



HAL
open science

Atelier DEGELS 2011 : DEfi GEste Langue des Signes

Leila Boutora, Annelies Braffort

► **To cite this version:**

Leila Boutora, Annelies Braffort. Atelier DEGELS 2011 : DEfi GEste Langue des Signes. Boutora, Leila et Braffort, Annelies. Traitement Automatique du Langage (TALN), France. ATALA, 2011. hal-01803811

HAL Id: hal-01803811

<https://hal.science/hal-01803811>

Submitted on 1 Jun 2018

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.

DEGELS2011

Actes du premier DÉfi Geste Langue des Signes

Proceedings of the First DEGELS Workshop

The word "DEGELS" is rendered in a stylized, three-dimensional font. Each letter is a vibrant blue color with a glossy, reflective surface that catches the light, giving it a metallic or crystalline appearance. The letters are slightly shadowed underneath, making them look like they are floating or resting on a surface.

1er juillet 2011

Montpellier, France

DEGELS2011

Actes du premier DÉfi Geste Langue des Signes

1er juillet 2011
Montpellier, France

Préface

Le Défi GEste Langue des Signes (DEGELS) a été créé cette année, en 2011, par Leïla Boutora, maître de conférence au Laboratoire Parole et Langage (LPL, UMR 6057 CNRS-Université de Provence), et Annelies Braffort, chargée de recherche au Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI, UPR 3251 CNRS).

La création de ce défi est née du besoin de rassembler les communautés scientifiques étudiant la gestualité coverbale et la langue des signes afin d'échanger sur des problématiques communes portant sur l'annotation de corpus. L'idée d'animer cette rencontre sous la forme d'un défi a été largement inspirée du défi de fouille de données DEFT, un atelier associé à la conférence TALN, créé en 2005 et qui est organisé chaque année depuis.

L'objectif du défi consiste à proposer une campagne d'annotation d'un corpus vidéo comparable en deux langues – français oral (avec sa gestualité) et langue des signes française (LSF) – avec une thématique partagée de manière à encourager les échanges sur les méthodologies et les problématiques liées à l'annotation du geste. Nous fournissons aux équipes participantes des données vidéo extraites d'un corpus réalisé pour l'occasion. Les équipes doivent les annoter, puis fournir leur annotation ainsi qu'un article décrivant leur méthodologie à des dates fixées à l'avance. Les organisateurs comparent les annotations sur une thématique différente chaque année et élabore une synthèse des différentes approches. Cette synthèse ainsi que les différentes annotations sont présentées lors d'un atelier organisé après la conférence TALN.

Pour cette première édition, nous avons proposé aux participants la thématique des gestes de pointages. Les pointages, présents à la fois dans la gestualité coverbale et dans la langue des signes, sont des phénomènes décrits dans la littérature et font l'objet d'annotations et d'analyses relativement nombreuses en regard d'autres thématiques. Ce sont donc de parfaits candidats pour cette première édition.

Ce recueil contient l'ensemble des articles soumis, ainsi qu'un article présentant plus en détail les principes et l'organisation de ce défi et une synthèse des différentes approches proposées par les participants.

Leïla Boutora et Annelies Braffort, *coorganisatrices de DEGELS 2011*

Comités

Coordinatrices

Leïla Boutora (LPL, Aix-en-Provence)

Annelies Braffort (LIMSI–CNRS, Orsay)

Comité d’organisation

Leïla Boutora (LPL, Aix-en-Provence)

Annelies Braffort (LIMSI–CNRS, Orsay)

Mathieu Lafourcade (LIRMM, Montpellier)

Comité de lecture

Roxane Bertrand (LPL, Aix-en-Provence)

Leïla Boutora (LPL, Aix-en-Provence)

Annelies Braffort (LIMSI–CNRS, Orsay)

Constitution du corpus

Leïla Boutora (LPL, Aix-en-Provence)

Annelies Braffort (LIMSI–CNRS, Orsay)

Annick Choisier (LIMSI–CNRS, Orsay)

Thierry Legou (LPL, Aix-en-Provence)

Loundou Linganzi (LPL, Aix-en-Provence)

Roxane Bertrand (LPL, Aix-en-Provence)

Table des matières

Préface	iii
Comités	v
Table des matières	vii
Présentation et résultats	1
Présentation et premiers résultats du défi d'annotation DEGELS2011 sur un corpus bilingue de français oral et de langue des signes française. <i>Leïla Boutora, Annelies Braffort et Roxane Bertrand</i>	3
Partie 1. Gestualité	13
Tu pointes ou tu tires ?! Annotation des pointages d'un « entendant vocalo-gestualisant ». <i>Dominique Boutet, Marion Blondel, Stéphanie Caët, Pauline Beaupoil et Aliyah Morgenstern</i>	15
Annotation multimodale du français parlé. Le cas des pointages. <i>Gaëlle Ferré</i>	29
Types de gestes et utilisation de l'espace gestuel dans une description spatiale : méthodologie de l'annotation. <i>Marion Tellier, Mathilde Guardiola et Brigitte Bigi</i>	45
Partie 2. Gestualité et LSF	57
De l'espace métrique l'espace de signation, vos GP(e)S. <i>François Lefebvre-Albaret et Jérémie Segouat</i>	59
Partie 3. LSF	75
DEGELS2011 : Analyses d'annotations pour la reconnaissance. <i>Monia Ben Mlouka, Juliette Dalle et Patrice Dalle</i> 77	
Les pointages en Langue des Signes : une méthodologie d'annotation. <i>Nadège Devos, Flora Badin et Émilie Chételat-Pelé</i>	95
Typologie des pointages en Langue des Signes Française (LSF) et problématiques de leur annotation. <i>Brigitte Garcia, Marie-Anne Sallandre, Camille Schoder, Marie-Thérèse L'Huillier</i>	107
Index	121
Index des auteurs	121
Index des mots-clés	123

Présentation et résultats

Présentation et premiers résultats du défi d'annotation DEGELS2011 sur un corpus bilingue de français oral et de langue des signes française

Leïla Boutora¹ Annelies Braffort² Roxane Bertrand¹

(1) Laboratoire Parole et Langage, UMR 6057 CNRS/Univ. Provence, 13100 Aix-en-Provence

(2) LIMSI- CNRS, Campus d'Orsay Bt. 508, BP133, 91403 Orsay cx

leila.boutora@lpl-aix.fr, annelies.braffort@limsi.fr, roxane.bertrand@lpl.aix.fr

Résumé. Dans cet article, nous présentons la première édition du défi d'annotation de gestes et de langue des signes (DEGELS). L'objectif est d'organiser une campagne d'annotation dans le but de comparer des méthodologies d'annotation et d'analyse de corpus de gestes coverbaux en français oral et de langue des signes française (LSF) en soumettant aux chercheurs linguistes et informaticiens de ces domaines un corpus constitué pour l'occasion. L'édition 2011 se propose d'étudier le phénomène de pointage, objet de recherche partagé par les communautés gestualiste et LSF. Après avoir présenté les objectifs, les enjeux scientifiques et le déroulement de cette manifestation scientifique, nous proposons une comparaison des méthodologies d'annotation adoptées par les sept équipes participantes, en précisant les critères de comparaison retenus. Nous concluons cette dernière partie par une synthèse qui met en évidence les apports d'une telle démarche méthodologique à l'étude même des phénomènes linguistiques. Dans le cas présent nous montrons l'influence directe que peut avoir la définition du pointage, selon les cas en termes formels ou en termes fonctionnels, sur l'identification et l'inventaire des pointages dans les deux corpus.

Abstract. In this paper, we present the first edition of the gesture and sign language annotation challenge (DEGELS). The goal is to organise an annotation campaign in order to compare methodologies for annotation and analysis of coverbal gestures in spoken French and French Sign Language (LSF) corpora. For that, we have submitting to linguists and computer scientists a corpus that has been specially created for this challenge. The 2011 DEGELS edition is dedicated to the study of pointing, a research topic shared by gestural and sign languages scientific communities. After presenting the objectives, challenges and organisation of this event, we propose a comparison of the annotation methods adopted by the seven participating teams, specifying the comparison criteria. We conclude this last part with a summary that highlights the contributions of such a methodological approach to study the linguistic phenomena. In this case we show the direct influence that can have the definition of pointing, its forms or its functions, for the identification and inventory of pointing in the videos.

Mots-clés : Méthodologie d'annotation, schéma d'annotation, gestualité, multimodalité, Langue des Signes Française, pointage

Keywords : Annotation methodology, annotation scheme, gesture, multimodality, French Sign Language, pointing

1 Introduction

1.1 Présentation et objectifs de DEGELS

DEGELS (DEfi Geste Langue des Signes) est un atelier de comparaison d'annotation de corpus de gestes coverbaux et de langue des signes. L'objectif de cet atelier est de rassembler les communautés scientifiques étudiant la gestualité coverbale et la langue des signes autour des problématiques communes portant sur l'annotation de corpus. Cet atelier vise à ouvrir de nouvelles perspectives de collaboration entre les gestualistes des communautés Gestes et Langue des Signes et à moyen terme permettra la constitution d'un réseau national puis francophone Geste et LS.

Cet atelier prend la forme d'un défi d'annotation : nous fournissons aux équipes participantes des données vidéo extraites d'un corpus comparable de langue des signes et de français oral (voix et gestes). Les équipes doivent les annoter, fournir leur annotation ainsi qu'un article décrivant leur méthodologie (choix théoriques, schéma d'annotation, critères de choix...). Les organisateurs comparent les annotations entre elles sur une thématique différente chaque année et élabore une synthèse des différentes approches. Cette synthèse ainsi que les différentes annotations sont présentées lors de l'atelier, avec une évaluation des atouts de chaque approche. Cette journée sera l'occasion d'échanger sur les méthodes et les schémas d'annotation.

Le partage des corpus et des annotations passe par une bonne connaissance des méthodes et une description claire des schémas d'annotation utilisés ainsi que des critères appliqués (segmentation, choix parmi les catégories, etc.), au travers de guides d'utilisation par exemple, tel que celui élaboré par Johnston (2011) pour l'annotation de la langue des signes australienne (AusLan). Cela profitera à l'avancée des descriptions linguistiques pour les domaines encore peu dotés que sont les études gestuelles et des langues des signes.

Pour cette première édition, nous proposons d'annoter les pointages. Les pointages, présents à la fois dans la gestualité et dans la langue des signes, sont des phénomènes décrits dans la littérature et font l'objet d'annotations et d'analyses relativement nombreuses en regard d'autres thématiques. Ce sont donc de parfaits candidats pour cette première édition.

1.2 Enjeux scientifiques

Peu d'études mettent en relation l'organisation gestuelle de la langue des signes avec celle de la langue parlée. Ces domaines sont peu dotés, manquent de standards au niveau des méthodologies et des annotations. Le contexte international est pourtant dynamique et propose des conférences qui comportent des workshops thématiques sur la multimodalité, les études gestuelles ou les langues des signes¹. Cependant, ces workshops internationaux n'offrent pas de véritables lieux de rencontre communs aux communautés scientifiques Gestes et Langue des Signes. Dans le cadre de DEGELS, nous souhaitons provoquer une rencontre entre ces communautés qui partagent nombre de problématiques de recherche. Le fait d'organiser ce type de manifestations scientifiques afin de permettre la rencontre des acteurs des domaines concernés apparaît donc comme une nécessité. Le nombre des soumissions et des inscriptions, au-delà de nos attentes, confirme d'ailleurs l'adéquation de cette manifestation aux besoins des chercheurs. Cette manifestation qui répond donc à un fort besoin de la communauté sera renouvelée tous les ans, associée à la conférence TALN.

En prévision de cet atelier, une journée de formation pratique a été organisée le 21 janvier 2011 sous l'égide de l'ATALA². L'objectif de cette journée était de présenter et faire pratiquer deux logiciels d'annotation de vidéos de productions multimodales et de langue des signes, Anvil et Elan, ainsi que les outils de requêtes et d'analyse complémentaires, afin d'inciter les jeunes chercheurs et doctorants à participer à l'atelier DEGELS2011 en leur fournissant les outils et les compétences nécessaires en amont.

¹ *Language Resources and Evaluation Conference (LREC)* et ses ateliers associés, en particulier « Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality » et « Workshop on Representation and Processing of Sign Languages (RPSL) »

Egalement *Sign Language Corpora Network (SLCN)*
² <http://tals.limsi.fr/jatala2011.html>

2 Déroulement

L'atelier DEGELS 2011 a lieu le 1er juillet 2011 à Montpellier, dans les locaux du Centre Régional de Documentation Pédagogique (CRDP) à Montpellier. Des chercheurs sourds interviendront, en présence d'interprètes français-Langue des Signes Française.

2.1 Constitution du corpus

Le corpus DEGELS1, qui est proposé pour l'atelier DEGELS2011, est référencé sous l'identifiant oai:crdo.fr:crdo000767 au CRDO d'Aix-en-Provence. Il s'agit de dialogues en français oral (son + geste) et de dialogues en LSF comparables. La thématique des échanges porte sur une proposition de lieux à visiter à Marseille qui permet d'éliciter des pointages dans les deux langues. Nous avons filmé quatre diades d'entendants et quatre diades de sourds. Chaque diade était composée d'un modérateur et d'un locuteur n'étant pas impliqué dans le projet. Le modérateur était le même pour chaque diade, mais il s'agissait d'une modératrice entendante pour le français et d'un modérateur sourd pour la LSF. Les deux modérateurs avaient préparé ensemble le protocole d'élicitation afin de garantir deux corpus comparables.

La captation a été réalisée dans la chambre sourde du LPL, avec deux micros externes (pour le français oral) reliés à deux caméras mini DV semi pro CANON XM2, et une caméra HD Sony hdr-cx 550 VE pour la vue d'ensemble (figure 1).



Figure 1 : Extrait du corpus DEGELS1 pour le français (à gauche) et la LSF (à droite)

Nous avons fourni aux participants deux extraits de DEGELS1. Ces extraits ont été sélectionnés de manière à comporter un nombre important de pointages, tout en restant de courte durée, étant donné le temps limité dont disposaient les participants pour annoter les corpus. Pour les deux langues, les extraits portent sur la même portion d'itinéraire : la corniche en bord de mer qui mène du vieux port aux plages de Marseille en passant par le David, statue emblématique d'un homme nu.

Le corpus, sous forme de fichiers vidéo et audio, était téléchargeable ainsi qu'un fichier texte comportant la traduction approchée en français écrit pour l'extrait en LSF. Les fichiers vidéo étaient disponibles en différents formats de compression (mp4, IV5.1, non-compressé), afin d'assurer la compatibilité avec les logiciels d'annotation Elan et Anvil. Les fichiers son permettaient de disposer de deux pistes (une par locuteur), séparées mais alignées entre elles et sur la vidéo.

2.2 Aspects juridiques

Avant de procéder à l'enregistrement des données vidéo et audio, nous avons exposé aux informateurs la finalité du recueil de corpus, ainsi que l'objectif et le contenu du formulaire de consentement éclairé qu'ils ont ensuite accepté de signer. Ce consentement précise les droits et les devoirs des responsables de la

constitution du corpus et des informateurs. En particulier, il précise dans quel cadre le corpus peut être diffusé et utilisé.

Afin de garantir le respect des droits des personnes filmées, les participants à l'atelier ont dû signer les accords de restriction d'usage des vidéos avant de pouvoir y accéder. Cet accord précise les limitations des droits d'utilisation et de diffusion du corpus par les participants.

2.3 Calendrier

L'ensemble du processus s'est déroulé en cinq étapes. Les participants ont eu accès au corpus à partir du 17 février. Ils ont eu 5 semaines pour fournir leur annotation (25 mars), puis 2 semaines pour soumettre leur article (11 avril). La notification a été envoyée aux auteurs le 22 avril. Les auteurs ont eu ensuite 2 semaines pour fournir la version finale de leur article (6 mai).

2.4 Les équipes participantes

Les acteurs des communautés encore restreintes de la gestualité et de la langue des signes ont répondu très largement à l'appel DEGELS. Sur 9 équipes inscrites au départ, 7 sont allées jusqu'au bout du processus. Ces équipes regroupent 21 intervenants, dont 9 gestualistes et 12 Langue des Signes. Elles présentent leur étude sur le corpus le jour de l'atelier.

Les communautés scientifiques aussi bien Gestes que Langue des Signes qui s'intéressent à l'annotation de corpus sont très récentes (années 2000). De fait ces communautés sont composées de jeunes chercheurs et de chercheurs en formation, en doctorat mais également en Master (traitement automatique du langage, linguistique, traitement d'image, physiologie du mouvement, entre autres). La 1^{ère} édition de cette manifestation vise tout particulièrement à attirer ces jeunes chercheurs afin de soutenir un réseau en cours de création.

Les sept équipes participantes sont constituées de la manière suivante :

- Équipe n°2 : D. Boutet (SFL Paris), M. Blondel (SFL Paris), S. Caët (Sorbonne Nouvelle Paris), P. Beaupoil (Sorbonne nouvelle Paris), A. Morgenstern (Sorbonne nouvelle Paris)
- Équipe n°3 : N. Devos (LIMSI Orsay), F. Badin (INIST Nancy) et E. Chételat-Pelé (Université de Provence Aix en Provence)
- Équipe n°4 : G. Ferré (LLING Nantes)
- Équipe n°5 : M. Ben Mlouka (IRIT Toulouse), J. Dalle (IRIT Toulouse), P. Dalle (IRIT Toulouse)
- Équipe n°7 : B. Garcia (SFL Paris), M-A. Sallandre (SFL Paris), C. Schoder (SFL Paris), M-T. L'Huillier (SFL Paris)
- Équipe n°8 : M. Tellier (LPL Aix en Provence), M. Guardiola (LPL Aix en Provence), B. Bigi (LPL Aix en Provence)
- Équipe n°9 : F. Lefebvre-Albaret (WebSourd), J. Segouat (WebSourd)

3 Comparaison des méthodologies et des annotations

L'objectif de l'atelier est donc d'exposer les différentes méthodologies adoptées par les chercheurs afin de les confronter les unes aux autres ; de mettre en évidence les spécificités de l'annotation propres aux objectifs de chaque équipe, leurs atouts en termes scientifiques ; également les trucs et astuces transférables à l'ensemble de la communauté scientifique. L'atelier est aussi l'occasion de faire émerger ou de confirmer des points de discussion qui seront abordés lors de la table ronde, ainsi que des pistes de réflexion pour les prochains ateliers.

Nous présentons ici les critères de comparaisons des annotations et proposons une synthèse sur la manière dont a été abordée la thématique de cette année, le pointage, en montrant « ce qu'en disent les annotations ».

3.1 Les critères de comparaison des annotations

Le tableau 1 synthétise la répartition des études selon la langue étudiée (colonne 1) et précise les modalités vocale ou gestuelle (colonne 2) et locuteurs annotés (colonne 3 : A est le modérateur et B le locuteur produisant le discours élicité), les logiciels utilisés (colonne 4) et si l'annotation a été réalisée de manière automatique ou manuelle (colonne 5), ainsi que les ressources déployées pour l'annotation (colonnes 6 et 7 : durée et nombre d'annotateurs).

Num équipe Langue	Modalité	Locuteur annoté	Logiciels	Mode d'annotation	Durée	Nb annotateurs
2 Français	Vocale Gestuelle	B B	Easy Align Elan	Auto+manuel Manuel	90h	4
3 LSF	Gestuelle	A et B A et B B	Anvil	Manuel	24h	3
4 Français	Vocale Vocale Gestuelle	A et B A et B B	Praat Easy Align Anvil	Manuel, Auto+manuel Manuel	25h	1
5 LSF	Gestuelle	A et B	Logiciel IRIT, Elan, VIES	Manuel	4h	3
7 LSF	Gestuelle	A et B	Elan	Manuel	30h	4
8 Français	Vocale Vocale Gestuelle	A et B A et B B	Logiciel LPL Praat et logiciel Loria Elan	Auto+manuel Manuel+auto Manuel	8h15	3
9 FR+LSF	Vocale Gestuelle	B	Anvil	Manuel	6h	1

Tableau 1 : Répartition des études, logiciels et ressources utilisés

Parmi les sept équipes qui sont allées jusqu'au bout du défi, trois d'entre elles ont annoté le corpus de français oral, et trois autres le corpus de LSF. Une septième a annoté les deux corpus. Cette répartition parfaitement équilibrée permet des comparaisons entre les pratiques des deux communautés scientifiques.

Les logiciels utilisés pour annoter le signal acoustique sont Praat, Easy Align, ainsi que deux logiciels de laboratoires (LPL et Loria). L'annotation du vocal comporte une phase automatique et une phase manuelle (qui précède ou qui suit la phase automatique). Les logiciels utilisés pour annoter les gestes sont Elan et Anvil, ainsi que deux logiciels développés en laboratoire (IRIT). Les annotations des gestes sont uniquement manuelles.

Les équipes, composées de 1 à 4 annotateurs, ont consacré un temps très variable à l'annotation, allant de 4 h à 90 h pour l'ensemble du travail (définition du schéma d'annotation, annotation, analyse). Les annotations fournies possèdent donc des degrés de description allant de simple à très fin. Selon le nombre d'annotateurs et l'organisation du travail (annotations en parallèle ou successives), le travail a pu impliquer des phases de discussion supplémentaires.

3.2 Synthèse des différentes approches

Avant de pouvoir identifier les comparaisons qu'il est possible de faire entre les annotations, il nous faut mettre en évidence les variations de toute nature (linguistique ou pas) qui peuvent avoir une incidence sur les résultats. Cela débute évidemment par la définition de ce qu'est un pointage, la description de sa forme et de sa fonction linguistique.

La description formelle des pointages

Nous avons constaté une grande variété quant aux moyens mis en œuvre pour décrire la forme du pointage ou ses interactions avec d'autres composantes corporelles. Cela concerne :

- le type de gestes pris en compte : uniquement manuel (Tellier *et al.*, 2011), associant manuel et regard (Boutet *et al.*, 2011 ; Ferré, 2011, Devos *et al.*, 2011 ; Garcia *et al.*, 2011), distinguant pointage manuel, par le regard et par le buste (Ben Mlouka *et al.*, 2011), considérant un autre type de gestes – les signes relocalisés- (Ben Mlouka *et al.*, 2011), ou par ailleurs limité aux gestes en lien avec les informations géographiques (Lefebvre-Albaret, Segouat, 2011)
- le nombre de supports considérés : seulement les mains (Tellier *et al.*, 2011), les mains et le regard (Boutet *et al.*, 2011 ; Ferré, 2011, Garcia *et al.*, 2011, Ben Mlouka *et al.*, 2011 ; Lefebvre-Albaret, Segouat, 2011) et des éléments supplémentaires tel que les bras (Boutet *et al.*, 2011) ou le clignement des yeux (Devos *et al.*, 2011).
- la granularité des descriptions pour les pointages manuels : uniquement l'emplacement (Tellier *et al.*, 2011 ; Ben Mlouka *et al.*, 2011), la composante corporelle employée et la direction pointée (Garcia *et al.*, 2011, Devos *et al.*, 2011), un ensemble de paramètres perceptifs (Ferré, 2011), un ensemble de paramètres productifs (Boutet *et al.*, 2011), une surface dans l'espace et la forme de la main (Lefebvre-Albaret, Segouat, 2011)
- la structure temporelle : non spécifiée (Garcia *et al.*, 2011 ; Lefebvre-Albaret, Segouat, 2011), une délimitation de la phase principale du geste (Ben Mlouka *et al.*, 2011 ; Boutet *et al.*, 2011), une description plus détaillée des différentes phases (Ferré, 2011 ; Tellier *et al.*, 2011 ; Devos *et al.*, 2011). Les critères formels concernant la segmentation des gestes ne sont que rarement spécifiée (Ben Mlouka *et al.*, 2011).
- La direction pointée : pas toujours annotée. Quand elle est annotée, les catégories sont assez proches entre gestualité et LSF.

L'incidence directe de cette variabilité dans l'annotation de la forme des pointages est bien sûr une différence importante dans le nombre d'unités gestuelles répertoriées sous la catégorie « pointage » : pour le corpus de LSF, le nombre de pointages manuels varie du simple au double (11, 16 et 20). Sur le corpus de français oral, cette variation est encore plus marquée (28, 39 et 108). Ce dernier chiffre très élevé s'explique par la nature physiologique de la définition formelle proposée du pointage (alignement de trois segments adjacents terminaux), le fait que seule la forme a été étudiée (il n'y a pas eu d'annotation de la fonction) et peut-être le temps d'annotation bien plus long (donc plus de précision, un niveau de granularité plus fin).

La fonction des pointages

Les aspects fonctionnels n'ont pas été traités par toutes les équipes. En particulier, celle ayant annoté la forme de manière très fine n'a pas travaillé sur le niveau fonctionnel (Boutet *et al.*, 2011). Les autres équipes ont au minimum cherché à identifier les unités gestuelles par rapport à leur fonction principale dans le discours, avec parfois une analyse plus fine basée sur l'utilisation d'autres typologies : par exemple, dans (Garcia *et al.*, 2011), les fonctions du langage (Jakobson, 1963), l'interaction, la construction de références (Trevisiol *et al.*, 2010) et la détermination (Rinfret, 2008).

Pour les équipes qui ont annoté le corpus de français oral, les catégories manipulées pour l'identification de la fonction principale des unités gestuelles sont issues des travaux de Kendon (2003) et McNeill (1992, 2005) (Ferré, 2011 ; Tellier *et al.*, 2011). Des sous-catégories ont été créées pour ce corpus dans (Tellier *et al.*, 2011). La typologie de McNeill semble encore à l'heure actuelle la plus utilisée par la communauté des gestualistes qui dispose ainsi d'une base commune pour mener des comparaisons entre les études.

Pour les équipes qui ont annoté le corpus de LSF, les catégories sont basées sur les travaux de Johnston (2011) (chez Devos *et al.*, 2011), de Cuxac (2000) pour l'identification de la fonction principale des unités gestuelles (Devos *et al.*, 2011 ; Garcia *et al.*, 2011 ; Ben Mlouka *et al.*, 2011) et pour l'interprétation des résultats (Devos *et al.*, 2011 ; Lefebvre-Albaret, Segouat, 2011). Si la typologie de Cuxac est largement répandue et utilisée en France, elle l'est beaucoup moins au niveau international. Il sera particulièrement intéressant d'étudier dans quelle mesure le choix du modèle théorique influence l'identification des unités gestuelles, en mettant en parallèle les annotations des pointages réalisées selon la typologie de Johnston ou celle de Cuxac.

Le pointage à la lumière du son

Nous avons vu précédemment que le parallèle entre les deux types d'études (geste / LSF) mettait en évidence des pratiques communes et dans le même temps des approches propres à chaque domaine qui faisaient envisager un même phénomène, le pointage, de manière différente selon ces approches. L'annotation du corpus de français oral comprend en plus de l'étude des productions gestuelles, celle des productions sonores. L'étude des deux modalités est menée en partie indépendamment, en partie en regard l'une de l'autre. Les annotations du gestuel et du son se déploient dans le temps en parallèle, ce qui permet de mettre en relation leurs éléments respectifs. Différentes stratégies sont envisageables : l'annotation du son et du geste indépendamment l'un de l'autre ou l'alignement du gestuel sur le son. Se pose ici la question de l'impact sur l'annotation elle-même de l'ordre de l'annotation des éléments sonores et gestuels et de la dépendance des pistes d'annotation concernées.

Le découpage temporel de la production sonore peut reposer sur des unités d'étendue variable : le mot, la syllabe, le phonème. Les trois études menées ici procèdent à un alignement temporel des gestes sur la base d'une segmentation en mots, voire en phonèmes. L'annotation des productions sonores à des niveaux d'analyse et de granularité différents peut être menée (semi)automatiquement ou manuellement, les méthodologies et les outils utilisés pour chaque niveau de l'analyse linguistique n'ayant pas reçu les mêmes développements logiciels. On dispose par exemple pour la syntaxe de parseurs automatiques (même si la plupart concernent encore souvent du texte écrit) alors que les domaines prosodiques ou discursifs sont encore largement annotés manuellement. Enfin, le lien établi entre le sonore et le gestuel peut avoir une incidence variable sur l'analyse du gestuel : le son peut constituer un simple repère temporel (Boutet *et al.*, 2011) ou faire partie intégrante d'une approche « multimodale » qui ne pose pas de hiérarchie entre les dimensions gestuelles, verbales et vocales (Ferré, 2011 ; Tellier *et al.*, 2011), jusqu'à avoir une incidence forte sur la catégorisation des pointages (Ferré, 2011).

Boutet *et al.* adoptent une démarche « gestuo-centrée » dans le sens où les auteurs prennent le parti de ne pas annoter la partie sonore en elle-même et se concentrent sur le gestuel qu'ils annotent à un degré de granularité extrêmement fin. Le son n'occupe ici qu'un rôle de repère temporel.

Dans les deux autres équipes, la TOE³ s'est avérée extrêmement utile dans la mesure où elle a permis d'obtenir un alignement phonémique d'une grande qualité sur des données non contrôlées (et sur lesquelles on ne disposait d'aucune information). Étant donné que l'alignement phonétique découle de cette TOE, et qu'il sert de référence pour les autres annotations (syllabes, mots, etc.), il était crucial qu'il soit le meilleur possible. Après un retour d'expérience, la TOE pourrait être désormais simplifiée (car certains phénomènes pourraient être spécifiés automatiquement tels que les élisions par exemple). En revanche, les prononciations particulières qui demeurent non prédictibles, nécessitent d'être notées. Finalement, quelle que soit la transcription utilisée, l'alignement phonétique doit être le meilleur possible car il permettra d'étudier très précisément les liens de synchronisation entre les différents niveaux linguistiques, depuis les phonèmes jusqu'aux phases gestuelles notamment (voir Tellier *et al.* ; Ferré).

Tellier *et al.* soulignent l'importance de disposer du niveau des tokens pour favoriser des études relatives à la synchronisation geste/mot. Si les auteurs précisent qu'ils n'utilisent pas directement la TOE pour établir des liens entre les annotations gestuelles et verbo-vocales inférieures au mot, elle leur permet toutefois d'obtenir un bon alignement des mots (puisque celui-ci est complètement dérivé de l'alignement phonétique lui-même dérivé de la TOE). L'annotation en token permet également de calculer le taux gestuel qui s'avère très important, selon les auteurs, lorsque l'on s'intéresse aux gestes coverbaux.

Ferré (2011) illustre l'intérêt de disposer d'annotations multimodales (geste~parole) notamment aux niveaux lexical, prosodique (unités prosodiques, accents) mais aussi discursif (mise en relation des gestes avec des énoncés de type « demande de confirmation » par exemple). Plusieurs perspectives d'analyse sont envisagées : relations temporelles entre ces différents niveaux et relations de cooccurrence de certains événements au sein de ces niveaux. Elle montre le lien ou parfois et de manière plus intéressante l'absence de lien entre certains types de gestes et certains lexèmes. Loin d'être redondants, les gestes peuvent donc être modalisateurs d'énoncés, et non pas simples illustrateurs de la parole. Plus généralement, la discussion autour de l'affiliation (affiliation lexicale, prosodique et gestuelle) paraît tout à fait intéressante notamment dans une perspective de comparaison avec la langue des signes (1 seule modalité).

³TOE : Transcription Orthographique Enrichie. utilisée dans l'annotation du corpus CID dans le cadre du projet OTIM, voir Bertrand *et al.*, 2008 et <http://aune.lpl.univ-aix.fr/~otim/>

Durant l'atelier, les pratiques et les catégories manipulées au sein des communautés gestualistes et LSF, ainsi qu'entre ces deux communautés seront mises en regard les unes des autres, en vue d'alimenter la discussion théorique et d'enrichir les pratiques de part et d'autre.

