

Annotation multimodale du français parlé Le cas des pointages

Gaëlle Ferré

LLING, Chemin de la Censive du Tertre, BP 81227, 44312 Nantes Cedex 3
Gaelle.Ferre@univ-nantes.fr

Résumé. L'objectif de cet article est de tester un schéma d'encodage existant, prévu pour l'annotation des gestes, sur un corpus de nature différente que celui pour lequel il a été conçu. Le schéma d'encodage présenté est une adaptation du schéma utilisé dans le cadre du projet ANR Otim. Ce dernier est prévu pour l'annotation des phénomènes verbaux, vocaux et gestuels sur des corpus de parole conversationnelle. Dans le cadre de cet atelier, nous y avons effectué des modifications relativement mineures pour qu'il puisse être utilisé dans l'annotation de l'extrait de corpus proposé par les organisatrices de l'atelier. Il s'agissait pour nous d'annoter les pointages produits en français parlé dans une tâche de description d'itinéraire. Nous présenterons donc dans le détail les parties du schéma d'encodage utilisées pour cette annotation, les ajouts par rapport au schéma d'encodage existant et montrerons comment les annotations réalisées peuvent être utilisées pour une analyse linguistique des pointages.

Abstract. This work aims at testing an existing coding scheme for the annotation of gestures on a different corpus than the one for which it was originally created. The coding scheme presented here has been adapted from the one used in the nationally funded OTIM project. The latter was created to annotate verbal, vocal and gesture phenomena produced in conversational speech. For this specific workshop, we made minute changes to the OTIM coding scheme so that it can be used to annotate the corpus excerpt proposed to participants by the organizers. We noted pointing co-verbal gestures in a short corpus of spoken French in which the speaker is giving route directions (although not in a map task experiment). The coding scheme will be presented extensively, as well as the changes which were made. We will then show how one can use the annotations of pointing gestures in linguistics.

Mots-clés : Multimodalité, pointages, schéma d'encodage.

Keywords: Multimodality, pointing gestures, coding scheme.

1 Introduction

Depuis quelques années, le champ de la pragmatique en linguistique, qui se consacre à l'étude des interactions, a pris conscience de l'importance de la production gestuelle dans certaines situations de communication. S'est alors développé un grand nombre d'études dans des domaines variés : on pense aux travaux en plein essor sur les langues des signes, intimement liées à la gestualité, mais aussi à un certain nombre d'études portant sur la multimodalité du langage, dont le but est de montrer les liens qui existent entre la gestualité et la parole en situation de communication. Le langage est alors considéré comme un système linguistique qui intègre parole et gestualité de manière organisée et il apparaît que, comme chaque élément de ce système, la gestualité possède plusieurs fonctions intra- et inter-personnelles, ainsi que le formulent Morsella & Krauss (2004 : 421). Il existe cependant peu d'études qui mettent en relation l'organisation gestuelle de la Langue des Signes avec celle de la langue parlée. Nous espérons que cet atelier ouvrira de nouvelles perspectives de collaboration entre les gestualistes.

En effet, l'atelier repose sur un défi d'annotation. Les organisatrices, A. Braffort et L. Boutora, proposent deux extraits vidéos, construits de la même manière, l'un en Langue des Signes Française, l'autre en français parlé, où les locuteurs décrivent le même itinéraire. Afin de circonscrire les annotations, le thème fixé est l'annotation des pointages et tous les éléments qui peuvent être en lien avec ces gestes. McNeill (2005 : 12) définit les pointages comme des gestes codant la directionnalité de manière plus ou moins conventionnelle. A cette définition, nous ajoutons que la directionnalité du geste doit participer au sens du message linguistique (car tout geste dynamique possède une directionnalité, mais celle-ci n'est pas systématiquement significative sur le plan linguistique).

Ce type d'atelier thématique est particulièrement intéressant, non seulement parce qu'il permettra peut-être de nouvelles collaborations entre les gestualistes spécialisés dans la Langue des Signes et les gestualistes travaillant sur la langue parlée, mais aussi parce qu'il permet de faire un point sur le type d'annotation dont on peut disposer actuellement et de comparer différentes expertises dans le domaine de la multimodalité qui manque encore de données aujourd'hui, même si un énorme effort a été produit dans l'annotation de corpus. C'est dans cette optique que nous avons décidé de participer à l'atelier et de relever le défi. Malheureusement, il ne nous a pas été possible de travailler sur les deux vidéos, car l'annotation de la Langue des Signes suppose de maîtriser un temps soit peu cette langue. Une annotation purement descriptive n'est pas possible car elle exige de pouvoir au moins déterminer des unités gestuelles dans une chaîne gestuelle ininterrompue. Nous avons néanmoins annoté le corpus de français parlé et espérons que cette annotation pourra permettre des discussions intéressantes lors de l'atelier et par la suite. L'annotation de l'extrait a été effectuée par une annotatrice expérimentée (environ 25 heures d'annotation en comptant la modification et la création des schémas d'encodage), sur les logiciels Praat et Anvil, à partir du schéma d'encodage mis en place dans le projet ANR OTIM (Bertrand et al., 2008). Il s'agit d'un schéma qui permet l'annotation (toujours en cours) du corpus CID, un corpus de français conversationnel. Le schéma est relativement général ce qui permet de l'utiliser avec d'autres types de corpus et a vocation à être partagé avec le reste de la communauté (c'est d'ailleurs pourquoi il est intégralement conçu en anglais). Nous le présentons donc dans cet article, ainsi que les légères modifications qui y ont été apportées dans les parties 2 et 3, qui décrivent les annotations réalisées sur la parole et celles qui ont été effectuées sur la modalité visuelle respectivement. Enfin, dans la partie 4, nous montrons comment nous établissons des liens entre les différentes modalités du langage.

2 Transcription et annotation de la parole

Toutes les annotations verbales ont été réalisées par un annotateur directement sous Praat (Boersma, Weenink, 2009) à partir du fichier Wave qui nous était fourni. Dans un premier temps, une transcription manuelle orthographique a été effectuée sur le fichier en utilisant une piste séparée par locuteur, et en adoptant une segmentation en unités intonatives supérieures (IP dans la théorie de Selkirk, 1978)¹. On peut définir une unité intonative supérieure comme une unité prosodique présentant un contour intonatif sans rupture et qui comporte une prééminence accentuelle majeure de type 'nucléaire'. Le choix de cette unité nous semble plus adapté à l'oral que des unités syntaxiques de type proposition ou phrase. Il est souvent très difficile de définir des phrases syntaxiques à l'oral à cause des nombreux emboîtements de propositions et des faux-départs. Un découpage en propositions syntaxiques est souvent possible, mais il pose le problème des nombreux marqueurs de discours employés à l'oral et que l'on rattache à une proposition plutôt qu'à une autre sur la base de l'écoute (et donc de l'intonation).

¹Une segmentation en 'groupes' est nécessaire au logiciel EasyAlign (Goldman, 2010), car il permet au logiciel d'ignorer les passages problématiques pour la transcription phonétique (présence de bruit) et de continuer à transcrire les groupes qui suivent.

