



**HAL**  
open science

# Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés

Marion Tellier

## ► To cite this version:

Marion Tellier. Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés. *Discours - Revue de linguistique, psycholinguistique et informatique*, 2014, 15, 10.4000/discours.8917. hal-01306805

**HAL Id: hal-01306805**

**<https://amu.hal.science/hal-01306805>**

Submitted on 25 Apr 2016

**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.

---

Marion Tellier

## Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés

---

### Avertissement

Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur.

Les œuvres figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage soit personnel, soit scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document.

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

**revues.org**

Revues.org est un portail de revues en sciences humaines et sociales développé par le Cléo, Centre pour l'édition électronique ouverte (CNRS, EHESS, UP, UAPV).

---

### Référence électronique

Marion Tellier, « Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés », *Discours* [En ligne], 15 | 2014, mis en ligne le 19 décembre 2014, consulté le 10 février 2015.

URL : <http://discours.revues.org/8917> ; DOI : 10.4000/discours.8917

Éditeur : Laboratoire LATTICE, UMR 8094 ENS/CNRS

<http://discours.revues.org>

<http://www.revues.org>

Document accessible en ligne sur : <http://discours.revues.org/8917>

Ce document PDF a été généré par la revue.

© Discours - Presses universitaires de Caen





Revue de linguistique, psycholinguistique et informatique

<http://discours.revues.org/>

## Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés

---

Marion Tellier

Aix Marseille Université, CNRS, LPL UMR 7309  
13100, Aix-en-Provence, France

.....  
Marion Tellier, « Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés », *Discours* [En ligne], 15 | 2014, mis en ligne le 19 décembre 2014.

.....  
URL : <http://discours.revues.org/8917>

.....  
Titre du numéro : *Varia*  
Coordination : Catherine Bolly et Benjamin Fagard

**revues.org**  
CENTRE POUR L'ÉDITION ÉLECTRONIQUE OUVERTE  
CENTRE FOR OPEN ELECTRONIC PUBLISHING

 discours

 Presses  
universitaires  
de Caen



# Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés

---

Marion Tellier

Aix Marseille Université, CNRS, LPL UMR 7309  
13100, Aix-en-Provence, France

.....

Cet article dresse un panorama méthodologique en vue de l'étude de la parole dans une perspective multimodale. Il ne s'agit donc pas de donner ici des résultats mais de montrer un peu les coulisses d'une étude multimodale en révélant les questionnements méthodologiques que tout chercheur doit se poser avant d'entreprendre une telle étude. Ainsi, cet article abordera la question du recueil des éléments constituant un corpus multimodal, de l'annotation (et notamment de la conception d'un schéma d'annotation) et enfin de la méthodologie du contre-codage en vue de valider les annotations. Il présentera donc diverses études et se fera le témoin de plusieurs expériences méthodologiques. L'accent sera mis en particulier sur les études de la gestuelle coverbale mais tout en envisageant le transfert de ces approches méthodologiques à d'autres modalités telles que les expressions faciales, le regard ou encore les postures.

**Mots clés :** geste coverbal, méthodologie, corpus, annotation, contre-codage

.....

*This paper establishes a methodological panorama for the study of speech in a multimodal perspective. Therefore, the aim is not to give results but to uncover what lies behind the scenes of a multimodal study by revealing methodological questions that any researcher should raise before undertaking such a study. Thus, this paper addresses the issue of the collection of a multimodal corpus, coding techniques (including the development of an annotation scheme) and finally the methodology to set up an inter-coder agreement to validate the annotations. Therefore, this paper presents several studies and relates several methodological experiences. The focus will be placed on studies about co-speech gestures but considering the transfer of these methodological approaches to other modalities such as facial expressions, gaze or postures.*

**Keywords:** co-speech gestures, methodology, corpus, coding, inter-coder agreement

## 1. Introduction

- 1 Lorsque nous parlons, nous produisons spontanément des gestes coverbaux, c'est-à-dire en lien avec la parole. Depuis une quarantaine d'années, le domaine des études de la gestuelle (*Gesture Studies*) s'est penché sur ces gestes et leur relation avec la parole et la pensée. La plupart des études en gestuelle s'inscrivent dans une perspective descriptive (gestes du discours et notamment de la narration) ou développementale (acquisition L1, L2). Les corpus étudiés peuvent être constitués d'interactions ou de discours libres (discours politiques, corpus de télé-réalité) ou plus contrôlés (narration à partir d'un stimulus donné, description d'images, etc.). Dans une discipline aussi récente que les études de la gestuelle, les cadres conceptuels, méthodologiques et typologiques évoluent rapidement et se forment au fil des études et des rencontres entre gestualistes mais

donnent rarement lieu à des publications méthodologiques. Or, tout chercheur qui souhaite se pencher sur l'analyse du geste coverbal est souvent dérouté par le manque de guidage méthodologique. À une époque où l'on se focalise trop souvent sur les résultats, s'interroger sur la démarche méthodologique peut apparaître comme un luxe mais s'avère pourtant nécessaire et n'en constitue pas moins un travail de recherche. Le présent article a donc pour vocation de donner des pistes de réflexion sur la constitution du corpus ainsi que sur la méthodologie de l'annotation. Il se focalise sur la gestuelle coverbale bien que les propositions ci-dessous soient également transférables à d'autres modalités kinésiques comme le regard, les expressions faciales ou encore les postures ainsi qu'aux langues à modalité gestuelle comme la langue des signes.

- 2 Après un rapide exposé théorique sur la gestuelle coverbale, nous évoquerons, dans une seconde partie, les différents types de corpus pertinents pour l'étude de la gestuelle. Puis, dans une troisième partie, nous aborderons la question des annotations pour terminer enfin, sur la problématique de la validation de ces annotations. Cet article présente donc un panorama méthodologique, nullement exhaustif, mais ayant plutôt pour objectif de faire le point sur les questions à se poser avant d'entreprendre une étude de corpus dans une perspective multimodale.

## 2. Le geste coverbal : quelques éléments théoriques

### 2.1. Qu'est-ce qu'un geste coverbal ?

- 3 Les coverbaux sont des gestes produits spontanément avec la parole (McNeill, 1992, 2005; Goldin-Meadow, 2003). Nous évoquerons ici essentiellement les gestes des mains à visée communicative (c'est-à-dire en excluant les gestes nerveux, les grattements, et autres gestes ludiques et autocentrés). Cependant, d'autres éléments kinésiques pourront être évoqués en raison de leur lien étroit avec la parole tels que les mouvements de tête, de sourcils ou encore les regards. Le cadre théorique de McNeill postule que le geste et la parole forment un seul et même système cognitif et plusieurs arguments permettent de soutenir cette théorie. Premièrement, le geste et la parole sont produits en synchronie dans le discours oral. Deuxièmement, il y a une certaine harmonie sémantique entre les deux. Les gestes permettent souvent de véhiculer les aspects visuels des référents comme la taille, la forme, la trajectoire, le positionnement des éléments dans l'espace, plus facilement que la parole (Goldin-Meadow, 2003). Troisièmement, le développement langagier de l'enfant se construit par une succession harmonieuse d'étapes gestuelles et verbales (Goldin-Meadow, 2000; Colletta, 2004). Enfin, certaines pathologies qui affectent la parole et l'organisation du discours (aphasie, schizophrénie...) semblent affecter également la gestuelle (McNeill, 1992; De Ruiter et De Beer, 2013).

### 2.2. Pourquoi produit-on des gestes ?

- 4 La première réponse qui peut venir à l'esprit pour répondre à cette question est : « pour aider notre interlocuteur à mieux comprendre notre discours ». Plusieurs études ont en effet montré que l'interlocuteur prend en compte les informations véhiculées par

le geste (Beattie et Shovelton, 1999) dans l'accès au sens. Par ailleurs, Bavelas *et al.* (1995) ont mis en évidence certains gestes propres à la situation dialogique, appelés *gestes interactifs*. Ils sont en général orientés vers l'interlocuteur et utilisés pour faire référence à du savoir partagé entre les interactants, pour gérer les tours de parole ou encore pour solliciter l'aide de l'interlocuteur. Certains gestes sont donc produits spécifiquement à l'intention de l'interlocuteur et n'apparaissent qu'en face-à-face. En outre, on remarque une adaptation gestuelle en fonction de l'interlocuteur, que ce soit au niveau de l'orientation spatiale des gestes (Özyürek, 2002) ou par rapport à son niveau de compréhension de la langue de l'échange (Tellier et Stam, 2012).

5      Toutefois, lorsque notre interlocuteur n'est pas visible, nous produisons des gestes quand même. C'est le cas lorsque nous parlons au téléphone ou lorsque des aveugles de naissance communiquent entre eux (Iverson et Goldin-Meadow, 1998). Plusieurs études comparant des sujets effectuant la même tâche avec un interlocuteur visible et avec un interlocuteur invisible ont montré que des gestes sont produits dans les deux conditions (voir Alibali *et al.*, 2001, pour une synthèse). D'autres études ont mis en évidence l'importance de la gestuelle coverbale dans la production langagière notamment pour faciliter l'accès lexical (Krauss *et al.*, 2000) ou l'organisation spatiale des éléments du discours (Morsella et Krauss, 2004). En somme, le geste aide aussi le locuteur à produire son discours en allégeant la charge cognitive.