Conclusion

Dans cet article, nous avons présenté les objectifs, les enjeux scientifiques et le déroulement de DEGELS 2011, première édition d'une manifestation scientifique, qui consiste à comparer des méthodes d'annotation et d'analyse de corpus de gestes coverbaux en français oral et de langue des signes française (LSF). L'édition de cette année s'est centrée sur le pointage, objet de recherche partagé par les communautés gestualiste et LSF.

Après avoir comparé les méthodologies d'annotation adoptées par les sept équipes participantes, nous avons montré l'influence directe que peut avoir la définition du pointage, selon les cas en termes formels ou en termes fonctionnels, sur l'identification et l'inventaire des pointages dans les deux corpus. Plus largement, cette première comparaison a permis de mettre en évidence des pratiques propres à chaque communauté. Ainsi, les gestualistes appréhendent la structure temporelle plus finement. Les chercheurs LS eux possèdent une expertise plus avancée de l'étude des phénomènes spatiaux et de l'agencement des mains entre elles (mono vs bimanuel, (a)symétrie largement exploitée) ou avec une autre partie du corps. Ces observations ne doivent rien au hasard. Les gestualistes, amenés à appréhender le geste régulièrement dans son rapport au sonore, s'appuient sur un cadre de référence éminemment temporel. A l'inverse, les chercheurs LS, par la nature même des LS qui exploitent de manière maximale un repérage dans l'espace ou par rapport au corps dans leur grammaire, sont dans la nécessité de manipuler des représentations spatiales tridimensionnelles pour modéliser finement la langue.

L'enjeu de cet atelier est également de montrer l'impact des outils et des méthodologies sur l'analyse elle-même, pour prendre conscience de son existence et favoriser à terme des pratiques qui ne se laissent pas contraindre par les outils existants plus ou moins adaptés, mais au contraire de formuler les besoins des chercheurs afin d'aller vers le développement d'outils qui soutiennent réellement l'exploration linguistique.

La journée de l'atelier qui comporte une session de présentations orales, une table ronde et une session de démonstrations, fera une large place à l'échange. Elle mettra en exergue les points qui semblent centraux dans l'ensemble des méthodologies et transférables d'une communauté à l'autre, tels que la segmentation des unités gestuelles ou l'annotation des informations de nature tridimensionnelles. Elle est envisagée comme une première étape dans la structuration d'un réseau, national et francophone dans un premier temps, qui permettra des échanges riches et soutenus dans les années à venir entre chercheurs LSF et gestualistes. Une série de questions fondamentales qui sous-tend toute la linguistique moderne et que l'inclusion des LS permet d'éclairer d'un nouveau jour sera sans doute au centre des débats de ce réseau, comme par exemple : la modalité seule permet-elle de définir ce qui est verbal ou non verbal ? le verbal est-il la linguistique ou le sonore ? à quoi correspond le coverbal en LS et comment peut-on le définir ?

Remerciements

Les auteurs remercient le modérateur, les informateurs et le personnel technique du LPL et du LIMSI ayant participé à la constitution du corpus, les relecteurs et les organisateurs de TALN et de DEFT pour leur aide à tous les niveaux (spécial dédicace à Cyril Grouin pour les scripts du site web et des actes) et bien sûr tous les participants à DEGELS 2011 !

Cette manifestation a reçu le soutien financier du CNRS (FIPHFP), du LPL, du LIMSI, de l'UFR LACS de l'Université de Provence et de l'Atala.

Références

Corpus DEGELS1 oai:crdo.fr:crdo000767

BEN MLOUKA M., DALLE J., DALLE P. (2011). DEGELS2011 : Analyses d'annotations pour la reconnaissance. Actes du *1er DEfi Geste Langue des Signes (DEGELS 2011)*, 79-95.

- BERTRAND R., BLACHE P., ESPESSER R., FERRE G., MEUNIER C., PRIEGO-VALVERDE B., RAUZY S. (2008). Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement automatique des langues (TAL)* 49 (3), 105-134.
- BOUTET D., BLONDEL M., CAËT S., BEAUPOIL P, MORGENSTERN A. (2011). Tu pointes ou tu tires ?! Annotation sous ELAN des pointages d'un 'entendant vocalo-gestualisant'. Actes du *1er DEfi Geste Langue des Signes (DEGELS 2011)*, 15-27.
- CUXAC C. (2000). La langue des Signes Française; Les voies de l'iconicité. *Fait de Langues*, 15-16, Oprys, Paris.
- DEVOS N., BADIN F., CHÉTELAT-PELÉ E. (2011). Annotation d'une vidéo en Langue des Signes Française : Étude des pointages. Actes du *1er DEfi Geste Langue des Signes (DEGELS 2011)*, 97-108.
- FERRÉ G. (2011). Annotation multimodale du français parlé : Le cas des pointages. Actes du *1er DEfi Geste Langue des Signes (DEGELS 2011)*, 29-43.
- GARCIA B., SALLANDRE M.-A., SCHODER C., L'HULLIER M.-T. (2011). Typologie des pointages en Langue des Signes Française (LSF) et problématiques de leur annotation. Actes du *1er DEfi Geste Langue des Signes (DEGELS 2011)*, 109-121.
- JAKOBSON R. (1963). *Essais de linguistique générale*. Paris, Editions de minuit.
- JOHNSTON T. (2011). *Auslan Corpus Annotation Guidelines*, March 2011.
- KENDON A., VERSANTE L. (2003). Pointing by Hand in Neapolitan. In S. Kita (ed.) *Pointing, Where language, culture and Cognition meet*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 109-138.
- LEFEBVRE-ALBARET F., SEGOUAT J. (2011). De l'espace métrique à l'espace de signation, à vos GP(e)S. Actes du *1er DEfi Geste Langue des Signes (DEGELS 2011)*, 61-74.
- MCNEILL D. (1992). *Hand and Mind : What Gestures reveal about Thought*. Chicago, London : University of Chicago press.
- MCNEILL D. (2005). *Gesture and Thought*. Chicago, London : University of Chicago press.
- RINFRET J. (2008). *L'association spatiale du nom en Langue des Signes Québécoise : Formes, fonctions et sens*. Thèse de doctorat en linguistique, Université du Québec, Montréal.
- TELLIER M., GUARDIOLA M., BIGI B. (2011). Types de gestes et utilisation de l'espace gestuel dans une description spatiale : méthodologie de l'annotation. Actes du *1er DEfi Geste Langue des Signes (DEGELS 2011)*, 45-56.
- TREVISIOL P., WATOREK M., LENART E. (2010). Topique du discours / topique de l'énoncé – réflexions à partir de données en acquisition des langues. M. Chini ed. *Topic, information structure and acquisition*. Franco Angeli.

Gestualité

Tu pointes ou tu tires ?! **Annotation des pointages d'un « entendant vocalo-gestualisant »**

Dominique Boutet^{1,2}, Marion Blondel¹, Stéphanie Caët³, Pauline Beaupoil³, Aliyah Morgenstern³

(1) Structures Formelles du Langage, UMR 7023 CNRS-Université Paris 8, 59/61 rue Pouchet, 75017 Paris

(2) Université d'Evry Val d'Essonne, bd François Mitterrand - 91025 Evry cx

(3) Sorbonne Nouvelle-Paris3, 17, rue de la Sorbonne, 75231 Paris cx 05

dominique_jean.boutet@orange.fr, marion.blondel@sfl.cnrs.fr, stephanie.caet@gmail.com,
pauline.beaupoil@gmail.com, aliyah.morgenstern@gmail.com

Résumé Le défi présenté ici vise à montrer que la forme canonique d'un pointage représente un type particulier de pointage, mais qu'on peut étendre la définition formelle à tout alignement de trois segments distaux adjacents. Cette hypothèse est à tester et par delà les éléments formels gestuels que nous nous donnons, le regard, la direction de la tête et le verbal sont évidemment convoqués. Ces deux modalités verbale/vocale et gestuelle sont vues comme un contexte lié qui devrait permettre la mise en évidence des prédicats psychologiques (McNeill). Après avoir donné les éléments du protocole d'annotation, nous exposerons une catégorisation possible de ces pointages mais également leurs inclusions dans des gestes iconiques. Nous aborderons la fonction anaphorique et la spécialisation formelle des gestes utilisés à cet usage. Ceci amène à mieux définir le cadre de référence associé aux pointages (principal vs secondaire, spécialisation éventuelle des degrés de liberté) et montre l'importance d'une micro-annotation gestuelle.

Abstract The aim of the challenge we are presenting in this study is to show that the canonical shape of pointing gestures is representative of a specific type of pointing gesture, but that the formal definition could include all types of alignment of three proximal adjacent segments. This hypothesis needs to be tested and not only do we use formal elements to describe gestures, but gaze, the direction of the head and verbal productions are here included in our analysis. The visual and vocal modalities are viewed as linked and enable us to underline psychological predicates (McNeill). We first present the various elements of our annotation procedure, and our categorization of the various pointing gestures we coded and their inclusion in iconic gestures. We tackle the anaphoric function of pointing gestures and the formal specialization of the gestures used for that function. This study helps us redefine the reference framework associated with pointing gestures (main vs. secondary, possible specialization of the degrees of freedom) and shows the importance of using micro-annotations for gestures.

Mots-clés : gestuelle coverbale, annotation allocentrée, pointage, anaphore

Keywords: coverbal gesture, allocentred annotation, pointing, anaphora

1 Contexte et objectifs de l'étude

Les études portant sur la gestualité coverbale reprennent majoritairement une classification qui différencie fonctionnellement le pointage, les gestes dits iconiques, d'autres appelés métaphoriques et enfin les *beats*¹ (McNeill 1992). Ces derniers viennent scander le discours selon un rythme correspondant aux pulsations de la parole. Les gestes iconiques montrent une relation formelle étroite avec le contenu sémantique de la production verbale. Pour McNeill, les gestes métaphoriques se distinguent des gestes iconiques par le fait que l'image qu'ils donnent présente une idée abstraite plutôt qu'elle ne représente un objet ou un événement. Ce type de geste décrit un concept par une métaphore concrète. Nous ne nous attarderons pas sur ces derniers et considérerons pour la suite qu'ils composent plutôt une sous-catégorie des gestes iconiques. Les gestes de pointage permettent d'indiquer des objets ou des événements dans un monde concret comme dans un espace relevant d'éléments abstraits ou absents. Une littérature abondante existe sur les pointages tant pour les fonctions, que pour la forme qu'ils revêtent (voir Kita 2003 pour une revue, mais aussi Kendon et Versante 2003, Clark 2003). Toutefois il n'existe pas de définition formelle stable et unanimement partagée du pointage. Nous adoptons la définition de l'acte de pointer d'Eco (1976) : « mouvement vers ». Tout « mouvement vers » qu'il soit effectif ou relayé par la présence d'un vecteur² peut être considéré comme un pointage. Ainsi, la forme que prend le pointage dépend donc des segments mis en mouvement ; les variations de support relevées – souvent ethniques – sont nombreuses : lèvres, majeur (Wilkins 2003), tête, pouce (Calbris 1990 p.127), voire auriculaire³, regard (Cuxac 2000). Cette variété d'utilisation des supports révèle la difficulté d'une définition unifiée de la forme du pointage. Les ressources formelles de son expression augmentant en proportion du nombre de supports convoqués pour pointer, cette variété accroît les formes possibles du pointage pour un support comme les gestes (index tendu, main plate doigts alignés, pouce tendu doigts repliés, majeur ou auriculaire tendus avec ou sans mouvement...). En outre, de nombreux exemples abondent de pointage d'une main pendant que l'autre effectue un geste iconique, de localisation précise et intentionnelle d'un geste iconique d'une seule main, de pointages et de rythmes battus par des *beats* de manière simultanée ; dans tous ces cas de figure, il s'agit de la composition d'un type de geste avec un autre. Kendon 2003 bâtit même une classification des pointages qui empruntent et surajoutent à ceux-ci – en un mot composent – des valeurs linguistiques (singularité vs exemplarité, déixis vs anaphore). Des bâtons cohabitent et coexistent avec des pointeurs, des gestes iconiques ou métaphoriques se fondent dans des pointages, des *beats* scandent plusieurs rythmes à la fois. Toutes les compositions semblent possibles. Pourquoi n'y aurait-il pas de pointages dans des pointages ? Ainsi, l'importance du pointage est telle que tous les supports sont mis à contribution, y compris ceux chargés pour nous d'une demande explicite d'échanges de baisers comme l'avancée de la lippe. En outre, l'étendue des formes du pointage sur le membre supérieur offre une gamme variée de possibilités (index étendu ou majeur, index et majeur, pouce, doigts avec des orientations différentes de la paume). Ainsi, si l'index est tendu, les autres doigts repliés, des petits mouvements latéraux ou perpendiculaires à l'alignement avant-bras/dos de la main/index sont toujours possibles ; ces derniers sont bien « des mouvements vers » et sont dirigés ailleurs que vers le pointage principal. L'ensemble des degrés de liberté de la main et des doigts peuvent être convoqués pour ajouter des minis pointages à celui, principal et tenu, qui voit un alignement général d'au moins deux jointures.

Il faut donc partir d'une définition formelle de ce que peut être un pointage principal pour pouvoir explorer la présence d'éventuels pointages secondaires. En effet, partir uniquement d'une définition fonctionnelle poserait des difficultés de distinction entre les pointages principaux et les secondaires.

Un premier point de vue formel pour définir le pointage gestuel circonscrit des conditions statiques. On définit *a priori* comme porteur d'un pointage gestuel tout alignement d'au minimum trois segments adjacents distaux. Ainsi quel que soit le doigt, un pointage est possible dès lors que les trois phalanges sont alignées pendant un laps de temps suffisant⁴. D'autres segments peuvent s'ajouter à cet alignement : le dos

¹ Nous n'aborderons pas ici les gestes dits cohésifs puisque ces derniers dans la terminologie de McNeill relèvent plus d'une fonction particulière que tous les types gestuels peuvent servir.

² Pour le membre supérieur, le vecteur est composé de l'alignement d'au minimum trois segments adjacents. Il est orienté dans la mesure où il présente une flèche distale libre de toute attache à un autre segment.

³ Observation personnelle chez plusieurs peintres mis en situation de commenter un tableau.

⁴ Dans les faits nous avons retenu ici une durée minimale de 120 ms (durée de 3 images). La question de la durée suffisante reste ouverte. Théoriquement cette durée peut être de 40ms (1 image) non pas tant pour des raisons de standard de captation, mais à cause de l'absence de distinction de l'oeil en deça de cette durée.

de la paume et l'avant-bras voire le bras. Dans certains cas, que nous aborderons, la direction du regard passe par le bout des doigts pointés vers le haut tout comme la main, et vise une zone vers l'avant comme on le ferait d'une cible avec un fusil muni d'une mire verticale dont l'extrémité est orientée vers le haut : la mire verticale est donnée par le bout des doigts étendus. Le corpus étudié montre des exemples de ces pointages/mires pour lesquels la direction du pointage ne dépend pas de l'alignement des segments gestuels. Ainsi, pour analyser le pointage, on peut dissocier alignement gestuel, directionnalité et mouvement.



Figure 1: La mire (compteur 1.06)

Le deuxième point de vue formel pour définir le pointage gestuel concerne le mouvement et son alignement temporel avec l'alignement spatial des segments. Si ce mouvement est effectué dès la mise en place de l'alignement des segments, il sera catégorisé comme pointage principal. Lorsque le mouvement intervient après la mise en place, il sera alors analysé comme un pointage secondaire, à moins qu'il ne s'agisse d'un geste iconique (voir *infra*). La direction du pointage principal est conférée par l'alignement des segments, celle du pointage secondaire est donnée par la direction du mouvement. Néanmoins, un mouvement latéral dès lors qu'il est co-occurent à l'alignement spatial d'au moins trois segments, peut composer le pointage principal : il s'agit d'un tracé, par exemple celui des « ronds-points », particulièrement présent dans le corpus. Les pointages/tracés dessinent une figure et rassemblent dès lors plusieurs mouvements dans un même motif. En l'absence de ces distinctions entre les paramètres d'alignement spatial et temporel, de mouvement et de direction, nous ne pourrions pas avoir de définition formelle englobante pour ce type de pointage/tracé. Les pointages latéraux qui ne sont pas pris dans des figures sont néanmoins susceptibles de composer des gestes iconiques appelés Unités Gestuelles (UG désormais, Boutet 2001, 2008 et 2010). Il convient de s'assurer que ces pointages latéraux ne s'inscrivent pas dans une UG. Nous verrons qu'une part significative des mouvements éligibles au statut de pointage secondaire relève en fait de gestes de type iconique.

Les pointages secondaires relèvent-ils du même type d'espace que ceux du pointage principal ? Autrement dit pour un mouvement manuel redevable d'un pointage secondaire – un mouvement de flexion de la main par exemple –, quand peut-on le considérer comme un mouvement dans un espace égocentré (peu importe qu'il s'agisse d'une flexion, ce qui compte est que le mouvement oriente la main vers la gauche) et quand déterminer que ce pointage a lieu dans un espace allocentré sur la main, dans lequel la flexion garde une fonction propre indépendamment de l'orientation manuelle ? Un des objectifs que nous nous sommes fixés dans ce défi consiste à mieux cerner cette différence.

Une des cibles possibles du pointage secondaire vise l'énoncé à l'instar de pointages principaux d'ailleurs, y compris dans des renvois anaphoriques du geste vers le verbal. Ainsi, lorsque le locuteur pointe vers une direction marquant le chemin à prendre, il peut simultanément, par un ou des pointages secondaires (lors de mouvements latéraux) renvoyer vers un ou des référents verbaux. Pour la gestuelle, nous reprenons la différenciation temporelle qui prévaut entre un pointage à valeur déictique dans une portée *hic et nunc* et un pointage à valeur anaphorique qui déplace la cible dans le temps de part et d'autre du moment du pointage cata- ou anaphorique dans la même modalité discursive. Rappelons qu'un pointage à valeur déictique est transmodal : une valeur sémantique langagière ou para-langagière (la gestuelle ici) est ancrée vers la situation référentielle. Cette distinction – renvoi vers la situation ou vers le discours – s'applique indifféremment au fait gestuel et à l'accompagnement verbal. Toutefois pour la gestuelle, plan de l'expression para-langagière et plan de la situation (référence) partagent le même espace. Le statut déictique

ou anaphorique d'un pointage gestuel vers l'espace est donc beaucoup plus brouillé que ne l'est la même distinction pour du verbal-vocal. En outre, un investissement corporel orienté (gestuel, mimique, postural, regard) accompagnera plutôt le mot « demain » dans un énoncé qu'il ne le fera pour « lendemain ». Kendon (2003) signale qu'un pointage fait avec le pouce renvoie de manière anaphorique et par synchronie avec un mot à valeur anaphorique, vers un élément déjà évoqué dans le discours, redoublé d'ailleurs d'un pointage fait avec l'index. Au même titre, il est possible que la répercussion verbale/vocale soit différente lorsque pour le plan para-langagier gestuel, un pointage a valeur déictique ou anaphorique. Nous n'examinerons pourtant pas ce dernier point ici. Ainsi, un pointage peut être anaphorique d'un geste (*ag*) qui a été effectué ou d'un mot déjà prononcé (anaphore du verbal, *av*) ou bien il peut être déictique d'un geste en train d'être fait (par l'autre main par exemple) ou appartenant au même *catchment* (McNeill 2005) (*dg*) ou d'un mot ou d'un syntagme co-occurent (*dv*). La limite entre valeur déictique et anaphorique (*dg*, *dv* versus *ag*, *av*) est rendue plus ténue encore par le fait que le décalage temporel entre le pointage et son référent dépend du *catchment* qu'a défini McNeill composé de traits gestuels qui se répètent au moins dans deux gestes. Le principe réside dans la récurrence de l'image dans l'esprit du locuteur qui génère ces traits gestuels (http://mcneilllab.uchicago.edu/writing/growth_points.html). L'épaisseur temporelle d'expression d'un *catchment* interdit donc qu'à l'intérieur de celui-ci, les pointages renvoient à des anaphores dont le référent serait gestuel para-langagier. Pour qu'il y ait geste anaphorique, il faut qu'il y ait un geste en dehors du même *catchment* entre lui et le référent verbal, que ce geste apparaisse en l'absence de verbal co-occurent ou avec un syntagme qu'il n'illustre pas. On peut dire qu'il y a geste anaphorique vers un référent verbal (*av*) décalé dans le temps d'énonciation sans *catchment* commun ni chevauchement avec ce référent verbal lorsqu'à partir d'une reprise ou d'une anticipation il y a inclusion de ce geste dans le discours verbal. Un geste anaphorique renvoyant à un référent gestuel (*ag*) répond à la même définition pour la même modalité : on peut dire qu'il y a geste anaphorique vers un référent gestuel décalé dans le temps d'énonciation sans *catchment* commun ni chevauchement avec ce référent gestuel lorsqu'à partir d'une reprise ou d'une anticipation il y a inclusion d'un geste dans un début de discours gestuel (enchaînement de *catchments*).

L'étude porte donc sur la représentation des ces types d'anaphores et de déictiques par les pointages secondaires : leur nombre comme les éventuelles spécialisations gestuelles qui les portent. Un renvoi anaphorique ou cataphorique apparaît-il avec gestes de directions différentes, voire opposées ? A quels cadres de référence, ces directions possibles appartiennent-elles ? Y a-t-il une séparation dans l'utilisation des cadres de référence : égocentré pour les gestes de pointage à valeur déictique redevables de la situation *hic et nunc*, d'une part, et allocentré sur la main et les doigts pour les gestes à valeur anaphorique, d'autre part ?

2 Description de la méthodologie d'annotation

Nous traiterons dans cette partie des aspects techniques et humains de l'annotation. Le schéma d'annotation ainsi que le détail des pistes seront abordés. Nous avons tenu à utiliser autant que possible des vocabulaires contrôlés pour les pistes afin de limiter les sources d'erreurs de saisie ainsi que d'extension et de compréhension des items.

2.1 Aspects techniques

Quatre annotateurs ont participé à l'annotation de ce défi. Ils ont utilisé l'outil ELAN et ont consacré 90 heures au total à l'annotation manuelle, réparties de la façon suivante : 15 heures pour l'annotation du regard, 25 heures pour l'annotation manuelle du verbal-vocal, 15 heures pour l'annotation des directions alignement et mouvement, 35 heures pour toutes les autres pistes, et intégration de la segmentation automatique via Easy Align (cf. ci-dessous).

2.2 Objets de l'annotation

Seules les productions gestuelles et verbales du locuteur ont été annotées.

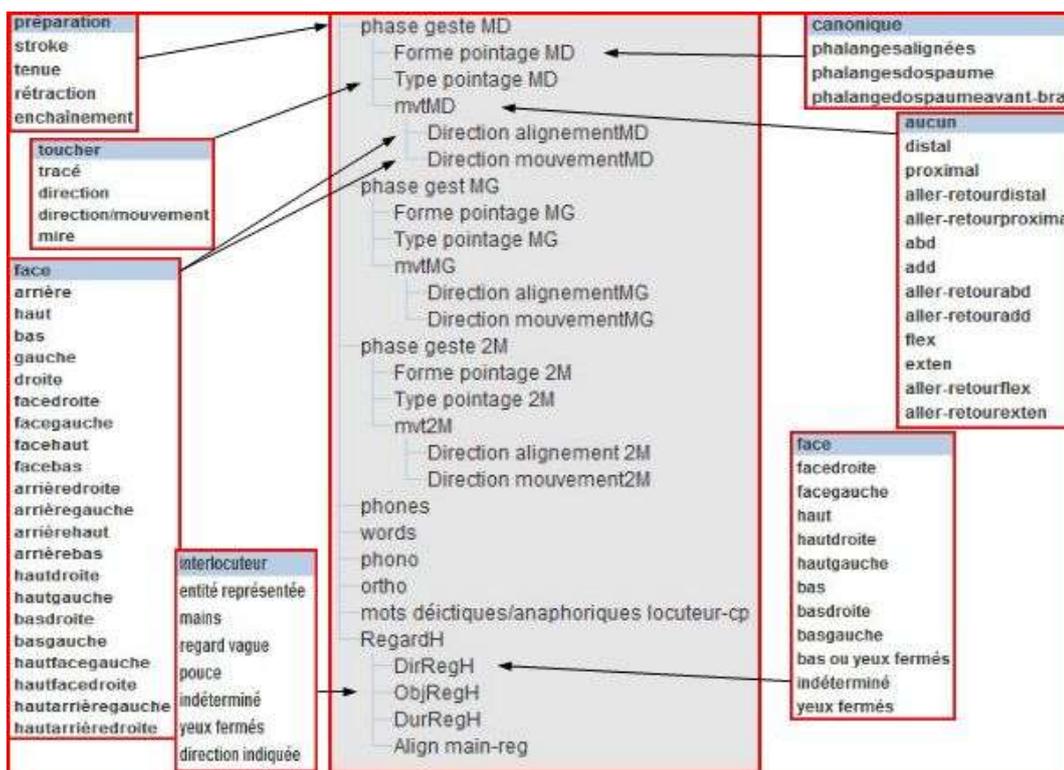


Figure 2 : Acteurs, dépendance dans le schéma et Vocabulaires contrôlés

L'annotation concerne principalement les gestes de pointage. En premier lieu l'ensemble des gestes a été noté selon une terminologie *phasique* (*préparation, stroke, tenue, rétraction*) sur la piste (*acteur*) « phase-geste ». Seuls les alignements d'au moins trois segments adjacents distaux ont été considérés plus avant et labellisés selon trois pistes principales : « *forme_pointage* », « *type_pointage* » et « *mouvement* ». Pour la première piste, « *Phase_geste* », étant donnée l'analyse très précise des mouvements que nous souhaitons faire, nous avons estimé que des changements de configurations ou de direction de mouvements marquaient le passage d'une phase signifiante à une autre (*stroke*). Dans certains cas, deux phases de *stroke* peuvent donc être enchaînées. En revanche aucune préparation, ni rétraction, ni tenue n'en précède une autre. La piste « *forme_pointage* » décrit les alignements de segments y compris celui d'une forme considérée comme canonique du pointage (index tendu, les autres doigts repliés dans la paume), cette piste est associée à un *vocabulaire contrôlé*. Le type de pointage répertorie cinq catégories que nous avons déjà évoquées plus haut : la catégorie *pointage/direction* donnée par un alignement de segments en l'absence de mouvement aligné et coordonnée par une figure, l'association *direction/mouvement* avec un mouvement allant dans la même direction, le *pointage/toucher* marqué par un mouvement concordant avec l'alignement et en aller-retour, le *pointage/tracé* dont le mouvement est perpendiculaire à l'alignement, complexe (plusieurs degrés de liberté en jeu) et formant une figure, enfin le *pointage/mire* qui montre une association du regard et d'au moins un bout de doigt le tout pris dans un alignement dirigé vers l'entité visée. L'ensemble des pistes sous la dépendance du mouvement (*mvtMD, mvtMG, mvt2M*⁵) annotent les degrés de liberté qui président au mouvement (repère allocentré sur la main : *abduction, aller-retour abduction*, etc.) et les directions de ceux-ci dans un espace égocentré (*droite, gauche*, etc.) ainsi que la direction des alignements de segments.

Le regard a été annoté sur 4 pistes. La première « *RegardH*⁶ » relève d'une piste technique de segmentation à l'image de la piste « *Phase_Geste* ». A cet effet, cette piste constitue le parent des autres pistes du regard afin de fixer et de limiter l'activité de bornage de la segmentation pour les acteurs dépendants. La direction du regard fait l'objet d'annotations (piste « *DirRegH* »), elle inclut un *vocabulaire contrôlé* basé sur des items égocentrés. L'entité sur laquelle le regard est posé (piste « *ObjRegH* ») a également fait l'objet d'annotations. Ici aussi un vocabulaire contrôlé est associé à la piste pour cadrer la liste des possibles et afin d'éliminer les sources d'erreurs de saisie. L'alignement du regard et de la main (piste « *Alignement_regard-main* ») a été noté dans la quatrième piste de la catégorie « *Regard* ». Toutes les annotations qui concernent

⁵ Main Droite, Main Gauche, Deux Mains.

⁶ *RegardH* : Regard Homme.

cette catégorie ont été faites par une personne différente de celles qui ont annoté les autres centres d'intérêts de ce corpus : la gestuelle et le verbal.

La composante verbale/vocale du corpus a été décrite en deux temps et selon deux procédés. Le premier état d'annotation réalisé ('à la main' par un chercheur) a permis d'établir une transcription orthographique précise des deux locuteurs sur la base d'un découpage en énoncés et a donné lieu à une *tokenisation* par mot. La deuxième phase à partir de la transcription orthographique (piste « *ortho* ») a été faite à l'aide du plug-in Easy Align sous Praat et a permis d'obtenir automatiquement un découpage et un alignement par mot (piste « *word* ») sur le signal sonore, une transcription phonologique en Sampa (piste « *phono* ») et lorsque c'était possible un découpage par phonème (piste « *phones* »). Le signal sonore initial utilisé en entrée présentait deux pistes, l'une isolant la locutrice et l'autre mixant les deux locuteurs. Cette absence d'individualisation des signaux a nuï à la segmentation et à l'alignement avec Easy Align pour les segments au cours desquels il y avait un chevauchement de parole. Ceci explique les lacunes des pistes « *phones* » et « *phono* ». Ces difficultés ont été contournées par une segmentation et un alignement faits à la main pour la piste mot (« *word* »). Enfin une piste « *mots_déictiques_anaphoriques* » contient les mots ayant l'une de ses valeurs en incluant d'ailleurs les anaphores associatives. Cette ligne n'a fait l'objet d'aucune vérification ou d'aucun double codage.

A l'exception des pistes qui concernent le verbal/vocal et des acteurs « *RegardH* » et « *Alignement_regard-main* », toutes les autres ont un vocabulaire contrôlé ; soit 20 sur 27. Les pistes 'enfants' ont un type linguistique associé, le plus généralement avec un stéréotype « *Time Subdivision* » (subdivisible, bornes internes déplaçables). Les pistes 'enfants' « *forme-pointage* », « *DirRegH* », « *ObjRegH* » ont un stéréotype « *Symbolic Association* » (bornes non déplaçable, non subdivisible). La piste « *Alignement_regard-main* » a quant à elle un stéréotype « *Included In* » (possibilité de placer une absence de segment à l'intérieur de l'annotation parente). Les pistes « *Phase_Geste* », « *RegardH* », ainsi que celles du verbal/vocal n'ont pas de stéréotype particulier associé.

La granularité des annotations a à voir avec une micro-analyse. A titre d'illustrations pour l'ensemble du corpus, les valeurs médianes des annotations *flex*, *exten*, *abd* et *add*⁷ pour la piste « *MvtMD* » sont respectivement de 0.193 0.155 0.165 0.189 secondes. Rappelons que 0.040s équivalent à une seule image sur les 25 images prises par seconde. On a donc ici des valeurs médianes de durée de segment de l'ordre de 4 images. La piste « *Direction_mouvement* » recouvre des durées similaires, bien que le nombre d'annotations soit supérieur de 20% par rapport aux pistes parentes « *Mvt* ». L'expérience de l'annotation nous a montré qu'on ne peut pas atteindre d'emblée cette finesse-là : une, voire deux étapes intermédiaires de segmentation sont indispensables. Cette granularité très fine est obtenue en cascade lors d'annotations successives de pistes parentes « *Phase_Geste* » puis la piste « *Mvt* » et enfin de la piste « *Direction_mouvement* ». Approximativement à la même échelle de durée figurent les découpages en mots, la valeur médiane pour les 369 mots est de 0,170s. Bien évidemment le découpage en phonèmes est plus fin encore : la valeur médiane des durées est de 0,06s.