Nous nous sommes ensuite concentrée sur la piste correspondant au locuteur principal (l'interlocutrice produisant essentiellement du feedback). Il va de soi que dans un extrait plus long, la même procédure aurait été adoptée pour les deux participants de manière symétrique, alors que nous avons ici limité les annotations à tout phénomène pouvant être lié aux pointages, ce qui ne justifiait pas une transcription fine de la parole de l'interlocutrice.

A partir de cette première transcription pour le locuteur principal (Piste 0), nous avons lancé le script d'EasyAlign (Goldman, 2010), un outil d'aide à l'annotation phonétique. EasyAlign a généré 4 nouvelles pistes sur le fichier Textgrid de Praat :

- Piste 1 : transcription des groupes intonatifs initiaux en SAMPA² (Wells, 1997) ;
- Piste 2 : transcription en mots orthographiques alignés sur le signal ;
- Piste 3 : transcription des syllabes en SAMPA alignées sur le signal ;
- Piste 4 : transcription en phonèmes (SAMPA) alignés sur le signal.

Chacune de ces pistes a été corrigée manuellement (correction de caractères et/ou d'alignement) après quoi, dans une Piste 5, nous avons annoté, toujours manuellement, les différents types de préminences sur les syllabes : S, pour un accent lexical secondaire ; P, pour un accent lexical primaire ; N pour un accent de groupe intonatif nucléaire.

Les pistes 6 et 7 comportent la transcription en groupes intonatifs du feedback de l'interlocutrice, ainsi qu'une transcription en mots en caractères orthographiques alignée sur le signal. Pour la comparaison avec les gestes, nous avons exporté dans Anvil (Kipp, 2001) exclusivement les pistes 2 & 7 (transcription en mots pour chaque interlocuteur) et la piste 5 (accents), alors que les autres pistes ont été exportées dans un fichier Anvil distinct, à l'exception de la piste 1. Cela ne signifie pas que ces autres pistes sont inutiles, mais plutôt que nous ne les avons pas utilisées directement dans ce travail sur les gestes de pointage (mais elles ont servi indirectement dans la mesure où les accents, par exemple, ont comme support les syllabes qui, elles, sont basées sur un alignement phonétique).

3 Annotation de la gestualité

L'intégralité des gestes a ensuite été annotée sous Anvil (Kipp, 2001) qui permet une visualisation image par image. La méthodologie générale d'annotation de chaque unité gestuelle a été la même dans le Corpus DEGELS1 (Braffort, Boutora, 2011) que dans le corpus CID, dans le cadre d'OTIM. Toutes les annotations gestuelles se font sur la base de vocabulaires contrôlés, ne comprenant que des chaînes de caractères ASCII. La délimitation des unités se fait entre deux images, juste avant la perception d'un mouvement pour le début du geste (Gesture onset) et juste après le dernier mouvement pour la fin du geste. Ainsi, sur la figure 1, où un geste enchaîne directement sur la tenue d'un autre geste, on voit que la main droite du locuteur ne bouge pas sur les deux premières vignettes alors qu'une légère modification de la position de la main est perceptible sur la troisième vignette. Par conséquent, le début du geste se situe entre la vignette 2 et la vignette 3. Le principe est le même pour la fin de l'unité gestuelle, et ceci quel que soit le type de geste annoté (geste manuel, direction du regard, haussement des sourcils, etc.).

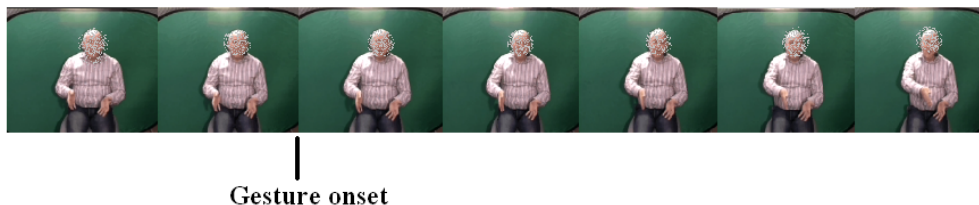


FIG. 1 – Début d'un geste dans le corpus DEGELS1, Braffort, Boutora, 2011)

²SAMPA est une convention de conversion des caractères de l'API en caractères lisibles par l'ordinateur. Cette convention est utile pour la mise en relation de ces informations avec les informations gestuelles car Anvil, le logiciel utilisé dans l'annotation des gestes, ne lit pas les polices phonétiques.

3.1 Schéma d'encodage pour les gestes manuels

Le schéma d'encodage adopté pour les gestes manuels est une adaptation du schéma d'encodage Mumin (Allwood et al., 2005) et a été légèrement modifié depuis la présentation du schéma d'encodage OTIM³ dans Bertrand et al. (2008), utilisé pour l'annotation du corpus CID. La méthodologie a également été sensiblement différente : en effet, le CID étant un corpus de trois heures de parole conversationnelle, nous avons été contraints à certains choix pour gagner du temps dans une annotation qui est extrêmement longue. Ainsi, lorsqu'un geste est réalisé avec deux mains parfaitement symétriques, nous n'annotons qu'une seule main (la main gauche par défaut) sachant qu'une piste "Symétrie" indique que le geste est réalisé avec les deux mains. Ici, dans la mesure où l'extrait est très court, nous avons annoté la configuration de chaque main. En revanche, comme pour le CID, nous avons annoté dans des fichiers séparés les gestes réalisés par chaque participant. Nous avons décidé d'annoter systématiquement tous les gestes manuels, et non exclusivement les pointages, car, comme nous le verrons plus loin, certains pointages font référence à un autre geste manuel réalisé plus tôt dans l'échange.

Le schéma d'encodage OTIM comporte deux premières pistes "GestureUnit" et "Symmetry". C'est essentiellement sur ces deux pistes qu'ont porté les modifications réalisées dans cette expérience. La piste "GestureUnit" permet d'annoter comme une seule unité gestuelle des gestes qui s'enchaînent sans que les mains retournent à la position de repos. Nous l'avons jugé inutile ici et l'avons supprimé du schéma d'encodage. En revanche, nous avons conservé la piste "Symmetry", qui dans OTIM permet d'annoter le fait que les gestes sont réalisés avec une seule main ou deux mains, symétriques ou non. Un problème concernant cette piste a été observé lors de l'annotation du CID et nous l'avons corrigé pour DEGELS1 : une seule piste "Symmetry" ne permet pas d'annoter les chevauchements de gestes manuels. Pourtant, un geste peut très bien être réalisé avec deux mains symétriques ou non, puis un deuxième geste à une seule main est produit alors que l'autre main reste en position de tenue du geste initial. On peut aussi envisager le cas où un premier geste est réalisé avec une main, et où un deuxième geste est réalisé avec l'autre main qui commence avant la rétraction totale du premier geste. Enfin, et ceci est particulièrement intéressant dans le cas des pointages comme nous le verrons plus loin, une main peut servir de référent à l'autre main et il n'est pas du tout certain que dans ce cas, il n'y ait qu'une seule unité gestuelle. En résumé, pour corriger les problèmes que pose l'annotation des chevauchements, nous avons ajouté une piste "Overlaps" dont les valeurs sont les mêmes que pour la piste "Symmetry". Ces pistes 1 et 2, qui sont donc liées au sens linguistique du geste, sont primaires et totalement indépendantes l'une de l'autre, ainsi que des pistes suivantes. Leur intérêt est de pouvoir annoter les mouvements de main perçus comme une seule unité gestuelle, y compris lorsqu'ils sont produits en chevauchement.