6      De ce fait, beaucoup d'études en gestuelle cherchent soit à (1) analyser l'articulation geste/parole dans le discours spontané, (2) étudier le rôle du geste dans l'interaction (notamment son rôle sur la coconstruction du sens), (3) examiner le rôle du geste pour celui qui le produit (notamment dans des tâches cognitivement et langagièrement complexes), (4) comprendre comment un geste est perçu par l'interlocuteur. Le choix du corpus d'étude est donc déterminant en fonction de ce que l'on veut étudier.

### 3. Quel type de corpus ?

7      La majorité des études sur la parole se basent sur un corpus, qu'il s'agisse de quelques phrases créées artificiellement ou d'un corpus vidéo de plusieurs heures d'interaction spontanée. Tout dépend de l'étude que l'on veut mener, de ce que l'on cherche à démontrer et, le cas échéant, des hypothèses que l'on souhaite valider. Tognini-Bonelli (2001) distingue les recherches sur corpus menées avec une approche déductive (*data-based*) de celles menées avec une approche inductive (*data-driven*). Dans le premier cas, le corpus est utilisé pour explorer, confirmer ou infirmer des théories et/ou des hypothèses alors que dans le second cas, les hypothèses et théories émergent de l'analyse du corpus sur lequel le chercheur n'a pas d'*a priori*. Les différents types de données recueillies peuvent être présentés sur un continuum (voir figure 1) allant des données contrôlées (recueillies dans un cadre expérimental) aux données écologiques (recueillies sur un terrain naturel). Nous n'aborderons pas ici la démarche expérimentale qui, bien que tout à fait intéressante pour traiter de nombreux aspects (notamment des caractéristiques prosodiques, phonétiques et même gestuelles), s'avère peu adaptée pour analyser l'interaction.

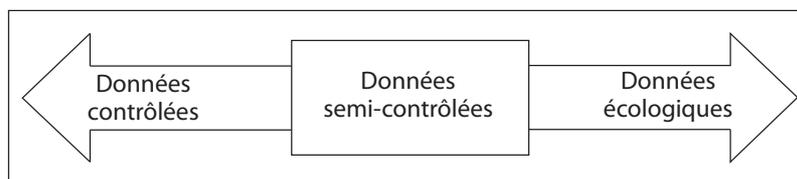


Figure 1. Continuum de recueil de données (adapté de Tellier, 2013)

### 3.1. Le corpus écologique

- 8 Le recueil de données en situation dite écologique vise à observer une situation naturelle et authentique. *Corpus écologique* est entendu ici au sens de Gadet *et al.* (2012 : 40) en tant qu'« ensemble de données langagières recueillies dans leur milieu naturel d'occurrence, un milieu non créé par le linguiste ». Il s'agit donc de placer un enregistreur ou une caméra dans un lieu particulier (salle de classe, interaction familiale, visite guidée, etc.) afin de recueillir le déroulement des faits et les diverses manifestations (voir par exemple Gadet *et al.*, 2012, pour la description du « Corpus international écologique de la langue française » [CIEL-F], ou encore Balthasar et Bert, 2007, pour le « Corpus de langues parlées en interaction » [CLAPI]). On considère également comme écologiques les enregistrements vidéo qui n'ont pas été constitués pour une recherche (corpus d'interviews et de discours politiques, par exemple, voir Calbris, 2003). Ce type de corpus produit des données denses et riches avec divers aspects multimodaux (aspects verbaux, prosodiques, kinésiques, etc.). La transcription puis l'annotation de ces données sont si complexes et coûteuses en temps qu'il n'est souvent pas possible de multiplier les enregistrements. En outre, la qualité sonore et vidéo des enregistrements est rarement optimale, ce qui peut gêner la transcription et l'annotation. Par exemple, pour la gestuelle, il arrive fréquemment que le geste soit produit hors champ ce qui complique la segmentation (trouver le début et la fin du geste) (voir Azaoui, 2014, au sujet du film de classe) et tout simplement sa visibilité. Quoi qu'il en soit, ce type de corpus nous offre une vision de l'activité réelle (ou quasi réelle) d'une situation d'interaction définie et peut être analysé finement dans ces différents aspects. À titre d'exemple, l'étude de Bigi *et al.* (2012) porte sur les débats à l'Assemblée nationale française pendant le Grenelle II de l'environnement<sup>1</sup>. L'analyse concerne un extrait de 4 minutes de l'intervention du député vert Yves Cochet (figure 2). Durant l'extrait choisi, l'orateur est très fréquemment interrompu par les députés de l'opposition et adopte différentes stratégies (allant du mépris à la réponse frontale). Bien qu'étant à l'origine un discours écrit et préparé, les différentes interruptions et les réponses qu'Yves Cochet y apporte en font un discours où la parole spontanée prend le relais de la déclamation. L'originalité de cette étude réside dans le fait que les chercheurs

1. Débats relatifs à un projet de loi sur l'environnement ayant eu lieu en 2010 et ayant suscité de vives controverses.

impliquées avaient toutes une qualification particulière dans un domaine lié aux sciences du langage (analyse des interactions, prosodie, gestuelle et traitement automatique de la parole). Ainsi, le passage étudié par Bigi *et al.* (2012) a été transcrit et annoté en fonction des contours prosodiques utilisés, des gestes coverbaux effectués et des stratégies de réponse apportées aux interruptions. Un certain nombre d'annotations automatiques ont également été générées (phonétisation, annotations morphosyntaxiques et détection des autorépétitions). L'objectif de cette étude était d'annoter un corpus de manière très complète afin de prendre une interaction dans son ensemble multimodal. Elle a également permis de générer des requêtes automatiques dans le corpus.



Figure 2. Corpus Grenelle II – Intervention d'Yves Cochet  
(d'après Bigi *et al.*, 2012)

- 9 Le travail sur un corpus écologique mène donc principalement à la réalisation d'une analyse descriptive fine. Si on souhaite aller plus loin dans la compréhension du phénomène de la gestuelle, par exemple en testant des hypothèses, il est nécessaire de l'aborder avec d'autres approches, quitte à perdre un peu de naturel.

### 3.2. Le corpus semi-contrôlé

- 10 Une solution qui permet de contrôler quelques paramètres tout en laissant une certaine liberté de production aux locuteurs est le *corpus semi-contrôlé*. Selon Tellier (2013 : 42) :

Le principe du corpus semi-contrôlé est que l'on recueille les productions (orales ou écrites) de différentes personnes à qui l'on donne les mêmes consignes. On peut donc contrôler certaines variables, telles que les personnes participant à l'étude et les conditions de passation. Les productions recueillies sont ainsi comparables entre elles. Les données récoltées peuvent mener à des analyses à la fois qualitatives et quantitatives. Le corpus semi-contrôlé est recueilli selon une méthodologie bien précise qui doit être décrite en détail par les chercheur(e)s afin qu'elle puisse être reproduite par d'autres.

- 11 Les consignes données aux participants peuvent plus ou moins contrôler l'interaction. Ainsi, dans le «Corpus of Interactionnal Data» (ci-après CID; voir Bertrand *et al.*, 2008), les participants (des collègues de travail) se trouvaient par paires installés dans une chambre d'enregistrement avec pour consigne de discuter librement de situations insolites dont ils avaient été témoins. Ce corpus de huit heures, incluant huit paires d'interactants, a également donné lieu à une annotation multimodale extrêmement dense (analyse des niveaux prosodique, pragmatique, discursif, syntaxique, phonétique et gestuel, ainsi que des phénomènes mimogestuels tels que le sourire, la posture, les mouvements de tête ou encore le regard). Une fois encore, c'est la collaboration de plusieurs chercheurs, experts dans leur domaine, qui a permis une annotation aussi précise et diversifiée. Outre la réalisation d'un travail très pertinent et utile sur les schémas d'annotation multimodaux (notamment grâce au projet ANR «Outils de traitement d'information multimodale» [OTIM] – Blache *et al.*, 2010), les annotations effectuées sur le CID ont permis d'étudier l'interface entre différents niveaux de la langue, impliquant notamment les phénomènes mimogestuels (Tan *et al.*, 2010; Ferré, 2011).
- 12 D'autres études, plus contrôlées, permettent de répondre à des questions et hypothèses de recherche plus spécifiques et rendent également possible une comparaison intra- et interindividuelle. Ainsi, dans le domaine des études de la gestuelle, la narration d'un stimulus est souvent utilisée pour recueillir une production orale (McNeill, 1992). Le stimulus (un dessin animé ou une bande dessinée en général) est présenté aux participants qui doivent ensuite le raconter à un tiers. Le stimulus choisi contient beaucoup d'actions, de déplacements et/ou d'autres éléments complexes car leur description élicite de nombreux gestes lors du récit. On peut donc, par la suite, comparer les gestes produits par différentes personnes pour évoquer un même événement. Nous sommes alors dans une perspective de comparaison interindividuelle (la tâche peut être proposée à des locuteurs de différentes langues, différents âges, différentes origines socioculturelles, etc.). Ainsi, dans une perspective développementale, on peut demander à des enfants de différents âges de narrer un même stimulus (Reig Alamillo *et al.*, 2012) afin de comparer l'évolution de la production gestuelle en lien avec l'acquisition langagière. Par ailleurs, il est également possible d'effectuer des analyses intra-individuelles, par exemple en demandant à des apprenants de langue étrangère de narrer le même stimulus en L1 et en L2 afin d'analyser leur gestuelle dans les deux langues et de mieux comprendre leur organisation cognitive (Yoshioka, 2005; Stam, 2007; Gullberg, 2009).
- 13 Cependant, le corpus semi-contrôlé ne se limite pas à la narration. D'autres chercheurs ont élaboré des tâches beaucoup plus interactives mettant en scène deux participants. Dans une étude visant à analyser la spatialisation du geste (visible ou non visible par l'interlocuteur) dans des séquences «normales» et lors de séquences de réparation dans le cadre d'une interaction, Saubesty (2013) a élaboré une tâche qui non seulement suscitait de nombreux gestes mais également distinguait deux espaces de production des gestes (l'un visible par l'interlocuteur, l'autre invisible).