Le nombre total d'annotation est 2378. La répartition par catégorie montre un équilibre entre le pourcentage d'annotations verbal/vocal et gestuel respectivement de 47,27% et 43,78%. Une troisième catégorie, celle du regard, est moins représentée puisque regroupant 8,96% du nombre total d'annotations. Si la composante gestuelle s'équilibre avec la composante verbale/vocale du point de vue du nombre d'annotation, le fait est à souligner doublement : d'une part, cet équilibre est plutôt rare dans les annotations pour lesquelles généralement la gestuelle est sous représentée, d'autre part, il ne provient pas d'un effet démultiplicateur dû à la présence des deux mains puisque 88,52% des annotations gestuelles relèvent de la main droite.

Cet équilibre quantitatif entre les composantes *gestuelle* et *verbale/vocale* est d'autant plus important qu'il relève pour ces deux composantes d'annotations *formelles* et non pas *fonctionnelles*. Aucune catégorie grammaticale ni aucune fonction ne sont annotées, il en va de même de la composante gestuelle pour laquelle la fonction du pointage n'est pas prise en compte, aucune catégorisation fonctionnelle en termes de

⁷ Le mouvement de flexion est défini pour la main comme le rapprochement de la paume vers l'intérieur de l'avant-bras, l'extension consiste en un éloignement de la face intérieure et, conséquemment, en un rapprochement de la face externe de l'avant-bras. L'abduction manuelle, mouvement du côté du pouce, fait basculer légèrement le plan de la paume dans le plan constitué par l'avant-bras et la main lorsqu'ils sont alignés. Le mouvement d'adduction opère un mouvement dans le même plan, du côté de l'auriculaire. Les valeurs de ces annotations concernent dans ce corpus les mouvements de la main, des doigts, mais également ceux de l'avant-bras voire du bras. Ces mouvements sont alors repérés par les directions reportées sur la main.

type de gestes n'apparaît ; seul ce qui correspond à une définition formelle du pointage est annoté. On se situe donc sur un même plan d'analyse. Il en va de même des annotations concernant le regard.

2.3 Vocabulaires Contrôlés

Les vocabulaires contrôlés concernent pour l'essentiel des directions que ce soit pour le regard, comme pour la gestuelle. Trois types de directions ressortent des annotations qui concernent l'espace. On trouve pour le regard (piste « *ObjRegard* ») un vocabulaire en lien direct avec la situation d'interlocution qui détaille les entités susceptibles d'être regardées. Ce vocabulaire a été bâti de manière *ad hoc* : une entrée concerne le pouce, longtemps regardé au cours d'une séquence du corpus. À côté de cette annotation spatiale sans repère fixe, on trouve une autre piste « *DirRegH* » pour laquelle l'espace est découpé en directions égocentrées possibles pour le regard (la direction *arrière* est proscrite). On peut d'ailleurs se demander s'il s'agit vraiment ici d'une 'égocentration' ou si on n'est pas en présence d'un repère allocentré sur/par le regard. Pour la gestuelle, les pistes « *Direction_alignement* » et « *Direction_mouvement* » partagent un vocabulaire contrôlé commun qui découpe l'espace égocentré de manière relative en rapport avec la direction des mouvements visibles sur la main (dont la piste parente est « *Mvt* ») et suivant la flèche de l'alignement des trois segments adjacents formant la composante statique du pointage. Le nombre d'entrées est 27. Il s'agit bien de direction et non pas de localisation aussi chaque catégorie simple (*gauche*, *face*, *arrière*) est-elle suffisante sur sa propre flèche. Nous n'avons pas codé la position dans ce schéma d'annotation. Il nous a paru cohérent pour le pointage – objet principal de l'annotation – de nous limiter aux directions. Le troisième type d'annotation pour l'espace concerne les mouvements associés au pointage ; on peut aussi parler des deux composantes du pointage que nous codons ici : le mouvement et la direction. Ils sont appréhendés non pas d'un point de vue égocentré mais de manière allocentrée sur la main. Les mouvements de cette dernière comme ceux des doigts et des segments plus proximaux (avant-bras, bras) sont rabattus sur le déplacement de la main. L'appréciation de la direction de ces derniers est composite et reprend à la fois des directions de degrés de liberté existant sur la main (*flexion*, *extension*, *abduction*, *adduction*) mais aussi des directions davantage en rapport avec l'alignement des segments adjacents lorsque cela inclut le dos de la paume. Ces deux directions, *distale* et *proximale*, quoique composites par rapport aux directions précitées n'ont pas posé de problème de discrimination à l'exception des séquences où le pouce est dressé, ses trois phalanges alignées : l'abduction est alors confondue avec une direction distale pour de faibles amplitudes. Un indice visuel a permis de distinguer les annotations à affecter au mouvement. Par exemple lorsque la main bougeait seule alors l'entrée *abduction* a été choisie, lorsque le mouvement provenait de l'avant-bras ou du bras, alors l'entrée *distale* était préférée. Un type de mouvement n'a pas été retenu, celui de la *pronosupination* (mouvement rotatoire de la main), il ne nous a pas semblé relever d'un mouvement susceptible de porter un pointage.

Le vocabulaire contrôlé associé à la piste « *Type_pointage* » comporte cinq entrées que nous avons déjà détaillées : *toucher*, *tracé*, *direction*, *direction/mouvement* et *mire*. Elles concernent toutes un alignement d'au moins trois segments adjacents (soit deux jointures) et distaux et se différencient entre elles par la composante de mouvement du pointage dont le début est synchrone avec la composante statique (alignement adjacent d'au moins trois segments). Le mouvement aligné avec la flèche du pointage peut être en aller retour (*toucher*), perpendiculaire à cette flèche et coordonné selon un motif (*tracé*), aligné simplement (*direction/mouvement*), aligné également avec une tendance à un mouvement tangentiel accompagné d'un regard dont la direction passe par le bout des doigts (*mire*) ou sans mouvement aligné (*direction*).

L'alignement des trois segments adjacents distaux qui est renseigné dans le vocabulaire contrôlé associé à la piste « *Forme_pointage* » est jugé à l'œil nu. Nous n'avons aucune mesure biomécanique en entrée pour décider de manière précise et constante de la fourchette angulaire acceptable pour l'alignement. Au demeurant, la discrimination visuelle est moins bonne pour les angles des segments courts (les phalanges) que pour ceux compris entre la paume et l'avant-bras. Toutefois le jugement de l'alignement est fait sur l'ensemble des segments et non pas pour chaque jointure. Une contrainte de mesure angulaire à chaque jointure à partir des données biomécaniques par exemple amènerait à accroître la courbure à mesure de l'augmentation du nombre de segments jugés alors comme alignées. Ainsi, pour un angle de 10° de part et d'autre d'un alignement parfait et compte tenu de l'addition à chaque segment supplémentaire, on aurait *in fine* un cumul possible de 60° en remontant jusqu'au bras. Ceci ne constitue pas précisément un alignement. Il faudrait donc adjoindre à des contraintes angulaires une composante de longueur de segments. Au demeurant, la prise en compte de la composante anatomo-physiologique du membre supérieur semble indispensable : les degrés de liberté des doigts offrent des butées articulaires propres à favoriser un alignement (extension 0° des phalanges) ainsi que le bec de l'olécrane au niveau du coude qui bloque peu ou prou l'avant-bras dans un alignement au bras (extension 0°), il en va différemment du poignet pour

lequel aucun élément anatomique ne stabilise une position d'alignement entre la main et l'avant-bras. Du point de vue formel, un alignement sur une butée articulaire ne correspond pas à un alignement en l'absence de butée. On peut incriminer la laxité, voire chez certaines personnes l'hyper-laxité des jointures d'exercer une influence sur la fourchette angulaire pour définir l'alignement. Cela signifie concrètement que la mesure de ce qui relève ou non d'un alignement serait idiosyncrasique. Sans être complet sur ces questions définitoires, un dernier point retient tout de même l'attention, celui d'une possible asymétrie de latitude angulaire encadrant l'alignement entre la butée articulaire d'une jointure favorisant l'alignement des deux segments adjacents (extension 0 en butée articulaire) d'un côté, et le pôle opposé du même degré de liberté ouvert quant à lui sur une amplitude telle que rien ne stabilise un alignement (flexion 90°). En l'absence d'indices contextuels verbaux et visuels et en partie pour ces raisons d'asymétries anatomiques, un bornage angulaire automatique de ce qui relève d'un alignement ne serait peut-être pas le même selon les degrés de liberté, selon les segments et selon les personnes.

3 Quelques résultats

Afin de déterminer le nombre de pointages secondaires par rapport aux pointages principaux, il faut tout d'abord connaître le nombre d'annotations pour la piste « *MvtMD* »⁸. En effet, le nombre d'annotations de cette piste y compris pour l'entrée 'aucun' (mouvement) donne l'ensemble des mouvements effectués pendant les alignements d'au moins trois segments adjacents. Pour déterminer le nombre de pointages secondaires il faut retrancher le nombre de pointages principaux, à savoir tous les mouvements *distaux*, *proximaux*, *ARdistaux*, *ARproximaux* ainsi que les pointages *mires* et les mouvements associés temporellement à un type de pointage *tracé*. Tous ces mouvements relèvent d'un pointage principal. Il y a plusieurs mouvements par pointage/tracé puisqu'il s'agit pour l'essentiel de traçage de ronds-points dans le corpus.

$$Pt\ second = \sum MvtMD - (MvtMD\ tracé + MvtMd\ (dist,\ prox,\ ARdist,\ ARprox) + MvtMd\ mire)$$

$$90 = 196 - (44 + (28 + 8 + 10 + 1) + 15)$$

On a donc à peu près autant de pointages principaux que de pointages secondaires : une centaine. Une part de ces pointages secondaires et principaux relève pourtant d'Unités Gestuelles (gestes iconiques) comme nous allons le voir. Une question théorique se pose néanmoins, celle d'un empilement possible des espaces et de la composition sémantique qui en résulte. Ainsi, des Unités Gestuelles peuvent être 'situées' au sens où si leur signification redevable d'un schéma d'action des degrés de liberté dessine un geste dont la direction est dépendante du schéma d'action, elles n'en sont pas moins placées dans l'espace fût-il allocentré. Une séquence l'atteste au cours de laquelle le locuteur utilise une Unité Gestuelle composée de mouvements de supination, d'*extension* et d'*abduction* de la main, qu'on peut étiqueter comme 'présenter' (processus). L'énoncé : « le musée d'art contemporain qui est intéressant à voir »⁹ présente cette UG en face et légèrement à la droite du locuteur, alors que juste avant dans l'énoncé (accompagnant « contemporain ») un pointage principal avait été effectué à sa gauche situant le musée à cet emplacement. Le fait de 'présenter' gestuellement en face porte l'intérêt à voir ce musée à l'attention de l'interlocutrice ; à ce titre l'UG est bien située de manière allocentrée (l'interlocution) et ajoute donc un pointage à l'Unité Gestuelle 'présenter', pointage qui est secondaire étant donné que l'élément principal est ici une UG. Cet allocentrage sur l'interlocution s'effectue par l'intermédiaire d'un mouvement de l'avant-bras qui appartient bien à l'alignement à ce moment-là ; toutefois au lieu d'avoir le décompte de deux pointages secondaires (*extension* et *abduction*) nous n'en avons qu'un seul : celui de l'interlocution, orienté vers la locutrice, porté par l'alignement dont la pointe est constituée par les doigts.

3.1 Durée moyenne des annotations de la piste « *MvtMD* »

Quelques récurrences apparaissent dans les entrées de la piste « *MvMD* ». Celles qui ont été étudiées à ce stade concernent les entrées simples d'un côté et les entrées en aller-retour de l'autre pour les degrés de liberté *flexion*, *extension* et *abduction*. Les durées moyennes ne sont pas très différentes entre les mouvements simples et ceux en aller-retour ce qui tend à montrer qu'on a un sorte de tempo commun peu

⁸ Nous nous concentrerons sur la main droite, le nombre de pointages de la main gauche et faits par les deux mains étant négligeable.

⁹ Le mot souligné prononcé entre 00:48.248 et 00:48.676 accompagne les mouvements de cette UG.

soumis à des déplacements significatifs redevables de localisations égocentrées. On pourrait s'attendre à ce que la durée moyenne soit deux fois plus longue pour les mouvements en aller-retour, il n'en est rien. Ceci tend à montrer que ces mouvements fonctionnent selon un schéma directionnel plutôt que localisateur. Un autre argument allant dans le même sens a trait aux amplitudes maximales différentes entre la flexion (90°) et l'extension (90°), d'une part, et à celles de l'abduction (30°) et de l'adduction (45°), d'autre part : elles ouvrent sur un angle entre 2 à 3 fois moins important pour cette dernière paire. Pourtant ne remarquant pas d'impact sur la durée moyenne, le mouvement n'est pas même lié à une localisation de type allocentré par rapport à l'espace que couvre chaque degré de liberté.

Entrée de la piste MvtMD (nbre)	Durée moyenne des annotations en s.
FLEXION (26)	0,218
A-R FLEXION (13)	0,292
EXTENSION (26)	0,229
A-R EXTENSION (5)	0,261
ABDUCTION (27)	0,209
A-R ABDUCTION (11)	0,420
ADDUCTION (25)	0,217
A-R ADDUCTION (6)	0,290

Tableau 12 : Durée moyenne d'annotations relevant de la piste « MvtMD »

On remarque pourtant que l'entrée *A-R abduction* dure en moyenne deux fois plus que le mouvement simple équivalent. Ceci est dû au fait que dans la moitié des cas (6 sur 11) le mouvement d'abduction en aller retour est impliqué dans le schéma d'action d'une UG dont 5 de ces 6 relèvent de l'étiquette 'aller'. Le schéma d'action de cette étiquette est ici réduit au seul mouvement d'abduction. Ce dernier couvre donc la durée de ce que normalement 2 mouvements couvrent (synchronie gestuel-verbal/vocal).

4 Discussion

Nous allons montrer l'existence de pointages secondaires associés la plupart du temps à des bâtons qui viennent scander le rythme parolier. Nous aurions à faire à des *beats* dans la classification de McNeill. Il ne fait pas de doute que bon nombre d'indices poussent vers cette interprétation : la durée similaire à celle des mots, la durée moyenne approximativement identique quel que soit le type de mouvement (simple ou en aller-retour), la décomposition binaire de ces mouvements et non pas ternaire (préparation, *stroke*, rétractation) des gestes iconiques.

Pourtant contrairement aux pulsations musicales, on distingue ici des directions qui semblent avoir été oubliées dans la classification. Certes, des mouvements en aller-retour surgissent mais du moins le font-ils selon des directions. A titre d'exemple le pointage secondaire sous forme de mouvement de flexion accompagnant le mot « hiver » (01:21.683) orienté vers l'arrière ressort d'un pointage principal tourné vers l'expression de l'antériorité. Ce mouvement scande bien le mot « hiver », sa direction exprime néanmoins doublement l'antériorité par un mouvement de flexion d'une part (Calbris 1990, p90) et par la direction arrière du mouvement (Calbris & Montredon 1986, ill.140). La scansion par déplacement de segments dans l'espace ne peut faire fi des directions, à moins de penser que le corps n'est pas orienté. Il semble d'ailleurs que l'interprétation usuelle des axes du temps (Calbris 1990, Cuxac 2000) même inversé (futur en arrière, passé devant, entre autres Nuñez & Sweetser 2006) puisse être générée par un cadre de référence non pas égocentré mais allocentré sur la main pour lequel la flexion est rectrice du passé et le mouvement d'extension du futur. Cette interprétation pourrait convenir au changement d'orientation de l'axe temporel (passage d'un axe frontal [gauche correspondant au passé, droite au futur, pour une position intermédiaire de la main droite] à un axe sagittal [arrière-passé, avant-futur, pour une position de supination de la main]).

Elle pourrait également rendre compte d'un axe sagittal pour lequel le passé est figuré vers l'avant et le futur vers l'arrière : une flexion de la main signifiant toujours le passé en position de pronation l'orienté vers l'avant, tandis qu'un mouvement d'extension de la main signifiant le futur en position de pronation l'orienté vers l'arrière.

Nous avons assigné une valeur à l'ensemble des annotations des entrées *flex*, *exten*, *abd*, *add*, *ARflex*, *ARexten*, *ARabd* et *ARadd* de la piste MvtMD (mais également *distal*, *proximal*, *ARdistal* et *ARproximal*). Ces valeurs peuvent être multiples par annotation. Elles relèvent a/du pointage, b/d'une valeur iconique ou c/de la préparation d'un geste en plus d'une valeur de bâton qui ne figure pas ici mais qui sous-tend *a priori* tous ces mouvements. Pour le pointage on distingue deux types d'items : ceux qui marquent une relation co-occurrence (au *catchment* près) avec ce à quoi renvoie chaque mouvement, on a alors des valeurs déictiques, et ceux qui marquent une relation non alignée temporellement avec ce à quoi renvoie le mouvement et hors *catchment* commun, on a à faire alors à des valeurs anaphoriques (voir *infra*, partie 1). Les référents répondent aux deux canaux vocal/verbal et gestuel. Ces référents sont désignés par *dv* (déictique verbal), *dg* (déictique gestuel), *av* (anaphore verbale), *ag* (anaphore gestuelle). Pour les déictiques nous avons noté le référent et la durée de chevauchement entre le pointage et le référent auquel il renvoie. Pour l'anaphore, nous avons noté la durée entre le début du pointage et le début du référent, certaines durées dépassent 2 secondes. Nous avons également noté la valeur iconique d'un regroupement de pointages secondaires lorsqu'un schéma d'action émerge d'un ou de plusieurs mouvements latéraux consécutifs de la main droite. Nous avons vu que malgré l'accaparement par une Unité Gestuelle de plusieurs mouvements latéraux, il est parfois possible de considérer qu'un pointage secondaire apparaît dans l'emplacement particulier dans lequel se situe cette UG. Dans ce cas, plusieurs mouvements latéraux regroupés au sein d'une UG définissent un pointage secondaire. Lorsqu'on a un alignement digito-manuel, pouce orienté vers le haut, l'auriculaire vers le bas, par exemple en 01:16.774 (« là au bout de la route ») pour lequel on a un mouvement d'abduction répété (mouvement vers le pouce) alors la valeur est très nettement iconique et marque à la fois l'action (UG d'étiquette 'aller'), la direction (pointage principal) et la situation de la route par rapport à un état antérieur dans le cheminement (la main est à hauteur de visage, pointage secondaire). On a également des schémas d'action (coordination du mouvement de plusieurs degrés de liberté faisant sens) notamment en 00:45.138 (« avec le musée d'art heu ») avec une association de mouvements *add* et *exten* pour une position de la main en supination redevable d'une Unité Gestuelle étiquetée 'constat'. Enfin, et c'est assez rare, nous avons quelques mouvements de préparation comme des rétractions sous forme de mouvement d'adduction (vers l'auriculaire) qui précèdent des mouvements d'abduction (vers le pouce) qui marquent la direction à prendre.

Chaque segment annoté de la piste MvtMD peut avoir plusieurs valeurs, en fait de 1 à 3. Ces pourcentages figurent dans le tableau 2. Ceci permet de comprendre que la somme des pourcentages par entrée est supérieure ou égale à 100%.

Dans ce corpus, il n'y a guère de déictiques renvoyant à un référent gestuel (*dg* : 2 cas). La cohésion médiante gestuelle ne semble pas assurée par des pointages si tant est qu'elle en ait besoin. Rappelons ici le choix privilégié par le locuteur d'une utilisation mono-manuelle. Il en va différemment des anaphores gestuelles (*ag* : geste renvoyant à un geste) composant environ 14% des cas. Une cohésion gestuelle coverbale est donc bien en place. La présence d'anaphores verbales (*av*, renvoyant vers un mot ou un syntagme vocal/verbal) est différenciée selon le type de mouvements : proportionnellement très présentes pour les mouvements en A-R, elles renvoient alors dans 50% des cas à une recherche de mots (A-Rflex) et rejoint par là les constats sur l'élaboration cognitive commune entre une composante vocale/verbale et une composante gestuelle. En outre, on peut remarquer que les mouvements simples présentent proportionnellement peu de fonctions renvoyant à des *av*. Même si on ne peut pas parler de spécialisation formelle pour ce type d'anaphores – il y a tout de même 14 occurrences d'*av* pour les mouvements simples – la proportion de ce type d'anaphores rapportée à chaque type de mouvement montre bien une préférence formelle du locuteur pour les mouvements en A-R (62,8%) au détriment des mouvements simples qui ne remplissent que dans 13,6% des cas la fonction d'anaphore à référent vocal/verbal. Les renvois gestuels vers des référents vocaux/verbaux semblent plus relever d'une conception temporelle que spatiale. Tout semble se passer comme si le moment du renvoi était signalé. Il l'est spécifiquement par l'utilisation de mouvements en A-R, bon candidats pour exprimer un instant. Il ne s'agit pas d'une spatialisation de la ligne du temps qui transposerait dans l'espace la durée séparant le geste anaphorique du moment où a été émis le référent vocal/verbal, mais plutôt du signalement du moment où pourrait s'insérer le référent vocal/verbal dans le discours. Ce signalement est effectué dans 41,6% des cas par un mouvement de flexion (en A-R et simple) contre 25% attendu dans une distribution aléatoire. Aucune égocentration dans la direction de ces mouvements ne transparait : les anaphores sont faites dans toutes les directions *arrière*, *haut*, *bas*, *avant*, *gauche* ou *droite*, bien évidemment dans des proportions différentes et qui respectent les orientations

privilegiées des mouvements. Il semble donc qu'une allocation prévale dans l'expression gestuelle des anaphores à référents verbaux, celle d'une flexion manuelle.

	i (iconic)	dv	dg	av	ag	Prépa	simple	double	tripl
Flex (n : 25)	52,00%	44,00%	0,00%	16,00%	12,00%	12,00%	44,00%	48,00%	8,00%
Exten (n : 26)	65,38%	46,15%	0,00%	19,23%	26,92%	11,54%	23,08%	50,00%	26,92%
Abd (n : 27)	66,66%	51,85%	3,70%	11,11%	11,11%	3,70%	22,22%	62,96%	14,81%
Add (n : 25)	52,00%	40,00%	4,00%	8,00%	4,00%	20,00%	40,00%	56,00%	4,00%
ARflex (n :13)	14,29%	21,43%	0,00%	78,57%	14,29%	0,00%	85,71%	0,00%	14,29%
ARexten (n :5)	0,00%	40,00%	0,00%	60,00%	0,00%	0,00%	100,00%	0,00%	0,00%
ARabd (n :11)	54,55%	45,45%	0,00%	54,55%	18,18%	0,00%	45,45%	36,36%	18,18%
ARadd (n : 6)	66,67%	66,67%	0,00%	33,33%	16,67%	0,00%	33,33%	33,33%	33,33%

Tableau 2 : Valeurs attribuées aux segments de la piste « MvtMD »

Au demeurant, nous avons codé la place du référent par rapport au geste de pointage secondaire : soit le référent vocal/verbal est émis avant le pointage (codage négatif de la durée par rapport au pointage) soit le référent vocal/verbal suit le pointage (codage positif de la durée par rapport au pointage). Dans 60% des cas d'anaphores à référents vocaux/verbaux (*av*), la durée entre le pointage et le référent est positive (pointage précède le référent) et dans 40% des cas, la durée est négative (référent vocal/verbal précède le pointage). Pourtant la répartition entre mouvements simples et mouvements en A-R n'est pas homogène : 57% des *av* négatifs sont faites avec des pointages simples, tandis que seulement 28% le sont avec des pointages en A-R. Ceci tend à montrer que le locuteur utilise plutôt des mouvements simples pour les pointages anaphoriques dont le référent verbal précède le geste comme si la fonctionnalité temporelle d'instantanéité du mouvement en A-R ne pouvait suffire à placer l'anaphore. Un référent vocal/verbal placé avant le geste anaphorique appelle majoritairement un pointage dont l'acmé dure, marquant plus le lien de référence que l'insertion ponctuelle dans le discours voco/verbo/gestuel que produit un mouvement en A-R.

Les déictiques verbaux (colonne *dv*) semblent davantage relever de mouvements simples que de ceux en A-R. Deux biais méthodologiques ont été évités dans ce décompte : a) celui du nombre effectif de déictiques dans les pointages/tracés figurés par les tours de « ronds-points » ; b) ceux des UG qui rassemblent 2 mouvements. De la même manière les gestes iconiques (*i*) apparaissent à chaque mouvement latéral décrivant un tour, aussi étaient-ils surreprésentés dans un premier décompte (le tableau 2 tient compte de ces biais).

On remarque un déficit de déictique verbal (*dv*) pour le mouvement ARflex (21,43%) et au contraire une surreprésentation de ce même mouvement pour l'anaphore verbale (*av*, 78,57%). Il apparaît que ce mouvement ressort plus d'un renvoi anaphorique que d'un déictique. La spécialisation va même plus loin puisque 100% des 11 cas de Mvt ARflex d'anaphore à référent verbal (*av*) relèvent d'une recherche de mot. Ceci pourrait expliquer en partie le déficit de ce mouvement à représenter un déictique. Les autres mouvements associés à une recherche de mots sont soit une flexion simple (1 cas) soit deux cas d'extension simple et un d'extension en A-R. Les orientations dans l'espace de ces mouvements sont certes

majoritairement à *gauche* et en *arrière*, mais les directions *avant*, *droite*, *arrièrebas* sont aussi présentes. Cette recherche de mots pourrait être associée sur une ligne du temps (frontale ou sagittale) à des mouvements de faibles amplitudes voire des oscillations des doigts ; ceci correspond aux cas mentionnés ici. Il semble même que ces mouvements peu amples soient effectués de part à d'autre d'un présent d'énonciation majoritairement par des flexions pourtant plutôt redevables de l'expression du passé, tandis que des mouvements d'extension, très minoritaires dans la recherche discursive, figureraient le futur. Cette apparente inversion de la flèche du temps qui accompagne la recherche d'un mot n'a pas vraiment lieu d'être pour des mouvements de si faible amplitude (épaisseur du présent ?), d'une part, et est grandement facilitée du côté de la flexion (digitale et/ou manuelle) tandis qu'elle est bloquée du côté de l'extension (butées articulaires digitales des phalanges), d'autre part.

Nous avons montré que le déplacement de la définition usuelle d'un pointage (index pointé tandis que les autres doigts sont fléchis, vers l'alignement d'au moins trois segments adjacents distaux) élargit considérablement le nombre d'unités de pointage dans le corpus, mais également accroît le champ des possibilités d'indexation puisque des pointages secondaires plus brefs apparaissent, susceptibles d'utiliser un tempo davantage en rapport avec le débit parolier. L'annotation des mouvements adoptée ici sur la base d'une allocentration sous forme d'annotation par les degrés de liberté manuels apporte quelques éléments de recadrage sur la description plus classiquement égocentrée de la gestuelle. L'une des premières conséquences de cette description allocentrée est la précision qu'elle offre : on a ici 6 items que l'on peut suivre même dans des orientations en diagonale, voire changeantes au cours du mouvement. La description classique égocentrée est moins robuste, moins précise parce qu'extérieure au segment qui bouge. Sa granularité est plus grossière (3 axes *fixes*) et les annotations plus indéfinissables autour des bissectrices, en sorte que, particulièrement pour des mouvements de faible amplitude, la marge d'erreur est telle qu'*in fine* seuls des mouvements de grande amplitude sont notés. Les conséquences sont massives et enchaînées : a) distorsion temporelle entre une annotation gestuelle grossière et une finesse d'annotation vocale/verbale ; b) annotation globale de gestes plus que de la diversité des mouvements les composant ; c) forte tendance à une réduction d'annotation des directions des gestes aux axes cartésiens égocentrés (avant/arrière, haut/bas, gauche/droite) au détriment d'une annotation dynamique des mouvements. Ces éléments 'techniques' obèrent les comparaisons entre le vocal/verbal et la gestuelle co-occurrence. Ils limitent le passage d'un type d'annotation à l'autre (égocentré vers allocentré) à cause des différences de granularités. Finalement, ils interdisent de fait l'émergence d'hypothèses sur la composition/décomposition des gestes. *A contrario*, nous avons montré ici qu'une annotation allocentrée permet un passage vers l'égocentration, qu'elle offre une précision descriptive et une finesse poussée jusqu'à la décomposition des gestes en mouvements successifs.

Cette précision qu'elle offre une annotation allocentrée nous a permis de montrer ici : a) le nombre de pointages secondaires et d'aborder leur distribution en catégories ; b) une description récurrente d'Unités Gestuelles (iconiques) sous la forme d'association de degrés de liberté indépendamment de la direction dans l'espace de ces mouvements ('aller', 'présenter', 'constater', 'partir') ; c) une spécialisation du mouvement de flexion simple ou en A-R vers le pointage secondaire anaphorique à référent vocal/verbal ; d) la composition/décomposition importante des valeurs au sein d'un même geste.

Références

- BOUTET, D., CUXAC, C. (2001). 'Approche Morpho-dynamique Du Sens Dans La Gestuelle Conversationnelle', Paris : Université de Paris VIII.
- BOUTET, D. (2008). 'Une Morphologie De La Gestualité : Structuration Articulatoire', *Cahiers De Linguistique Analogique*, N°5, Dijon : Abell, 80-115.
- BOUTET, D. (2010). 'Structuration Physiologique De La Gestuelle : Modèle Et Tests', *Lidil, Multimodalité De La Communication Chez L'enfant*, Jean-Marc Colletta, Agnès Millet et Catherine Pellenq Eds, 77-96.
- CALBRIS, G. (1990). *The Semiotics of French Gestures*, Advances in Semiotics, Bloomington: Indiana University Press.
- CALBRIS, G., MONTREDON, J., ZAÛ. (1986). *Des Gestes Et Des Mots Pour Le Dire*, Mini-dictionnaires, R. Galisson Eds, Paris : CLE International.
- CLARK, H. H. (2003). 'Pointing and Placing', dans *Pointing Where Language, Culture, and Cognition Meet*, Kita, Sotaro eds, Mahwah, London : Lawrence Erlbaum Associates, 243-268.
- CUXAC, C. (2000). *La Langue Des Signes Française (LSF) : Les Voies De L'iconicité*, Paris, Gap: Ophrys.
- ECO, U. (1977). *A Theory of Semiotics*, Critical Social Studies, London : Macmillan.
- KENDON, A., VERSANTE, L. (2003). 'Pointing by Hand in « Neapolitan »', *Pointing Where Language, Culture, and Cognition Meet*, Kita, Sotaro eds, Mahwah, London : Lawrence Erlbaum Associates, 109-138.
- MCNEILL, D. (1992). *Hand and Mind : What Gestures Reveal About Thought* Chicago, London : University of Chicago press.
- NUÑEZ, R. E., SWEETSER, E. (2006). 'With the Future Behind Them: Convergent Evidence from Aymara Language and Gesture in the Crosslinguistic Comparison of Spatial Construals of Time', *Cognitive Science*, 401-450.
- WILKINS, D. (2003). 'Why Pointing With The Index Finger Is Not a Universal', *Pointing Where Language, Culture, and Cognition Meet*, Kita, Sotaro eds, Mahwah, London : Lawrence Erlbaum Associates, 171-216.

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 effectué 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.