Les autres pistes sont divisées en deux groupes : un groupe pour la main gauche (LeftHand, 10 pistes) et un groupe pour la main droite (RightHand, 10 pistes identiques à celles de la main gauche). Chaque groupe contient donc les pistes suivantes :

- PISTE 3, PHASE : selon Kendon, 1980 (cité dans McNeill, 1992), chaque geste peut se décomposer en plusieurs phases comme la préparation (mise en place des articulateurs), puis vient une phase de réalisation (le geste lui-même), qui peut être précédée ou suivie d'une tenue (les articulateurs font une pause). Certains gestes sont statiques et non pas dynamiques, et comptent ainsi une préparation et une tenue, mais pas de phase de réalisation. Enfin, soit un deuxième geste s'enchaîne directement sur le premier, soit il y a une phase de rétraction (retour des articulateurs à une position de repos). Le geste peut également inclure un battement et s'achever par un léger rebond de la main. Cette piste est une piste primaire dans notre annotation et est liée au sens du geste de manière phonologique plutôt que discursive : en effet, il s'agit d'identifier quelle partie du geste est significative (phase de réalisation) et non pas quel est le lien entre le geste et la parole.
- PISTE 4, PHRASE : cette piste dépendante de la piste 3 permet de noter la/les dimension(s) gestuelle(s). Elle permet donc d'identifier la fonction linguistique du geste dans sa relation à la parole. Parmi les dimensions gestuelles, on compte les gestes iconiques, déictiques, métaphoriques, les battements, les adaptateurs (gestes non-verbaux mais que l'on a choisi de conserver dans OTIM car ils révèlent l'état de l'interaction et sont corrélés aux tours de parole), les emblèmes et les "butterworths" (gestes désordonnés que l'on rencontre lors de la recherche lexicale notamment). Certains gestes possèdent une dimension gestuelle dominante, mais aussi une dimension gestuelle secondaire qui est notée de manière booléenne dans cette piste. Ainsi, un geste iconique peut être réalisé dans un plan de l'espace significatif et ainsi acquérir une dimension déictique. Il est donc important de pouvoir noter les deux valeurs.

³Il s'agit d'un schéma d'encodage, plus que d'un schéma d'annotation, car dans le projet OTIM, nous nous sommes intéressés à la mise en relation d'informations encodées différemment selon les outils. Par exemple, PRAAT encode les données de manière non hiérarchisées, alors qu'ANVIL les encode au format XML.

- PISTE 5, LEXICON : cette piste dépendante de la piste 4 a pour but de préciser la valeur de certains gestes (iconiques, emblèmes, métaphoriques et déictiques) et est liée à la fois au sens et à la forme des gestes. En effet, le geste qui consiste à toucher sa montre peut être perçu comme un emblème lié à l’heure et au temps qui passe, ou comme un déictique s’il est lié à l’objet “montre” (que l’on rencontrerait dans un contexte où le locuteur mentionne par exemple qu’il a acheté une montre), et c’est le sens qui est ici décrit dans les deux cas. En revanche, certains métaphoriques sont par exemple notés “conduit-flip”, à savoir un retournement de la main et c’est le type de mouvement qui est noté ici, pas le sens. Il serait trop long de passer en revue l’ensemble des valeurs retenues pour le schéma d’encodage. Pour les déictiques qui nous intéressent plus particulièrement, il est possible de noter si le geste pointe vers soi, vers l’interlocuteur, vers un objet, ou vers un point dans l’espace. Pour cette annotation précise, nous avons ajouté vers une partie du corps. L’ensemble des valeurs pour les autres dimensions sont comprises dans un vocabulaire contrôlé qui évolue en fonction de l’annotation du corpus CID (et notamment pour les gestes emblématiques qui ne sont pas tous listés dans le schéma d’encodage d’origine). Nous espérons, en fin de projet ANR, pouvoir fournir un lexique assez complet des différentes valeurs rencontrées.
- PISTE 6, HAND SHAPE : cette piste décrit la configuration des doigts de la main, donc uniquement la forme du geste. Les valeurs de base sont reprises de Friedman (1977) pour la description de la Langue des Signes Américaine, dans un tableau cité dans McNeill (1992 : 87-88). Ces valeurs ont été complétées au cours de l’annotation du CID, mais aucune valeur supplémentaire n’a été ajoutée pour DEGELS1. Contrairement à la piste 5, cette piste est dépendante de la piste 3 “Phase”, car pour certains gestes, ce qui est significatif n’est pas tant la configuration manuelle à l’apogée du geste, mais le changement de configuration manuelle. Ainsi, pour un geste emblématique signifiant “lâcher quelque chose”, il est important de noter que le poing est fermé pendant la phase de préparation (configuration “A” dans notre annotation) et que la main est ouverte en fin d’apogée gestuelle (configuration “5” dans notre annotation). Seules les configurations perçues comme contrastives ont été distinguées dans la préparation du geste par rapport à la phase de réalisation. Enfin, le schéma d’encodage OTIM prévoit une valeur booléenne “lax”, très utile pour les gestes produits en parole spontanée. En effet, les configurations données dans McNeill sont souvent réalisées de manière très relâchée dans les gestes co-verbaux. Dans ce cas, nous notons la forme canonique la plus proche à laquelle nous associons la valeur “lax”.
- PISTE 7, HAND ORIENTATION : pour la même raison que pour la piste précédente, cette piste est dépendante de la piste 3 “Phase” et permet d’annoter l’orientation de la paume de la main en une liste fermée (vers le haut, le bas, vers soi, vers l’autre, vers l’espace, sur la tranche à l’intérieur/l’extérieur). Là aussi, il s’agit uniquement de la forme du geste.
- PISTE 8, GESTURE SPACE : cette piste dépendante de la piste 4 “Phrase”, décrivant également la forme du geste, permet d’annoter deux dimensions de l’espace de gestualisation par rapport au corps du locuteur pour la globalité du geste, données dans McNeill (1992 : 89, cf. Figure 3), à savoir la région (“center center, center, periphery, extreme periphery”) et les coordonnées (“left, right, upper, lower, upper right, etc.”, ainsi que des valeurs permettant d’annoter les gestes réalisés à deux mains en miroir “upper left-right, etc.”). Le point de référence fixe pour l’annotation est la position du poignet lors de l’extension maximale du geste.
- PISTE 9, CONTACT : dépendante de la piste 4 “Phrase”, cette piste décrivant la forme du geste permet d’annoter le type de contact s’il y en a un (“partie du corps, interlocuteur, objet”) et de préciser la partie du corps touchée (19 parties du corps sont listées). Cette piste permet l’annotation plus précise des adaptateurs, mais pas uniquement dans la mesure où certains gestes comme les applaudissements, par exemple, impliquent un contact. Dans le cas de gestes réalisés avec les deux mains, si les deux mains sont actives, le contact est noté pour chaque main, mais si une seule main est active, le point de contact n’est noté que pour cette main. Là encore, le contact est noté pour la globalité du geste car il ne nous a pas semblé significatif de noter quelle phase exactement comporte le contact. De plus, dans un geste qui consisterait à frapper dans ses mains par exemple, le contact n’a lieu qu’à la fin de la phase de réalisation et non sur toute la phase.
- GROUPE : MOVEMENT QUALITY (les 3 pistes suivantes décrivent uniquement la forme du geste)
- PISTE 10, MOVEMENT TRAJECTORY : dépendante de la piste 3 “Phase”, cette piste permet d’annoter la trajectoire du mouvement (“upper, lower, left, right, upper right, towards self, away from self, sideways, etc.”). Toutes les valeurs sont booléennes car il est possible qu’un geste soit réalisé à la fois vers son propre corps et vers la gauche, par exemple. Pour les gestes statiques, la trajectoire n’est notée que pour la phase de préparation et la rétraction s’ils en ont une.
- PISTE 11, MOVEMENT VELOCITY : également dépendante de la piste 3 “Phase”, cette piste permet de noter si un geste est perçu comme “rapide”, “normal” ou “lent”. Cette perception est pour l’instant subjective et nous n’avons pas encore déterminé de critère objectif pour l’annotation. La vitesse est annotée sur les phases de