La tâche proposée aux participants avait un format proche d'une *map task* (une tâche à réaliser avec une carte) : un participant (le directeur) possédait un dessin de base sur lequel différentes formes complexes et abstraites étaient disposées à des endroits précis. Son partenaire (le destinataire) possédait le dessin de base mais vierge car les formes complexes lui étaient données à part, sous forme de petites cartes à disposer sur le dessin en suivant les indications du directeur. L'objectif final de la tâche pour le destinataire était de reproduire à l'identique le dessin du directeur (qu'il ne pouvait évidemment pas voir) en suivant les indications de celui-ci. Les participants étaient installés face à face devant des pupitres. L'étude de Saubesty (2013) avait pour objectif d'analyser la production gestuelle du directeur lors de ses explications. Deux espaces gestuels ont été définis : un espace dit *égocentrique* (à hauteur du pupitre) et un espace *allocentrique* (au-dessus du pupitre). Dans le premier cas, les gestes produits ne sont pas visibles par le destinataire (la figure 3a montre un geste produit dans l'espace égocentrique et la figure 3b un geste produit dans l'espace allocentrique). La prise de vue est assurée par trois caméras, ce qui permet, d'une part, de restituer fidèlement ce que chaque partenaire voit, et, d'autre part, de donner une vue d'ensemble de la scène qui révèle les gestes cachés par le pupitre. Cette étude repose sur le fait, établi par de précédentes analyses (voir 2.2), que certains gestes sont produits pour soi (dans l'espace égocentrique) et d'autres pour l'interlocuteur (la majorité des gestes produits dans l'espace allocentrique). L'hypothèse de Saubesty (2013) était que dans les situations de réparation (lors d'une incompréhension manifestée par le destinataire), le directeur aurait tendance à produire davantage de gestes dans l'espace allocentrique afin de faciliter la compréhension de son partenaire. Les résultats montrent que sur les dix directeurs analysés, deux profils se dégagent. En effet, les participants ont tendance à utiliser majoritairement l'un ou l'autre des espaces sans que l'on puisse distinguer de différence entre les situations de réparation et le reste de l'interaction. Ainsi, un participant au profil dit « égocentrique » va produire majoritairement ses gestes dans l'espace égocentrique, et ce, même lorsque des incompréhensions surviennent.

14 Un autre exemple de corpus semi-contrôlé visant à susciter des gestes est le projet « Gestures in Teacher Talk » (ci-après GTT) de Tellier et Stam (2012), dont l'objectif était d'étudier la production de gestes coverbaux par de futurs enseignants de français langue étrangère (FLE) et de comparer cette production gestuelle lorsque l'interlocuteur est un natif de la langue et lorsqu'il est un apprenant (*i. e.* un non-natif). Pour qu'une telle comparaison soit possible, il faut donc enregistrer un même locuteur (futur enseignant de FLE) dans la même situation interactive mais avec deux interlocuteurs différents (un natif et un non-natif). La situation interactive choisie était une tâche d'explication lexicale : dix étudiants de master (FLE) devaient faire deviner douze mots de vocabulaire (tirés au sort aléatoirement) à un partenaire natif et à un autre non natif (ordre contrebalancé). L'hypothèse de cette étude était qu'avec un partenaire non natif, le futur enseignant aurait tendance à utiliser des gestes plus facilitants pour la compréhension (plus illustratifs, par exemple).



Figure 3 (a et b). Dispositif de l'étude de Saubesty (2013) pour l'analyse des gestes égo-centriques et allo-centriques



cond. native c'est un verbe/synonyme de/ de [escalader]



cond. non native c'est un verbe/ euh synonyme de [escalader]

Figure 4 (a et b). Étude de la variabilité intra-individuelle (d'après Tellier et Stam, 2012)

15 Les résultats quantitatifs montrent que les futurs enseignants de cette étude produisent significativement des gestes plus illustratifs, plus amples et plus longs en durée lorsqu'ils s'adressent à des non-natifs (Tellier et Stam, 2012). Le taux gestuel (nombre de gestes produits par rapport au nombre de mots) est également significativement plus élevé avec les interlocuteurs non natifs. Ainsi, dans l'exemple de la figure 4, on voit le futur enseignant à gauche faire deviner le verbe grimper à un partenaire natif (cliché a) et non natif (cliché b). Dans les deux cas, il dit la

même phrase : « c'est un verbe synonyme de [escalader]<sup>2</sup> ». Avec le partenaire natif, le geste produit est un métaphorique (selon la classification de McNeill, 1992, 2005, voir 4.2) qui ne véhicule pas le sens du verbe « escalader » mais met simplement en évidence le concept. En revanche, avec le partenaire non natif, un geste iconique est produit ayant pour objectif d'illustrer l'action d'escalader sous forme de mime.

- 16 Pour résumer ce qui a été exposé dans cette partie, on constate qu'une première décision méthodologique à prendre lorsque l'on souhaite travailler sur le geste (mais cela est valable pour toute autre modalité) est le type de corpus à recueillir. Cela va bien entendu dépendre de la question de recherche et des hypothèses éventuelles que l'on formule. L'étude peut porter sur un corpus constitué d'une unique personne dans une situation écologique et permettra donc une analyse qualitative fine, ou bien sur un corpus rassemblant plusieurs personnes dans un contexte déterminé et avec des consignes précises, permettant un traitement à la fois quantitatif et qualitatif ainsi que la validation ou l'invalidation d'hypothèses. Ce traitement des résultats ne peut se faire, bien entendu, qu'après un travail fin de transcription et d'annotation des données.

#### 4. Quelles annotations ?

- 17 Nous n'évoquons pas ici les conventions de transcription de la parole qui sont nombreuses et dont l'objectif est de restituer à l'écrit les spécificités de l'oral (voir Blanche-Benveniste, 1997, par exemple). Face à un corpus multimodal, il est fondamental de mesurer le caractère chronophage de la transcription et de l'annotation (pour une heure de corpus on peut compter entre 20 heures et 60 heures de travail en fonction du nombre d'éléments à annoter).

##### 4.1. Quel outil pour annoter ?

- 18 Pendant longtemps, une difficulté majeure pour l'annotation gestuelle était liée à la technique. En effet, les vidéos étant au format VHS, le chercheur devait les visionner sur un magnétoscope, souvent en ralentissant le film afin de saisir les subtilités du geste et sa relation chronologique avec la parole. La transcription de la parole comme l'annotation du geste (et d'autres phénomènes mimogestuels) se faisaient alors dans un logiciel de traitement de texte. Toute la difficulté étant de restituer des éléments visuels et non verbaux par écrit tout en maintenant la chronologie d'apparition de ces différents éléments. Sont alors apparues de très astucieuses techniques dont nous ne donnerons que deux exemples ici : celles de McNeill (1992) et de Calbris (2011). Le lecteur pourra également consulter les travaux de Kendon (2004) ou Streeck (2009) parmi de nombreux autres exemples. Ce type d'annotation sous format texte présente l'avantage de bien mettre en avant la partie verbale de l'interaction en montrant comment les gestes viennent coconstruire le discours oral. Cette pratique est cependant assez chronophage et demande d'acquérir une habitude de lecture des différents

2. La partie de la phrase entre crochets représente le moment où le geste est produit.

codes employés. Dans le cas de McNeill (1992 : 150), les paroles du locuteur sont transcrites et la partie gestualisée est mise entre crochets, accompagnée d'un numéro. Le numéro renvoie à une description du geste. Pour illustrer les exemples de l'ouvrage, l'image accompagne le texte afin de permettre au lecteur de mieux visualiser. Dans l'exemple de Calbris (2011) (voir la figure 5), les symboles sont utilisés pour décrire la configuration de la main, son orientation, sa localisation dans l'espace, la latéralisation du geste (mono ou bimanuel, quelle main, etc.) ainsi que le regard.