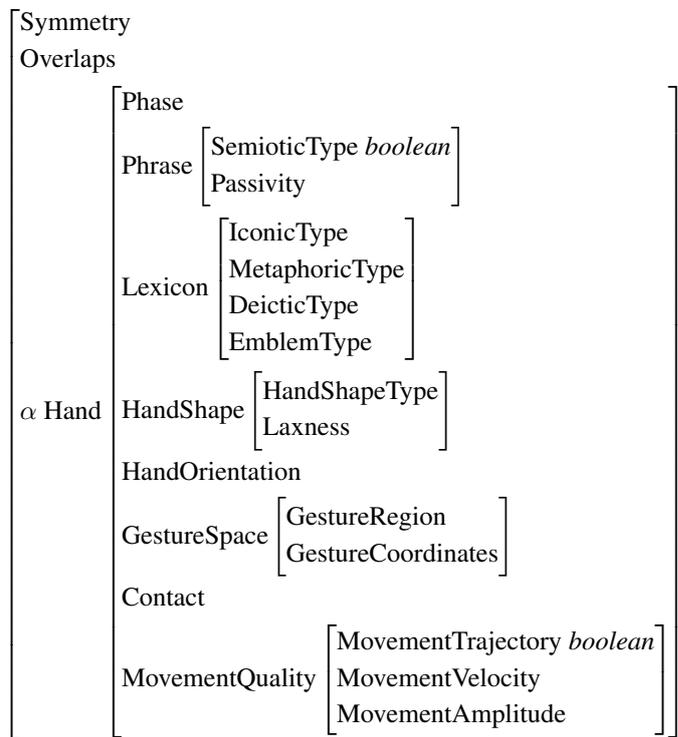


FIG. 2 – Schéma d’encodage des gestes manuels utilisé dans DEGELS1 (adapté d’OTIM)

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

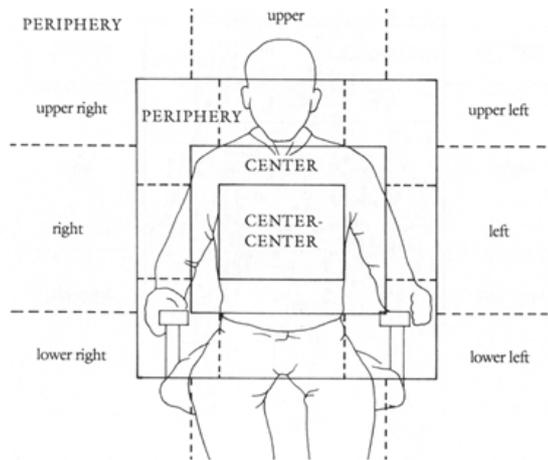


FIG. 3 – Espace de gestualisation (repris de McNeill, 1992 : 89)

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.

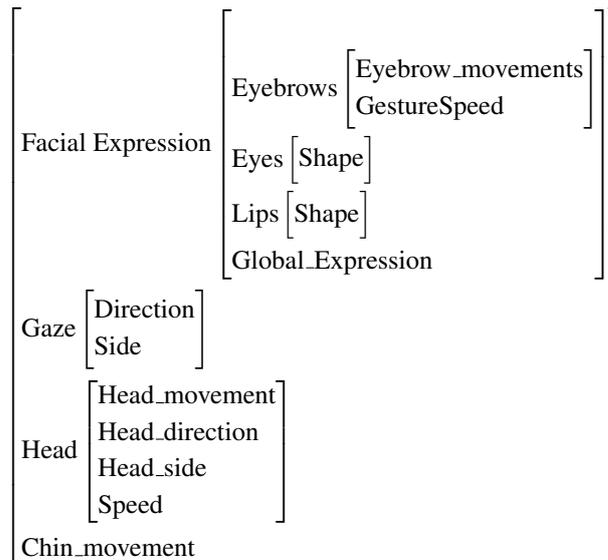


FIG. 5 – Schéma d’encodage pour la tête (adaptation du schéma d’encodage OTIM) utilisé dans DEGELS1

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”.

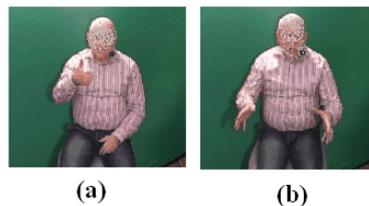


FIG. 6 – Deux gestes pouvant être mis en relation avec un affilié lexical dans DEGELS1

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 :

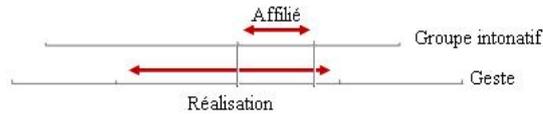


FIG. 7 – Alignement geste / parole dans le corpus CID

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.

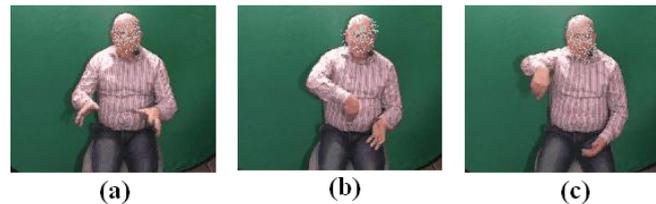


FIG. 9 – Affiliation d'une suite de trois gestes manuels produits par un locuteur

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

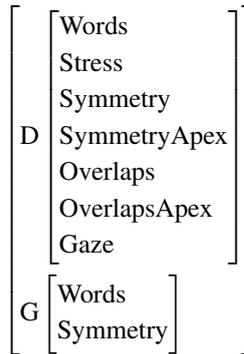


FIG. 11 – Schéma d’encodage pour l’annotation de l’affiliation dans DEGELS1

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.

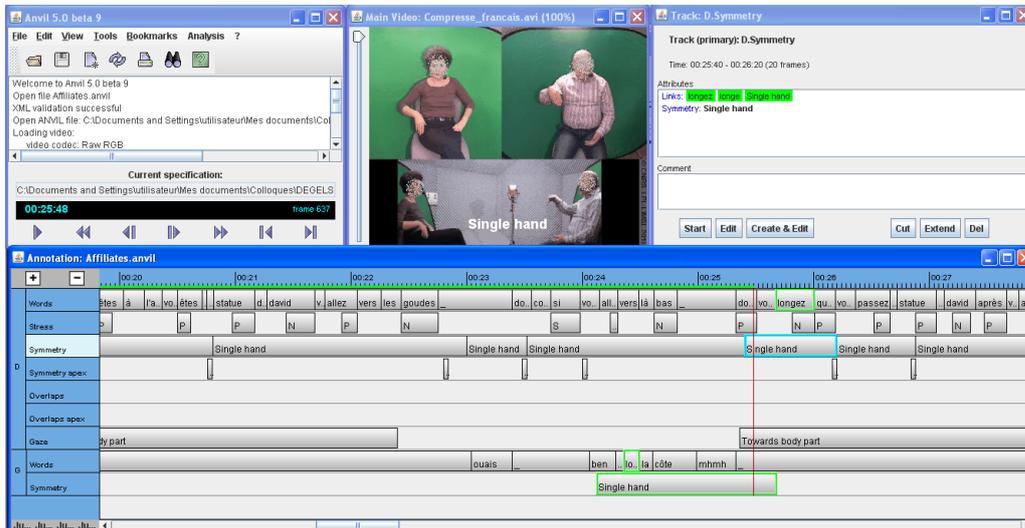


FIG. 12 – Fenêtre d’annotation d’Anvil pour le fichier “Affiliations”

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.

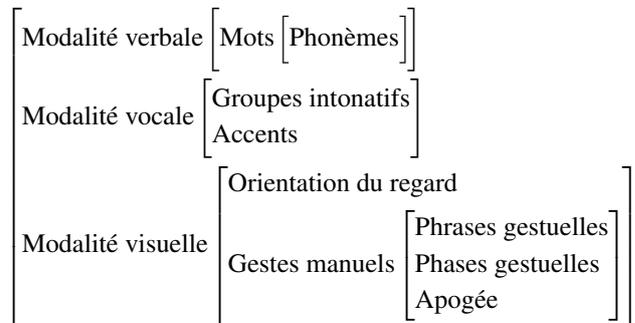


FIG. 13 – Unités annotées dans les différentes modalités

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 est 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 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

Types de gestes et utilisation de l'espace gestuel dans une description spatiale : méthodologie de l'annotation

Marion Tellier, Mathilde Guardiola, Brigitte Bigi

Laboratoire Parole et Langage, UMR 6057, CNRS & Université de Provence,
5 avenue Pasteur - 13604 Aix en Provence
marion.tellier@lpl-aix.fr , mathilde.guardiola@lpl-aix.fr , brigitte.bigi@lpl-aix.fr

Résumé Cet article présente la méthodologie utilisée par une équipe de 3 chercheuses du Laboratoire Parole et Langage pour annoter des corpus multimodaux. Nous décrirons à la fois le découpage automatique du signal audio en unités inter-pausales (IPU) pour segmenter la parole, les conventions de la Transcription Orthographique Enrichie (TOE) pour le verbal/sonore et enfin la typologie utilisée pour l'annotation du geste coverbal et l'espace gestuel. Ce travail nous permet de faire le bilan des techniques actuellement utilisées et développées dans notre laboratoire et au sein de l'ANR OTIM sur l'annotation multimodale de corpus. Il met également au jour un certain nombre de problèmes posés par l'annotation gestuelle.

Abstract This paper presents the methodology used by a team of 3 researchers of the Laboratory Parole et Langage to code multimodal data. We will first describe how we segmented the audio signal in Inter-Pausal Units (IPU), and then the conventions used to transcribe oral/verbal aspects with the Enriched Orthographic Transcription. We will also present the typologies used to code co-speech gestures and gesture space. This study enables us to introduce and discuss the techniques developed in our lab and in the OTIM project to code multimodal data. It also enables us to shed light on several current issues for gesture coding.

Mots-clés : multimodalité, geste, annotation, Transcription Orthographique Enrichie

Keywords: multimodality, gesture, coding, Enriched Orthographic Transcription

1 Objets de l'annotation

Nous avons transcrit et annoté le corpus en français. Préalablement à toute annotation, le signal de parole a été découpé automatiquement en unités inter-pausales (Interpausal Unit - désormais IPU -). Il s'agit de blocs de paroles bornés par des pauses silencieuses d'au moins 200 ms (durée variable selon les langues). Ensuite, la partie verbale a été transcrite avec les conventions de la Transcription Orthographique Enrichie (TOE) élaborées pour l'annotation du corpus CID (Corpus of Interactional Data) développées au LPL et approfondies dans le cadre du projet ANR OTIM (BLAN08-2_349062 ; <http://www.lpl-aix.fr/~otim/>). L'intérêt de cette transcription est qu'elle permet de rendre compte de manière orthographique de phénomènes typiques de l'oral tels que les pauses remplies (euh, hum...), les amorces, les mots tronqués, les répétitions, les faux-départs, les rires, les prononciations atypiques... cette transcription rend possible l'alignement temporel des phonèmes nécessaire aux analyses phonétiques et permet également une automatiser des annotations morpho-syntaxiques et syntaxiques (Bertrand & al. 2008). Ceci a été fait pour les deux locuteurs.

Au niveau du geste, notre attention s'est portée sur tous les gestes coverbaux produits uniquement avec les mains. Nous avons annoté les types de gestes ainsi que l'espace gestuel dans lequel chaque geste est produit. Pour ce corpus spécifique, il nous a semblé pertinent d'attribuer des valeurs d'indications spatiales pour tous les gestes (déictiques ou non) qui servaient à donner des indications spatiales. L'annotation des gestes a été faite pour le locuteur masculin uniquement (locuteur B).

2 Description de la méthodologie de l'annotation

2.1 Aspects techniques

La première étape a donc été la segmentation en IPU. Chaque locuteur étant enregistré sur un canal audio distinct, ce découpage est bien adapté. En raison de sa nature formelle, objective et repérable automatiquement, l'IPU est souvent utilisée sur des corpus de taille importante. Ce découpage vise à faciliter la transcription mais aussi les étapes ultérieures de conversion graphème-phonème et d'alignement temporel avec le signal audio.

La segmentation en IPU a été faite automatiquement, avec un logiciel que nous avons développé et que nous diffusons sous licence GPL (<http://www.lpl-aix.fr/~bigi/logiciels.html>). Ce programme prend en entrée un fichier son au format .wav. Il recherche les zones de silences dans une fenêtre de largeur fixe. La valeur de rms (root-mean-square) des trames de la fenêtre est comparée à une valeur seuil fixée par l'utilisateur. Ce programme fournit un répertoire contenant autant de fichiers sons que de zones de parole trouvées, et un fichier TextGrid indiquant les intervalles parole/silence. Ces derniers sont indexés : gpd_X pour les zones de parole, gpf_X pour les zones de silences (où X est la valeur d'index). Le temps passé pour cette annotation est de 15 minutes car la segmentation obtenue a été vérifiée manuellement.

Le format de fichier TextGrid a été choisi en raison de sa simplicité et de la portabilité qui en découle. C'est le format natif du logiciel Praat (Boersma, 2011), mais il peut être lu par d'autres logiciels d'annotation (notamment Elan), et par tous les éditeurs de texte. C'est un fichier ASCII qui contient une entête, suivie d'une liste d'intervalles avec : le temps de début, le temps de fin et le label de chaque intervalle. La Figure 1 présente le début du fichier obtenu après la segmentation en IPU du locuteur B.

Le fichier .wav du locuteur A a été segmenté en 15 IPU ; le fichier son du locuteur B en 14 IPU.

Les fichiers TextGrid des annotations de la parole ont ensuite été importés dans le logiciel Elan, logiciel gratuit développé au Max Planck Institute de Nijmegen. L'ensemble des annotations gestuelles a été réalisé avec ce logiciel par une troisième annotatrice. Ce travail a pris environ 5 heures.

7 pistes ont été définies : TOE locuteur A, TOE locuteur B, Tokens (= Mots du locuteurs B), Type de geste principal locuteur B, Type de geste secondaire locuteur B, Valeur indication spatiale et Espace gestuel. L'utilisation de deux pistes pour annoter les gestes permet de rendre compte des superpositions de deux types de gestes. Par exemple, lorsque un battement est superposé sur un métaphorique, la piste *geste principal* étiquette le métaphorique et la piste *geste secondaire* montre la durée pendant laquelle

apparaissent les battements superposés. Cette deuxième piste sert aussi à indiquer lorsqu'un geste a deux dimensions (une dimension iconique et une dimension déictique, par exemple). Pour faciliter la lecture, des couleurs pastel ont été attribuées aux pistes et des couleurs vives pour les types de gestes (voir Figure 1).

Les pistes « Type de geste principal » et « Type de geste secondaire » sont des annotations qui portent sur le sens du geste. La piste « valeur indication spatiale » donne une indication fonctionnelle et la piste « espace gestuel » apporte des informations relatives aux coordonnées spatiales du geste.

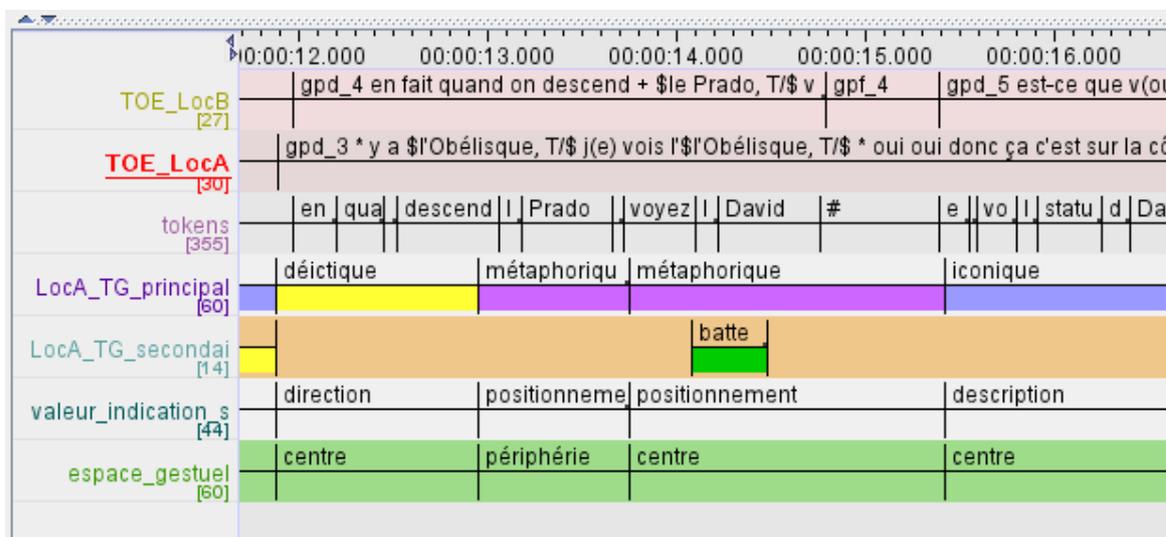


Figure 1 : Pistes définies pour l'annotation du geste

2.2 Aspects théoriques

Transcription des aspects verbaux et sonores

Pour la transcription de la parole, nous avons opté pour la Transcription Orthographique Enrichie (TOE) qui a été développée pour transcrire le corpus CID : Corpus of International Data (Bertrand & al, 2008). Elle permet de rendre compte de manière orthographique de la plupart des phénomènes typiques de l'oral ce qui facilite ensuite la détection de phonèmes par les aligneurs automatiques.

Les conventions de la TOE sont basées sur celles développées à l'origine par le GARS : Groupe Aixois de Recherche en Syntaxe (Blanche-Benveniste & Jeanjean 1987). Il n'y a ni ponctuation, ni correction de la syntaxe (les phénomènes du français parlé sont reproduits fidèlement). Le groupe OTIM (Blache & al, 2008 ; Blache & al, 2010) y a ajouté des précisions sur des spécificités de prononciation avec si besoin des transcriptions codées en pseudo-orthographe. Les caractéristiques principales de la TOE utilisées pour la transcription de ce corpus sont résumées ci-après (Tableau 1).

```
File type = "short ooTextFile"
Object class = "TextGrid"

0
87.92

<exists>
1

"IntervalPiste"
"TOE"
0
87.92
30
0
4.6012
"dummy"
4.6012
```

```

5.5458
"gpd_1 "
5.5458
6.8052
"gp_f_1 "
6.8052
8.0647
"gpd_2 "
8.0647
11.8957
"gp_f_2 "
    
```

Figure 2 : Extrait du textgrid obtenu après segmentation en IPU

Phénomène	Convention	Exemples
Elision	<i>On met entre parenthèses ce qui aurait dû être prononcé.</i>	<i>i(l)s sont, p(u)is, t(u) as □</i>
Prononciation particulière	<i>On utilise la forme [orthographe, prononciation].</i>	<i>[je suis, chui], [stats, stateu]</i>
Amorce	<i>C'est une disfluece, phénomène propre à l'oral, notées par un tiret '-' à la fin. □</i>	<i>co- co- comment tu veux que je fasse, i- ils</i>
Noms propres	<i>entre dollars suivi du type de nom propre □</i>	<i>\$la statue d(u) David, T/\$</i>
Séquences incompréhensibles	<i>par une astérisque</i>	<i>et le * il est parti</i>

Tableau 1 : Exemples de la convention pour la TOE

Il va de soi que certains de ces phénomènes peuvent être imbriqués. Ci-après, deux exemples d'IPU transcrites en TOE du locuteur B :

gpd_1 ben [di(s)ons, dzon] si on si on est à le à \$l'Escale Borély, T/\$

gpd_8 et là à [c(e), ss] rond-point sur la droite c'est plein d(e) restaurants euh ça s'appelle \$l'Escale Borély T/\$

La transcription de la parole a été faite manuellement selon les conventions de la TOE par une seconde annotatrice avec le logiciel Praat. Cette annotation a pris 3 heures environ. La transcription orthographique enrichie permet de générer, par une conversion graphème-phonème, une suite de phonèmes la plus proche possible de ce qui a réellement été prononcé par le locuteur. Ce sont les outils de phonétisation et d'alignement automatique qui déterminent l'annotation phonétique. La première étape consiste en une conversion graphème-phonème. Le convertisseur graphème-phonème (phonétiseur) permet de retranscrire des phrases de leur forme orthographique sous leur forme phonétique. Cette étape de conversion est destinée à produire la séquence de phonèmes nécessaire à l'aligneur. À partir de la TOE, la transcription orthographique truquée seule est fournie au convertisseur qui s'appuie sur un dictionnaire de formes fléchies et un ensemble de règles modifiables (exception, liaison, etc.). En sortie, le convertisseur fournit une suite de tokens phonétisés en SAMPA. Ensuite, l'aligneur employé, développé au LORIA par D. Fohr et Y. Laprie, est basé sur des modèles de Markov cachés (Hidden Markov Models). L'alignement est effectué IPU par IPU. À partir de la séquence de phonèmes et du signal audio, l'aligneur fournit en sortie la localisation temporelle de chaque phonème sur le signal. Il fournit également l'alignement temporel des mots. Ce dernier est particulièrement intéressant pour l'annotation du geste car il permet de voir aisément la synchronie entre geste et mot et il facilite le travail d'annotation gestuelle.

Étant donné le découpage en IPU, les transcritteurs n'ont pas dû noter les pauses silencieuses, exceptées celles internes aux IPU (donc inférieures à 200 ms mais perceptibles). Les chevauchements de parole, qui

peuvent être repérés automatiquement grâce aux IPU, n'ont pas été transcrits non plus. La transcription spécifie par ailleurs certaines réalisations phonétiques particulières. A priori ces réalisations phonétiques en parole conversationnelle apparaissent plus étendues, en forme et en type, que celles habituellement recensées sur des corpus de parole lue ou contrôlée. Pour ces derniers, une grande partie des phénomènes phonétiques particuliers (type élision ou réduction par exemple) est automatisable par implémentation de règles puis constitution d'un lexique de variantes systématiques testées lors de la phase d'alignement (« je suis » réalisé /SHi/ ; « je sais » réalisé /Se/).

En revanche, à côté de ces réalisations phonétiques particulières habituelles, on peut constater des élisions non standards, des substitutions et/ou ajouts de phonèmes, etc. (« expérience » réalisé /perja~s/ ; « demande » réalisé /ma~/), qui peuvent rendre la constitution d'un lexique inefficace en raison d'un trop grand nombre d'entrées et de variantes possibles par entrée. De plus, les schwas réalisés (en particulier en finale) ou les sons épenthétiques sont aussi spécifiés (« rappelle quand » réalisé /rapel2ka~t2/).

Ces différentes contraintes ont conduit à opérer une désambiguïsation manuelle préalable au stade de la transcription, pour tenter de pallier les limites des outils actuellement disponibles et davantage adaptés à de la parole lue ou contrôlée sur du français standard. C'est en ce sens que nous parlons de transcription orthographique enrichie, à partir de laquelle deux transcriptions sont automatiquement dérivées :

- une transcription orthographique standard, dont sont dérivés les « tokens orthographiques », est destinée aux modules d'analyse textuelle.
- une transcription orthographique « truquée », dont sont dérivés les « tokens phonétiques », est destinée au convertisseur graphème-phonème.

L'usage de la TOE n'a pas d'influence directe sur l'annotation gestuelle et son analyse. En revanche, elle permet de générer automatiquement des annotations syntaxiques et phonétiques permettant une analyse multi-niveaux du discours. Cette analyse n'a pas été faite pour le présent corpus.

Transcription des aspects gestuels

Les types de gestes ont été annotés selon la typologie de McNeill (1992, 2005) c'est-à-dire *déictique* (Figure 3), *iconique* (Figure 4), *métaphorique* et *battements*. Elle a été enrichie par 4 autres types de gestes : les *emblèmes*, les *Butterworth*, les *interactifs* et les *gestes avortés* (voir Tableau 2). Cette typologie a déjà été utilisée par Tellier et Stam (2010). Un vocabulaire contrôlé a été utilisé pour l'annotation des types de gestes principaux et secondaires, il est constitué des 8 types de gestes mentionnés dans le Tableau 2.

Déictique	<i>Geste de pointage</i>
Iconique	<i>Geste illustratif d'un concept concret</i>
Métaphorique	<i>Geste illustratif d'un concept abstrait</i>
Battement	<i>Geste rythmant la parole, sans contenu sémantique</i>
Emblème	<i>Geste culturel, conventionnel</i>
Butterworth	<i>Geste de recherche lexicale</i>
Interactif	<i>Geste adressé à l'interlocuteur pour la gestion de l'interaction</i>
Avorté	<i>Geste esquissé mais avorté</i>

Tableau 2 : Typologie des gestes utilisée pour l'annotation



Figure 1 : [vous allez vers les Goudes]



Figure 2 : [y a un rond point]

Nous avons également annoté l'utilisation de l'espace gestuel en utilisant les 4 zones principales définies par McNeill, 1992 (voir Figure 5) à savoir *centre centre*, *centre*, *périphérie* et *extrême périphérie*. L'annotation de l'espace gestuel rend possible l'évaluation de la taille des gestes (amples ou non) et permet de déterminer si un locuteur utilise particulièrement une zone plutôt qu'une autre. Afin de rendre compte fidèlement de la taille des gestes, nous avons choisi d'étiqueter l'espace gestuel d'un geste en prenant le point le plus ample du mouvement (quelque soit la phase). Ainsi si un geste démarre dans le *centre*, se déplace dans *l'extrême périphérie* et se termine dans le *centre centre*, l'espace gestuel codé sera *extrême périphérie* car c'est le point le plus ample. Cela correspond la plupart du temps au « stroke » du geste mais ce n'est pas systématique. Autre choix méthodologique, c'est la main qui sert de point de référence pour l'espace gestuel sauf dans les cas où les doigts sont signifiants et qu'ils sont dans une autre zone que le reste de la main. La Figure 6 exemplifie ce type de situation : la majeure partie de la main est dans la zone *centre centre* alors que le pouce est dans le *centre*, or dans ce cas le pouce est l'élément signifiant du geste, c'est pourquoi l'espace gestuel est codé *centre*. L'annotation de l'espace gestuel pose aujourd'hui beaucoup de problèmes car la grille de McNeill se révèle insuffisante principalement parce qu'elle est en 2 dimensions alors que le geste est produit dans un espace en 3 dimensions. Ainsi, un geste produit dans l'espace *centre centre* avec la main proche du ventre est codé comme un geste produit dans le même espace mais avec le bras tendu alors que ces deux gestes n'ont pas la même amplitude. Voilà pourquoi nous avons rajouté une catégorie *bras tendu devant* dans le vocabulaire contrôlé pour coder l'espace gestuel.

TYPES DE GESTES ET UTILISATION DE L'ESPACE GESTUEL DANS UNE DESCRIPTION SPATIALE :
MÉTHODOLOGIE DE L'ANNOTATION

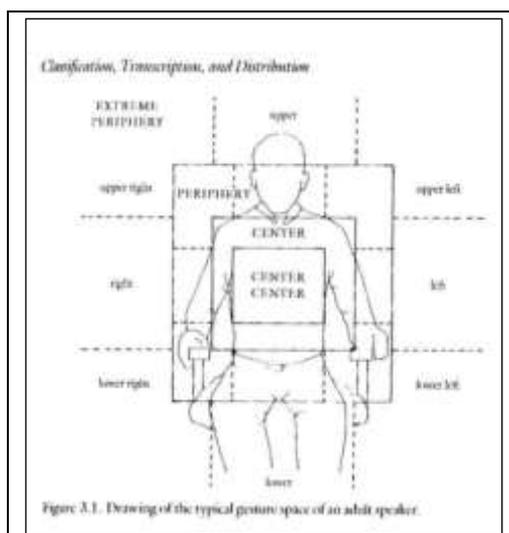


Figure 5 : L'espace gestuel (McNeill, 1992 : 89)

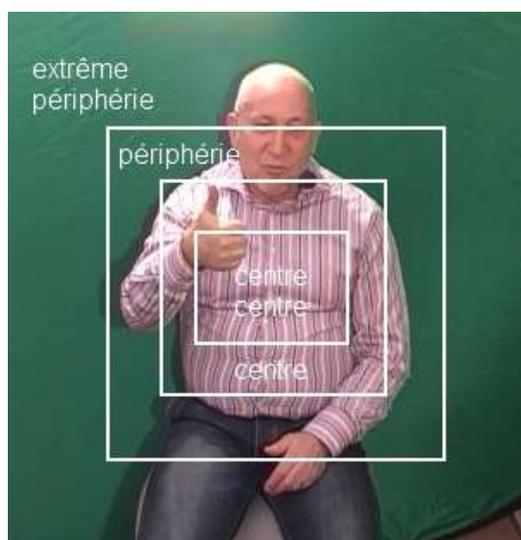


Figure 6 : [vous avez la statue du pouce là]

Pour les besoins de ce corpus, il nous a semblé intéressant d'attribuer une valeur d'indication spatiale aux gestes qui étaient utilisés pour donner des indications géographiques. Nous avons créé un vocabulaire contrôlé avec les quatre valeurs suivantes : *anaphorique* (lorsqu'un geste fait référence à un geste précédemment produit), *description* (lorsque le geste sert à décrire un point de repère, dans sa forme par exemple), *direction* (lorsque le geste donne un sens ou une direction) et *positionnement* (lorsque le geste place les éléments les uns par rapport aux autres). Cette typologie ne préexistait pas à l'annotation mais a été élaborée pendant le travail sur ce corpus.

Le Tableau 3 synthétise les vocabulaires contrôlés utilisés dans le schéma d'encodage pour l'annotation du geste coverbal.

Type de geste	<i>avorté</i> <i>battement</i> <i>Butterworth</i> <i>déictique</i> <i>emblème</i> <i>iconique</i> <i>métaphorique</i>
Espace gestuel	<i>centre centre</i> <i>centre</i> <i>périphérie</i> <i>extrême périphérie</i> <i>bras tendu devant</i>
Valeur indication spatiale	<i>anaphorique</i> <i>déscription</i> <i>direction</i> <i>positionnement</i>

Tableau 3 : Synthèse du schéma d'encodage pour le geste

3 Résultats

3.1 Au niveau du verbal sonore

La TOE du corpus pour les locuteurs A et B durant la minute transcrite rend compte d'un certain nombre de phénomènes. Ils sont résumés dans le tableau 4.

Phénomène	Locuteur A	Locuteur B
Elision	5	12
Prononciation particulière	0	2
Amorces	0	1
Pauses courtes	2	5
Noms propres (toponymes)	3	20
Inaudible	2	3

Tableau 4 : TOE du corpus pour les locuteurs A et B

On remarque que le locuteur B utilise un grand nombre de noms propres, tous des toponymes, ce qui est fortement lié à la tâche de maptask qui élicite ce genre d'items lexicaux. Il y a également 12 élisions, phénomène propre à l'oral. La qualité de l'enregistrement permet d'avoir très peu de séquences inaudibles ou inintelligibles (notées par « * » dans la transcription, afin de signaler à l'aligneur de ne pas aligner ces séquences sur le signal).

3.2 Au niveau du gestuel

Soixante gestes principaux ont été annotés et 14 gestes secondaires. Parmi ces derniers, on trouve 10 séries de battements superposés ce qui donne un total de 70 gestes. On remarque donc que la très grande majorité des battements sont superposés à d'autres gestes, principalement à des déictiques, iconiques et

TYPES DE GESTES ET UTILISATION DE L'ESPACE GESTUEL DANS UNE DESCRIPTION SPATIALE :
MÉTHODOLOGIE DE L'ANNOTATION

métaphoriques. Vingt-huit gestes du locuteur masculin sont des déictiques, ce qui est bien sûr lié à la maptask. On remarque également dans le Tableau 5 une quantité importante d'iconiques utilisés principalement pour décrire des lieux.

Types de geste	Nombre gestes principaux	Nombre gestes secondaires
Avorté	2	
Battement	1	10
Butterworth	3	
Déictique	28	2
Emblème	1	
Iconique	16	2
Métaphorique	11	

Tableau 5: Nombre de gestes par type

Au niveau de l'espace gestuel, sur les 70 gestes (60 principaux + 10 battements superposés), 23 sont produits dans le centre et 25 dans la périphérie (voir Tableau 6).

Espace gestuel	Nombre de gestes
Bras tendu devant	0
Centre centre	8
Centre	23
Périphérie	25
Extrême périphérie	14

Tableau 6 : Répartition des gestes dans l'espace gestuel

En regardant la répartition *type de geste/espace gestuel*, on constate que le *centre centre* est assez peu utilisé et que n'y sont produits que des Butterworth et des iconiques. L'ensemble des iconiques est principalement réparti sur les 3 plus petits espaces. Les déictiques, quant à eux, sont répartis entre les 3 plus grands espaces. Il est intéressant de noter que la majorité des gestes produits dans l'extrême périphérie (9 sur 12) sont des déictiques (voir Figure 7). Une fois encore cela peut être dû à la tâche.