préparation et de rétraction dans le cas des gestes statiques.

- PISTE 12, MOVEMENT AMPLITUDE : cette piste qui compte trois valeurs (“small, medium, large”) est dépendante de la piste 4 “Phrase” et s’appuie sur les annotations de la piste 8 “Gesture Space”. En effet, il a été décidé que si le poignet, dans la réalisation du geste, ne franchit aucune frontière de l’espace gestuel, la geste est petit. S’il franchit une frontière, le geste est moyen. Et enfin, s’il franchit plus d’une frontière, le geste est large. Se pose cependant le problème de la dimension : le graphique donné par McNeill pour l’espace gestuel s’inscrit dans deux dimensions (“haut-bas, droite-gauche”), mais exclut la troisième dimension de la profondeur. Donc, un geste qui s’éloigne du corps à l’horizontale est noté comme “petit” puisqu’aucune zone gestuelle n’est franchie, ce qui ne correspond pas nécessairement à l’extension du bras. Les membres du projet OTIM réfléchissent actuellement à cette question qui devrait donc être résolue prochainement mais ne l’était pas encore pour l’annotation de DEGELS1.

L’ensemble des pistes est résumé dans la figure 2.

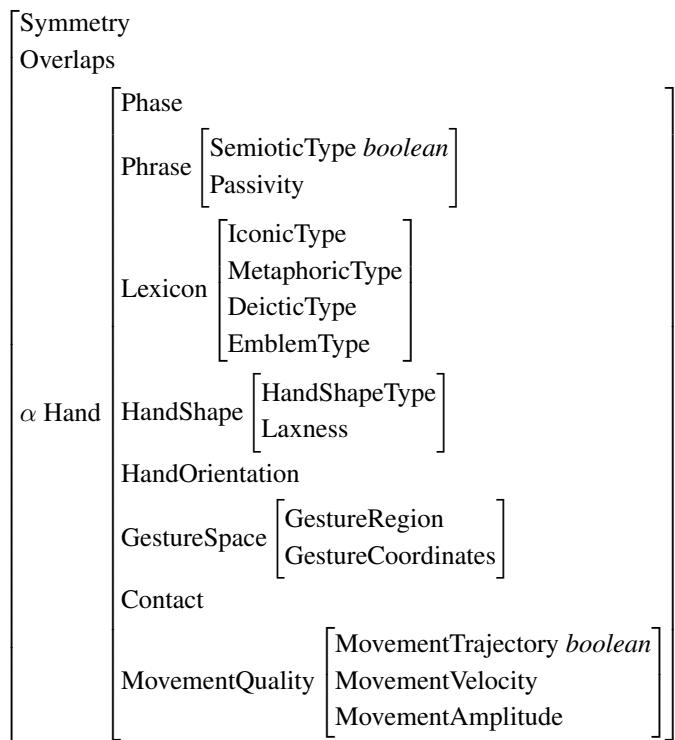


FIG. 2 – Schéma d’encodage des gestes manuels utilisé dans DEGELS1 (adapté d’OTIM)

3.2 Les gestes manuels dans DEGELS1

Sur le corpus DEGELS1, 68 gestes manuels ont été annotés : 1 geste pour la locutrice située à gauche sur la vidéo (qui se trouve être un pointage) et 67 gestes pour le locuteur situé à droite sur l’image, dont nous allons maintenant faire la description. Parmi ces 67 gestes, deux seulement ont été produits en chevauchement ; 55 ont été réalisés avec la main droite, 1 avec la main gauche et 11 avec les deux mains. Ils se répartissent dans les dimensions suivantes, le Type 1 correspondant à la dimension gestuelle principale (nombre indiqué dans la colonne de gauche) et le type 2 à la dimension secondaire pour certains gestes bi-dimensionnels :

Les chiffres donnés pour le type dominant dans le tableau 1 peuvent être comparés à ceux obtenus pour les gestes manuels dans le corpus CID. Pour la comparaison, nous n’avons pas pris en compte les adaptateurs, car ils sont réalisés essentiellement lorsque le participant écoute l’interlocuteur. Ainsi, comme il apparaît dans le tableau 2 ci-dessous, le pourcentage de métaphoriques est bien plus élevé dans le corpus CID que dans DEGELS1 avec 32.2 % contre seulement 14.9 % pour DEGELS1, qui compte en revanche un pourcentage beaucoup plus élevé de

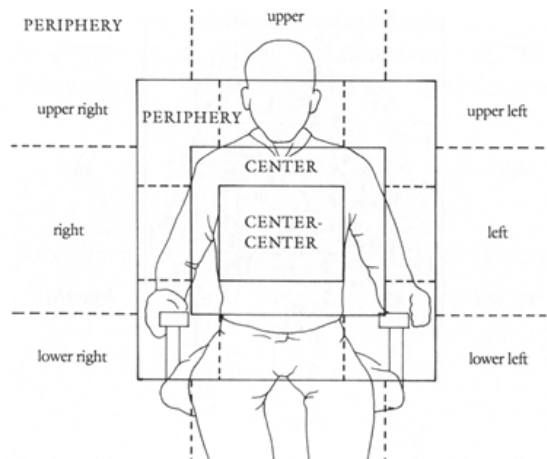


FIG. 3 – Espace de gestualisation (repris de McNeill, 1992 : 89)

Nombre	Type 1	Type 2
1	Battement	
39	Déictique	6 iconiques
4	Enblème	1 iconique
13	Iconique	1 battement, 5 déictiques, 2 emblèmes
10	Métaphorique	1 iconique

TAB. 1 – Nombre de gestes pour Loc D dans DEGELS1 par type de geste

déictiques (58.2 % contre 12.3 % pour le CID). Le type de tâche verbale donnée aux locuteurs a donc bien une influence sur leur production gestuelle.