LJ: Mais si on m'explique... Ben d'abord je je, j'hérite d'une situation, donc je la traite telle qu'elle est. Ensuite, je | décide | [OO] de ne [↔] masquer | aucune réalité, c'est quand même une méthode | de vérité | que je dois aux Français. [• on L Ω.l] *S'il y a plus de trois millions de chômeurs, je peux pas dire qu'ils sont pas là*; [• on R Ω.r (estimated)] *s'il y a un déficit de la protection sociale, j'peux pas faire semblant de ne pas le voir*; si [• sur II] *je ne veux pas casser la croissance* parce que | les Françaises et les Français | [¶¶ b] *veulent* [.f] *un dynamisme de l'économie*, [• on II] *je ne peux pas prendre des mesures* | qui cassent cette croissance. En même temps | [R vir + •ir segments] *j'ai le rendez-vous de la monnaie unique, nous avons le rendez-vous de la monnaie unique et* | **nous voulons le faire**

Figure 5. Exemple d'annotation gestuelle sous format texte : utilisation de symboles (Calbris, 2011 : 109)

19 Le développement et la diffusion de logiciels dédiés à la transcription de la parole et même à l'annotation multimodale ont considérablement fait évoluer les pratiques. Tout d'abord, la fenêtre vidéo et le support où se font la transcription et l'annotation se trouvent dans le même logiciel et la navigation dans la vidéo est facilitée (ralentissement de la vidéo, visionnage image par image, etc.). La transcription verbale et l'annotation gestuelle (mais également discursive) se font sous forme de partition et les différentes pistes utilisées (une par dimension annotée) sont disposées les unes sous les autres : il suffit d'y placer des étiquettes qui correspondent aux annotations. Ces étiquettes peuvent être préenregistrées dans le modèle de fichier (ou *template*) conçu par le chercheur et constituent ce que l'on appelle un *vocabulaire contrôlé*. Ainsi, lorsque l'on annote on peut choisir une étiquette dans un menu déroulant (voir figure 9) ce qui permet, outre un gain de temps dans l'annotation, de limiter les risques d'erreur et les fautes de frappe dans les étiquettes. Cette méthode facilite grandement l'étape de requêtage une fois l'annotation terminée (c'est-à-dire l'exploration du corpus et la recherche d'étiquettes ou de phénomènes spécifiques).

20 La lecture de ces annotations est similaire à la lecture de la partition du chef d'orchestre : sur un axe syntagmatique, de gauche à droite, pour le déroulement temporel de la parole et sur un axe paradigmatique, de haut en bas, pour visualiser les différents éléments qui cooccurrent. Parmi les différents logiciels possibles, citons les deux plus souvent utilisés à notre connaissance (et développés spécifiquement pour l'analyse multimodale) : Elan (Sloetjes et Wittenburg, 2008) (voir figure 6) et Anvil (Kipp, 2001) (voir figure 7). Ces logiciels présentent, en outre, deux avantages majeurs,

de notre point de vue : d'une part, la précision temporelle est considérablement accrue puisque l'on peut donner la durée d'un geste à 40 ms près (la durée d'une image vidéo, 25 images par seconde) et annoter très précisément la synchronie avec la parole ; d'autre part, la possibilité de requêtes automatiques et d'extraction de données rend la « fouille » de corpus beaucoup plus précise et rapide.

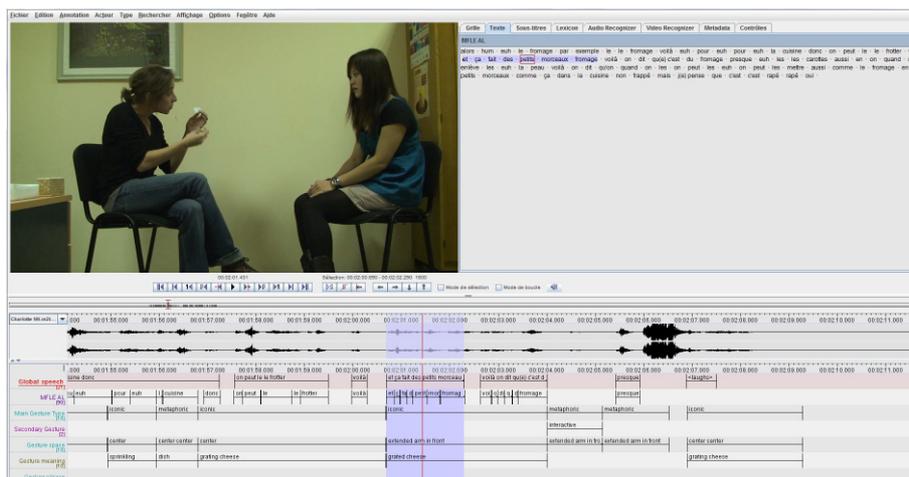


Figure 6. Logiciel Elan (extrait du corpus GTT)<sup>3</sup>

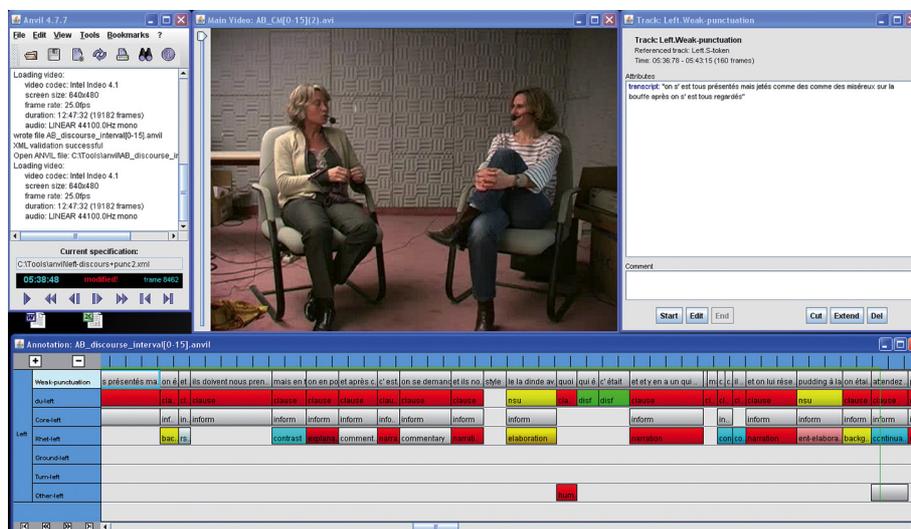


Figure 7. Logiciel Anvil (extrait du corpus CID)<sup>4</sup>

3. Archivé sur le « Speech and Language Data Repository » (SDLR) : <http://sldr.org/sldroo012> ; identifiant pérenne : hdl:11041/sldroo012.
4. Archivé sur le SLDR : <http://sldr.org/sldroo027> ; identifiant pérenne : hdl:11041/sldroo027.

#### 4.2. Construire un schéma d'annotation

21 Il n'existe pas de *schéma d'annotation* universel, comme le rappelle Calbris (2011 : 102) :

[Coding] must be exhaustive and, if possible, common to all the researchers carrying out the analysis. In practice, each researcher adopts a coding system, preferably economic, adapted to the particular framework of his research.

22 Cependant, on dénombre plusieurs tentatives de conception et de diffusion, dans la communauté scientifique, de schémas d'annotation très détaillés et généralistes : à titre d'exemple le schéma élaboré au sein de l'ANR OTIM pour les gestes coverbaux, les postures et les regards (Tan *et al.*, 2010). Un peu plus complet, le schéma créé pour le réseau MUMIN (« Multimodal Interfaces ») (Allwood *et al.*, 2007) permet d'annoter à la fois les fonctions des gestes et les formes. Pour les fonctions, il se focalise sur le feed-back, l'aménagement des tours de parole et le séquençage (Allwood *et al.*, 2007 : 277). Pour les formes, il permet d'annoter les gestes manuels (types de gestes, manua-lité, trajectoire), les expressions faciales (bouche, yeux, sourcils, tête, etc.) et prévoit également les postures. Kipp *et al.* (2007) proposent un autre schéma (et dressent aussi un état de la question sur les schémas d'annotation) qu'ils présentent comme peu coûteux en temps d'annotation (90 minutes d'annotation pour une minute de corpus, tout de même). Enfin, les participants à l'ANR « Multimodalité » ont mis en ligne un manuel de codage<sup>5</sup> très précis pour transcrire et annoter la parole et la gestuelle des enfants (entendants et sourds) (Colletta *et al.*, 2011). Ces schémas d'annotation sont relativement exhaustifs et leur utilisation demeure très coûteuse en temps, mais, tout comme pour le recueil de corpus, on constituera le schéma d'annotation en fonction des questions de recherche. Inutile donc d'annoter les configurations de la main si l'on s'intéresse uniquement à l'espace gestuel ; on peut se limiter à certains aspects, quitte à revenir sur le corpus avec de nouvelles annotations lorsque le besoin s'en fait sentir.

23 Au niveau fonctionnel, pour donner d'autres exemples de schémas plus spécifiques, liés à des questions de recherche bien précises, on pourra regarder pour l'analyse de l'interaction, la typologie des feed-back utilisée dans le projet MUMIN (Allwood *et al.*, 2007), la typologie pour la relation geste-parole de Colletta *et al.* (2011) (*i. e.* redon-dance, complément, élaboration, supplément, contradiction et substitution) ou encore, pour les interactions didactiques, la typologie fonctionnelle de Tellier (2008) pour les gestes des enseignants (gestes pour informer, animer, évaluer).

24 Au niveau de la forme du geste, de nombreuses études en gestuelle s'intéressent aux types de gestes employés. Il existe plusieurs typologies de gestes (voir Kendon, 2004, pour un historique). Une typologie relativement utilisée ces dernières années est celle de McNeill (1992), même si elle demeure controversée car parfois difficile à manipuler. McNeill (2005) lui-même a suggéré de considérer ces catégories davantage comme des dimensions que comme des types restrictifs (un même geste pouvant avoir plusieurs dimensions). Il distingue ainsi les *déictiques* (gestes de pointage), les

5. Projet ANR « Multimodalité » 2005-2009 ; manuel en ligne à l'adresse suivante : <http://lidilem.u-grenoble3.fr/IMG/pdf/anrmultimodalite-manueldecodage.pdf>.