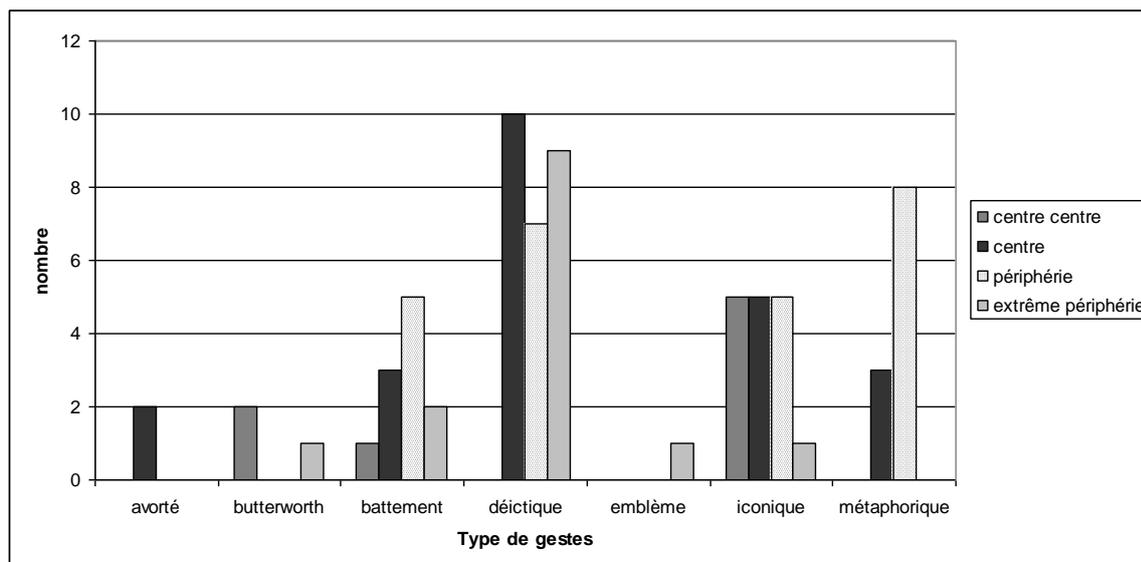


Figure 7 : Répartition des types de gestes dans l'espace gestuel

Il est difficile de tirer des conclusions sur un corpus si réduit. Cependant, nous pouvons faire quelques comparaisons avec la répartition des types de gestes dans l'espace gestuel étudiée par McNeill (1992 : 88). Ce dernier remarque en effet que les iconiques sont surtout produits dans les régions *centre centre* et *centre* et que les déictiques s'étendent dans la *périphérie* et l'*extrême périphérie*. Il note également que les battements sont répartis dans l'espace. En revanche, la plupart des métaphoriques qu'il a analysés sont produits dans le *centre* alors que dans le présent corpus, ils sont plus nombreux en *périphérie*. Cela dit, il convient de ne pas pousser la comparaison trop loin dans la mesure où la tâche proposée par McNeill à ses sujets est une tâche de narration (qui élicite des gestes et une utilisation de l'espace différents de la *maptask*). De plus, la variation interindividuelle dans la production gestuelle est reconnue pour être très importante. En somme la production gestuelle est tâche dépendante et sujet dépendante.

Enfin, nous avons attribué des fonctions à 44 gestes utilisés pour indiquer des directions. Le graphique de la Figure 8 montre la répartition statistique de l'attribution de ces fonctions selon les types de geste. La fonction *anaphorique* permet de faire référence à un élément déjà mentionné dans le discours et concerne 7 gestes (5 déictiques et 2 iconiques), la *description* sert à décrire un lieu et concerne également 7 gestes (6 iconiques et 1 métaphorique). La fonction de *direction* permet d'indiquer des sens de déplacements et concerne 15 gestes (dont 11 déictiques) et enfin la fonction de *positionnement* qui permet au locuteur de localiser des éléments dans l'espace gestuel est représentée par 15 gestes (dont 10 déictiques). Le lien entre type de gestes et fonction apparaît très cohérent. On s'attend en effet à ce que les déictiques apparaissent dans la formulation de directions et des positionnements. Ils servent également à faire des références anaphoriques. Enfin, on s'attend à ce que des iconiques soient privilégiés dans la description d'items.

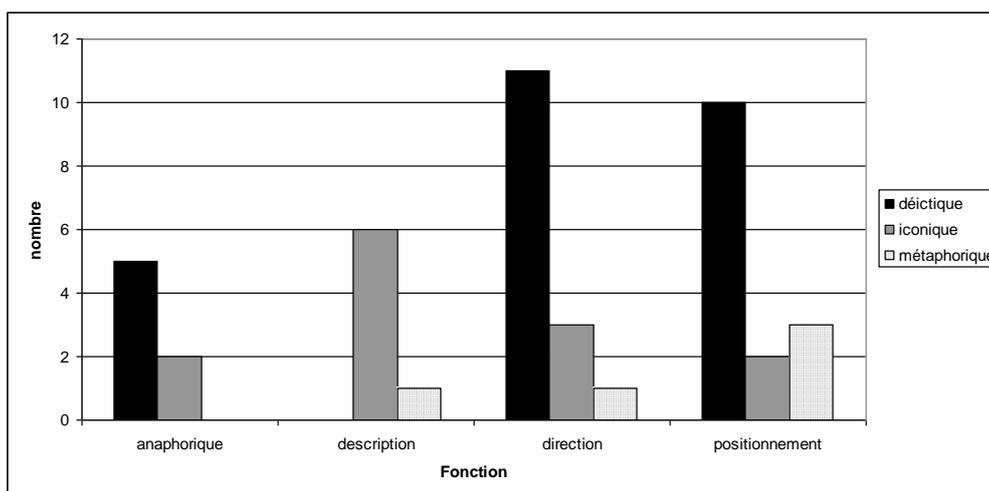


Figure 8 : Attribution des fonctions en fonction des types de geste

3.3 Interaction verbale/gestuelle

L'annotation en token (mots) facilite comme nous l'avons déjà évoqué l'annotation gestuelle, notamment pour la synchronisation geste/mot. Elle facilite également le calcul du taux gestuel qui est souvent nécessaire dans les études sur les co-verbaux. Il se calcule en divisant le nombre total de gestes par le nombre total de mots (donné par Elan dans les statistiques). Dans ce corpus le taux est de 0,2 ce qui est assez élevé. Bien sûr, sur ce corpus, calculer le taux gestuel ne présente pas vraiment d'intérêt mais dans le cas où l'on voudrait comparer le taux gestuel de différents locuteurs engagés dans une même tâche langagière, ce serait une donnée pertinente.

Le fait d'avoir une piste présentant les différents mots sur la ligne temporelle rend plus précise la recherche de certaines co-occurrences gestes/mots. Par exemple, si l'on regarde dans les statistiques produites par Elan, on voit que le nom « David » apparaît 4 fois. On peut ensuite rechercher s'il est accompagné de gestes et, le cas échéant, de quels types il s'agit. La première mention du « David » (à 14.22) est accompagnée d'un métaphorique car le locuteur suppose que son interlocutrice connaît cette référence. Comme ce n'est pas le cas, il accompagne sa seconde occurrence du nom « David » en la replaçant dans un contexte explicité « la statue du David » (15.91) et en produisant un geste iconique avec cette expression qui illustre le référent de manière plus concrète. Ceci fait, les deux mentions suivantes de la statue (21.6 et 27.1) sont accompagnées d'un geste déictique car il ne s'agit plus de décrire un élément du décor mais de s'en servir pour localiser le parcours.

On peut aussi imaginer qu'une étude sur la synchronisation entre gestes et parole serait facilitée par ce type de transcription en token-mots.

4 Conclusion

Ce travail sur corpus nous a permis de mettre à l'épreuve différentes techniques utilisées au sein du Laboratoire Parole et Langage et au sein de l'ANR OTIM. Des techniques et outils tels que la segmentation automatique en IPU et la TOE facilitent l'annotation semi-automatique de corpus et permettent un traitement plus rapide de certaines modalités notamment vocales (non abordées ici). Il reste que l'annotation du geste demeure un processus long et instable car les différentes grilles et typologies ne font pas toujours consensus et demandent une large part d'interprétation du codeur. Même si la typologie de McNeill (1992) est encore à l'heure actuelle la plus usitée par la communauté, elle laisse de côté certains gestes et les frontières entre certains types demeurent encore poreuses. Cela dit, le fait qu'une grande part de la communauté scientifique des gestualistes utilise cette typologie permet de faire des comparaisons entre différentes études et différents résultats. Une base commune est donc indispensable.

Le travail sur l'espace gestuel demeure encore très peu satisfaisant et présente trop d'imprécisions. L'utilisation de capteurs de mouvements pour enregistrer des coordonnées spatiales du geste devrait pouvoir

apporter plus de précision sur ce type de données mais on peut se demander si cette technique n'est pas trop invasive et n'affecte pas la production gestuelle.

Références

BERTRAND R., BLACHE P., ESPESSE R., FERRÉ G., MEUNIER C., PRIEGO-VALVERDE B., RAUZY S. (2008). Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement automatique des langues (TAL)* 49 (3), 105-134.

BLACHE P., BERTRAND R., BIGI B., BRUNO E., CELA E., ESPESSE R., FERRÉ G., GUARDIOLA M., HIRST D., MAGRO E.-P., MARTIN J.-C., MEUNIER C., MOREL M.-A., MURISASCO E., NESTERENKO I., NOCERA P., PALLAUD B., PRÉVOT L., PRIEGO-VALVERDE B., SEINTURIER J., TAN N., TELLIER M., RAUZY S. (2010). Multimodal Annotation of Conversational Data. *Proceedings of Linguistic Annotation Workshop* (2010 juillet 15-16 : Uppsala, SWEDEN), 186-191.

BLACHE P., BERTRAND R., FERRÉ G. (2008). Creating and exploiting multimodal annotated corpora. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)* (Marrakech, MOROCCO). Marrakech: ELDA.

BOERSMA P., WEENINK D. (2011). *Praat: doing phonetics by computer*. <http://www.praat.org>

BLANCHE-BENVENISTE C., JEANJEAN C. (1987). *Le français parlé, Transcription et édition*. Paris :Didier-Erudition/ InaLF, 2e éd.

MCNEILL D. (1992). *Hand and Mind : What gestures reveal about thought*. Chicago : The University of Chicago Press.

MCNEILL D. (2005). *Gesture & thought*. Chicago: The University of Chicago Press.

SLOETJES H., WITTENBURG, P. (2008). Annotation by category - ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)* (Marrakech, MOROCCO). Marrakech: ELDA.

TELLIER M., STAM G. (2010). Découvrir le pouvoir de ses mains : La gestuelle des futurs enseignants de langue. In Actes du Colloque *Spécificités et diversité des interactions didactiques : disciplines, finalités, contextes*. INRP, 24-26 juin 2010, Lyon.

Gestualité et LSF

De l'espace métrique à l'espace de signation, à vos GP(e)S

F. Lefebvre-Albaret J. Segouat
WebSourd, 99 route d'Espagne, 31 100 Toulouse
francois.lefebvre-albaret@websourd.org
jeremie.segouat@websourd.org

Résumé. L'utilisation intensive de l'espace est l'une des spécificités des langues des signes (LS). Au travers de différents types de pointage, il est possible de placer des concepts dans l'espace ou d'y faire référence. Certaines des stratégies de positionnement en LS peuvent également être identifiées dans la gestuelle co-verbale. L'étude présentée dans cet article propose plusieurs hypothèses sur la manière dont l'espace de signation ou l'espace gestuel peuvent être utilisés pour décrire une structure de l'espace réel. En particulier, nous nous intéressons à la manière dont la relation entre ces espaces évolue au cours d'une description d'itinéraire. Nos résultats seront mis en regard de la théorie des visées de C. Cuxac (Cuxac, 2000) et nous montrerons en conclusion, en quoi notre analyse peut s'inscrire plus largement dans une réflexion sur la génération automatique d'énoncés en langue des signes.

Abstract. The intensive use of space is one of the sign languages (SL) main feature. It is possible to place concepts in the space or to refer to them, in using several types of pointing gestures. Some of those pointing strategies can also be identified in coverbal gestures. The article presents several hypothesis on the way to use the signing or the gestural spaces to describe some structures in the real space. In particular, our analysis focus on the variation of the association between those spaces during an itinerary description. Our results will be interpreted in the light of the C. Cuxac theory (Cuxac, 2000). We show as a conclusion, how our analysis takes place in a larger reflexion on sign language utterance automatic generation.

Mots-clés : Langue des Signes Française, geste coverbal, espace de signation.

Keywords: French Sign Language, coverbal gesture, signing space.

1 Introduction

Les Langues des Signes (LS) se distinguent des langues vocales par de nombreuses caractéristiques parmi lesquelles l'utilisation intensive de l'espace (Coletta & Millet, 2002; Liddell, 2003). La mise en espace, ou spatialisation est présente dans la gestuelle coverbale en tant qu'illustration ou complément de la parole, mais est réellement structurante pour les LS. Les relations spatiales entre les éléments peuvent représenter par exemple des relations de temps, de hiérarchie, de composition, de parallélisme ou d'espace. C'est sur cette dernière dimension que s'axe notre réflexion.

L'analyse que nous présentons s'appuie sur des annotations de deux vidéos où deux locuteurs décrivent un itinéraire. Ces explications sont extraites de corpus de dialogue entre deux locuteurs s'exprimant en français pour la première vidéo et en Langue des Signes Française (LSF) pour la seconde. Nous disposons à la fois d'une vue de face de chacun des intervenants, ainsi que d'une vue de profil. Les noms de lieux contenus dans ces deux extraits d'une minute environ permettent de se référer à la cartographie réelle.

Nous sommes ainsi en présence de deux corpus où un espace métrique en trois dimensions (dans lequel se trouve l'itinéraire à décrire) est projeté dans un espace gestuel par l'utilisation de signes ou de gestes plus ou moins standardisés qui s'inscrivent dans un espace à quatre dimensions (3D+t). Il existe donc une application de l'espace métrique vers l'espace des gestes coverbaux ou l'espace de signation permettant de présenter l'information topologique de manière efficace à l'interlocuteur. Cette application n'est ni continue, ni linéaire, ni bijective et permet pourtant de décrire les lieux sans ambiguïté. Existe-t-il des règles de "projections" entre ces

espaces à un instant donné ? Ces règles sont-elles similaires pour un énoncé en LS et pour une production vocale accompagnée de gestes co-verbaux ? Que se passe-t-il lors d'un changement de l'application de projection, qu'il soit continu ou discontinu ?

Pour répondre à ces questions, nous avons relevé les positions pointées dans l'espace gestuel¹. Nous prenons dans cet article, le mot pointage dans son acception la plus large, en désignant toute opération permettant d'activer directement ou indirectement une portion de l'espace gestuel. Nous prenons donc en compte les localisations directes des signes dans l'espace, l'utilisation de pointés, le pointage par le regard, la position et l'orientation du tronc ainsi que l'utilisation de proformes spatialisés. Nous détaillons dans un premier temps les conventions d'annotations utilisées, tant pour localiser les pointages que pour les nommer et énumérons les autres paramètres annotés. Dans un deuxième temps, nous analysons les résultats en focalisant notre attention sur la manière dont les combinaisons de signaux corporels et vocaux permettent d'opérer des modifications de l'application de passage de l'espace réel à l'espace gestuel. Nous procédons ensuite à une comparaison des stratégies adoptées dans le cas des langues vocales et des langues des signes. Dans un troisième temps, nous mettons nos résultats en regard de la théorie des visées de C. Cuxac en montrant comment les observations peuvent être analysées sous forme d'une alternance de transferts qui induisent des projections spécifiques de l'espace réel dans l'espace gestuel. Nous achevons notre article par une réflexion plus large sur la manière dont ces opérations de projection pourraient être réutilisées dans un système de génération automatique d'énoncés en LSF.

2 Espaces conceptuel et gestuel

La structure de la LSF fait appel à l'utilisation intensive de l'espace (Cuxac, 2000) qui permet à la fois de positionner les différents concepts et d'y faire référence. Cette caractéristique semble être commune à toutes les LS, y compris les langues des signes émergentes (Fusellier-Souza, 2004). Les positionnements peuvent être réalisés à l'aide des différentes parties du corps du signeur et prennent différentes formes : pointages (à l'aide de la tête, du buste, du regard ou de la main), positionnement de proformes dans l'espace de signation, positionnement de signes standards. Il est important de souligner que ces stratégies présentes de manière systématique dans les LS sont également observables pour les gestes coverbaux (Calbris, 2000) et que les mains et le regard permettent là aussi de donner un complément d'information portant notamment sur l'emplacement relatif et l'étendue d'entités. Cela est en particulier valable pour la description d'itinéraires (Cosnier, à paraître), qui nous intéresse dans le cadre de cette étude.

Pour passer de l'espace réel à sa description à l'aide de gestes, le locuteur crée une représentation mentale à partir du réel. Il existe dès cette étape des biais systématiques (Tversky, 1993) relatifs à l'estimation des distances, des angles ou des alignements des éléments topologiques. Ensuite, cette représentation sera projetée dans différents espaces mentaux (Fauconnier, 2004) puis restituée dans l'espace gestuel. Une des spécificités des LS est le fait de pouvoir utiliser simultanément plusieurs projections (Perniss, 2007) qui peuvent être chacune associée à une visée dans le cadre de doubles transferts (Sallandre, 2003).

La modélisation de l'Espace de Signation (ES) a naturellement une place très importante dans le Traitement Automatique des LS (TALS). Plusieurs travaux (Braffort *et al.*, 2005; Huenerfauth, 2006) l'utilisent pour effectuer une modélisation d'énoncé respectant la syntaxe des LS. Dans le domaine de l'analyse automatique de vidéos, l'ES peut également être utilisé pour interpréter des signes comme les pointages ou les actions (Lenseigne, 2004). Rares sont en revanche les études de TALS qui considèrent la succession des différentes projections des espaces conceptuels dans l'ES. Or il s'agit, selon nous, d'une étape décisive pour parvenir à une génération ou une interprétation automatique d'énoncés plus complexes. L'objectif de notre étude est justement de proposer plusieurs pistes pour modéliser ces changements de projection en nous basant à la fois sur l'annotation de corpus en français vocal et gestuelle coverbale, et sur l'annotation de productions comparables en LSF.

3 Annotation

Notre étude est basée sur le corpus DEGELS1, qui est référencé sous l'identifiant oai:crdo.fr:crdo000767 au CRDO d'Aix-En-Provence. Pour les deux vidéos (gestuelle coverbale et LSF), nous nous intéressons aux paramètres

¹Pour rendre l'article plus lisible, nous utiliserons à partir de maintenant l'appellation *espace gestuel* pour nommer indifféremment l'espace dans lequel sont effectués les signes ou les gestes coverbaux.

manuels participant au référencement spatial. Nous notons systématiquement le rôle de chaque configuration manuelle qui modifie l'espace de gestuel. Quand celle-ci désigne une forme précise dont l'étendue peut être inférée à partir de la forme de la main, nous mesurons les coordonnées de la boîte englobante du volume décrit. Nous décrivons également pour chaque corpus la direction du regard.

Pour le corpus de gestuelle coverbale, nous notons les répliques qui sont associées à un geste structurant l'espace gestuel ou à un changement d'espace gestuel. Pour le corpus en LSF, nous notons les transferts (Cuxac, 2000).

L'ensemble de ces annotations a nécessité environ 3H pour chaque vidéo d'une minute. Nous n'avons cependant annoté que les répliques des locuteurs qui effectuaient la description du trajet.

Les différentes annotations ont été réalisées dans l'ordre suivant :

- Mesure de la boîte englobante des différents pointages² et inventaire des différents proformes à l'aide d'Excel,
- Annotation des différents proformes de pointage dans Anvil. Le critère de segmentation des pointages utilisé est la stabilisation de la configuration manuelle.
- Annotation de la direction du regard dans Anvil.
- Annotation des paroles associées à une rotation de l'espace gestuel ou à un pointage dans Anvil (pour le corpus en français seulement).
- Annotation du type de transfert dans Anvil (pour le corpus en LSF seulement).

L'ensemble des pointages annotés est récapitulé en annexes pour chaque vidéo.

3.1 Mesures de position

Le corpus a été enregistré à l'aide de trois caméras. Les deux locuteurs sont assis face à face, une caméra filme les deux locuteurs de profil, et deux caméras filment chaque locuteur de face. Comme nous ne disposons d'aucune information sur la calibration (position et réglages optiques) des caméras, nous mesurons les positions des mains de manière approximative. Nous ne prenons en compte dans notre étude que les coordonnées (x, y) de la main en projection dans le plan horizontal. La figure 1 montre de quelle manière ces positions sont estimées : X est la distance entre le bord gauche de la vidéo et le point dont les coordonnées sont à estimer sur la vue de face du locuteur, Y est la distance entre le bord gauche de la vidéo et ce même point sur la vue de profil, la vidéo a une largeur L . Nous utilisons ensuite simplement les formules $x = X/L$ et $y = Y/L$ en émettant l'hypothèse que les caméras se trouvent suffisamment loin des sujets pour négliger la calibration.

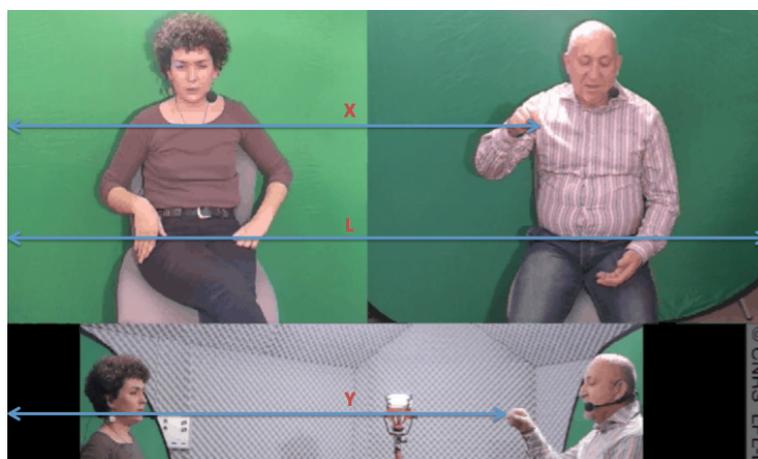


Figure 1: Principe d'approximation des coordonnées de la main dans la vidéo

²Pour chaque pointage, nous notons également quelle est l'entité réelle désignée

Les positions des mains sont mesurées manuellement en cliquant sur la vidéo à l'aide d'un logiciel créé pour l'occasion qui permet de relever à la fois le timecode de l'image courante et les coordonnées relatives des positions pointées.

3.2 Formes de la main et boîte englobante

Pour arriver à connaître précisément la projection courante de l'espace réel dans l'espace gestuel, il est utile de noter le volume des entités décrites à l'aide de proforme. Nous effectuons ces mesures à l'aide de boîtes englobantes. Pour chaque pointage, nous notons donc le proforme utilisé ainsi que la boîte englobante sur les axes x et y lorsque le volume peut être estimé à partir d'une ou des deux mains. Soulignons que le nom du proforme correspond à configuration manuelle de la (ou des) main ET à la forme décrite. L'ensemble des configurations est présenté dans les figures 2, 3 et 4.



Figure 2: Arc1 Cercle1 Cercle2 Cercle3 Cercle4 Cercle5



Figure 3: Ligne1 Point1 Point2 Point3 Port



Figure 4: Route1 Route2 Zone1 Zone2 Zone3

3.3 Autres paramètres annotés

Nous annotons également les paramètres suivants que nous savons être aussi en rapport d'une manière ou d'une autre avec l'espace gestuel :

Le regard Les valeurs utilisées sont **ES** quand le regard est dirigé vers le centre de l'espace gestuel, **Haut-gauche** quand le regard est dirigé en haut à gauche de la vidéo³, **Haut-droit** quand le regard est dirigé en haut à droite, **Haut** quand le regard est dirigé vers le haut, **D** quand le signeur regarde à droite de la vidéo, **MD** quand le signeur regarde sa main droite et **MG** quand le signeur regarde sa main gauche. Lorsque le regard n'est pas annoté, cela signifie que le locuteur regard en face de lui ou que la direction de son regard n'a pas pu être déterminée de par la résolution de la vidéo⁴.

³Nous adoptons par défaut le point de vue de la personne qui visionne la vidéo pour nommer les côtés *droite* et *gauche*, sauf quand nous précisons explicitement le point de vue.

⁴Nous sommes toutefois conscients qu'une étude plus poussée nécessiterait de bien distinguer le regard de face du regard de direction non évaluable.

Les transferts Nous utilisons la typologie des visées décrites dans (Cuxac, 2000) soit **TTF** pour les transferts de taille et de forme, **DT** pour les doubles transferts (situationnel et personnel), **TP** pour les transferts personnels et **TS** pour les transferts situationnels.

Les informations verbales Seules les informations verbales associées à un geste structurant l'espace gestuel sont annotées. Nous les notons sous forme textuelle en transcrivant uniquement les segments de phrase.

4 Estimation des projections de l'espace réel dans l'espace gestuel

Une fois les différents paramètres annotés, nous les utilisons pour tenter de retrouver à chaque instant les paramètres de passage de l'espace réel à l'espace gestuel. Dans cette section, nous exposons d'abord la formalisation théorique que nous faisons du problème, puis nous essayons de retrouver les applications de passage entre les différents espaces en utilisant les indices laissés par les locuteurs dans l'espace gestuel. Nous comparons ensuite les stratégies utilisées dans le cadre de la LSF et d'un discours accompagné de gestuelle coverbale pour effectuer des modifications dans l'application de projection de l'espace réel à l'espace gestuel.

4.1 Formulation mathématique du problème

Nous émettons dans un premier temps l'hypothèse que l'application de passage Π de l'espace réel à l'espace gestuel conserve les rapports de longueur, les angles et le sens de rotation. Dans ce cas, à chaque instant t , $\Pi(t)$ peut être décrite grâce à quatre paramètres (donc quatre degrés de liberté) :

la translation (dx, dy) entre un point de référence de l'espace réel et un point de référence de l'espace gestuel,

le rapport d'agrandissement Γ entre une longueur de l'espace réel et la longueur correspondante dans l'espace gestuel,

la rotation α entre un segment de l'espace réel et le segment correspondant de l'espace gestuel.

Chaque réalisation dans l'espace gestuel faisant référence à l'espace réel vient apporter une contrainte sur cette application, en restreignant le nombre de degrés de liberté (DDL) de l'application de projection au moment où elle est réalisée :

- Un pointage dans l'espace de signation correspondant à un point localisable dans l'espace réel supprime 2 DDL.
- Un pointage dans l'espace de signation situé sur une droite localisable dans l'espace réel (par exemple, un pointage sur une portion de route rectiligne) supprime 1 DDL.
- Une configuration manuelle indiquant une direction qu'on peut estimer également dans l'espace réel supprime 1 DDL.
- Une indication d'échelle à l'aide d'un proforme indiquant un volume dont on connaît la taille dans l'espace réel supprime 1 DDL.

A chaque instant t , nous disposons donc d'un certain nombre de mesures qui vont contraindre l'application $\Pi(t)$. Nous remarquons que dans certaines situations, le locuteur fixe trop de contraintes pour que le problème d'estimation de $\Pi(t)$ ait une solution. Prenons l'exemple de la situation décrite figure 5. Nous y voyons la figuration simultanée de l'île du château d'If à l'aide de la main gauche du locuteur indiquant à la fois l'emplacement à la taille de l'île, et la position pointée de la statue du David⁵ à l'aide de la main droite du locuteur. Cette situation est courante lors de description de scènes en LSF comme à l'aide de geste coverbaux. En raison de cette contradiction apparente entre la distance et l'échelle locale, nous choisissons de ne pas prendre en compte dans

⁵Nous émettons ici l'hypothèse que l'index touche la position de la statue du David. Une autre hypothèse serait que l'index pointe dans la direction de la statue du David et il n'y aurait alors plus de surcontrainte.

un premier temps les contraintes relatives à l'échelle provenant des proformes.

La recherche de l'application optimale de passage de l'espace réel à l'espace gestuel peut être assimilée à un problème de minimisation de la somme des déplacements de tous les points de l'espace gestuel sous les contraintes que nous venons d'énumérer plus haut. Une telle approche de problème, même si elle est intéressante d'un point de vue théorique, se révèle difficile à mettre en pratique car la résolution fait appel à un ensemble d'équations non linéaires. Nous choisissons donc une résolution approchée du problème en proposant une application Π pour chaque intervalle temporel $]t_1, t_2]$. Notre solution approchée ne fait donc pas figurer les variations continues qui ont lieu au cours de ces intervalles. Nous reviendrons tout de même sur ce point dans la section 5.

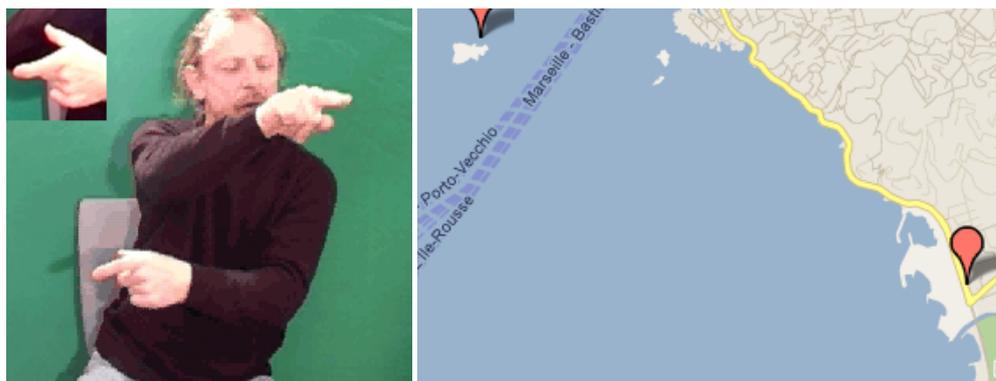


Figure 5: Exemple de simultanéité de contraintes d'associations de plusieurs points de l'espace gestuel avec l'espace réel auxquelles s'ajoute une contrainte d'échelle.

4.2 Résolution pour chaque vidéo

Dans cette section, nous faisons le parallèle entre l'espace de gestuel et la carte de Marseille via une superposition de graphiques. Cela nous permet de représenter de manière intuitive les opérations de rotation, de translation et d'agrandissement qui ont lieu entre les différentes étapes de description. En rouge, sont dessinés les éléments qui nous ont permis de recalibrer l'espace gestuel sur l'espace réel. Les résultats qui sont exposés dans les figures 6 à 10 sont à prendre avec la plus grande précaution pour plusieurs raisons :

- Les positions ont été mesurées sans tenir compte de la calibration de la caméra qui nous est inconnue.
- Nous partons du principe que l'application Π est constante sur chaque intervalle temporel.
- Lors des pointages, nous supposons que le signeur touche l'entité désignée (mais il serait également possible que la main indique uniquement sa direction).

4.2.1 Résolution pour le corpus en LSF

Le tableau 4.2.1 expose la manière dont ont été déterminés les paramètres des différentes projections. Il permet de connaître pour chaque intervalle temporel, les indices qui nous ont permis de trouver les 4 paramètres de l'application de l'espace réel à l'espace de signation.