Type	DEGELS1	CID
Battements	1.4	13.9
Déictiques	58.2	12.3
Emblèmes	5.9	13.4
Iconiques	19.4	25.1
Métaphoriques	14.9	32.2

TAB. 2 – Pourcentages de types gestuels dans DEGELS1 et dans le CID

Les déictiques dans DEGELS1 (et c’est d’ailleurs aussi le cas dans le corpus CID), qui accompagnent la parole, sont produits avec non moins de 10 configurations manuelles différentes, les plus fréquentes étant “B” (17 occ.), que l’on note “relâchée” car le pouce n’est en général pas replié, “B spread” (9 occ.) et “G” (8 occ.). Ces trois configurations sont données dans la figure 4 ci-dessous :

Les autres annotations sur les gestes manuels sont moins spécifiquement liées aux pointages, mais il est important de les avoir, à la fois pour permettre des comparaisons avec les pointages que l’on rencontre dans les langues signées, mais aussi afin de pouvoir rapidement interpréter les gestes en discours, ce que nous présenterons dans la section 4 sur l’affiliation.

3.3 Regard : schéma d’encodage et annotations sur DEGELS1

Le schéma d’encodage utilisé ici repose sur une simplification du schéma d’encodage créé pour OTIM, et est présenté dans la figure 5. Dans un premier temps, nous pensions que tous les types d’information pouvaient s’avérer utiles pour des mises en relation avec les pointages. Cependant, seul le regard s’est révélé avoir des liens



FIG. 4 – Configurations manuelles les plus fréquentes pour les pointages dans DEGELS1

avec les gestes produits.

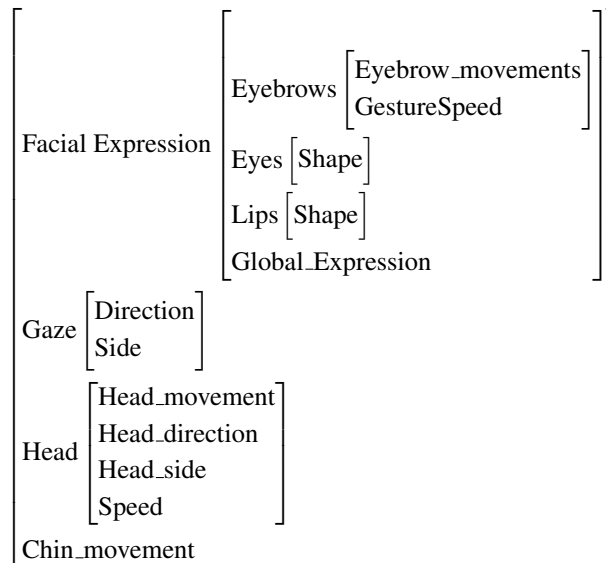


FIG. 5 – Schéma d’encodage pour la tête (adaptation du schéma d’encodage OTIM) utilisé dans DEGELS1

Nous avons donc annoté la direction du regard du locuteur principal. La méthodologie consiste à ne pas annoter lorsque le locuteur regarde l’interlocuteur assis en face de lui (dans la mesure où la direction du regard est alors dans l’alignement de la tête en position de repos). Tous les regards dans une direction autre que cette position de repos ont été annotés et se répartissent comme indiqué dans le tableau 3.

Regard	21 occ.
Sideways	4
Towards Body Part	12
Up	5

TAB. 3 – Direction du regard du locuteur principal dans DEGELS1

4 Affiliation

4.1 Affiliation lexicale

L’intérêt d’une annotation multimodale de la langue parlée est de pouvoir mettre en relation des informations provenant de modalités différentes. Ainsi, on conçoit que certains gestes sont redondants avec la parole (alors que d’autres sont en relation de complémentarité avec la parole) et ces gestes peuvent donc être mis en relation avec des unités verbales, que l’on nomme “affiliés lexicaux”. McNeill (2005) semble concevoir l’affiliation de manière

très large comme toute unité verbale produite de manière concomitante avec les unités gestuelles, alors que Schegloff (1984) décrit l’affilié lexical comme “the word or words deemed to correspond most closely to a gesture in meaning”. Il y a donc selon lui une relation sémantique entre le geste et son affilié lexical. Nous allons plus loin en pensant que tout geste ne possède pas nécessairement un affilié lexical. En effet, certains gestes (notamment des métaphoriques) constituent des “modalisateurs d’énoncés”, au sens où ils apportent une modalité à l’énoncé. Dans ce cas, les gestes “n’illustrent” pas la parole et ne peuvent être mis en relation directe avec aucun mot du discours. D’autres gestes en revanche, majoritairement les iconiques et les pointages, ont la plupart du temps un sémantisme proche d’un mot du discours, même si la redondance n’est jamais parfaite puisque la simple production d’un geste apporte nécessairement des informations visuelles que la parole n’apporte pas.

Par exemple, dans la figure 6, l’image (a) correspond à l’apogée du geste (extension maximale du geste) produit alors que le locuteur dit “ la statue du pouce”. Le fait de lever le pouce permet au locuteur d’illustrer un pouce (geste iconique) tout en fournissant des informations spatiales sur la position du pouce (geste déictique). Ce geste bi-dimensionnel peut donc être mis en relation directe avec le mot “pouce” du discours. En revanche, on ne peut établir de relation directe avec “statue” sauf si l’on considère que “statue du pouce” forme une seule unité lexicale, ce qui n’a pas été notre choix ici (d’ailleurs, c’est un autre geste iconique qui est réalisé pour “statue” et qui accompagne l’expression “statue du David” (voir plus bas). L’image (b) correspond à l’apogée du geste produit alors que le locuteur dit “Y a un rond-point”. Là-encore, la forme sphérique de la configuration des deux mains permet de déterminer que ce geste a pour affilié lexical “rond-point”.

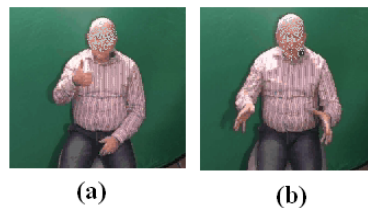


FIG. 6 – Deux gestes pouvant être mis en relation avec un affilié lexical dans DEGELS1

Cependant, la notion d’affiliation va bien plus loin que la simple affiliation lexicale dans la parole. Ainsi, dans la partie 2, nous avons présenté les annotations prosodiques réalisées sur le corpus. Or, il est tout à fait possible de mettre ces informations en relation avec des informations tirées de la modalité visuelle qui s’organise elle aussi en une “prosodie gestuelle” (Gonzales et al., 2007). Il faut cependant prendre garde de mettre en relation des unités comparables dans les deux modalités.

4.2 Affiliation prosodique

Dans sa thèse, Loehr (2004) met en relation trois types d’unités entre les modalités orales et visuelles :

1. L’unité gestuelle dans son intégralité avec le groupe intonatif de niveau supérieur ;
2. La phase dynamique du geste (“gesture stroke”) avec l’affilié lexical (mot) ;
3. L’apogée du geste avec l’accent lexical.

Les résultats qu’il obtient vont dans le sens d’une synchronisation entre geste et voix, mais plusieurs questions se posent : l’étude repose sur tous les types de geste, or, comme nous l’avons vu en 4.1, certains gestes ne sont pas du tout en relation de redondance par rapport à la parole, par conséquent, certains choix d’affiliés lexicaux ne sont-ils pas forcés ? Pour la mise en relation de l’apogée du geste avec un accent lexical, Loehr a choisi d’adopter une fourchette temporelle de plus ou moins 275 ms, ce qui semble très large pour noter des synchronisations au niveau de la syllabe et risque de retourner des synchronisations systématiques, là où il n’y en a peut-être pas. Enfin, Loehr ne distingue pas les différents types d’accents.