*iconiques* (gestes illustrant des propriétés concrètes et visuelles d'un référent comme montrer la taille ou la forme de quelque chose par exemple), les *métaphoriques* (gestes qui représentent abstraitement un concept) et les *battements* (gestes rythmant la parole, sans contenu sémantique). À cela on peut ajouter les *emblèmes* (gestes culturels à la forme conventionnalisée comme la main qui passe à plat au-dessus de la tête pour dire «Ras-le-bol!»). Calbris (2011) propose également une méthodologie très détaillée du codage de la forme du geste en s'intéressant notamment à la configuration de la main, son orientation et sa localisation dans l'espace.

25 L'utilisation de l'espace gestuel demeure une composante encore peu analysée dans les études de la gestuelle car difficile à annoter. Les nouvelles technologies (notamment les capteurs de mouvements) ne résolvent que partiellement le problème. En effet, si les données obtenues sont précises, elles demeurent difficiles à analyser. En outre, le placement de capteurs sur les mains et les bras des interactants risque d'inhiber la spontanéité du geste et les mouvements des mains masquent parfois certains capteurs, ce qui entraîne la perte de données. L'analyse manuelle demeure donc d'actualité. On peut citer à titre indicatif le schéma proposé par McNeill (1992) (figure 8).

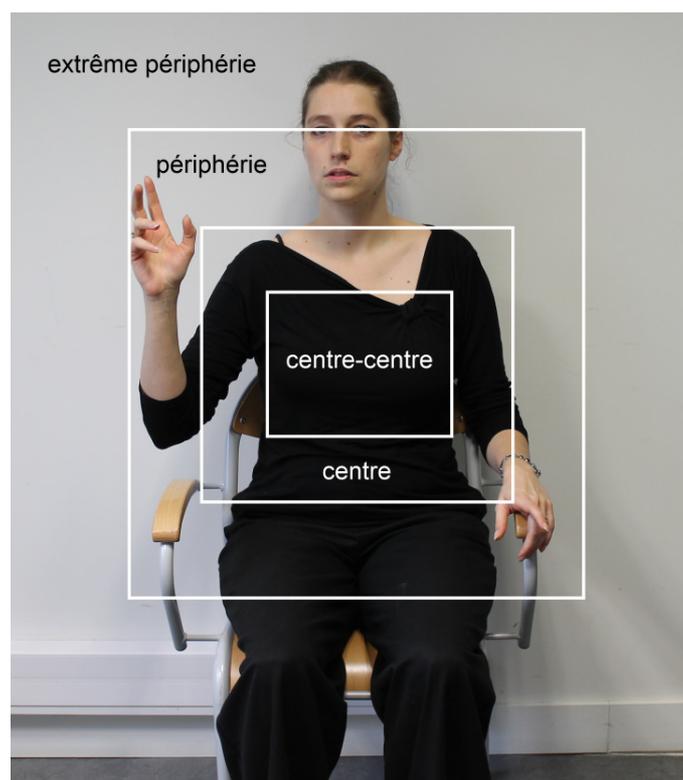


Figure 8. Espace gestuel (adapté de McNeill, 1992)<sup>6</sup>

6. Le schéma de l'espace gestuel proposé ici est une version simplifiée.

### 4.3. Autres aspects à analyser

- 26 Outre la forme du geste, sa fonction et sa position dans l'espace, un autre aspect qui peut être étudié est la quantité de gestes produits. Lorsque l'on compare différentes interactions entre locuteurs et que l'on veut calculer le nombre de gestes produits, il faut prendre en considération le fait que les interactions peuvent être de durées très différentes. Il faudra donc prendre en compte le *taux gestuel*. Il existe plusieurs méthodes pour le calculer. Une première possibilité est de calculer le nombre de gestes par minute, mais ce mode de calcul n'est valable que s'il se base sur le temps de l'élocution et non sur le temps de l'interaction, car cela inclut les pauses qui peuvent être plus ou moins nombreuses et de durée variable selon les interactants. C'est donc un mode de calcul possible pour la narration (Chu et Kita, 2009, par exemple) mais pas pour l'interaction. Une seconde méthode est le calcul du nombre de gestes par proposition (méthode utilisée dans McNeill, 1992, par exemple), mais elle pose le problème de la définition d'une proposition (est-ce que l'on considère une proposition subordonnée ou coordonnée comme une proposition principale lorsque l'on calcule le nombre de propositions utilisées? Une clause syntaxique minimale est-elle considérée comme équivalente à une proposition complexe?). Enfin, la solution de calculer le nombre de gestes par mot semble plus précise (utilisée par Tellier et Stam, 2012, par exemple; mais on trouve aussi une version «nombre de gestes pour 100 mots» dans Alibali *et al.*, 2001), mais pose quand même un certain nombre de questions méthodologiques telles que la simple définition de ce qu'est un mot à l'oral. Par exemple, lorsque l'on dit «j'ai faim», on compte 3 mots au niveau syntaxique mais 2 sur un plan plus phonologique. Par ailleurs, comment prend-on en compte les disfluences typiques de l'oral ainsi que les phénomènes vocaux comme les «euh» et les «hum»? Méthodologiquement, il convient de bien définir ce que l'on comptera comme mot et de s'en tenir à ces principes pour assurer une comparaison fiable de différents corpus ou de différentes conditions analysées.
- 27 Un second aspect intéressant est la durée du geste. Pour cela on peut diviser le temps passé à gestualiser (facile à obtenir lorsque l'annotation a été réalisée sur un logiciel dédié) par le nombre de gestes produits. Ainsi, dans l'étude de Tellier et Stam (2012) mentionnée plus haut, qui compare la production gestuelle d'un même individu avec un partenaire natif et un partenaire non natif, on constate qu'avec les non-natifs, les gestes durent significativement plus longtemps qu'avec les natifs. Cela est lié sans doute au débit de parole, lui aussi plus lent.

## 5. Quelle validation des annotations?

- 28 Quiconque s'est penché sur l'annotation<sup>7</sup> (qu'elle soit gestuelle, prosodique ou discursive) a pu constater le caractère subjectif d'un tel travail. Les questionnements,

7. Annotation et codage sont ici utilisés comme synonymes, de même que leurs dérivés : schéma d'annotation/codage, contre-annotateur/contre-codeur, etc.

les doutes envahissent fatalement le chercheur face à ses données. Ai-je toujours été régulier/ère dans mes annotations? Une autre personne aurait-elle annoté comme moi? Quelle est la fiabilité du traitement de mon corpus sur lequel va reposer toute mon analyse? À ce sujet, Calbris (2011: 102) écrit :

*Note that achieving the ideal of scientific objectivity when coding a corpus is a delusion, because coding depends on perception, an essentially pre-interpretative and therefore subjective activity. Objectivity lies more in recognizing the subjective perceptual biases that may possibly enter into the interpretation of spontaneous and non-conscious co-speech gestures.*

29 Quelles solutions méthodologiques sont possibles? Cela passe, bien évidemment et avant tout, par un *guide d'annotation* précis et univoque. La conception du schéma d'annotation a été évoquée plus haut; ce schéma doit être accompagné d'un guide qui précise la définition de chaque catégorie d'annotation avec des exemples pour illustrer. Le lecteur pourra se référer au manuel de codage de l'ANR «Multimodalité» (Colletta *et al.*, 2011) disponible en ligne.

30 Une autre solution méthodologique (complémentaire) possible afin de réduire le caractère subjectif des annotations (il s'agit bien de réduire et non d'éradiquer) est le *contre-codage*. Ce terme semble être issu du français québécois. Les autres appellations possibles ou notions corrélées sont, selon Royer *et al.* (2009: 74): la notion de «procédures de validation du codage», celle d'accord «intercodeurs», d'accord «interjuges», ou encore celle de «codeurs indépendants». Tous ces termes ne renvoient pas aux mêmes choses. Le contre-codage désigne le processus par lequel on demande à une tierce personne (que l'on peut appeler «juge» ou «codeur indépendant» ou «contre-annotateur») d'annoter notre corpus en aveugle (c'est-à-dire sans avoir accès aux annotations initiales), à l'aide du guide qu'on lui aura fourni. Il s'agit ensuite de confronter nos propres annotations (initiales) à celles du codeur indépendant. On procède enfin au calcul d'un taux d'accord interjuges ou accord intercodeurs, en calculant un *kappa* ou en donnant le pourcentage d'accord entre annotateurs, bien que ces deux processus soient un peu différents (Allwood *et al.*, 2007: 283-284):

*Percent agreement is generally considered too liberal a measure for intercoder agreement, and indeed, the results shown above concerning the Finnish case study show that it gives consistently higher values than the k-scores. However, it is interesting to ponder upon the differences between the two agreement measures since kappa can often be very low, even negative, while percentage agreement is still quite high.*

31 Cette confrontation, parfois douloureuse (les écarts pouvant être importants) est extrêmement enrichissante car elle permet: (1) de s'assurer que le guide d'annotation est clair, (2) de réduire la subjectivité de l'annotation, (3) de mettre au jour certaines imprécisions du guide d'annotation. Cependant, faire un contre-codage ne va pas de soi et de nombreuses questions méthodologiques surgissent. Les articles basés sur des études ayant eu recours au contre-codage n'étant pas toujours très explicites sur la marche à suivre, il convient de formaliser un peu les choses.