DE L'ESPACE MÉTRIQUE À L'ESPACE DE SIGNATION

ID	intervalle (sec.)	Indices
A	12.04 - 12.20	Position (2DDL) orientation (1DLL) et échelle (1DDL) du port
B	13.08 - 13.48	Position (2DDL) orientation (1DLL) et échelle (1DDL) du port
C	13.92 - 15.16	Position du port (2DDL) orientation de la route (1DDL), portion de la route contenant des virages
D	18.24 - 18.52	Position (2DDL) orientation (1DLL) et échelle (1DDL) du port
E	19.12 - 19.36	Position de l'église (haut de la côte) (2DDL) et position du port (2DDL)
F	20.32 - 20.84	Position de l'église (haut de la côte) (2DDL) et position du port (2DDL)
G	20.84 - 21.16	Position (2DDL) orientation (1DLL) et échelle (1DDL) du port
H	21.84 - 22.04	Position (2DDL) orientation (1DLL) et échelle (1DDL) du port
I	22.36 - 23.76	Position du port (2DDL) orientation de la route (1DDL), portion de la route contenant des virages
J	24.92 - 25.24	Position de l'Île d'If (2DDL), orientation de la route (1DDL) et un point de la route (1DDL)
K	30.87 - 30.87	Position de l'Île d'If (2DDL), échelle de l'île (1DDL)
L	31.20 - 31.79	Position de l'Île d'If (2DDL), orientation de la route (1DDL) et un point de la route (1DDL)
M	34.84 - 36.27	Position de l'Île d'If (2DDL), position du David (2DDL)

Dans plusieurs intervalles (C, I, K), nous ne disposons pas d'informations suffisantes pour estimer les valeurs des paramètres de Π . Nous choisissons alors une solution pour que les variations de Π soient faibles. Ajoutons que dans le cas de l'intervalle K, l'occultation d'une des mains du seigneur ne nous permet pas de connaître avec précision sa position.



Figure 6: Corpus LSF, espaces de signation A, B et C



Figure 7: Corpus LSF, espaces de signation D, E et F



Figure 8: Corpus LSF, espaces de signation G, H et I

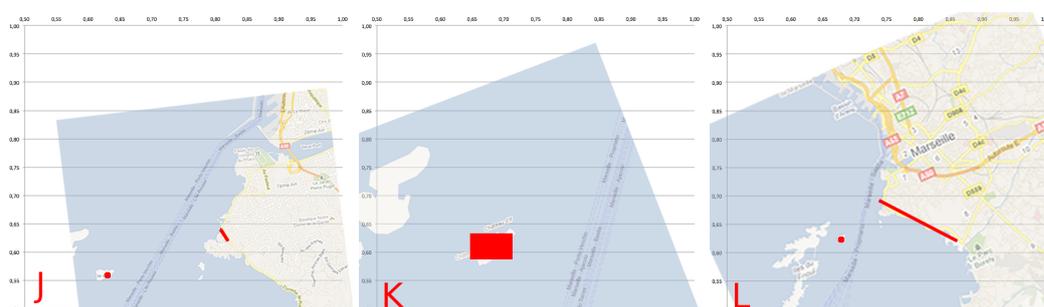


Figure 9: Corpus LSF, espaces de signation J, K et L



Figure 10: Corpus LSF, espace de signation M

4.2.2 Résolution pour le corpus en français et gestuelle coverbale

Nous avons tenté, en vain, d'effectuer le même procédé pour le corpus en français. Les vidéos permettent de connaître approximativement le paramètre de rotation, mais les autres paramètres sont très difficilement accessibles pour plusieurs raisons :

- Il n'est pas toujours facile de déterminer si les gestes utilisés véhiculent de l'information topologique. C'est en particulier valable pour la forme de la main *cercle1* utilisée à $t = 5s$,
- La plupart du temps, le locuteur n'utilise qu'une main. Dans un tel cas, il est extrêmement difficile d'avoir accès simultanément à des contraintes qui figent 4 DDLs,

Malgré tout, nous avons tenu à effectuer de la même manière que pour le corpus en langue des signes un découpage temporel en montrant le nombre d'indices spatiaux réellement disponibles à chaque étape de la description.

ID	intervalle (sec.)	Indices
N	1.72 - 11.96	Position (2DDL) et taille du rond point de l'escale Borelli (EB) (1DDL)
O	16.28 - 16.28	Position de la statue du David (2DDL)
P	19.00 - 19.00	Position d'un point de la côte (1DDL) et direction (1DDL)
Q	19.48 - 29.00	Position de la statue du David (2DDL) et position de l'escale Borelli (2DDL)
R	29.20 - 30.52	Position (2DDL) et taille du rond point de l'EB (1DDL), orientation des restaurants (1DDL)
S	37.12 - 38.12	Orientation de la route (1DDL), position de l'hippodrome (2DDL), distance route / hippodrome (1DDL)
T	40.04 - 41.07	Position de l'escale Borelli (2DDL), Position et orientation de la route (2DDL)
U	47.69 - 47.69	Position du musée d'art contemporain (2DDL)
V	50.72 - 50.72	Position du musée d'art contemporain (2DDL)
W	64.72 - 70.33	Position et orientation d'un point de la route (2DDL)

4.3 Stratégies

4.3.1 Stratégies utilisées en LSF

Il suffit de regarder les différentes superpositions entre les espaces de signations successifs et l'espace réel figuré par la carte pour s'apercevoir des variations importantes de $\Pi(t)$ au cours de la description. Nous avons tenté d'inventorier à chaque étape les moyens utilisés par le locuteur pour manipuler l'espace de signation. Nous noterons systématiquement la traduction des signes entre crochet et les proformes en italique.

Entre A et B Utilisation de la répétition du signe [PORT] pour effectuer une translation et une dilatation.

Entre B et C Utilisation d'un proforme *port* par une seule main au lieu des deux utilisées précédemment pour effectuer un changement d'échelle, déplacement du proforme *port* pour effectuer une translation, rotation de la tête vers la droite de l'image et rotation du proforme *port* pour effectuer une rotation, utilisation d'une faible vitesse de déplacement de la main pour indiquer la longueur du chemin parcouru.

Entre C et D Utilisation de la répétition du signe [PORT] pour revenir à la situation identique à l'intervalle B, insertion d'une pause pour marquer un changement de repère.

Entre D et E Utilisation d'un proforme *port* par une seule main au lieu des deux utilisées précédemment pour effectuer un changement d'échelle, déplacement du proforme *port* pour effectuer une translation.

Entre E et F Reprise d'un même tracé avec des proformes différents (*zone2* au lieu de *route2*) qui permet d'indiquer simultanément la translation, la rotation et le changement d'échelle.

Entre F et G Utilisation de la répétition du signe [PORT] pour revenir à la situation identique à l'intervalle B.

Entre G et H Utilisation d'un proforme *port* par une seule main au lieu des deux utilisées précédemment pour effectuer la translation, tracé de la forme du port avec le proforme *cercle3*.

Entre H et I Utilisation d'un proforme *port* par une seule main au lieu des deux utilisées précédemment pour effectuer un changement d'échelle, déplacement du proforme *port* pour effectuer une translation, rotation de la tête vers la droite de l'image pour effectuer une rotation du repère, utilisation d'une faible vitesse de déplacement de la main pour indiquer la longueur du chemin parcouru.

Entre I et J Conservation de la position de la route par un double pointage⁶, changement du proforme de *route2* en proforme *Point2*, utilisation d'un mouvement de pointage en cloche en parallèle avec un soulèvement du bassin, soulèvement des sourcils pour indiquer l'éloignement. L'ensemble de ces indices permet de coder un changement d'échelle.

Entre J et K Création d'une nouvelle projection liée à la description de l'île d'IF découlent directement de la description qui précède,

⁶Double pointage : point de l'espace pointé successivement par une main, puis par l'autre.

Entre K et L Conservation d'un seul proforme *zone2* par une seule main au lieu des deux utilisées précédemment pour coder un changement d'échelle. Utilisation d'un mouvement antagoniste à la situation dans la transition entre I et J.

Entre L et M Froncement des sourcils pour indiquer l'éloignement, bras en hyperextension. Ces deux événements sont des indices permettant d'inférer l'échelle.

Le tableau 4.3.1 résume les différentes stratégies utilisées par le signeur pour effectuer des changements de projection de l'espace réel dans l'espace de signation.

Type d'opération	Stratégies mises en œuvre
Grande discontinuité	Insertion d'une pause dans l'expression
Restauration d'une projection antérieure	Utilisation d'un même signe placé à un endroit proche Description d'une même forme positionnée au même endroit
Translation	Déplacement de la main représentant un référent spatial immobile : proforme ou partie d'un signe standard
Rotation	Rotation de la tête ou du buste Rotation de la main représentant un référent spatial immobile : proforme ou partie d'un signe standard
Changement d'échelle	Utilisation du regard plus ou moins haut pour indiquer une distance Sourcils levés ou froncés, yeux plissés, pour renforcer l'éloignement Changement de proforme pour une même forme Effectuation d'un signe à une ou à deux mains Utilisation de l'hyperextension du bras Pointage avec un mouvement en cloche pour indiquer l'éloignement Utilisation d'un temps important pour effectuer un itinéraire Re-description d'une forme en changeant ses proportions

4.3.2 Stratégies pour les gestes coverbaux

Bien que les données dont nous disposons dans le cadre du corpus en gestuelle coverbale ne permettent pas de reconstituer à chaque instant l'espace gestuel, nous appliquons, lorsque c'est possible, la même méthode que pour le corpus en langue des signes afin de pouvoir comparer les stratégies de manipulation de l'espace gestuel mises en oeuvre.

Entre N et O Insertion d'une pause pour marquer le changement d'espace gestuel.

Entre O et P Aucun lien apparent avec les espaces précédents, ce geste semble être un illustrateur du mot "côte".

Entre P et Q Maintien de la position du David entre l'espace O et l'espace Q, prononciation du mot "David" en même temps que le pointage ; de la même manière, prononciation du mot "rond point" en même temps que le pointage de l'escalé Borelli. Ces deux indications permettent de déduire une rotation de 90° de l'espace de signation.

Entre Q et R Indication de la rotation de l'espace gestuel grâce à l'indication "sur la droite" .

Entre R et S Indication de changement de repère par l'emploi du passé "vous avez longé l'hippodrome" qui signifie qu'on se trouve à nouveau sur la route qui vient d'être empruntée.

Entre S et T Indication de changement d'échelle et de translation en décrivant la forme du rond-point en même temps que les mots "escalé Borelli" sont prononcés.

Entre T et U Indication de changement de repère en prononçant le mot "musée d'art contemporain" en même temps qu'une nouvelle zone de l'espace est pointée.

Entre U et V Utilisation d'une configuration manuelle identique aux intervalles U et V pour désigner le musée d'art contemporain en même temps qu'est prononcé le mot "musée d'art contemporain".

Entre V et W Rotation de l'espace gestuel qui peut être inférée grâce à l'indication verbale "en continuant tout droit". De même que précédemment, nous récapitulons les différentes stratégies de modification de la projection de l'espace réel dans le tableau qui suit.

Type d'opération	Stratégies mises en œuvre
Grande discontinuité	Insertion d'une pause dans la prosodie Indication vocale seule Placement et indication vocale sur le nom de l'entité
Restauration d'une projection antérieure	Utilisation de la même configuration placée à un endroit proche Indication vocale seule Placement et indication vocale sur le nom de l'entité
Translation	Utilisation de la même configuration placée à un endroit proche Indication vocale seule
Rotation	Utilisation de la même configuration placée à un endroit proche Indication vocale seule
Changement d'échelle	Re-description d'une forme en changeant ses proportions avec une indication vocale

Dans la majorité des exemples précédents, il apparaît clairement que les gestes coverbaux sont uniquement en appui ou en illustration de la parole. Ils ne sont en aucun cas interprétables sans la parole, mais contiennent tout de même des informations qui ne sont présentes que dans l'espace gestuel (ex : le tournant à droite à l'escale Borelli). Le phénomène de mobilisation simultanée des deux mains pour situer les entités les unes par rapport aux autres est ici tout à fait marginal. Le rôle du regard qui permettait d'indiquer un éloignement ou une rotation semble ici bien moindre même s'il est présent à certains moments (ex : pour indiquer la position de la statue du pouce). Par contre, nous relevons le fait que le locuteur regarde ses mains lorsqu'il effectue des descriptions.

Nous avons donc bien à faire à deux manières différentes d'utiliser l'espace. Contrairement aux gestes co-verbaux, la langue des signes optimise réellement la gestion de l'espace de manière à pouvoir dire et montrer en même temps.

5 Changements de repères, transferts et optimisation de l'espace

Concentrons nous maintenant sur le corpus en LSF. Nous avons mentionné en section 4.1 un problème de surcontrainte si on tient compte simultanément de toutes les informations spatiales véhiculées par les paramètres manuels et non-manuels. Nous avons fréquemment relevé dans notre corpus l'utilisation simultanée de trois échelles :

- Une échelle Γ_1 liée au corps du signeur dans le cas où celui joue un rôle. Il s'agit dans notre cas de l'échelle 1:1 quand le signeur joue le rôle d'une personne qui sillonne la route, et d'une échelle d'à peu près 1:2 lorsqu'il prend le rôle du David.
- Une échelle Γ_2 liée à un espace qui sert à placer les éléments les uns par rapport aux autres. C'est cet espace que nous avons figuré dans les représentations graphiques du 4.2.
- Une échelle Γ_3 liée à la représentation des entités elles-mêmes qui correspond à la dimension du volume délimité par le proforme ou le signe.

Il est intéressant de mettre cette constatation en regard de la théorie des visées de C. Cuxac (Cuxac, 2000) qui distingue également trois manières de dire en donnant à voir. L'échelle Γ_1 correspond selon nous à la situation de transfert personnel. Il serait toutefois erroné d'affirmer que Γ_2 et Γ_3 correspondent respectivement aux transferts de situation et aux transferts de taille et de forme. L'échelle Γ_2 est celle des transferts situationnels, mais est également impliqué dans le cadre de transferts de taille et de forme dès le moment où une forme est déployée par rapport à une autre. Quelle interprétation donner à cette rupture d'échelle locale entre Γ_2 et Γ_3 dans la mesure où elles ne sont pas associées chacune à une catégorie de transfert différente ?

Cette rupture d'échelle peut résulter d'une contrainte sur la production. D'une part, les différents articulateurs dont disposent un signeur ne permettent parfois pas de représenter forcément avec précision les tailles des différentes entités placées dans l'espace de signation. D'autre part, il est fréquent qu'un même signe placé dans l'ES ne puisse pas être effectué avec différentes amplitudes.

Du point de vue perceptif, cette rupture d'échelle, loin d'entraîner une incompréhension dans l'appréhension de l'espace a de nombreux avantages. A la manière d'une carte touristique sur laquelle les bâtiments sont figurés en représentation surdimensionnée (cf. figure 11⁷), le zoom sur les entités permet d'utiliser l'ES de manière plus intensive et économise la combinaison *signe standard + pointage*. Cela présente également un avantage du point de vue du changement d'échelle, fréquemment observé dans l'analyse de notre corpus. Nous rappelions en section 4.3.1 qu'une des manières pour effectuer un changement d'échelle était de redessiner une forme ou d'effectuer une fois de plus un signe standard, déjà présent dans l'ES. Si les proportions devaient être respectées, une entité serait forcément réduite à un point de l'Espace de Signation avant ou après un changement d'échelle important, et ce point ne pourrait plus véhiculer aucune information de forme, de taille et de rotation.

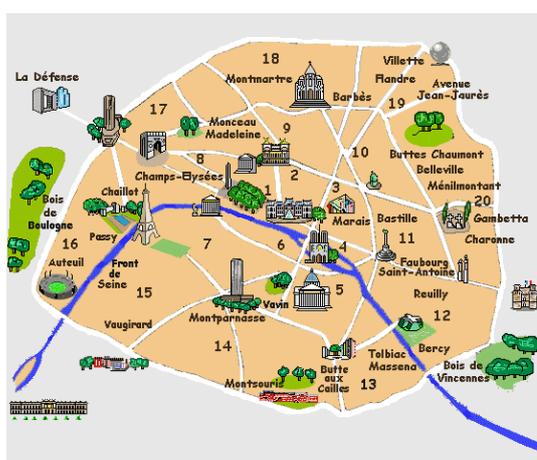


Figure 11: Exemple de rupture d'échelle locale pour montrer l'allure de bâtiments.

6 De la description à la synthèse automatique

Dans le cadre de cette étude, nous avons relevé différentes stratégies qu'utilisent des locuteurs en LSF pour effectuer des opérations de rotation, de translation et de zoom dans l'Espace de Signation. L'ensemble des stratégies dégagées n'est pas forcément exhaustif et doit encore être confirmé en traitant des corpus impliquant davantage de locuteurs et des corpus de durée plus importante. Ces stratégies présentes en LSF ont été observées pour partie dans le corpus en gestuelle coverbale, mais elles ne sont pas utilisées assez systématiquement pour que la gestuelle et la voix permettent de déterminer une projection de l'espace réel dans l'espace gestuel. De plus, nous observons que les indications les plus cruciales comme les changements ou maintien d'orientation sont le plus fréquemment communiqués verbalement.

Du point de vue de la synthèse d'énoncés en Langue des Signes, l'analyse que nous exposons s'inscrit en complément des travaux déjà existants prenant en compte un espace de signation fixe (Gianni *et al.*, 2007). Toutes les stratégies que nous décrivons pour effectuer des rotations, des translations et des changements d'échelles constituent autant d'éléments nécessaires à implémenter pour effectuer une synthèse d'un énoncé complexe en LS. En effet, les énoncés "complexes" de LSF impliquent la projection simultanée de différents espaces sémantiques dans un même espace de signation, ainsi que des changements fréquents de projection. Il reste encore un grand travail d'investigation à mener pour arriver à prédire comment un contenu sémantique sera séquencé puis projeté pas à pas dans l'ES, et quelles stratégies pourront être sélectionnées pour maintenir une cohérence de l'ES tout au long de l'expression.

⁷illustration tirée du site <http://www.pariserve.tm.fr/quartier/decouv.htm>

References

- BRAFFORT A., BOSSARD B., SEGOUAT G., BOLOT L. & F. L. (2005). Modélisation des relations spatiales en langue des signes française. In *Atelier LS, TALN*, Dourdan.
- CALBRIS G. (2000). Espaces symboliques révélés par l'étude sémantique du geste coverbal. *Cahier du CIEL*.
- COLETTA J. & MILLET A. (2002). *Gestualité et syntaxe*, volume 26. Lidilem.
- COSNIER J. (à paraître). Les gestes de la transmission d'itinéraires entre piétons. *J.-M. Barberis et M.-C. Manes Gallo (eds.) Verbalisation de l'espace et cognition située. La description de l'itinéraire piéton. Université Montpellier III*, p.10.
- CUXAC C. (2000). *La langue des signes française. Les voies de l'iconicité*. Faits de langue.
- FAUCONNIER G. (2004). Mental spaces, language modalities, and conceptual integration. *S. Davis and B. Gillon, eds. Semantics: A Reader*.
- FUSELLIER-SOUZA I. (2004). *Étude de langues des signes primaires pratiquées par des sourds brésiliens*. PhD thesis, Université Paris 8.
- GIANNI F., COLLET C. & LEFEBVRE-ALBARET F. (2007). Modèle et méthode de traitement d'image pour l'analyse de la langue des signes. *Traitement Automatique des Langues, Modélisation et traitement des langues des signes*, **48**(3), 175–200.
- HUENERFAUTH M. (2006). *Generating American Sign Language Classifier Predicates For English-To-ASL Machine Translation*. PhD thesis, Université de Pennsylvanie.
- LENSEIGNE B. (2004). *Intégration de connaissances linguistiques dans un système de vision, application à l'étude de la Langue des Signes*. PhD thesis, Université Paul Sabatier, Toulouse.
- LIDDELL S. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge University Press.
- PERNISS P. (2007). Locative functions of simultaneous perspective constructions in german sign language narratives. In *Vermeerbergen, M., Leeson, L., and Crasborn, O. (Eds.), Simultaneity in Signed Languages: Form and function*, p. 27–54, Amsterdam.
- SALLANDRE M. (2003). *Les unités du discours en Langue des Signes Française. Tentative de catégorisation dans le cadre d'une grammaire de l'iconicité*. PhD thesis, Université Paris 8.
- TVERSKY B. (1993). Cognitive maps, cognitive collages and spatial mental models. In A. FRANK & I. CAMPARI, Eds., *Spatial Information Theory: A Theoretical Basis for GIS – Proceedings of COSIT'93*, p. 14–24, Berlin: Springer.

Annexes

Nous consignons ici les différents pointages observés dans nos deux vidéos. La présentation un peu différente de celle de l'article permet de voir combien de fois les proformes ont été utilisés pour localiser les différentes entités. Par convention, nous comptons un pointage deux fois lorsqu'il mobilise les deux mains⁸.

⁸Cet inventaire ne fait pas figurer les pointages par le regard et les pointages par le buste.

	Cercle1	Cercle2	Cercle3	Cercle5	Cercle4	Pt1	Pt2	Pt3	Zone1	Route1
E. Borelli	4	3	2	3				1		
Musée				2						
David					1	1	2	1		
RP Musée								1		
Hippodrome								1	1	
Restaurants									1	
Desc. du Prado										1
Rte vers musée										5
Rte vers Sormiou										1

Figure 12: Tableau récapitulatif des pointages annotés dans la vidéo du dialogue en français et gestuelle coverbale.

	Ligne1	Route2	Arc1	Port	Ile (S. Standard)	Pt2	Zone2	Zone3
Corniche	1	4				1	1	
Route église		1						2
David						1		
Ile d'If					2	4	2	
Eglise						1		
Vieux port			1	8		3		
Plage							1	1

Figure 13: Tableau récapitulatif des pointages annotés dans la vidéo du dialogue en LSF.

LSF

DEGELS2011 : Analyses d'annotations pour la reconnaissance

Monia Ben Mlouka, Juliette Dalle, Patrice Dalle
IRIT-UPS, 118, route de Narbonne 31062 Toulouse Cedex 9
mlouka@irit.fr, jdalle@irit.fr, dalle@irit.fr

Résumé. En reconnaissance de la langue des signes française, on utilise des corpus annotés de manière descendante par interprétation afin d'explicitier les résultats attendus du module de reconnaissance. D'autre part, on analyse automatiquement de manière ascendante des corpus 3D pour en extraire des modèles précis et des corpus vidéo pour mettre au point des stratégies de reconnaissance. Dans cet article, on aborde cette double démarche et les différentes représentations informatiques associées, en prenant comme exemple les pointages manuels.

Abstract. In Sign Language recognition, we annotate corpus, with top-down interpretation procedures, to explicit the outcomes of the recognition module. On the other hand, by an automatic bottom-up processing, we analyze 3D corpora, to extract accurate models, and video corpora, to develop recognition strategies. In this paper, we address this dual approach and we present the associated computer representations, taking the pointing gestures as example.

Mots-clés : Langue des signes, annotation manuelle, annotation automatique, pointage manuel, reconnaissance.

Keywords: Langue des signes, manual annotation, automatic annotation, pointage manuel, recognition.

1 Introduction

L'activité qui nous est proposée dans le cadre de DEGELS ne correspond que très partiellement au processus que nous suivons en reconnaissance de la LSF. Aussi nous expliquons cette démarche afin de situer les annotations que nous avons réalisées pour DEGELS dans ce processus. Puis nous présentons ces annotations et nous les commentons par rapports à nos objectifs.

En reconnaissance de la LSF, les corpus sont utilisés pour atteindre deux objectifs : pour modéliser la LSF (caractériser les signes, les structures grammaticales, les distributions des rôles, les modalités, ...) et pour apprendre des stratégies de reconnaissance (c'est-à-dire savoir exploiter l'aspect des personnes signantes pour retrouver des informations sur leur production en LSF, reconnaître les signes et éventuellement comprendre le sens de l'énoncé). On va ainsi être amené à utiliser différents types de corpus pour ces deux tâches. De même l'annotation de ces corpus devra être descriptive et factuelle s'il s'agit de décrire la LSF et elle fera appel à l'interprétation lorsqu'elle visera à indiquer au module de reconnaissance les résultats qu'il doit obtenir. Nous présentons cette double démarche et les différentes représentations informatiques associées, en prenant comme exemple l'analyse des pointages.

2 Exploitation des corpus en reconnaissance de la LSF

2.1 Le processus de reconnaissance de la LSF

Comme dans tout processus de reconnaissance, reconnaître un énoncé en LSF consiste à associer des éléments extraits d'une vidéo (segments de données) à des éléments du modèle de langue acquis par apprentissage (étiquettes), en exploitant des connaissances a priori.

Cette association est faite à un niveau qui dépend du centre d'intérêt de la reconnaissance. C'est ainsi qu'on peut chercher à reconnaître des caractéristiques des éléments qui entrent en jeu dans l'énoncé (configuration des mains, signature gestuelle), à identifier des formes syntaxiques particulières (pointage, structures de transferts), à reconnaître un signe (recherche d'un signe dans un index ou dans une vidéo, comparaison de signes), ou à comprendre l'énoncé (compréhension d'une commande ou d'une requête, traduction). Chaque niveau implique généralement le traitement des niveaux inférieurs. On s'intéressera ici plus particulièrement aux pointages, qui peuvent être réalisés de différentes façons et sont, le plus souvent, précédés par une focalisation du regard sur la zone pointée :

- geste de pointage manuel, le plus souvent avec la configuration index tendu,
- signe spatialisé,
- mouvement du buste ou d'une épaule,
- mouvement de la tête.



FIG. 1 – Pointages manuels

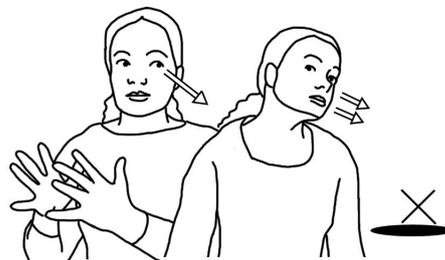


FIG. 2 – Pointages par le visage

Les connaissances a priori sont de deux natures : elles apportent des connaissances sémantiques (non visuelles) sur les éléments du discours et sont utilisées par les couches haut niveau du programme de reconnaissance pour construire du sens et faire le lien entre ce qui est dit et ce dont on parle. D'autre part, des informations plus contextuelles doivent être apportées pour limiter le champ de la recherche et rendre ainsi réalisable la tâche de

reconnaissance¹. Elles peuvent porter sur le contexte d'acquisition (informations sur la scène, sur le signeur, sur le système de prise de vue), sur des limites imposées dans la forme du discours (structures syntaxiques non autorisées, ...), sur l'étendue du champ lexical et donc aussi sur des connaissances extra linguistiques sur les éléments du discours. Ici par exemple, on pourrait donner des indications sur la disposition des personnages pour faciliter leur détection, fournir une liste des termes (lexique) pouvant figurer dans l'énoncé et donner des informations sémantiques sur la scène, comme les relations spatiales entre les principaux composants d'une scène de bord de mer (terre, ville, côte, plage, mer, île) ou sur la notion de trajet.

Dans ce processus, les corpus de LSF interviennent à deux niveaux : pour apprendre les modèles de langues et pour apprendre des stratégies de reconnaissance.

2.2 Analyse des corpus

2.2.1 Modèle de langue

Le premier corpus a pour but de modéliser informatiquement la langue étudiée. En principe ce corpus devrait comporter des informations suffisantes pour pouvoir déterminer comment ont été réalisés les signes. C'est le cas de données acquises par capture de mouvement, à partir de marqueurs fixés sur les articulateurs du buste, des membres, des mains et du visage. Fournissant les positions 3D réelles de ces articulateurs, ils peuvent être exploités automatiquement (Ben Mlouka *et al.*, 2010). On peut ainsi extraire des caractéristiques relationnelles, géométriques et dynamiques des signes et calculer des représentations statistiques des gestes, des signes et des structures grammaticales, tels qu'ils sont effectivement réalisés. La construction de ces représentations se fait dans un cadre linguistique donné, ici le modèle sémiologique de Cuxac (Cuxac, 2000), et suivant des choix de modélisation des signes, dans notre cas une modélisation mono-segmentale paramétrique (Lefebvre-Albaret & Dalle, 2010). Pour pouvoir exploiter ces modèles en reconnaissance, il faudra être capable de les reformuler en 2D, par projection ; ce qui suppose des connaissances sur le dispositif d'acquisition.

D'autre part, ces modèles ne portent que sur les positions et donc sur la géométrie et la dynamique des gestes et non sur l'aspect des composants corporels dans une image, pour lesquels il faudrait introduire les notions de volumes et de matière (peau, vêtements, fond), pour gérer les occultations, pour produire l'aspect du signeur dans la vidéo (aspect des mains, expression faciale...). Pour cela il faudrait disposer d'un deuxième jeu de données, sous la forme d'une ou de plusieurs vues du même personnage observé par une vidéo conventionnelle fournissant des données du même type que celles qui seront traitées dans la phase de reconnaissance. Pour pouvoir être exploitée en phase d'apprentissage de modèle, ces vidéos doivent pouvoir être traitées automatiquement : les caractéristiques doivent être calculées à partir de données numériques de positions et de vitesses ou à partir de la couleur et de la texture des composants corporels dans la vidéo. On doit donc, au minimum, être capable d'effectuer la détection et le suivi des articulateurs et de segmenter les principaux composants (mains, visage).

Dans la pratique, on dispose rarement de ces deux données, les conditions d'acquisition par capture de mouvement rendant difficile une acquisition simultanée par caméra conventionnelle. Une solution intermédiaire consiste à fabriquer le corpus à partir de vidéos 2D provenant de plusieurs vues de la même scène, comme dans le cas du corpus DEGELS (vue de face et vue latérale). Mais là aussi, les données doivent être exploitables automatiquement. De plus, actuellement on ne sait pas analyser et caractériser correctement les configurations manuelles du fait des occultations ; les mains seront donc représentées par des régions (blobs) ou par leur centroïde.

Enfin, l'approche purement statistique consisterait à apprendre les modèles de LSF directement à partir de gros corpus de données 2D, de manière analogue à ce qui est fait pour la parole. Il y a cependant deux obstacles : il est difficile de constituer des corpus de grande taille en LSF et du fait de la grande latitude de création de l'espace de signation, le même énoncé peut produire un grand nombre de réalisations différentes et donc d'aspects différents dans les vidéos, sans parler des différences inter-personnelles. Cette approche n'est donc pas réalisable actuellement. Elle doit être combinée avec une modélisation symbolique de la LSF faisant intervenir par exemple des règles de création de l'espace de signation ou des informations de nature sémantique.

On est donc amené, comme nous l'avons fait ici, à apprendre le modèle de LSF uniquement sur un corpus de capture de mouvement et donc à exclure les caractéristiques d'aspect.

¹Ces informations peuvent être données explicitement ou être contenues implicitement dans le choix des données à traiter : dans ce cas on conçoit un programme de reconnaissance spécialisé et on ne l'appliquera que dans un contexte particulier.

2.2.2 Stratégies de reconnaissance

Le deuxième corpus est utilisé pour l'apprentissage des stratégies de reconnaissance. Il s'agit d'un corpus vidéo analogue à ce qui sera utilisé en phase d'exploitation. Son annotation fournit ce que le module de reconnaissance devrait produire. Le corpus est annoté de manière descendante, c'est-à-dire à partir de la compréhension du sens de l'énoncé, puis de sa segmentation en signe, de l'ancrage dans l'espace de signation et enfin des actions intervenant sur cet espace (création, référencement, pointage). L'annotateur exploite donc ce qu'il voit mais aussi ce qu'il sait du domaine (plus exactement ce que le système de reconnaissance est censé connaître du domaine). Dans le cas de DEGELS, ce corpus est utilisé pour évaluer les résultats de la détection des pointages et faire évoluer les stratégies de reconnaissance des types de pointage.

L'espace, support de l'énoncé, doit être explicitement représenté, notamment pour les pointages. Les outils ANVIL et Elan n'étant pas bien adaptés à l'annotation 3D de cet espace, nous avons utilisé notre logiciel VIES (Lenseigne & Dalle, 2005) pour indiquer la date de création et l'emplacement des entités utilisées par les signeurs dans ce corpus (fig. 3.3.2). C'est donc le résultat d'une interprétation. L'espace de signation étant un volume divisé en cubes, nous avons reporté dans l'annotation Elan les coordonnées (Horizontal, Vertical) des emplacements de chaque entité.