Dans deux études (Ferré, 2010a et b), nous avons reproduit les niveaux de comparaison 1 et 2 de Loehr sur le corpus CID pour les gestes iconiques en relation de redondance avec la parole. Nous avons ainsi mis en relation le début de la phase de préparation avec le début du groupe intonatif et le début de la phase de réalisation avec le début de l’affilié lexical. Nous avons aussi mesuré la durée de l’affilié lexical et du groupe intonatif d’une part,

que nous avons comparée avec la durée du geste dans son ensemble et de la phase de réalisation d'autre part. Sur le plan de la durée, il est apparu que le geste dans son ensemble est plus long que le groupe intonatif comportant l'affilié attaché au geste. De la même manière, la phase dynamique du geste ("stroke") est plus longue que l'affilié lexical auquel elle correspond. Sur le plan de l'alignement temporel, il est apparu que le geste dans son ensemble commence avant le groupe intonatif et se termine légèrement après le groupe intonatif. De même, la phase dynamique du geste commence avant l'affilié lexical auquel elle correspond et se termine légèrement après celui-ci. Nous pouvons ainsi proposer l'organisation temporelle suivante :

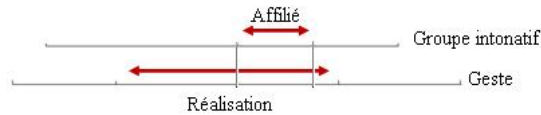


FIG. 7 – Alignement geste / parole dans le corpus CID

Ce type de comparaison geste/parole est également très possible dans le corpus DEGELS1, puisqu'il est également possible de déterminer un affilié lexical précis pour un grand nombre de pointages. De plus, la comparaison entre la position de l'accent lexical et l'apogée du geste n'est pas possible à l'heure actuelle sur le CID puisque l'apogée n'a pas été annotée, alors qu'elle l'a été sur DEGELS1 comme nous le verrons plus loin.

4.3 Affiliation gestuelle

Le point a été relativement peu évoqué dans la littérature portant sur la gestualité co-verbale dans les langues orales, mais le corpus DEGELS1 fait particulièrement bien ressortir le fait que l'affiliation peut être strictement gestuelle. Cette affiliation peut aussi être de trois natures différentes :

1. Une unité gestuelle produite avec une partie du corps peut être affiliée à une autre unité gestuelle produite avec une autre partie du corps (tête vs. main(s), regard vs. main(s)) ;
2. Une unité gestuelle produite avec les mains peut être affiliée à une autre unité gestuelle produite antérieurement avec les mains (soit par un autre locuteur, soit par le même locuteur) ;
3. Une unité gestuelle réalisée avec une seule main peut être affiliée à une autre unité gestuelle produite avec l'autre main de manière simultanée par le locuteur.

Dans DEGELS1, nous pouvons illustrer le point 1 par l'exemple suivant : tout en disant "son moulage de pouce là", le locuteur tient le pouce de sa main droite avec sa main gauche. Ce faisant, il regarde son geste. Ce regard orienté vers le geste permet d'activer le geste et de le distinguer d'un simple adaptateur. Il y a donc un lien d'affiliation entre le regard et le geste.

Deux exemples permettent d'illustrer le point 2. Au cours de l'échange, la locutrice demande confirmation de la direction en disant "ben je longe la côte mhmh", ce que confirme immédiatement le locuteur principal en répondant "donc vous longez". Lors de sa demande de confirmation, la locutrice produit un geste de la main droite vers la droite et ce geste est repris avec un léger chevauchement par le locuteur lors de la confirmation. Les deux gestes apparaissent dans la figure 8.

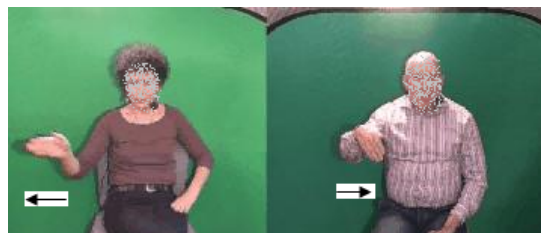


FIG. 8 – Affiliation de deux gestes manuels produits par deux locuteurs

Dans la figure 9, cette fois, c'est le même locuteur qui produit trois gestes. Le discours du locuteur est : “^(a)[y a un rond-point euh il s'appelle l'Es] ^(b)[cale Borély c'est un rond-point où y a tout plein de restau] rants (...) ^(c)[euh ça s'ap]pelle l'Escale Borély”. Il apparaît dans cette séquence que le locuteur pose tout d'abord l'objet “rond-point” par un geste des deux mains (image (a)), puis, avec la main droite en configuration “B-spread”, il décrit un cercle qui reprend le geste initial (image (b)), et enfin, plus tard dans l'échange, avec l'index de la main droite en configuration “G”, il décrit à nouveau le même cercle (image (c)). Or, dans ce dernier geste qui est lui aussi affilié au premier geste donné dans l'image (a), il n'est plus du tout fait mention du mot “rond-point” dans le discours. Ce geste ne peut se comprendre que par référence aux gestes qui précèdent et auquel il est affilié.

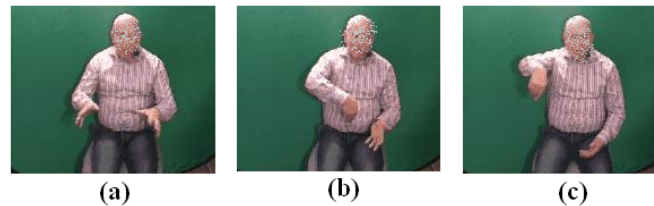


FIG. 9 – Affiliation d'une suite de trois gestes manuels produits par un locuteur

Le point 3 est illustré par la figure 10 : il est produit alors que le locuteur dit “et en fait vous avez longé l'hippodrome” et la main droite du locuteur sert de référent “l'hippodrome”, alors que la main gauche reproduit dans son mouvement l'action de “longer”. Les deux mains n'ont donc pas le même affilié lexical et l'on peut donc raisonnablement compter les mouvements comme deux unités gestuelles et non pas une.



FIG. 10 – Affiliation de deux gestes manuels produits simultanément par un locuteur

4.4 Création du schéma d'encodage et annotation des affiliations

Dans le projet OTIM cité précédemment, aucun schéma d'encodage ne permet à l'heure actuelle d'annoter des affiliations aussi complexes que celles que nous venons de décrire, et les affiliations n'ont donc pas été annotées de manière systématique sur le corpus CID. Pour DEGELS1, nous avons en revanche trouvé intéressant la réflexion autour de la référence construite grâce aux pointages et avons créé un schéma d'encodage permettant d'exprimer des liens divers entre les annotations réalisées. Ce schéma reprend donc certaines des pistes créées pour les autres niveaux de l'analyse en les incluant comme pistes primaires, indépendantes les unes des autres. Le schéma idéal reprendrait la quasi totalité des pistes des autres niveaux, mais nous pensons que cela charge inutilement la lecture. Aux pistes retenues, nous avons ajouté deux nouvelles pistes permettant de noter l'apogée des gestes (deux pistes sont nécessaire pour les gestes produits en chevauchement). Nous obtenons donc le schéma d'encodage donné dans la figure 11, où “D” signifie “locuteur à droite de l'image” et “G”, “locuteur à gauche de l'image”.