32 Les réflexions présentées ici sont le fruit à la fois d'expériences de contre-codage (exposées notamment dans Allwood *et al.*, 2007; Tan *et al.*, 2010; Tellier *et al.*, 2012, 2013) et d'échanges au sein d'un groupe de travail<sup>8</sup>.

33 Les questions centrales que l'on peut se poser sur le contre-codage sont : (1) À quel moment de la recherche proposer un contre-codage ? (2) Quelle proportion du corpus contre-annoter ? (3) Quels annotateurs (naïfs ou non, experts ou non) ? (4) Que leur donner (schéma et guide d'annotation, étiquettes vides ou non) ? (5) Que fait-on après le contre-codage ?

### 5.1. À quel moment ?

34 Deux moments sont possibles. Une première possibilité peut être à la fin du travail : lorsque tout le corpus a été annoté, on peut donner un échantillon à des annotateurs indépendants afin de vérifier si le codage que l'on a effectué est similaire à celui d'une tierce personne. Ce processus, à ce moment-là de l'étude, valide les annotations sur lesquelles l'analyse va être faite. Il nous semble que c'est le cas le plus fréquent (voir par exemple Chu et Kita, 2009; Tan *et al.*, 2010; Colletta *et al.*, 2011; Tellier *et al.*, 2012).

35 Une seconde possibilité est de le faire plutôt au début. Une fois que le schéma et le guide d'annotation sont conçus et qu'ils ont été testés sur une petite partie du corpus, il est possible de faire intervenir un autre codeur en aveugle. Cela permet dès le début de faire émerger d'éventuelles imprécisions dans le guide d'annotation, d'éventuelles lacunes aussi, et de les corriger avant d'annoter tout le corpus. C'est ce qui est évoqué par Allwood *et al.* (2007), mais également Tellier *et al.* (2013).

36 Ces deux moments sont intéressants pour des raisons différentes : au début de l'étude, cela permet de rectifier le tir et de préciser le schéma et le guide d'annotation. À la fin de l'étude, cela permet de valider le travail fait. On peut donc en conclure que deux contre-codages sont nécessaires, encore faut-il trouver les volontaires pour le faire car le travail est coûteux en temps.

### 5.2. Quelle proportion du corpus contre-annoter ?

37 Il n'y a pas de réponse unique à cette question. Tout dépend, d'une part, de la taille du corpus et, d'autre part, de la complexité des annotations. À titre d'exemple, dans l'étude de Tellier *et al.* (2013) sur les fonctions des gestes produits dans les pauses du discours, l'examen se faisait sur 20 séquences comprenant plusieurs explications lexicales d'environ 1 minute chacune (à partir du corpus GTT). Le contre-codage a

8. Je tiens ici à remercier les participants à ce groupe de travail et tout particulièrement Brahim Azaoui, Erika Dussire, Charlotte Bouget, Aurélie Goujon, Alexandra Gadoni et Jan Gorish du laboratoire Parole et langage et de l'équipe de recherche Praxiling, ainsi que Cathy Cohen, Nicolas Guichon et Ciara Wigham du laboratoire ICAR (Interactions, corpus, apprentissages, représentations), pour leur participation et leurs remarques pertinentes.

porté sur huit séquences. Dans l'étude de Tellier *et al.* (2012) sur le corpus DEGELS<sup>9</sup>, celui-ci étant très court (1 minute/60 gestes), il a ainsi été contre-annoté dans sa totalité par trois annotateurs indépendants.

38 Lorsque le corpus est plus long, il semble que 15 % à 20 % du corpus constituent un échantillon raisonnable, mais ceci n'est donné qu'à titre indicatif. Sur le corpus CID évoqué précédemment, Tan *et al.* (2010) ont réalisé un contre-codage à trois annotateurs sur 3 minutes d'une interaction qui durait 75 minutes, soit 4 %. Dans l'étude de Kita et Özyürek (2003), les récits de 51 participants appartenant à trois groupes linguistiques différents (anglais-américain, turc et japonais) ont été codés et analysés. Il s'agissait de raconter un extrait d'un épisode du dessin animé *Tweety Bird*. Le contre-codage a été réalisé par deux annotateurs indépendants sur trois récits dans chaque langue de l'étude (soit neuf récits ou encore 17 % du corpus). On peut aussi sélectionner des éléments précis sur lesquels on souhaite un contre-codage. Ainsi, dans cette étude, ce qui intéresse les chercheurs ce sont les gestes utilisés pour transcrire le mouvement de trajectoire d'un personnage du stimulus, et donc le codage comme le contre-codage se sont limités à ces gestes précis.

### 5.3. Quels annotateurs (naïfs ou non, experts ou non) ?

39 Qui peut faire le contre-codage ? Les annotateurs doivent-ils être experts dans le domaine ? Doivent-ils connaître ou non l'objectif de l'étude ? Ces questions méthodologiques ne sont pas triviales mais il n'existe hélas pas de réponse satisfaisante. Dans le projet MUMIN, le schéma d'annotation a été testé sur plusieurs corpus de différentes langues d'Europe du Nord. Pour le corpus danois, deux annotateurs indépendants avec un profil de linguiste mais une expertise limitée dans l'annotation gestuelle ont réalisé l'annotation. Ils se sont d'abord entraînés ensemble sur une partie du corpus pour se familiariser avec la tâche (Allwood *et al.*, 2007 : 280). Allwood *et al.* (2007 : 280) argumentent en faveur de l'entraînement des annotateurs avant de pratiquer l'annotation réelle et expliquent que c'est une pratique courante dans les projets d'annotation. L'annotation réelle se fait ensuite indépendamment. Pour le corpus finnois, cinq annotateurs ont participé : un expert et quatre naïfs. Ces derniers ont suivi un rapide cours introductif ainsi qu'un entraînement à l'annotation individuelle. Ensuite, ils ont formé deux paires et ont donc annoté en ne produisant qu'un seul fichier d'annotation par binômes. Les calculs du taux d'accord interjuges comparent les annotateurs par groupes de deux : le premier binôme de naïfs avec le codeur expert, le second binôme de naïfs avec le codeur expert et les deux binômes de naïfs ensemble. On remarque que le premier binôme et l'expert ont un fort taux

9. Les ateliers DEGELS (Défi, geste, langue des signes) ont été organisés en France dans le cadre de la conférence TALN (Traitement automatique des langues naturelles) en 2011 et 2012. Ils ont réuni gestualistes et spécialistes de la langue des signes française, invités à comparer leurs approches méthodologiques en termes d'annotation à partir d'un même corpus. L'objectif de ces ateliers sous forme de défis est de « mutualiser les questions de recherche autour de l'annotation des corpus de français oral (voix et gestes) et de langue des signes française, dans le but d'élaborer un guide d'annotation portant sur les aspects communs et ainsi permettre une meilleure accessibilité aux annotations des différents chercheurs » (<http://degels.limsi.fr/>).

d'accord alors que le second binôme et l'expert ont un fort taux de désaccord (et de fait, les deux binômes entre eux sont en désaccord). Que peut-on en conclure ?

40 Un codeur naïf peut être fiable à condition de bien l'entraîner et de lui donner à faire un travail qui ne requiert pas un niveau de technicité trop accru (voir la question suivante). Le codeur expert présente l'avantage d'être habitué à l'annotation gestuelle et à la lecture d'un guide d'annotation et il maîtrise le logiciel et les cadres théoriques ; il peut cependant être très tatillon sur certaines annotations voire être accoutumé à des schémas d'annotation un peu différents et avoir du mal à trancher sur certains phénomènes. Le non-expert, en revanche, se fie plus au schéma d'annotation et se pose peut-être moins de questions liées à des cadres théoriques. Après un apprentissage rapide du logiciel et grâce à l'utilisation de la méthode des étiquettes vides (voir ci-dessous), il peut facilement annoter du corpus. On pourra aussi se reporter au travail de Fort (2012) qui propose une typologie des annotateurs (« annotateur-analyste » et « annotateur-exécutant »), notamment dans le cadre de campagnes d'annotation par collaboration en ligne (*crowdsourcing*).

41 Dans l'étude de Tellier *et al.* (2012) sur le corpus DEGELS, les trois annotateurs étaient des gestualistes et ont annoté les différentes phases d'une soixantaine de gestes (ceci concerne le découpage de gestes en étapes de production). Cela s'est avéré très complexe et le taux d'accord s'est révélé extrêmement faible, mais ce résultat décevant a conduit à des discussions très pertinentes sur la façon dont ces trois chercheurs envisageaient la notion théorique de segmentation du geste en phases (McNeill, 1992 ; Kendon, 2004).

#### 5.4. Que donner aux contre-annotateurs ?

42 Comme il a été évoqué précédemment, il est préférable de former un peu les contre-annotateurs avant. Ils doivent s'appropriier, en particulier s'ils sont naïfs, à la fois le schéma d'annotation (et le guide qui l'accompagne) et le logiciel d'annotation. En fonction de l'étude que l'on mène, la segmentation du geste (trouver le début et la fin) peut être de peu d'importance au niveau du contre-codage. En plus d'être complexe et coûteuse en temps, elle est souvent moins fondamentale que le contenu même de l'annotation que l'on pose sur le geste (*i. e.* de quel type de geste il s'agit, quelle fonction, dans quel espace gestuel, etc.). Une approche méthodologique utilisée par Tellier *et al.* (2012) consiste à proposer aux contre-annotateurs un fichier pré-segmenté. En somme, les différents gestes du corpus sont déjà repérés dans le fichier d'annotation, mais les étiquettes correspondant à ces gestes du corpus sont données vides au contre-codeur, qui n'a plus qu'à choisir, parmi les items du schéma d'annotation, ce qui correspond selon lui à chaque geste (voir figure 9).