Inversement, de manière ascendante, le corpus est analysé automatiquement à partir des données fournies par le logiciel de détection et de suivi de la tête et des mains. Dans le cas de DEGELS, la surexposition du visage ne permet pas d'en analyser le contenu et il y a encore trop d'erreurs de suivi des mains ou de confusion droite/gauche, ce qui nous a amenés à réaliser un outil de relevé manuel des positions des mains et de la tête. En effet, une annotation manuelle des premiers niveaux ne nous apporterait aucune information sur les stratégies de reconnaissance automatique.

Nous détaillons dans la suite ce traitement automatique et l'évaluation des résultats sur la détection des pointages et leur segmentation.

3 Apprentissage pour la reconnaissance

3.1 Annotation manuelle

3.1.1 Objet de l'annotation

L'annotation manuelle est basée sur la compréhension globale de l'énoncé. Il s'agit d'une transcription descendante structurée sur plusieurs niveaux. On part du niveau sémantique où on délimite l'énoncé sous forme de phrases. Dans un second niveau, on transcrit l'énoncé sous forme de gloses. A ces deux niveaux, s'ajoute un troisième où on transcrit les gestes de pointages réalisés par la main, le buste et le regard. Cette structure est marquée par une dépendance dans la délimitation temporelle entre les deux premiers niveaux.

3.1.2 Définition d'un pointage

On retrouve dans (Bras *et al.*, 2004) un inventaire des procédés déictiques dont les formes de pointages manuels et non manuels. Dans notre étude, nous allons adopter les définitions suivantes :

Pointage manuel : Index tendu ou main plate. Le mouvement de pointage se termine par une décélération importante.

Pointage par le regard : Cet événement représente une fonction linguistique déduite à la fois à partir du changement de la direction du regard chez le signeur et des unités gestuelles produites avec un certain décalage temporel par rapport au regard. Par exemple, l'annotateur déduit un pointage par le regard quand il observe une déviation du regard vers une direction donnée suivie d'un déplacement de la main vers cette même direction

Pointage par le buste : Il s'agit d'un mouvement du buste qui est impliqué dans la fonction de pointage. Quand on observe une rotation du buste suivie d'un pointage manuel, l'annotateur attribue au mouvement du buste le rôle de pointage. Il faut noter que repérer le mouvement du buste est une tâche difficile car, dans une vidéo, le mouvement du buste est marqué par un faible déplacement des épaules. De plus, l'annotateur doit faire la distinction entre un mouvement des épaules dépendant du buste et un autre qui ne l'est pas.

Signe spatialisé : Il s'agit de la transcription d'un mouvement manuel impliqué dans la production d'un signe localisé dans l'espace de signation. Dans la plupart des cas, on observe une fixation du regard sur la zone où le signe sera construit. Il s'agit d'un indice qui peut montrer que le signe occupe une zone particulière de l'espace. La posture de la main dominante (s'il s'agit d'un signe à une seule main) est un deuxième élément qui informe sur la nature du signe. En effet, plus la position de la main dominante est loin du corps du signeur, moins on aura d'hésitation pour attribuer au signe la propriété de signe spatialisé.

3.1.3 Aspect technique

Logiciels : On a utilisé le logiciel "Elan" pour l'annotation textuelle, "VIES" pour transcrire la structure de l'espace de signation. On a également développé un outil de détection manuelle, afin de récupérer les coordonnées cartésiennes des composants corporels (les mains et la tête) en s'affranchissant d'éventuelles erreurs du logiciel de suivi automatique de ces composants.

Transcriptions : On a annoté le dialogue en langue de signes française. Pour cela, on a utilisé la même structure d'annotation (figure 3) pour chacun des deux locuteurs. Chaque annotateur a utilisé un outil différent de transcription. Les annotations ont été réalisées en parallèle sur une durée totale de quatre heures :

- Annotation sous Elan : deux heures.
- Transcription VIES : une heure
- Détection graphique des positions des mains et de la tête : une heure.

3.1.4 Aspect théorique

Cadre : L'annotation s'est faite dans un cadre linguistique donné, celui du modèle sémiologique de Cuxac. La transcription des énoncés a été faite séparément pour chaque signeur (nom des pistes en double) car on a constaté qu'à un certain moment les deux personnes signaient en même temps ce qui est difficile d'annoter sur une seule piste.

Structure : Il s'agit d'une analyse descendante (à partir du sens). La figure (3) montre les noms des pistes ainsi que leurs dépendances. Le stéréotype utilisé dans la majorité des pistes est la dépendance de l'acteur parent sur les bornes sauf pour la piste "Niveau de compréhension" qui ne possède pas de stéréotype particulier.

3.1.5 Délimitations des segments des pistes "Niveau de compréhension" et "Gloses"

L'annotateur décompose l'énoncé en phrases. L'annotateur procède à la délimitation temporelle quand il juge que la phrase est complète. Les figures (4) et (5) sont deux exemples de délimitations de phrases dans une première piste et de sa retranscription en un ensemble de gloses dans une seconde piste.

3.1.6 Délimitations des segments de la piste "Signe spatialisé"

Il s'agit de reprendre certaines annotations de la piste "Gloses" dont le signe occupe une zone particulière de l'espace de signation. Il y a des exceptions : dans l'exemple (6), on remarque que cette dépendance temporelle n'est pas respectée. L'annotateur a inclus le mouvement de préparation pour mettre en évidence la zone occupée par le début du signe [ROUTE EN PENTE].

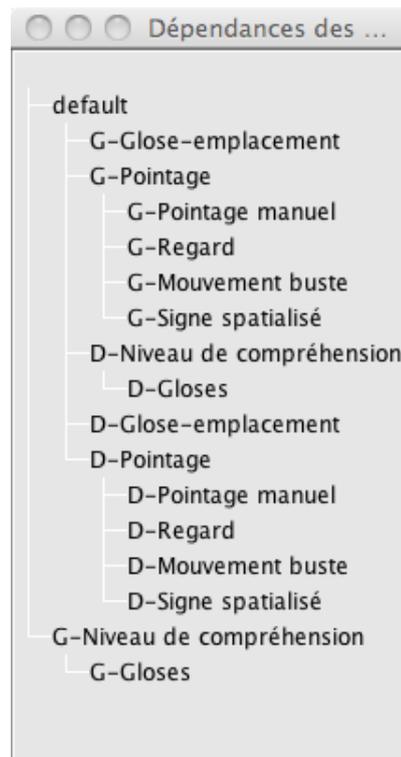


FIG. 3 – Hiérarchie de la structure



FIG. 4 – Exemple de délimitation d'un segment de la piste "Niveau de compréhension"

3.1.7 Signification et délimitations des segments de la piste "Pointage manuel"

Il s'agit d'une délimitation en unités syntaxiques. Le pointage manuel inclut les pointages vers l'espace de signation, les pointages vers une partie du corps et les marqueurs de personne.

Stratégie de délimitation :

- Début : Configuration stable (index tendu)².
- Fin : Changement de configuration.

On retrouve cette méthode de délimitation de pointage manuel dans la majorité des segments annotés. La figure (8) présente un exemple de trace de pointage manuel. Les trois positions de la main droite sont les positions clés qui marquent (i) le début de la phase de préparation au mouvement de pointage, qui correspond au segment dynamique selon le modèle de Johnson (Liddell & Johnson, 1989), (ii) le début du mouvement de pointage et (iii) la phase de tenue, qui correspond au segment statique. Les flèches symbolisent le sens du mouvement entre chaque position clé.

Bien que cette méthode représente la stratégie de l'annotateur pour repérer le début et la fin d'un pointage manuel, on a pu relever une délimitation différente où l'annotateur a déterminé le début du pointage quand la main est en mouvement alors que la configuration index tendu n'est pas encore visible. Le figure (7) retrace cette délimitation.

²Il existe des variantes de flexion d'index dans la configuration manuelle [index tendu]

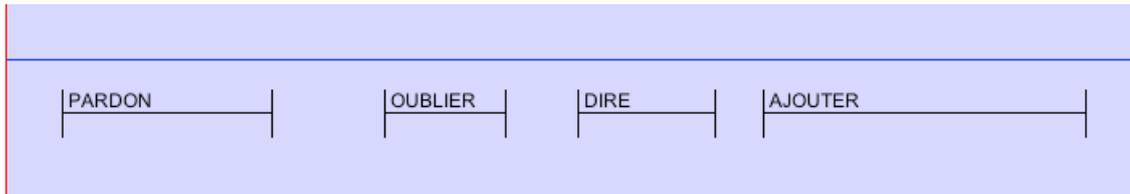


FIG. 5 – Deuxième exemple de délimitation d'un segment de la piste "Niveau de compréhension"



FIG. 6 – Exemple de délimitation d'un segment de la piste "Signe spatialisé"

3.1.8 Délimitations des segments de la piste "Regard"

Cette piste indique le changement de la direction du regard vers une zone spécifique de l'espace de signation. Par exemple si au départ le regard est dirigé vers l'interlocuteur, le début du segment correspond au début de changement de direction du regard. On prend aussi comme hypothèse que le regard cible une zone qui sera occupée par la main par la suite. Quand le regard ne cible plus la main, on considère que c'est la fin du segment.

3.1.9 Attributs et valeurs

On n'a pas utilisé de vocabulaire contrôlé pour la transcription des gestes. On a étiqueté les annotations de la piste "Emplacement-glose" avec une nomenclature spécifique au découpage de l'espace de signation selon VIES. Par exemple, l'étiquette "H3V2S3" désigne l'emplacement illustré dans la structure de l'espace de signation. La liste des positions VIES est présentée dans la section (3.3).

Le tableau 1 résume la nature des valeurs attribuées pour chaque piste.

Piste	Attributs	Valeurs
Niveau de compréhension	Description de la sémantique	Gloses
Glose-emplacement	Description de la forme	Positions VIES (voir section 3.3)
Pointage	-	Simple délimitation

TAB. 1 – Description des attributs

3.1.10 Pointage : Résultats d'annotation manuelle

Type de pointage	Nombre	Durée moyenne (ms)	Remarques
Manuel	16 (4 pour le locuteur "G" et 12 pour le locuteur "D")	200	3 pointages dont le signifié est un pronom [MOI] et [TOI]
Regard	23 (2 pour le locuteur "G" et 21 pour le locuteur "D")	≈ 278	-

TAB. 2 – Résultats d'annotation



FIG. 7 – Exemple de délimitation d'un segment de la piste "Pointage manuel" selon le mouvement



FIG. 8 – Exemple de délimitation d'un segment de la piste "Pointage manuel" selon la configuration manuelle

Commentaires sur le tableau (2) : On a remarqué que le nombre d'occurrences de pointage manuel, égal à 16, est faible par rapport au nombre de gloses, qui s'élève à 115, soit un taux de 14%. Ce taux est inférieur à celui d'une étude similaire à la nôtre, celle de (Fusellier-Souza, 2004). Le résultat de ses analyses a montré que la fréquence de pointage est de l'ordre de 25%. Il convient cependant de noter que les travaux de (Fusellier-Souza, 2004) regroupent plusieurs classes de pointages qui ne font pas partie du corpus DEGELS, telles que le référencement de l'espace physique et l'énumération. Si on ne comptabilise que les classes de pointages utilisés dans le corpus DEGELS, c'est-à-dire les marqueurs de pronom, la référence spatiale, le pointage vers une partie du corps, le taux sera de 16%, légèrement supérieur à celui du corpus de DEGELS. Cela nous permet de conclure que les deux taux d'occurrence de pointages sont très proches, alors que les énoncés du corpus de (Fusellier-Souza, 2004) sont en langues des signes émergente.

3.1.11 Evaluation

Les 3 participants ont annoté le corpus selon des points de vue particuliers qui ont des recouvrements, ce qui constitue une forme de contrôle partiel lors de la fusion de l'ES et des pointages annotés sous Elan. Une seconde méthode d'évaluation consiste à comparer les résultats d'annotation manuelle à ceux de l'annotation automatique.

3.2 Annotation automatique

3.2.1 Objectif de la méthode

L'objectif de la méthode est de détecter automatiquement les segments correspondants à un pointage manuel sur des vidéos.

3.2.2 Description de la méthode

Dans un premier temps, nous nous sommes focalisés sur des données de capture de mouvement, en particulier les positions tri-dimensionnelles de la main dominante. Nous avons quantifiés la dynamique du mouvement à l'intérieur des segments qui correspondent à des pointages manuels. Par la suite, nous avons pu extraire des motifs communs à certains segments³. Les signatures de vitesse instantanée des pointages manuels sont illustrées dans les graphes (9) et (10). Pour rejoindre notre objectif, celui de la détection de profils de pointages manuels dans des vidéos, donc en 2D, nous avons éliminé la composante de profondeur. Dans un deuxième temps, nous avons extrait les positions de la main dominante sur la vidéo en utilisant un outil d'annotation interactif image par image. Finalement, nous avons appliqué un filtre sur les données de la vitesse de la main dominante pour détecter les motifs de pointages manuels.

Le premier modèle a été élaboré à partir de 8 pointages manuels. Il comporte deux phases (Figure 9).

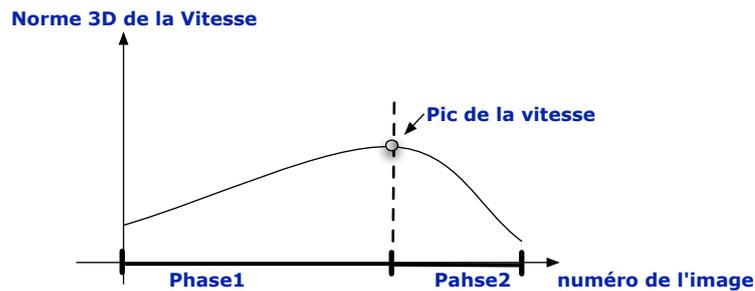


FIG. 9 – Signature d'un pointage manuel - S1

Le second modèle a été élaboré à partir de 4 pointages manuels. Il comporte une seule phase descendante (Figure 10).

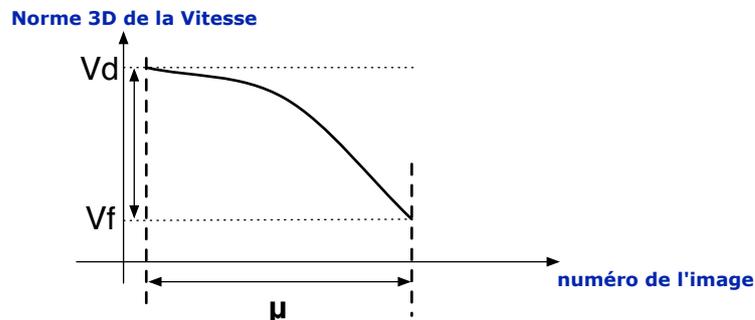


FIG. 10 – Deuxième signature d'un pointage manuel - S2

³Le corpus objet de l'analyse est produit en langue des signes québécoise et réalisé dans le cadre du projet franco-canadien "Marqspat".

3.2.3 Résultats et interprétations

Le résultat de détection s'avère insatisfaisant pour les deux profils car le nombre de faux pointages est trop élevé. Ceci met en cause, principalement, le passage de 3D à 2D, qui entraîne une perte d'information significative. On peut aussi supposer que les profils de vitesses obtenus ne caractérisent pas seulement les pointages. La détection des composants corporels ayant été faite manuellement, on considère que ces mauvais résultats ne proviennent pas d'erreurs de détection de leur position. Il convient aussi de noter que les niveaux de langue utilisés dans le corpus de DEGELS, tant pour l'énonciation que dans le déroulement du dialogue, sont trop complexes par rapport à ce qu'on est, actuellement, capable de traiter en reconnaissance de la LSF.

3.2.4 Stratégies de correction

Afin d'améliorer ces résultats, nous prévoyons d'introduire des paramètres supplémentaires comme la position relative des mains par rapport à la tête et d'inclure des opérateurs de traitement d'images pour extraire des informations sur la configuration manuelle. Il ne s'agit pas de reconnaître les configurations manuelles, mais de vérifier, lors d'une hypothèse de pointage, si les mesures obtenues sont compatibles par exemple avec la forme "index tendu" et de diminuer ainsi le nombre de faux pointages.

3.3 Annotation VIES

3.3.1 Fonctionnement :

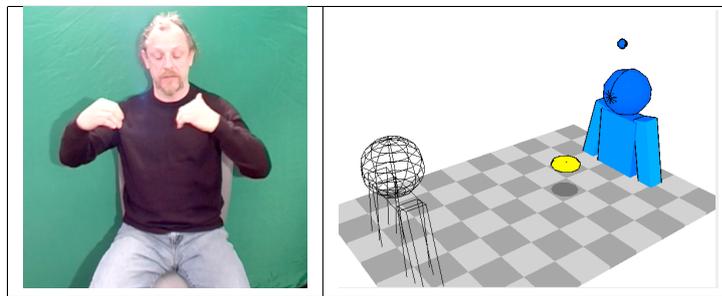
La construction incrémentale de l'espace de signation à l'aide de l'outil interactif VIES suit les étapes suivantes :

1. Déterminer les instants de début et de fin d'un signe, ex : [PORT].
2. Déterminer le type du signe : lieu, objet, personne, action, etc.
3. Déterminer visuellement la position géométrique approximative de la zone occupée par le signe.
4. Désigner la position (H, S) de la projection du signe sur le plan horizontal, puis son altitude (V) : l'entité est alors placée à l'emplacement 3D correspondant.

Dans le cas où le signe représente une action, désigner l'entité acteur puis le ou les patients ou bénéficiaires.

Les illustrations suivantes sont les résultats d'annotations obtenues selon cette démarche.

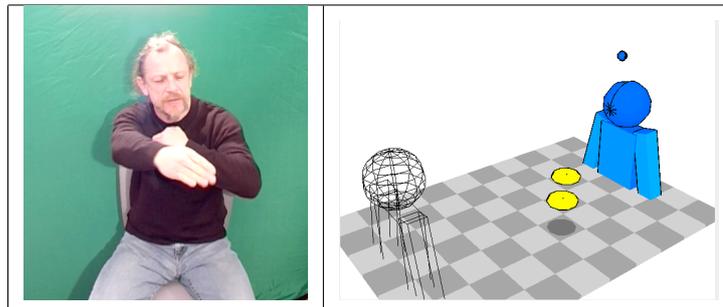
3.3.2 Transcription : Bâtiments en U (Vieux port)



TAB. 3 – Transcription : Bâtiments en U (Vieux port)

- Signe : [PORT]
- Début (sec) : 12.91
- Type : Lieu
- Position dans VIES : H3-V2-S3
- Constat par rapport au choix de l'annotateur : le lieu [PORT] est placé devant le buste du signeur.

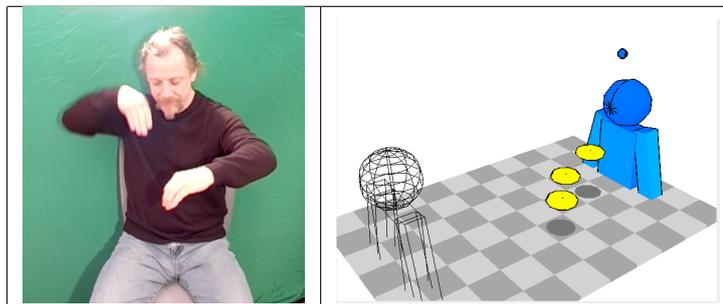
3.3.3 Transcription : Route sinueuse (corniche)



TAB. 4 – Transcription : Route sinueuse (corniche)

- Signe : [CORNICHE]
- Début (sec) : 13.88
- Type : Lieu
- Position dans VIES : H4-V2-S4
- Constat par rapport au choix de l'annotateur : le lieu [CORNICHE] est placé à coté de [PORT] ⁴

3.3.4 Transcription : Rue en pente (Canebière)

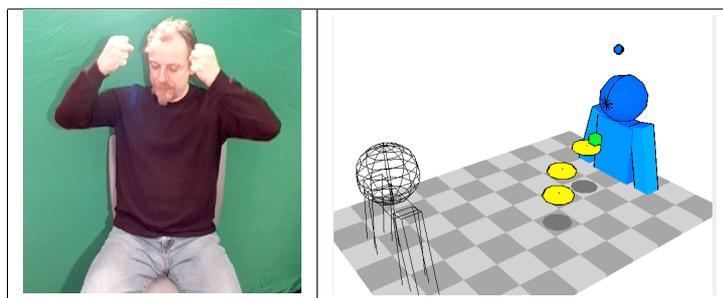


TAB. 5 – Transcription : Rue en pente (Canebière)

- Signe : [CANEBIERE]
- Début (sec) : 18.99
- Type : Lieu
- Position dans VIES : H3-V3-S2
- Constat par rapport au choix de l'annotateur : le lieu [CANEBIERE] est placé très près du buste, plus haut que le lieu [PORT].

⁴On voit ici une des limitations de VIES qui ne permet pas d'indiquer l'étendue spatiale des entités.

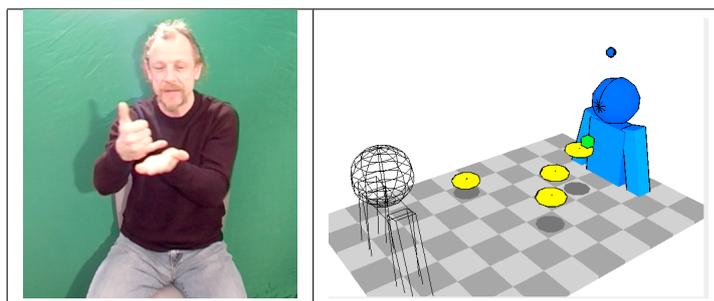
3.3.5 Transcription : Eglise (les Réformés)



TAB. 6 – Transcription : Eglise (les Réformés)

- Signe : [REFORMES]
- Début (sec) : 19.63
- Type : Objet
- Position dans VIES : H3-V3-S2
- Constat par rapport au choix de l'annotateur : l'objet [REFORMES] est placé au début de [CANEBIERE].

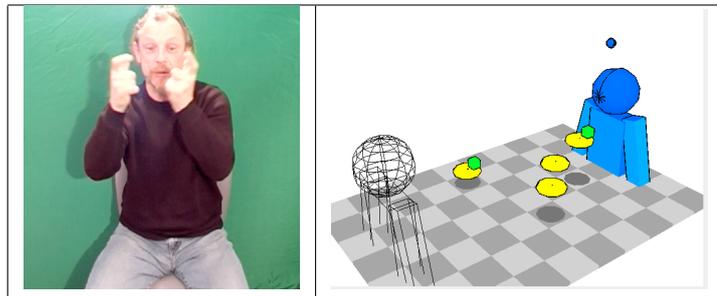
3.3.6 Transcription : "île"



TAB. 7 – Transcription : "île"

- Signe : [ILE]
- Début (sec) : 24.82
- Type : Lieu
- Position dans VIES : H1-V1-S5
- Constat par rapport au choix de l'annotateur : le lieu [ILE] est placé plus loin à droite du signeur.

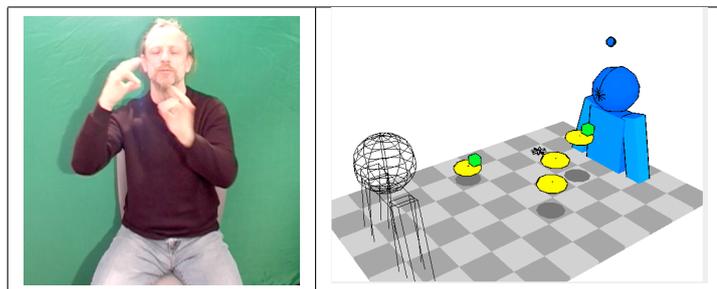
3.3.7 Transcription : Château (château d'If)



TAB. 8 – Transcription : Château (château d'If)

- Signe : [CHATEAU]
- Début (sec) : 26.2
- Type : Objet
- Position dans VIES : H1-V0-S5
- Constat par rapport au choix de l'annotateur : L'annotateur reporte le fait que le signeur a placé l'objet [CHATEAU] exactement dans le lieu [ILE].

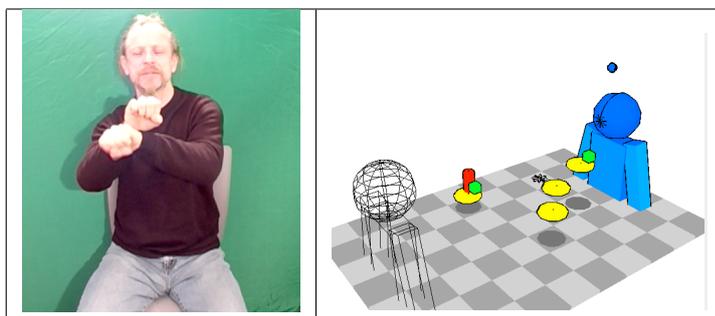
3.3.8 Transcription : "Date"



TAB. 9 – Transcription : "Date"

- Signe : [HISTOIRE]
- Début (sec) : 27.66
- Type : Date
- Position dans VIES : H1-V0-S2
- Constat par rapport au choix de l'annotateur : [DATE] est placée à droite du signeur près de son épaule droit.

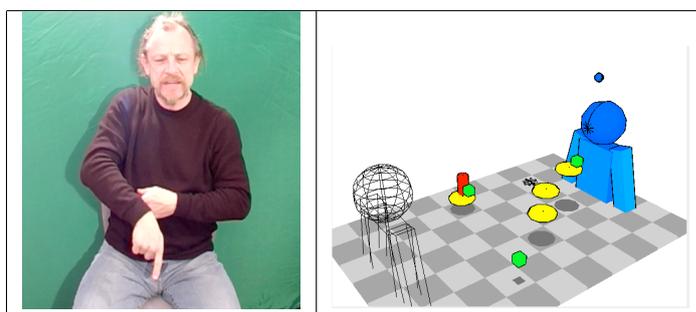
3.3.9 Transcription : "Prisonnier"



TAB. 10 – Transcription : "Prisonnier"

- Signe : [PRISONNIER]
- Début (sec) : 28.57
- Type : Personne
- Position dans VIES : H1-V1-S5
- Constat par rapport au choix de l'annotateur : L'annotateur reporte le fait que le signeur a placé l'objet [PRISONNIER] exactement dans le lieu [ILE]..

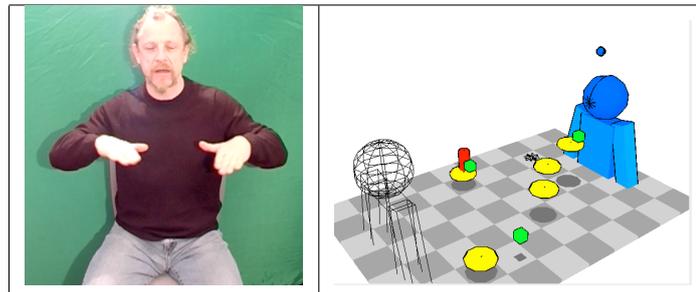
3.3.10 Transcription : "Statue David"



TAB. 11 – Transcription : "Statue David"

- Signe : [PENIS] (signe de la statue David)
- Début (sec) : 35.86
- Type : Objet
- Position dans VIES : H5-V1-S6
- Constat par rapport au choix de l'annotateur : l'objet [PENIS] est placé à gauche du signeur derrière [CORNICHE].

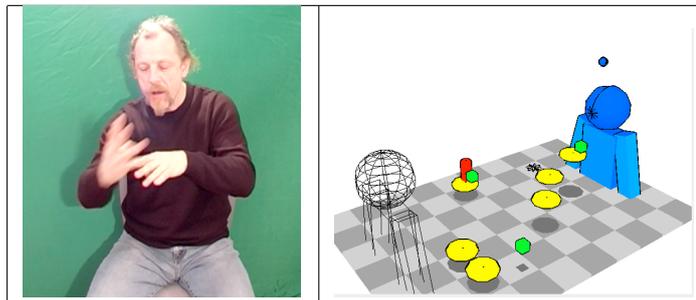
3.3.11 Transcription : "Plage"



TAB. 12 – Transcription : "Plage"

- Signe : [PLAGE]
- Début (sec) : 44.91
- Type : Lieu
- Position dans VIES : H5-V1-S7
- Constat par rapport au choix de l'annotateur : le lieu [PLAGE] est placé derrière [PENIS].

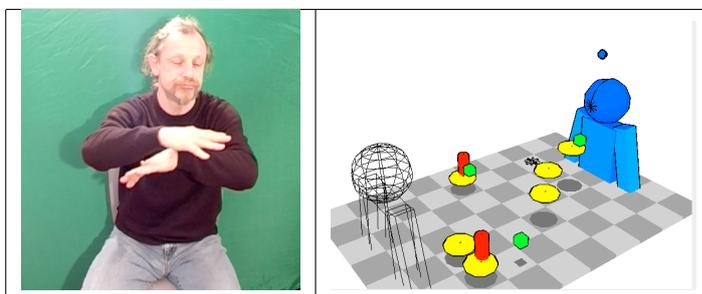
3.3.12 Transcription : "Prairie"



TAB. 13 – Transcription : "Prairie"

- Signe : [PRAIRIE]
- Début (sec) : 49.78
- Type : Lieu
- Position dans VIES : H4-V1-S7
- Constat par rapport au choix de l'annotateur : le lieu [PRAIRIE] est placé derrière [PLAGE].

3.3.13 Transcription : "Foule"



TAB. 14 – Transcription : "Foule"

- Signe : [FOULE]
- Début (sec) : 51.42
- Type : Personne
- Position dans VIES : H5-V1-S7
- Constat par rapport au choix de l'annotateur : L'annotateur reporte le fait que le signeur a placé [FOULE] exactement dans le lieu [PRAIRIE].

3.3.14 Commentaires sur les résultats de la transcription VIES

- Le placement des entités dans le schéma 3D est fidèle à leur position dans l'espace de signation, en termes de hauteur et de proximité.
- Bien que VIES ne permette pas de représenter l'étendue des entités, l'annotateur a mis en évidence un élément important en plaçant le lieu [CANEBIERE] plus haut que [PORT] et [CORNICHE] et très près du buste. Ceci est fait pour deux raisons :
 1. Privilégier la mise en évidence de la forme spécifique du début de [CANEBIERE].
 2. Tenir compte du contexte et en particulier de l'objet qui sera placé par la suite dans ce même lieu, c'est-à-dire l'objet [REFORMES].
- Pour des raisons de visibilité, l'annotateur était contraint d'attribuer le type "objet" à l'entité [CHATEAU] (au lieu du type "lieu") car VIES ne permet pas de rendre visibles deux lieux superposés.
- L'annotateur n'a pas introduit d'action dans cette transcription. Cela aurait pu être fait pour [VISITER], en traduisant la séquence (voir section 3.1.5) : [BIEN] [JOLI] [VISITER] [TOI], par "c'est joli, tu peux le visiter", ce qui aurait été réalisé par une entité "personne" localisée sur l'interlocuteur ([TOI]) et une action (représentée par une flèche) depuis cette entité vers l'entité [CHATEAU]. Ici, il a plutôt traduit cette séquence comme un attribut du château (château joli à visiter), ce qui n'apparaît pas dans l'espace de signation.

4 Conclusion

On a présenté deux stratégies d'analyses. La première, ascendante, qui part des mesures numériques pour caractériser et détecter des éléments syntaxiques. La deuxième, descendante, qui part du sens et en déduit des éléments syntaxiques structurés. Ce document présente une première mise en œuvre de cette démarche pour caractériser un pointage manuel. Les résultats de reconnaissance ne sont pas encore satisfaisants. Cela donne lieu à plusieurs pistes de réflexion, outre l'introduction de nouvelles mesures, celle de fusionner les deux stratégies pour mettre en place des méthodes de reconnaissance à différents niveaux de compréhension.