Chaque piste reprend donc les attributs qu'elle avait à l'origine et pour toutes les pistes (à l'exception des pistes “Words” et “Stress”), nous avons ajouté un attribut “Links” permettant d'exprimer directement dans Anvil les affiliations, grâce à la fonction “MultiLink” du logiciel. Nous avons ensuite créé un fichier d'annotation dans lequel nous avons copié les annotations de chaque piste à partir des fichiers précédents. Enfin, nous avons annoté l'apogée des gestes et une fois les annotations terminées, avons pour chaque unité gestuelle déterminé le ou les affilié(s). Ainsi, pour reprendre un exemple cité dans la section 4.3, la figure 12 montre l'étiquette correspondant à la figure 8. Les affiliés notés pour le geste produit par le locuteur D (encadré bleu) sont d'une part le geste et

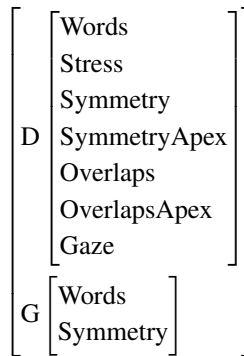


FIG. 11 – Schéma d’encodage pour l’annotation de l’affiliation dans DEGELS1

le mot “longe” produits juste avant par la locutrice G, et d’autre part, la propre reprise du verbe “longer” par le locuteur D. Les affiliés apparaissent encadrés en vert sur l’image.

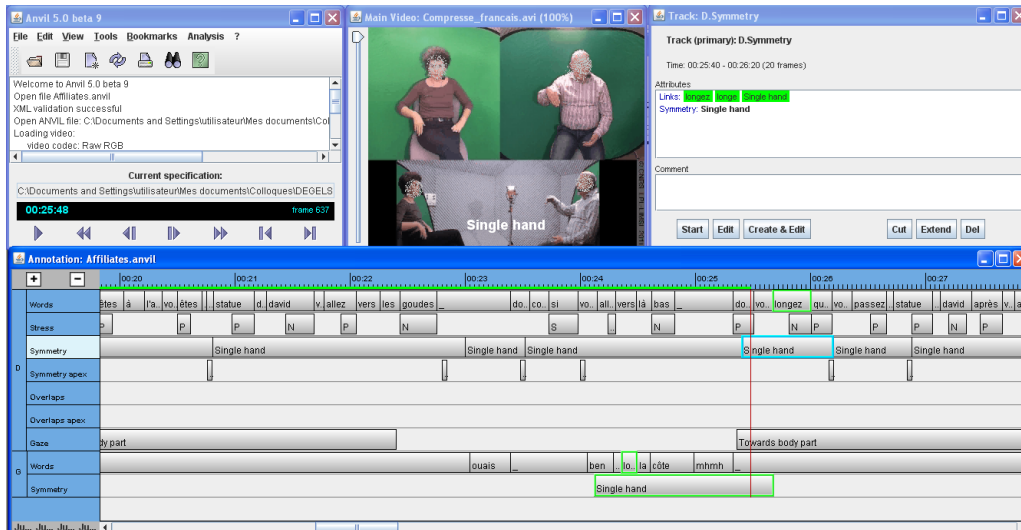


FIG. 12 – Fenêtre d’annotation d’Anvil pour le fichier “Affiliations”

Dans le corpus DEGELS1, le tableau 4 indique le nombre d’étiquettes qui ont été mises en relation avec un ou plusieurs affiliés. Le fait que la quasi totalité des étiquettes ait un ou plusieurs affiliés tient à la nature du corpus dans lequel le locuteur construit un itinéraire à partir de jalons situés à la fois dans le discours au niveau verbal et dans les gestes déjà réalisés et sur lesquels le locuteur appuie son discours.

Etiquette	Nb	Liens
Mains	73	54
Apogées	57	57
Regards	21	21

TAB. 4 – Nombre d’étiquettes pour lesquelles il a été possible de déterminer un ou plusieurs affiliés

Il serait vain de tenter une analyse quantitative complète sur 1.27 mn de signal, mais nous pouvons néanmoins donner un exemple du type de travail que rend possible cette annotation. Comme il est précisé dans le tableau 4, il a été possible de mettre en relation 57 apogées gestuelles avec les syllabes accentuées de leur affilié lexical, puis d’exporter la recherche sur cette affiliation dans Excel. Il apparaît que les apogées gestuelles correspondent à un accent nucléaire dans 46 cas et à un accent primaire dans 11 cas, ce qui signifie que dans la majorité des

cas, le geste accompagne le lexème le plus proéminent dans le groupe intonatif. On sait qu'en français, la syllabe accentuée est le plus souvent placée en fin de mot et que l'apogée du geste est le plus souvent la dernière image de la partie dynamique du geste. Or, cette apogée commence le plus souvent (dans 41 cas sur 57) après le début de la syllabe accentuée avec un décalage moyen de 1.05 s, ce qui confirme le graphique donné dans la figure 7.

L'annotation des affiliés gestuels concorde bien avec l'analyse discursive que l'on peut faire de l'extrait et qui est donnée en annexe en fin d'article. Cette analyse n'a pas été notée sous Anvil, ce qui nécessiterait une réflexion sur l'annotation des emboîtements, mais s'appuie sur la théorie des espaces focaux proposée par Grosz & Sidner (1986) : l'unité de discours large "description d'un itinéraire" se décompose en "segments de discours" (désormais SD) que constitue chaque étape de la description, segments de discours composés eux-mêmes d'actes de langage (notés AL dans l'annexe). Ainsi, dans cet extrait, le locuteur commence sa description en posant un premier point de repère, le rond-point de l'Escale Borély. Ce point de repère est accompagné d'un geste iconique sphérique comme nous l'avons vu plus haut (cf. figure 9a). A la suite de ce SD 1, le locuteur semble penser que ce 1er point de repère ne fait sans doute pas partie des connaissances partagées avec l'interlocuteur et il entame dans les SD 2 et 3 une négociation qui a pour but de s'assurer que l'interlocuteur visualise bien l'Escale Borély. En SD 2, il pose donc un second point de repère, la statue du David, avec un geste iconique qui trace un trait dans l'espace représentant une forme debout légèrement sur la droite du corps du locuteur. En DS 3, qui constitue une direction à partir de la statue, il effectue un pointage à l'endroit où il a réalisé le geste iconique du DS 2. En DS 4, il est désormais assuré que l'interlocuteur visualise le rond-point mentionné en DS 1 et effectue le geste de la figure 9b qui permet d'assurer un retour au DS 1. C'est ce que McNeill et al. (2001) nomment "gesture catchments", à savoir un geste qui présente des caractéristiques similaires à celles d'un ou plusieurs autres gestes produits antérieurement dans le discours, ce qui permet de révéler les liens cognitifs entre différents espaces focaux.