43 Cela permet un gain de temps considérable, facilite le travail du néophyte (qui n'a qu'à cliquer dans le menu déroulant qui s'affiche) et permet de se focaliser sur le contenu de l'étiquette plutôt que sur ses frontières. Encore une fois, cela n'est pertinent que dans le cas où le positionnement des frontières du geste (sa segmentation) n'est pas au centre du codage.

Figure 9. Méthode pré-segmentée utilisée par Tellier *et al.* (2012)

### 5.5. Que fait-on après le contre-codage ?

- 44 Comme évoqué plus haut, le contre-codage sert, si on y a recours au début de l'étude, à rectifier le guide d'annotation si celui-ci manque de précision voire à modifier le schéma d'annotation s'il apparaît que des catégories sont redondantes, superflues ou même manquantes. En fin d'étude, il permet de tester (dans la mesure du possible) la fiabilité de l'annotation. Dans un même fichier d'annotation, on fusionne les codages des différents participants, ce qui permet de regarder les points d'accord et de désaccord. Par exemple, dans la figure 10, extraite de l'étude de Tellier *et al.* (2012) sur le corpus DEGELS, trois annotateurs devaient travailler sur une soixantaine de gestes. Un type d'annotation demandé était de spécifier la dimension de chaque geste en se basant sur la typologie de McNeill (1992) (voir 4.2).

Figure 10. Comparaison des annotations (extrait de Tellier *et al.*, 2012)

- 45 Chaque annotateur devait donner une dimension primaire pour chaque geste et avait la possibilité de spécifier une dimension secondaire lorsqu'il lui semblait que cette double attribution était pertinente. En intégrant toutes les annotations dans le même fichier, on visualise aisément les accords et désaccords. Par exemple, on voit que sur le premier geste, le premier annotateur (A1) est d'accord avec le troisième annotateur (A3) pour dire que le geste a une dimension primaire iconique, alors que le deuxième annotateur (A2) considère que c'est la dimension déictique qui prime. On voit aussi qu'A1 et A2 attribuent deux dimensions au premier geste alors qu'A3 n'en donne qu'une seule. Sur le deuxième geste, les trois annotateurs sont en accord parfait pour classer ce geste dans la dimension déictique. Un *kappa* de Cohen (permettant de comparer deux juges) et un *kappa* de Fleiss (utilisé lorsqu'il

y en a plus de deux)<sup>10</sup> (Santos, 2010) ont été réalisés. Pour interpréter les résultats, on utilise le barème de Landis et Koch (1977; cité dans Santos, 2010) qu'il faut cependant relativiser car plus il y a de juges et plus il y a de catégories, plus le *kappa* aura tendance à être faible (Santos, 2010 : 2).

Ainsi, par exemple, un  $\kappa = 0.40$  pourra être considéré comme très médiocre si deux juges avaient seulement à choisir entre deux scores A et B, mais pourra être perçu comme relativement honorable s'ils devaient choisir entre 10 stades différents.

Juges	Taux d'accord
A1 et A2	0,520
A1 et A3	0,347
A2 et A3	0,281
A1, A2 et A3	0,406

Figure 11. Taux d'accord inter-annotateurs (extrait de Tellier *et al.*, 2012)

46 On constate globalement (figure 11) qu'A1 et A2 ont un taux d'accord plus important et que les trois annotateurs ensemble n'ont pas un taux très élevé (entre 0,21 et 0,40 est considéré comme un accord faible selon le barème de Landis et Koch, 1977). Cependant, selon Tellier *et al.* (2012 : 46) :

[S]i l'on considère le kappa de Fleiss entre les trois juges (0,406), on est très proche d'un accord moyen, ce qui, considérant le fait que les trois annotateurs devaient choisir entre 8 catégories, est plutôt honorable. De plus, en ce qui concerne les dimensions, le désaccord est à relativiser car très souvent, les annotateurs ont attribué les deux mêmes dimensions à un geste mais n'étaient pas forcément d'accord sur la dimension qui devait être considérée comme primaire.

47 En cas de désaccord, l'idéal est de réunir les différents annotateurs et de discuter afin de s'entendre sur l'annotation à garder pour l'analyse finale des données.

## 6. Conclusion

48 Cet article méthodologique avait pour objectif de faire le point sur la marche à suivre pour mener une étude sur corpus dans une perspective multimodale, en se focalisant tout particulièrement sur l'analyse du geste coverbal. Ainsi, de la constitution du corpus à son annotation, toutes les étapes de travail ont été discutées. Bien sûr, cet article présente certaines lacunes, faute de place. Il n'évoque pas, par exemple, les aspects techniques de l'enregistrement de corpus (combien de caméras? disposées comment? quel éclairage? quelle prise de son?). Il n'aborde pas non plus l'analyse

10. Pour un exposé sur les différentes méthodes de calcul du taux d'accord, voir Fort (2012).

quantitative et/ou qualitative des données, qui n'a pas été évoquée ici tant cet aspect de la recherche peut varier d'une étude à l'autre. Bien entendu, les propositions méthodologiques exposées ici ne sont que des propositions, basées sur l'expérience de l'auteure et celle d'autres chercheurs ayant pris le soin de détailler les coulisses techniques de leur recherche dans leurs publications. Cet article ne vise aucunement l'exhaustivité et reconnaît que d'autres approches méthodologiques sont tout à fait possibles. Par ailleurs, il pourrait être intéressant d'explorer comment ces questions méthodologiques sont traitées dans d'autres champs tels que celui de la langue des signes (le lecteur pourra consulter les actes des ateliers DEGELS, 2011 et 2012) et celui du traitement automatique des langues (voir Fort, 2012, ou encore Bigi *et al.*, 2012, parmi beaucoup d'autres).

49 L'approche multimodale d'un corpus est d'une grande richesse pour saisir la complexité de la parole, notamment en interaction. Tout ce qui se joue dans la communication passe par du verbal, des gestes, mais également d'autres éléments kinésiques comme le regard, les expressions faciales, les postures, etc. De même, la prosodie, non abordée ici, est un paramètre fondamental pour saisir tous les niveaux de la parole et comment ceux-ci coconstruisent le discours. Cependant, la maîtrise de tous ces paramètres est peu commune pour un même chercheur (sans compter l'aspect chronophage de la chose). Nous ne pouvons, par conséquent, qu'encourager des projets de recherche qui réuniraient des spécialistes de ces divers domaines, afin de travailler sur un corpus commun pour ensuite examiner précisément les interfaces entre les différentes modalités qui constituent la parole. De tels projets existent et ont été évoqués dans cet article, mais ils demeurent marginaux.

## Références

- ALIBALI, M. W., HEATH, D. C. et MYERS, H. J. 2001. Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen. *Journal of Memory and Language* 44 (2): 169-188.
- ALLWOOD, J. *et al.* 2007. The MUMIN Coding Scheme for the Annotation of Feedback, Turn Management and Sequencing Phenomena. *Language Resources and Evaluation* 41 (3-4): 273-287.
- AZAOU, B. 2014. Segmenter les gestes à partir d'un film de classe. In M. TELLIER et L. CADET (éd.), *Le corps et la voix de l'enseignant: théorie et pratique*. Paris: Maison des langues: 177-180.
- BALTHASAR, L. et BERT, M. 2007. La plateforme «Corpus de langues parlées en interaction» (CLAPI). *Lidil* 31: 13-33. En ligne à l'adresse suivante: <http://lidil.revues.org/139>.
- BAVELAS, J. *et al.* 1995. Gestures Specialized for Dialogue. *Personality and Social Psychology Bulletin* 21: 394-405. En ligne à l'adresse suivante: [http://www.web.uvic.ca/psyc/bavelas/1995InterAGestures\\_dialogue.pdf](http://www.web.uvic.ca/psyc/bavelas/1995InterAGestures_dialogue.pdf).
- BEATTIE, G. et SHOVELTON, H. 1999. Do Iconic Hand Gestures Really Contribute Anything to the Semantic Information Conveyed by Speech? An Experimental Investigation. *Semiotica* 123 (1-2): 1-30.

