repose sur un monitoring des positions réciproques ; enfin, sans une annotation synchronisée temporellement de ces détails multimodaux il n'aurait pas été possible de montrer comment ces pertinences s'ajustent mutuellement au fil de la temporalité et de la séquentialité émergente de l'action. Les exigences de la prise de vue, du traitement des données dans la transcription et de l'analyse sont ainsi fortement cohérentes et convergentes.

Conclusion : quels corpus pour une approche temporalisée et émergente de la parole en interaction

Les analyses développées dans ces pages montrent comment se constitue progressivement une collection, base pour une systématisation des observations effectuées sur les données. Les occurrences de la collection sont caractérisées par une série de propriétés qui en font des patterns associant ressources multi-modales, environnements séquentiels, actions accomplies. Plus ces occurrences sont caractérisées de manière complexe et plus elles expriment de contraintes sur la constitution de la collection — étant par conséquent plus exigeantes dans la description de la systématité du phénomène étudié. Le fait d'intégrer dans la description du phénomène les ressources multimodales — associant au langage les gestes, les regards, les postures du corps, les mouvements des participants... et de manière plus générale tous les détails que les participants ont traité comme étant pertinents pour l'organisation de leur action — introduit de nouveaux défis dans cette démarche. En effet, afin de pouvoir traiter de manière comparable des occurrences afin de décider si elles appartiennent ou non à la collection, il est nécessaire de pouvoir compter sur une disponibilité comparable des détails dans les données primaires et les données secondaires. Cela a des conséquences pour la manière de filmer l'événement — le même type de prise de vue, obéissant aux mêmes exigences analytiques, est requis — aussi bien que pour la manière de le transcrire. Les pertinences des phénomènes multimodaux ne sont pas établies d'avance : elles émergent de l'interaction en train de se dérouler dans le temps et *in situ*. Cela demande un dispositif de captation de l'action qui puisse s'ajuster à la contingence et à l'émergence des phénomènes interactionnels. On trouve ainsi les principes fondamentaux de pertinence, de temporalité, d'émergence, d'indexicalité dans l'exercice même de la constitution du corpus — dans ce que nous avons appelé une forme de proto-analyse sur le terrain — ainsi que dans toutes les opérations qui permettront finalement d'aboutir à une analyse et à sa systématisation.

Conventions de transcription

Voir Mondada (2008). Les conventions sont accessibles sur le site suivant : http://icar.univ-lyon2.fr/projets/corinte/bandeau_droit/convention_icor.htm

Références bibliographiques

- AUER P., 1993, « Ueber <=| », *Zeitschrift für Literaturwissenschaft und Linguistik* (Lili), 90/91, 104-138.
- BERGMANN J. R., 1985, « Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit : Aufzeichnungen als Daten der interpretativen Soziologie », in BONSS W. & HARTMANN H. (éd.), *Entzauberte Wissenschaft*, Göttingen, Schwarz.
- BIRKNER K. & STUKENBROCK A., 2010, « Multimodale Ressourcen für Stadtführungen », in COSTA M. & MÜLLER-JACQUIER B. (Hg.), *Deutschland als fremde Kultur : Vermittlungsverfahren in Touristenführungen*, München, Judicium Verlag, 214-243.
- DE STEFANI E., 2010, « Reference as an Interactively and Multimodally Accomplished Practice : Organizing Spatial Reorientation in Guided Tours », in PETTORINO M., GIANNINI A., CHIARI I. & DOVETTO F. (éd.), *Spoken Communication*, Newcastle upon Tyne, Cambridge Scholars Publishing, 137-170.
- GÜNTNER S. & HOPPER P., 2010, « Zeitlichkeit & sprachliche Strukturen : Pseudoclefts im Englischen und Deutschen », *Gesprächsforschung* 11, 1-28 (www.gespraechsforschung-ozs.de).
- HERITAGE J. C., 1984, *Garfinkel and Ethnomethodology*, Cambridge and New York, Polity Press.
- HINDMARSH J. & HEATH C., 2000, « Embodied Reference : A Study of Deixis in Workplace Interaction », *Journal of Pragmatics* 32, 1855-1878.
- HOPPER P., 2004, « The Openness of Grammatical Constructions », *Chicago Linguistic Society* 40, 153-175.
- MACBETH D., 1999, « Glances, Trances and Their Relevance for a Visual Sociology », in JALBERT P. (éd.), *Media Studies : Ethnomethodological Approaches*, Lanham, University Press of America, 135-170.