En DS 5, le locuteur reprend donc le point de repère "Escale Borély" (accompagné du geste produit dans la figure 9c) pour élaborer une direction à l'aide d'un autre point de repère, le rond-point du Pouce. Cet autre point de repère est lui aussi accompagné d'un geste iconique représenté dans la figure 6. Ce geste est à nouveau repris en DS 7 après la parenthèse qu'offre le DS 6, formant ainsi un autre "catchment" dans la gestualité du locuteur.

5 Conclusion

Dans cet article, nous avons présenté une adaptation du schéma d'encodage mis en place dans le projet ANR OTIM. Les parties de schéma utilisées permettent la transcription et l'annotation de la parole et d'événements prosodiques, ainsi que la transcription des unités gestuelles. Appliqué à DEGELS1, un corpus de description d'itinéraire en français parlé, le schéma d'encodage nous a permis de mettre en relation des informations relevant de trois modalités différentes : modalité verbale, modalité vocale et modalité visuelle, comme il apparaît dans la figure 13. Une énorme réflexion a été conduite autour de l'annotation des gestes manuels notamment, puisqu'il est possible pour chaque type de geste (suivant la typologie de McNeill) de noter aussi bien la configuration des doigts de la main et l'orientation de la main par rapport au corps du locuteur, que la direction, l'amplitude et la vélocité du mouvement. Ces annotations manuelles, réalisées avec le logiciel Anvil, sont certes très longues et le schéma d'encodage mérite encore quelques modifications, comme la prise en compte de trois dimensions par exemple, mais l'on peut espérer pouvoir constituer des bases de données à partir de corpus déjà annotés. Les études qui s'inscrivent dans la multimodalité enrichissent considérablement l'analyse linguistique, comme le montre cette expérience dont la thématique est le pointage.

En effet, nous avons montré que lors d'une interaction en vis-à-vis, le locuteur construit sa description d'itinéraire en prenant appui sur un certain nombre de gestes manuels. Il est vraisemblable que dans une situation où les deux participants ne se voient pas, le locuteur aurait construit sa description en utilisant d'autres procédés. En langue parlée, le fait de prendre appui sur la gestualité donne donc une organisation particulière au discours et il serait intéressant de comparer cette construction sur les plans syntaxiques et discursifs avec celle que l'on obtiendrait sans production de gestualité co-verbale. Le locuteur fixe des points de repère grâce aux gestes manuels, qu'il actualise avec le regard et construit son itinéraire autour de ces points. Nous avons vu que dans cette construction discursive particulière, il ne peut suffire d'établir des liens entre gestes manuels et contenu sémantique au niveau verbal (ce qui est certes déjà important), mais qu'il faut également mettre en relief les liens qui existent à l'intérieur d'une même modalité.

Il ne nous a malheureusement pas été possible de procéder à une annotation du corpus de Langue des Signes, car une telle annotation suppose d’avoir un accès direct au sens et de pouvoir segmenter la chaîne signée, mais nous espérons vivement que des points de comparaison existent entre les annotations réalisées en Langue des Signes et en langue parlée ce qui ferait certainement émerger des collaborations intéressantes.

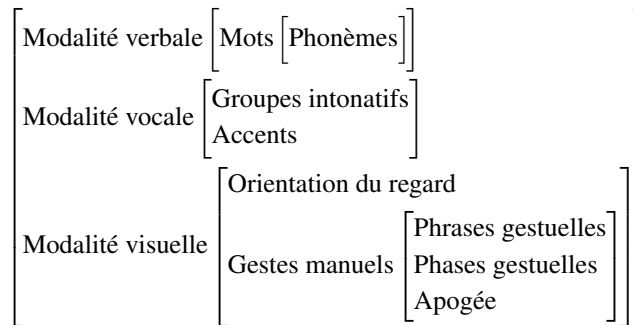


FIG. 13 – Unités annotées dans les différentes modalités

Remerciements

Nous tenons à remercier en particulier Leila Boutora et Annelies Braffort, qui, en organisant l’Atelier DEGELS, rendent possible une véritable discussion autour de l’annotation de la gestualité. Nous souhaitons remercier aussi les membres participant au projet ANR OTIM (ANR BLAN08-2_349062) qui ont contribué à l’élaboration de la méthodologie et du schéma d’encodage utilisé dans cet article. Le projet est référencé sur la page web : <http://aune.lpl.univ-aix.fr/~otim/>.

Références

Annexe : Analyse discursive de l’extrait

Segment de discours 1 : pose un 1er point de repère, l’Escale Borély

- AL1 : disons si on si on est à à l’Escale Borély hein
- AL2 : le rond-point qui est euh à côté de l’hippodrome
- AL3 : y a un rond-point euh il s’appelle l’Escale Borély
- AL4 : c’est un rond-point où y a tout plein de restaurants

Segment de discours 2 : négociation, pose d’un 2ème point de repère, statue du David

- AL1 : en fait quand on descend le Prado
- AL2 : vous voyez le David est-ce que vous voyez la statue du David
- AL3 : voilà donc c’est je parle de la côte

Segment de discours 3 : négociation, direction à partir du 2ème point de repère

- AL1 : donc là vous êtes à l’au- vous êtes à la statue du David
- AL2 : vous allez vers les Goudes
- AL3 : donc comme si vous allez vers là-bas

Segment de discours 4 : négociation réussie, retour au segment de discours 1

- AL1 : donc vous longez
- AL2 : quand vous passez la statue de David après vous arrivez à un rond-point

AL3 : et là à ce rond-point sur la droite c'est plein de restaurants euh ça s'appelle l'Escale Borély

AL4 : et en fait vous avez longé l'hippodrome parce que vous êtes euh cette route elle longe l'hippodrome

Segment de discours 5 : pose d'un 3ème point de repère, rond-point du pouce

AL1 : et donc à ce rond-point de l'Escale Borély vous remontez à gauche

AL2 : vous allez jusqu'en haut y a euh y a le rond-point du Pouce

Segment de discours 6 : emplacement du 3ème point de repère par rapport à un 4ème point, musée d'art contemporain

AL1 : avec le musée d'art euh là y a le musée d'art contemporain

AL2 (parenthèse) : qui est intéressant à voir si on s'intéresse à l'art contemporain

Segment de discours 7 : retour au segment de discours 5

AL1 : donc vous avez la statue du Pouce là

AL2 (parenthèse) : qui est la statue de l'artiste euh du fameux artiste euh qui a fait qui a mis qui qui a fait son sa sa sculpture de enfin sa son moulage de pouce là c'est un pouce géant

Segment de discours 8 : direction par rapport au rond-point

AL1 : et en continuant tout droit euh vous arrivez à un autre rond-point

AL2 : et là c'est la direction de Sormiou

Segment de discours 9 : direction par rapport à la prison des Baumettes

AL1 : donc vous allez jusqu'au bout en direction des Baumettes

AL2 (parenthèse) : la prison la fameuse prison des Baumettes

Segment de discours 10 : arrivée

AL1 : et là euh à un moment donné euh et y a et là au au bout de la route y a la route qui monte vers Sormiou

AL2 (parenthèse) : alors là en hiver vous pouvez y aller en voiture c'est ouvert sachant que quand vous arrivez à Sormiou il faut payer deux euros parce qu'il y a un petit parking euh bon