- BERTRAND, R. *et al.* 2008. Le CID – Corpus of Interactional Data : annotation et exploitation multimodale de parole conversationnelle. *TAL – Traitement automatique des langues* 49 (3) : 105-134.
- BIGI, B. *et al.* 2012. A Multimodal Study of Answers to Disruptions. *Journal on Multimodal User Interfaces* 7 (1-2) : 55-66.
- BLACHE, P. *et al.* 2010. Multimodal Annotation of Conversational Data. In *Proceedings of the 4th Linguistic Annotation Workshop (Uppsala, Sweden, 15-16 July 2010)*. Stroudsburg : Association for Computational Linguistics : 186-191. En ligne à l'adresse suivante : <http://aclweb.org/anthology/W/W10/W10-1829.pdf>.
- BLANCHE-BENVENISTE, C. 1997. *Approches de la langue parlée en français*. Paris – Gap : Ophrys.
- CALBRIS, G. 2003. *L'expression gestuelle de la pensée d'un homme politique*. Paris : CNRS Éditions.
- CALBRIS, G. 2011. *Elements of Meaning in Gesture*. Amsterdam : J. Benjamins.
- CHU, M. et KITA, S. 2009. Co-speech Gestures Do Not Originate from Speech Production Processes: Evidence from the Relationship between Co-thought and Co-speech Gestures. In N. TAATGEN et H. VAN RIJN (éd.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society (CogSci 2009)*. Austin : Cognitive Science Society : 591-595. En ligne à l'adresse suivante : <http://csjarchive.cogsci.rpi.edu/proceedings/2009/papers/105/paper105.pdf>.
- COLLETTA, J.-M. 2004. *Le développement de la parole chez l'enfant âgé de 6 à 11 ans. Corps, langage et cognition*. Sprimont : P. Mardaga.
- COLLETTA, J.-M. *et al.* 2011. *Transcription et annotation de données multimodales sous ELAN – Manuel de codage*. Projet ANR « Multimodalité » 2005-2009. En ligne à l'adresse suivante : <http://lidilem.u-grenoble3.fr/IMG/pdf/anrmultimodalite-manuel-decodage.pdf>.
- DE RUITER, J. P. et DE BEER, C. 2013. A Critical Evaluation of Models of Gesture and Speech Production for Understanding Gesture in Aphasia. *Aphasiology* 27 (9) : 1015-1030.
- FERRÉ, G. 2011. Annotation multimodale du français parlé. Le cas des pointages. In *Actes de la 18<sup>e</sup> conférence sur le Traitement automatique des langues naturelles (TALN 2011) – Atelier DEGELS (Montpellier, 27 juin-1<sup>er</sup> juillet 2011)*. 29-43. En ligne à l'adresse suivante : [http://degels2011.limsi.fr/actes/pdf/4\\_ferre.pdf](http://degels2011.limsi.fr/actes/pdf/4_ferre.pdf).
- FORT, K. 2012. *Les ressources annotées, un enjeu pour l'analyse de contenu : vers une méthodologie de l'annotation manuelle de corpus*. Thèse de doctorat non publiée. Université Paris XIII.
- GADET, F. *et al.* 2012. Un grand corpus de français parlé : le CIEL-F. *Revue française de linguistique appliquée* 17 (1) : 39-54.
- GOLDIN-MEADOW, S. 2000. Beyond Words: The Importance of Gesture to Researchers and Learners. *Child Development* 71 (1) : 231-239.
- GOLDIN-MEADOW, S. 2003. *Hearing Gesture: How Our Hands Help Us Think*. Cambridge : Harvard University Press.
- GULLBERG, M. 2009. Reconstructing Verb Meaning in a Second Language. How English Speakers of L2 Dutch Talk and Gesture about Placement. *Annual Review of Cognitive Linguistics* 7 (1) : 222-244.

- IVERSON, J. M. et GOLDIN-MEADOW, S. 1998. Why People Gesture When They Speak. *Nature* 396: 228.
- KENDON, A. 2004. *Gesture. Visible Action as Utterance*. Cambridge: Cambridge University Press.
- KIPP, M. 2001. Anvil – A Generic Annotation Tool for Multimodal Dialogue. In P. DALSGAARD, B. LINDBERG et H. BENNER (éd.), *Eurospeech 2001: 7th European Conference on Speech Communication and Technology, September 3-7, 2001, Aalborg*. Aalborg: Aalborg University: 1367-1370.
- KIPP, M., NEFF, M. et ALBRECHT, I. 2007. An Annotation Scheme for Conversational Gestures: How to Economically Capture Timing and Form. *Language Resources and Evaluation* 41 (3-4): 325-339.
- KITA, S. et ÖZYÜREK, A. 2003. What Does Cross-Linguistic Variation in Semantic Coordination of Speech and Gesture Reveal? – Evidence for an Interface Representation of Spatial Thinking and Speaking. *Journal of Memory and Language* 48 (1): 16-32.
- KRAUSS, R. M., CHEN, Y. et GOTTESMAN, R. 2000. Lexical Gestures and Lexical Access: A Process Model. In D. MCNEILL (éd.), *Language and Gesture*. Cambridge: Cambridge University Press: 261-283.
- LANDIS, J. R. et KOCH, G. G. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33 (1): 159-174.
- MCNEILL, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: The University of Chicago Press.
- MCNEILL, D. 2005. *Gesture and Thought*. Chicago: The University of Chicago Press.
- MORSELLA, E. et KRAUSS, R. M. 2004. The Role of Gestures in Spatial Working Memory and Speech. *The American Journal of Psychology* 117 (3): 411-424. En ligne à l'adresse suivante: <http://online.sfsu.edu/morsella/images/MorsellaRMKAJP.pdf>.
- ÖZYÜREK, A. 2002. Do Speakers Design Their Cospeech Gestures for Their Addressees? The Effect of Addressee Location on Representational Gestures. *Journal of Memory and Language* 46 (4): 688-704.
- REIG ALAMILLO, A., COLLETTA, J.-M. et GUIDETTI, M. 2012. Gesture and Language in Narratives and Explanations: The Effects of Age and Communicative Activity on Late Multimodal Discourse Development. *Journal of Child Language* 40 (3): 511-538.
- ROYER, C., BARIBEAU, C. et DUCHESNE, A. 2009. Les entretiens individuels dans la recherche en sciences sociales au Québec: où en sommes-nous? Un panorama des usages. *Recherches qualitatives – Hors-série* 7: 64-79. En ligne à l'adresse suivante: [http://www.recherche-qualitative.qc.ca/documents/files/revue/hors\\_serie/hors\\_serie\\_v7/HS7\\_Texte\\_Royer\\_Baribeau.pdf](http://www.recherche-qualitative.qc.ca/documents/files/revue/hors_serie/hors_serie_v7/HS7_Texte_Royer_Baribeau.pdf).
- SANTOS, F. 2010. Le kappa de Cohen: un outil de mesure de l'accord inter-juges sur des caractères qualitatifs. Article en ligne à l'adresse suivante: [http://www.pacea.u-bordeaux1.fr/IMG/pdf/Kappa\\_Cohen.pdf](http://www.pacea.u-bordeaux1.fr/IMG/pdf/Kappa_Cohen.pdf).
- SAUBESTY, J. 2013. *Les gestes coverbaux sont-ils marqueurs de la théorie de l'esprit?* Mémoire de master 2 en sciences du langage, non publié. Université d'Aix-Marseille.

- SLOETJES, H. et WITTENBURG, P. 2008. Annotation by Category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. Stroudsburg: Association for Computational Linguistics: 816-820. En ligne à l'adresse suivante : [http://www.lrec-conf.org/proceedings/lrec2008/pdf/208\\_paper.pdf](http://www.lrec-conf.org/proceedings/lrec2008/pdf/208_paper.pdf).
- STAM, G. 2007. Second Language Acquisition from a McNeillian Perspective. In S. D. DUNCAN, J. CASSELL et E. T. LEVY (éd.), *Gesture and the Dynamic Dimension of Language: Essays in Honor of David McNeill*. Gesture Studies 1. Amsterdam : J. Benjamins: 117-124.
- STREECK, J. 2009. *Gesturecraft: The Manu-facture of Meaning*. Amsterdam – Philadelphie : J. Benjamins.
- TAN, N. *et al.* 2010. Multi-level Annotations of Nonverbal Behaviors in French Spontaneous Conversation. In M. KIPP *et al.*, *Proceedings of Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality (18 May 2010, Malta)*. 74-79. En ligne : <http://embots.dfki.de/doc/MMC2010-Proceedings.pdf>.
- TELLIER, M. 2008. Dire avec des gestes. *Le français dans le monde. Recherches et applications* 44: 40-50.
- TELLIER, M. 2013. De l'usage du corpus semi-contrôlé dans la recherche en didactique des langues. *Cahiers de l'Asdifle* 24: 39-47.
- TELLIER, M., AZAOU, B. et SAUBESTY, J. 2012. Segmentation et annotation du geste : méthodologie pour travailler en équipe. In A. BRAFFORT, L. BOUTORA et G. SÉRASSET (éd.), *Actes des conférences JEP-TALN-RECITAL 2012 – Ateliers DEGELS 2012 (Grenoble, 4-8 juin 2012)*. Paris – Avignon : ATALA – AFCEP: 41-55. En ligne à l'adresse suivante : <http://www.jeptaln2012.org/actes/DEGELS2012/pdf/DEGELS201204.pdf>.
- TELLIER, M. et STAM, G. 2012. Stratégies verbales et gestuelles dans l'explication lexicale d'un verbe d'action. In V. RIVIÈRE (éd.), *Spécificités et diversité des interactions didactiques*. Paris: Riveneuve Éditions: 357-374.
- TELLIER, M., STAM, G. et BIGI, B. 2013. Gesturing While Pausing in Conversation: Self-oriented or Partner-oriented? In *Proceedings of the Tilburg Gesture Research Meeting (TIGER 2013)*. En ligne à l'adresse suivante : <http://tiger.uvt.nl/pdf/papers/tellier.pdf>.
- TOGNINI-BONELLI, E. 2001. *Corpus Linguistics at Work*. Amsterdam – Philadelphia: J. Benjamins.
- YOSHIOKA, K. 2005. *Linguistic and Gestural Introduction and Tracking of Referents in L1 and L2 Discourse*. Groningen: Rijksuniversiteit Groningen.