- MONDADA L., 2003, « Observer les activités de la classe dans leur diversité : choix méthodologiques et enjeux théoriques », in Actes du Colloque *Ensenyar llengües en la diversitat i per a la diversitat* (Barcelona, 30.5-1.6.2002), Barcelona, ICE Universitat de Barcelona, 49-70.
- MONDADA L., 2005 a, « L'analyse de corpus en linguistique interactionnelle : de l'étude de cas singuliers à l'étude de collections », in CONDAMINE A. (éd.), *Sémantique et corpus*, Paris, Hermès, 76-108.
- MONDADA L., 2005 b, « La constitution de l'origo déictique comme travail interactionnel des participants », *Intellectica* 2/3 (41-42), 75-100.
- MONDADA L., 2006, « Video Recording as the Reflexive Preservation-Configuration of Phenomenal Features for Analysis », in KNOBLAUCH H., RAAB J., SOEFFNER H.-G. & SCHNETTLER B. (éd.), *Video Analysis*, Bern, Lang.
- MONDADA L., 2008 a, « La transcription dans la perspective de la linguistique interactionnelle » et « Documenter l'articulation des ressources multimodales dans le temps : la transcription d'enregistrements vidéos d'interactions », in BILGER M. (éd.), *Données orales, les enjeux de la transcription*, Perpignan : Presses universitaires de Perpignan, 78-109 et 127-155.
- MONDADA L., 2008 b, « L'analyse de collections de phénomènes multimodaux en linguistique interactionnelle », *Cahiers de praxématique*, 50, 23-70.
- MONDADA L., 2009, « Emergent Focused Interactions in Public Places : A Systematic Analysis of the Multimodal Achievement of a Common Interactional Space », *Journal of Pragmatics* 41, 1977-1997.
- MONDADA L., 2012 a, « Garden Lessons : Embodied Action and Joint Attention in Extended Sequences », in NASU H. & WAKSLER F. C. (éd.), *Interaction and Everyday Life : Phenomenological and Ethnomethodological Essays in Honor of George Psathas*, Lanham, MD, Lexington Books, 293-311.
- MONDADA L., 2012 b, « Descriptions en mouvement : l'organisation systématique du déplacement dans une visite guidée », in DUFLET J.-P. (éd.), *Les visites guidées. Discours, interaction, multimodalité*, Trento, Università degli Studi di Trento, Dipartimento di Studi Letterari, Linguistici e Filologici.

- MONDADA L., à paraître, « Organisation multimodale de la parole-en-interaction : Pratiques incarnées d'introduction des référents », *Langue française* 175.
- PITSCH K., 2013,
« Museumsexponat, Alltagsobjekt oder Turngerät ? — Zur Konstitution von Objekten im Interaktion », in HAUSENDORF H., MONDADA L., & SCHMITT R. (éd.), *Raum als interaktive Resource*, Tübingen, Narr.
- SCHEGLOFF E. A.,
1987, « Analyzing Single Episodes of Interaction : An Exercise », in *Conversation Analysis, Social Psychology Quarterly* 50(2), 101-114.
- SCHEGLOFF E. A.,
1996, « Confirming Allusions : Toward an Empirical Account of Action », *American Journal of Sociology* 102(1), 161-216.
- ZOUINAR M., RELIEU M. & SALEMBIER P.,
2004, « Observation et capture de données sur l'interaction multimodale en mobilité », *Mobilité & Ubiquité'04*, Nice, A.C.M.