Références

- BEN MLOUKA M., ALBARET-LEFEBVRE F., DALLE J. & DALLE P. (2010). Annotation automatique d'une vidéo en lsf à partir de données de capture de mouvement. In *TALS*, Montréal, Canada.
- BRAS G., MILLET A. & RISLER A. (2004). Anaphores et deixis en lsf – tentative d'inventaire des procédés. In *journée d'études internationales, La linguistique de la LSF : recherches actuelles*, p. 57–64, Université de Lille : revue Silexicales.
- CUXAC C. (2000). *Faits de Langues - La langue des signes française (LSF) - Les voies de l'iconicité*. Faits Des Langues. Paris : Ophrys.
- FUSELLIER-SOUZA I. (2004). *Sémiogénèse des langues des signes : étude de langues des signes primaires (LSP) pratiquées par des sourds brésiliens*. Université PARS 8.
- LEFEBVRE-ALBARET F. & DALLE P. (2010). Requête dans une vidéo en langue des signes modélisation et comparaison de signes. In *RFIA*, Caen, France.
- LENSEIGNE B. & DALLE P. (2005). A tool for sign language analysis through signing space representation. In *Sign Language Linguistics and the Application of Information Technology to Sign Languages*, Università degli Studi di Milano, Milan, Italie.
- LIDDELL S. & JOHNSON R. (1989). American sign language : the phonological base. In *Sign Language Studies*, Vol. 64, p. 197–277.

Les pointages en Langue des Signes : une méthodologie d'annotation

Nadège Devos (1) Flora Badin (2) Émilie Chételat-Pelé (3)

(1) LIMSI-CNRS, Campus d'Orsay Bt 508, BP133, 91 403 Orsay cx

(2) INIST-CNRS, 2, allée du parc de Brabois, 54519 Vandoeuvre-lès-Nancy

(3) Université de Provence, 29 av. Schuman, 13100 Aix en Provence
nadege.devos@limsi.fr, flora.badin@inist.fr, emilie.chetelat@univ-provence.fr

Résumé Dans le cadre du premier DEfi GEste Langue des Signes (DEGELS) 2011, nous proposons une méthode d'annotation de corpus vidéo de langue des signes. Celle-ci permet une analyse des pointages aussi bien au niveau de leur structure que de leur interaction avec des éléments non manuels comme le regard et les clignements. Cette annotation met en évidence un schéma type propre au pointage en Langue des Signes Française (LSF).

Abstract For the French GESture and Sign Language challenge (DEGELS), we propose a corpora annotation methodology of sign language videos. It allows an analysis of pointing signs: their structure and their interaction with non-manuals events such as gaze or blinks. This annotation reveals one pattern of pointing signs in French Sign Language.

Mots-clés : Langue des signes française, méthodologie d'annotation, pointages, regard, clignements

Keywords: French sign language, annotation methodology, pointing signs, gaze, blinks

Introduction

Dans le cadre du premier DEfi GEste Langue des Signes (DEGELS) 2011 nous nous intéressons aux pointages en langue des signes française.

Nous définissons dans une première partie notre objet d'étude : les pointages.

Dans la seconde partie de cet article, nous présentons plus particulièrement le système d'annotation de la langue des signes de Johnston dont nous nous inspirons. Nous complétons son système pour proposer une nouvelle méthodologie d'annotation plus étendue.

Enfin dans une dernière partie nous présentons les résultats de notre annotation, avec une mise en évidence d'un schéma type de pointage.

1 Objet de l'étude

Selon Cochet et Vauclair (2010), le pointage est un geste de communication. Dans l'étude du développement du langage chez l'enfant, ils le définissent comme l'extension du bras et de l'index vers un objet, une personne ou un événement. Ils constatent que le pointage est en général associé à une vocalisation et un regard en alternance dirigé vers l'objet du pointage ou le visage de l'interlocuteur. Wilkins (2003) généralise un pointage comme l'utilisation d'une partie du corps (main, bouche, yeux) pour faire une référence déictique gestuelle.

Le pointage peut également être vu comme un geste temporel (Cooperrider, Núñez, 2009), qui dirige l'attention visuelle d'un interlocuteur vers des objets, des personnes ou des événements (Clark, 2003) (Pizzuto, Capobianco, 2008).

Le regard joue un rôle crucial dans le pointage aussi bien pour les langues vocales (Jorio, 2000 ; Enfield, 2001) que signées (Cuxac, 2000 ; Dalle, Lefebvre-Albaret, 2009). En général, lorsque l'on pointe vers une cible, le regard cherche d'abord le référent, puis l'élément qui pointe identifie la cible (Orie, 2009).

Les pointages en langue des signes peuvent être définis comme des gestes déictiques, utilisant l'index tendu ou la main, le plus souvent « associés à un arrêt marqué du mouvement de désignation ». Ils sont généralement associés à une rotation du buste ou du visage, ainsi qu'à un mouvement de l'épaule ou du menton. (Dalle, Lefebvre-Albaret, 2009). Ils peuvent être dirigés vers une personne comme un lieu, dans l'environnement proche ou lointain, réel ou virtuel du locuteur. L'emplacement de l'entité désignée peut être dans le passé, le futur ou encore être hypothétique (Liddell, 2003).

Nous considérons par la suite comme pointage tout événement manuel qui sert à désigner une entité par le biais d'au moins un doigt tendu vers une direction ou un point de l'espace.

Nous présentons ci-dessous notre méthodologie d'annotation.

2 Méthodologie d'annotation

Dans un premier temps, nous expliquons l'organisation mise en place entre les annotatrices ainsi que les choix techniques réalisés. Nous présentons ensuite notre méthodologie d'annotation et en détaillons chaque piste.

2.1 Aspects techniques

Notre équipe se compose de trois personnes de formations différentes et éloignées géographiquement. Nous avons choisi de travailler sur la vidéo en LSF. Pour annoter le corpus, nous utilisons le logiciel Anvil, version 5.0 beta 17 (Kipp, à paraître). Cet outil offre une grande liberté à l'annotateur quant à la structuration des éléments qu'il souhaite annoter, et à la façon de le faire (à l'aide de mots, de symboles, etc.). Il permet également la mise en relief des éléments choisis avec par exemple l'utilisation de couleurs.

Toutes ces spécifications sont définies dans un fichier XML qu'il est possible de modifier en partie grâce à l'interface graphique. De plus, Anvil peut être utilisé par tous les systèmes d'exploitation, sans problème de compatibilité ce qui nous a permis de travailler aussi bien sous Mac (OS X) que sous Linux (Ubuntu 10.4).

Les tâches ont été divisées entre les annotatrices. Une personne a d'abord annoté les éléments manuels (signes standards, signes iconiques¹ -dont les pointages-, gestes et dactylogie) et la cible du regard. Une seconde annotatrice a repéré les clignements et a segmenté les phases des pointages. Enfin, une troisième personne a vérifié les segmentations, la nomenclature utilisée et a fait ressortir les éléments pertinents au moyen de couleurs.

Nous ne disposons pas à l'heure actuelle de logiciels suffisamment précis pour réaliser des annotations automatiques d'éléments du visage. C'est pourquoi nous avons annoté manuellement. Par ailleurs, notre annotation est principalement subjective car elle fait appel à l'interprétation des corpus par les annotatrices.

L'annotation et sa vérification sont estimées en tout à 24h, réparties sur une semaine.

2.2 Aspects théoriques

Nous replaçons dans une première section le cadre théorique de notre étude puis, dans une seconde section, nous présentons notre méthodologie d'annotation.

2.2.1 Annotation des éléments du pointage

Nous distinguons ici les éléments manuels (les pointages à proprement parler), et les éléments non manuels (le regard et les clignements).

Les pointages selon Johnston

Afin d'élaborer notre schéma d'annotation nous avons tout d'abord étudié la littérature. Nous nous sommes particulièrement intéressées aux travaux de (Johnston, 2011), eux-mêmes inspirés de ceux de (Liddell, 2003), qui proposent une réflexion poussée sur les pointages en proposant une classification riche et détaillée. Nous expliquons ci-dessous cette méthodologie sur laquelle nous baserons notre étude.

Johnston répertorie les pointages comme des signes partiellement lexicaux et les signale sous l'abréviation « PT ». Il en répertorie actuellement neuf² :

- Si le pointage désigne un référent, Johnston utilise les pronoms personnels sujets (PRO) avec l'écriture PT:PROxSG/PL. Cela permet de préciser s'il s'agit d'une personne du singulier (SG) ou du pluriel (PL), x précisant laquelle (1, 2 ou 3). Par exemple, un pointage vers le buste de la personne qui pointe est défini par « PT:PRO1SG » qui correspond au je/moi (1^{ère} personne du singulier), tandis que « PT:PRO3PL » correspond à ils/elles/eux (3^{ème} personnes du pluriel).
- Si le pointage désigne un lieu, « PT:LOC³ » est utilisé.
- Si le pointage définit la localisation du référent, les deux précédentes catégories sont combinées en « PT:LOC/PRO » par défaut au singulier mais la terminaison « PL » peut être rajouté le cas échéant.
- Si le référent est déjà connu et qu'il est réalisé simultanément ou avant le pointage, il fait office

¹ Johnston définit une catégorie de signe qu'il nomme *Depicting signs*. Ces signes sont parfois aussi nommés *classifier signs* ou *classifier predicates* (Liddell, 1977). Nous les avons traduits ici par signes iconiques. Ils peuvent difficilement être listés dans un dictionnaire. Ce sont des signes illustratifs, qui peuvent varier selon le contexte. Ils sont différents des signes lexicaux, plus stables, dont le sens et la forme suivent des conventions (Johnston, 2011). Ils correspondraient aux transferts de situation ainsi qu'aux transferts de taille et de forme de Cuxac (2000).

² Johnston en aborde un dixième, actuellement en étude (TBUOY), qui traiterait des pointages abstraits

³ LOC pour *location*

de déterminant « PT:DET ».

- Si le pointage désigne celui qui possède un élément ou l'élément en possession, il est noté « PT:POSS ».
- Si le pointage est dirigé vers l'autre main dans le but de lister (cette main aura alors la configuration d'un chiffre), il utilise « PT:LBUOY⁴ ». Quand une main maintient la dernière configuration d'un signe tandis que l'autre continue de signer, elle est dénommée « fragment buoy⁵ » par Johnston. Un pointage vers une main dans cet état est appelé « PT:FBUOY ».
- Si le pointage est dirigé vers un élément de l'espace correspondant à une entité ou une idée, qu'il est maintenu tandis que l'autre main signe des choses relatives à ce qui est pointé, il est appelé « PT:BUOY ».
- Si le pointage est dirigé vers un signe iconique, il utilise « PT:BUOY ».

Les notations des catégories de pointages sont récapitulées dans le Tableau 5 en annexe.

Étude sur le regard et les clignements

Comme nous l'avons évoqué, les pointages sont souvent accompagnés d'éléments non manuels, comme par exemple le regard. Concernant celui-ci, Johnston (2011) l'annote uniquement quand le locuteur signe. Il utilise quatre valeurs : l'interlocuteur (addressee), la cible (target), autre (other) et non identifiable (cannot be coded).

Les études portant sur les clignements sont plus rares et récentes, mais ont montré leur importance dans les pointages (Chételat-Pelé, 2010). Nous utilisons ici une méthodologie mise au point précédemment (Chételat-Pelé, Braffort, Véronis, 2008) et qui reprend et adapte le découpage temporel de Johnson et Liddell (à paraître) en postures clé et transitions.

Pour les clignements⁶ cela se traduit par une décomposition en trois postures et deux transitions. Les postures sont les états œil ouvert, œil fermé puis œil ouvert. Les transitions sont les mouvements qui conduisent d'une posture à l'autre (Figure 1).

Les différentes phases du mouvement sont distinguées dans l'annotation à savoir la fermeture et l'ouverture de la paupière.



Figure 1: phases d'un clignement

Nous avons adapté l'ensemble de ces modèles afin de créer notre propre méthodologie d'annotation que nous présentons ci-dessous.

2.2.2 Schéma d'annotation

Sur la vidéo montée (Figure 2) du corpus mise à notre disposition, nous nommons A le locuteur de gauche et B le locuteur de droite.

⁴ LBUOY pour *list buoy*

⁵ Fragment de signe standard, réutilisable par la suite

⁶ Nous considérons comme clignement toute fermeture de l'œil qui n'excède pas une image (Chételat-Pelé, Braffort, 2010). En revanche, les transitions peuvent avoir une durée totalement variable.



Figure 2: Extrait de la vidéo montée

Annotation des éléments manuels

Nous annotons les événements manuels à l'aide de cinq pistes⁷. Dans une piste 'RH⁸' figure ce que le locuteur fait avec sa main droite (élément monomanuel), dans une piste '2H⁹' ce qu'il fait avec ses deux mains (élément bimanuel) et dans une piste 'LH¹⁰', avec sa main gauche (élément monomanuel). Dans une quatrième piste nommée 'Hands Interpretation¹¹' est notée une interprétation des annotateurs quand le locuteur fait autre chose que des signes lexicaux. Enfin, dans une piste 'Pt structure¹²' est précisé le découpage temporel des pointages.

Les pistes peuvent être indépendantes ou posséder des dépendances sur les bornes temporelles quand elles sont liées à une autre piste. Cette spécification des pistes est définie selon le type qui sera choisi. Chaque piste est constituée d'un ensemble d'éléments délimités par un temps de début et un temps de fin, nommés segments. Les segments possèdent une ou plusieurs valeurs selon le nombre d'attributs que l'on souhaite leur donner.

Pour la plupart des pistes, nous avons choisi deux types d'attributs. Le premier permet d'entrer n'importe quelle chaîne de caractères. Le second laisse à l'annotateur le choix d'une valeur dans une liste prédéterminée. Nous récapitulons dans le Tableau 1 les différentes pistes, toutes indépendantes, et les attributs utilisés pour cette annotation.

⁷ Les noms des pistes sont évoqués à l'aide de guillemets simples tandis que les valeurs des segments sont présentées entre guillemets doubles

⁸ *Right Hand* (Main droite)

⁹ *2 Hands* (Deux mains)

¹⁰ *Left Hand* (Main gauche)

¹¹ Interprétation des événements manuels

¹² *Pointing structure* (Structure des pointages)

Nom de la piste	Type des pistes	Type des attributs	Valeurs contrôlées des attributs
RH	Indépendante	Chaîne de caractères	
2H	Indépendante		
LH	Indépendante		
Hands interpretation	Indépendante	Chaîne de caractères	pt:loc pt:det pt:pro1sg
		Valeurs contrôlées	
Pt Structure	Indépendante	Valeurs contrôlées	Preparation Hold Retraction

Tableau 1: Pistes et attributs utilisés pour les éléments manuels

Nous annotons la configuration de l'élément qui pointe et la cible du pointage dans les pistes 'RH', 'LH' ou '2H' selon le format « doigt(s) en jeu:cible ». La configuration la plus fréquente de la main lors du pointage est un index tendu. Il est également possible de pointer avec l'ensemble des doigts, le pouce, etc. La cible permet de préciser l'élément de l'espace visé par le pointage et son identification demande une part d'interprétation de la part de l'annotateur. Les cibles sont en nombres finis. S'il s'agit de l'interlocuteur, on utilisera « Ad¹³ ». S'il s'agit d'une cible sur le corps du signeur, nous utilisons les sites corporels utilisés dans le langage Zebedee¹⁴ (Filhol, 2008). Nous écrivons par exemple « index:@ST » si le locuteur pointe son sternum. Tout pointage vers l'espace de signation¹⁵ est noté « SSp¹⁶ ». Si l'emplacement dans l'espace de signation n'a pas d'importance, on pourra utiliser « X ». Si la cible est un objet du réel, on note « real ». Enfin, si la cible ne peut pas être déterminée nous notons « ? ».

La piste 'Hands interpretation' reprend la convention d'écriture de Johnston, (2011). Elle contient en partie l'interprétation des pointages précisés dans les piste 'RH', 'LH' ou '2H'. L'analyse du corpus a montré la présence de trois des types de pointages (pt:loc, pt:det et pt:pro1sg) dans notre annotation. Ceci explique le peu de valeurs contrôlées pour cette piste. Cette dernière permet également de définir précisément la cible du pointage au moyen de l'attribut de type chaîne de caractères (exemple : pt:loc, île d'If). Cette précision est subjective et peut varier d'un annotateur à un autre. Elle permet tout de même une meilleure lecture de l'annotation, notamment lorsqu'elle est partagée par plusieurs personnes.

Les segments des quatre premières pistes sont découpés selon la méthodologie de Johnston (2011). Nous considérons qu'un segment (qu'il s'agisse d'un signe lexical, d'un signe iconique ou d'un pointage) commence quand la ou les mains changent de direction et/ou de configuration et d'orientation. De plus dans le cas de pointages, le sens du mouvement doit être celui de la cible. De la même manière, un segment se termine juste avant que la main semble changer de direction et/ou change de configuration et d'orientation. Ces bornes de début ou de fin de segment correspondent souvent à des images nettes. Dans le cas contraire, on choisira l'image suivante.

¹³ Adresse

¹⁴ Zedebée est un langage de modélisation géométrique de la langue des signes

¹⁵ Ensemble des emplacements possibles des réalisations gestuelles dans l'espace ou sur le corps (Liddell, 2003)

¹⁶ Signing Space

Dans la dernière piste, nous découpons les pointages en phases : préparation, maintien et rétraction, dont les valeurs sont respectivement « Preparation », « Hold » et « Retraction ». Nous considérons que la préparation commence au moment où la main se met en mouvement pour adopter la posture du pointage. Le maintien correspond à toutes images où la posture est adoptée complètement (par exemple quand le doigt est tendu) et où la cible est pointée. Enfin, nous estimons que la rétraction commence au moment où la posture se relâche, jusqu'à ce qu'on ne puisse plus distinguer le pointage. A noter également que certains pointages peuvent ne pas présenter de phase de maintien c'est à dire qu'à peine la posture est adoptée, elle est aussitôt relâchée. De plus, certains pointages comportent un mouvement arrière-avant dont l'intégralité du mouvement sera assimilée à la phase de maintien. En effet, lors de ces mouvements, la posture est adoptée et la cible est pointée. En revanche, l'intégralité du mouvement ne sera pas prise en compte dans la délimitation du segment pointage de la piste des éléments manuels correspondante, qui lui correspond à un sens précis du mouvement (vers la cible).

Annotation des éléments non manuels

Nous annotons le regard en deux pistes : 'Gaze' et 'Gaze interpretation'¹⁷. Dans la piste 'Gaze', est annoté ce que le locuteur regarde de manière générale. Quand le regard se porte sur l'espace de signation, ce que le locuteur regarde précisément est identifié dans la piste 'Gaze interpretation'.

Cette piste est dépendante de la piste 'Gaze'. Autrement dit, les deux pistes auront les mêmes tailles de segments.

La piste 'Gaze' peut prendre sept valeurs. Le locuteur peut regarder son interlocuteur « Ad » ou l'espace de signation « SSp ». Il peut activer une zone de l'espace de signation (Cuxac, 2000) « SSp activation » en le regardant avant d'utiliser cette zone pour y placer une entité. Son regard peut se diriger vers un élément du réel « real ». Il peut aussi décrocher, par exemple quand le signeur réfléchit ou que son regard quitte sa cible première pendant quelques images, on utilise alors « X ». Enfin les annotateurs peuvent ne pas savoir ce que le signeur regarde, et le préciser alors au moyen d'un « ? ». S'il n'est pas possible de voir où le signeur regarde, par exemple en raison d'absence de zoom, de qualité d'image ou de son positionnement, cela sera spécifié au moyen de « ### ». Il ne s'agit donc pas d'indiquer la cible du regard en terme de coordonnées géométriques mais d'interpréter la cible.

La piste 'Gaze interpretation', permet, comme 'Hands interpretation', une meilleure lecture de l'annotation en faisant figurer la cible du regard selon l'annotateur (exemple : île d'If). Si la cible ne peut être définie avec certitude mais que l'annotateur pense qu'il s'agit d'un point dans l'espace de signation, il l'indiquera grâce à la valeur « SSp supposé ».

Nous donnons dans le Tableau 2 les caractéristiques de ces pistes.

¹⁷ Interprétation du regard

Nom de la piste	Type des pistes	Type des attributs	Valeurs contrôlées des attributs
Gaze	Indépendante	Valeurs contrôlées	Ad SSp SSp activation X ### ?
Gaze interpretation	Dépendante de la piste Gaze	Chaîne de caractères	-

Tableau 2: Pistes et attributs utilisés pour l'annotation du regard

Enfin, nous annotons les clignements en deux pistes indépendantes (Tableau 4). Dans la piste 'Blinks'¹⁸ est noté l'ouverture (Op¹⁹) ou la fermeture (Cl²⁰) des paupières. Dans la piste 'Blinks interpretation'²¹ est défini le rôle du clignement (Chételat-Pelé, 2010). Par exemple, les clignements qui précèdent un regard « SSp activation » vont être identifiés par « Point act » pour activation du pointage. Les clignements qui suivent un pointage, eux, seront nommés simplement « Point » pour pointage.

Nom de la piste	Type des pistes	Type des attributs	Valeurs contrôlées des attributs
Blinks	Indépendante	Valeurs contrôlées	Cl Op
Blinks interpretation	Indépendante	Chaîne de caractères	-

Tableau 3: Pistes et attributs utilisés pour l'annotation des clignements

En raison d'un certain nombre de problèmes techniques nous n'avons pu procéder à l'ensemble de l'annotation initialement souhaitée. Sous Mac OS X, un problème de chargement des vidéos, signalé par un message d'erreur, a été rencontré. De plus, la modification du choix de la vidéo dans le fichier XML ne prenait pas effet. Les clignements ont donc été repérés seulement sur la vidéo non montée du locuteur B. Par ailleurs, la basse résolution des vidéos a uniquement permis une annotation des clignements les plus visibles. Il était en effet parfois difficile de savoir si le locuteur clignait des yeux ou les fermait. De même, pour cette raison, et à cause de l'absence de plan rapproché, nous n'avons finalement pas annoté les mouvements des sourcils, l'essai réalisé ayant été trop approximatif.

Une fois l'annotation réalisée, nous avons procédé à une analyse des pointages relevés. Nous en présentons les résultats dans la partie suivante.

3 Résultats de l'annotation

L'analyse de notre annotation a permis de mettre en évidence onze pointages : trois sont réalisés par le signeur A et huit par le signeur B. Nous les récapitulons dans le Tableau 4.

¹⁸ Clignements

¹⁹ *Open*

²⁰ *Close*

²¹ *Interprétation des clignements*

Locuteur : image (durée du pointage en nombre d'images)	Doigt(s) en jeu:cible	Durée de la phase de maintien = durée du pointage	Regard	Clignements
A : 14 (1)	index:@ST	oui	X	Non annotés
A : 102 (2)	index:SSp	oui	Ad	Non annotés
A : 809 (4)	index:SSp	non	Ad	Non annotés
B : 310 (1)	index:SSp	oui	Yeux fermés	Yeux fermés
B : 341 (3)	index:SSp	Pas de maintien	SSp activation	Avant
B : 554 (1)	index:SSp	Pas de maintien	SSp activation	-
B : 623 (8)*	index:SSp	oui	SSp activation	Avant et après
B : 834 (3)*	index:SSp	oui	Ad	Après (fermeture)
B : 856 (2)*	index:SSp	oui	SSp activation	Après
B : 901 (6)	index:SSp	non	SSp activation	Après
B : 1057 (1)*	index:@ABST	oui	Ad	Après (fermeture)

Tableau 4: Synthèse des pointages de notre annotation

Lors de la phase d'analyse, nous avons procédé à des ajustements de segments de la grille d'annotation. Les astérisques indiquent les endroits où l'annotation originale a été modifiée. Les éléments en gras correspondent aux pointages les plus longs.

À noter également que comme nous l'avons indiqué, nous considérons comme clignement toute fermeture de l'œil qui n'excède pas une image. Or, le locuteur B cligne souvent des yeux sur plusieurs images, ce qui n'a donc pas été considéré comme des clignements mais comme des fermetures (ils sont donc ignorés dans l'annotation). Nous les avons tout de même précisées par « Après (fermeture) » dans la dernière colonne du tableau qui indique la position des clignements par rapport au pointage.

L'observation de notre annotation nous a permis de mettre en évidence deux structures de pointage :

- Dans la première structure, la désignation d'un élément proche de l'espace de signation précède le pointage de l'élément (pt:det) ou de son emplacement (pt:loc). Cela correspond à six pointages.
- La deuxième structure est utilisée dans le cas de pointages d'éléments lointains de l'espace de signation ou d'auto pointages. Dans le premier cas, le pointage de l'élément précède la désignation de l'élément. Ceci correspond aux deux pointages en gras dans le tableau. Dans le second cas, le pointage précède la suite du discours. Ceci concerne quatre pointages.

Nous avons également fait ressortir un schéma de l'interaction entre les pointages et les éléments non manuels qui serait un pointage suivi d'un clignement (Figure 3).

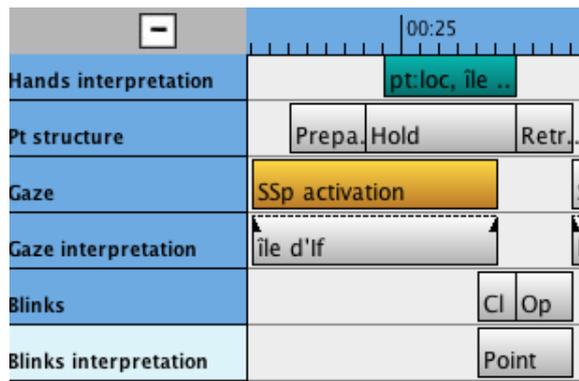


Figure 3: Illustration du schéma, tirée de notre annotation

Par exemple, dans la Figure 3, la première piste montre un pointage, qui est découpé en phases dans la seconde piste. Les deux dernières pistes montrent le clignement qui se situe donc à la fin du pointage. Concernant le regard, nous distinguons deux cas de figures²² :

- Le regard peut activer l'espace de signation juste avant le pointage (dans cinq des huit pointages analysés) comme le montre la troisième piste de la Figure 3. Le pointage pourra alors être suivi d'un clignement (dans trois cas).
- Le regard peut ne pas interagir avec le pointage : dans ce cas, le locuteur regarde le plus généralement son interlocuteur (dans trois pointages).

Nous rencontrons cependant une exception à ce schéma. En effet, lorsque le regard active l'espace de signation et qu'il n'y a pas de phase de maintien dans le pointage, il n'y a pas de clignement après (pour deux pointages).

Conclusion

Notre méthodologie permet d'annoter avec précision les pointages en interaction avec certains éléments non manuels. L'annotation est visuelle et facilite l'analyse. Elle nous a permis de mettre rapidement en évidence un schéma et des structures de pointages. Cependant, notre méthodologie ne tient pas compte du buste, de la tête ou encore des sourcils. Cela représente de potentielles perspectives à cette étude.

De plus, les éléments mis en évidence par notre annotation sont basés sur l'observation de onze pointages, il serait donc pertinent d'élargir notre étude sur un plus vaste corpus afin de valider et d'affiner nos schémas types.

De même, nous avons noté deux éléments manuels que nous avons nommés « pi », dans le corpus. Nous pourrions les catégoriser comme pointage, étant donné la configuration de la main. Nous les avons donc annotés comme tel. Cependant, ils diffèrent des autres pointages. En effet, ils ne sont pas dirigés vers un élément choisi de l'espace de signation, leur fonction linguistique semble mettre l'emphase sur ce qui précède dans le discours. On pourrait aisément leur attribuer un sens lexical. Il serait intéressant, sur un plus vaste corpus, de mener une analyse plus poussée afin d'observer les similitudes et les variances des « pi » par rapport aux pointages afin de les classer dans des signes lexicaux ou partiellement lexicaux.

Ce défi nous a permis de dégager des caractéristiques des pointages en LSF et de mettre en relation des schémas d'annotation et des critères de découpage temporel que nous avons testés séparément pour l'annotation des pointages et des clignements. Elles sont non seulement parfaitement compatibles mais

²² Les clignements n'étant pas annotés chez le locuteur A ses pointages ne sont pas comptabilisés dans cette partie.

également complémentaires. Nous pourrions envisager, comme validation de notre méthodologie, de la confronter avec d'autres, par exemple lors d'une journée centrée sur les pointages.

Références

CHÉTELAT-PELÉ E. (2010). Les Gestes Non Manuels en langue des signes française ; Annotation, analyse et formalisation : Application aux mouvements des sourcils et aux clignements des yeux. *Thèse de Doctorat*, Université de Provence, Aix-en-Provence.

CHÉTELAT-PELE E., BRAFFORT A. (2010). Investigation et analyse des Gestes Non Manuels impliqués en LSF: Le cas des clignements, *Actes de la conférence Traitement Automatique des Langues Naturelles (TALN)*, atelier *Traitement Automatique des Langues des Signes (TALS)*, 2010, Montréal.

CHÉTELAT-PELE E., BRAFFORT A., VÉRONIS J. (2008). Sign Language Corpus Annotation: Toward a New Methodology. *Proceeding of the Conference on Language Resources and Evaluation (LREC): Workshop Construction and Exploitation of Sign Language Corpora 2008*. Marrakesh.

CLARK H. (2003). Pointing and placing. *Pointing: where language, culture, and cognition meet*. Mahwah, London: Lawrence Erlbaum Associates.

COCHET H., VAUCLAIR J. (2010). Features of spontaneous pointing gestures in toddlers. *Gesture* 10 n°1, 86-107.

COOPERRIDER K., NÚÑEZ R. (2009). Accross time, accross the body: Transversal temporal gestures. *Gesture* 9 n°2, 181-206.

CUXAC C. (2000). La Langue des Signes Française : Les voies de l'iconicité. *Faits de Langues* 15-16. Paris : Ophrys.

DALLE P., LEFEBVRE-ALBARE F. (2009). Analyse des pointages en LSF par traitement automatique de vidéos. *Publication personnelle*²³.

ENFIELD N.J. (2001). Lip pointing: a discussion of form and function with reference to data from Laos. *Gesture* 1 n°2, 185-211.

FILHOL M. (2008). Modèle descriptif des signes pour un traitement automatique des langues des signes, *Thèse de doctorat*, Université Paris-11, Orsay.

JOHNSON R. E., LIDDELL S. K. (à paraître) *Sign Language Phonetics: Architecture and Description*.

JOHNSTON T. (2011). *Auslan Corpus Annotation Guidelines*. March 2011.

JORIO DE A. (2000). *Gesture in Naples and gesture in classical antiquity*. Bloomington & Indianapolis: Indiana University Press.

KIPP M. (à paraître). Multimedia Annotation, Querying and Analysis in ANVIL, in *M. Maybury (ed.) Multimedia Information Extraction*, Chapter 19, IEEE Computer Society Press.

LIDDELL S. K. (1977). An investigation into syntactic structure of American Sign Language. *Ph.D. dissertation*, University of California, San Diego.

LIDDELL S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge :

²³ Accessible à l'adresse http://www.irit.fr/~Patrice.Dalle/Publis/pointages09_PD.pdf

Cambridge University Press.

ORIE O. O. (2009). Pointing the Yoruba way. *Gesture* 9 n°2, 237-262.

PIZZUTO E.A., CAPOBIANCO M. (2008). Is pointing “just” pointing? Unraveling the complexity of indexes in spoken and signed discourse. *Gesture* 8 n°1, 82-103.

WILKINS D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). *Pointing: where language, culture, and cognition meet*, 171-215. Mahwah, London: Lawrence Erlbaum Associates.

Annexe

PT:PRO	Pointage vers un référent (moi, toi, lui...)
PT:LOC	Pointage vers un lieu (ici, là...)
PT:LOC/PRO	Pointage vers un référent locatif (lui ici, eux là-bas...)
PT:DET	Pointage à fonction de déterminant (ça, celui-là...)
PT:POSS	Pointage vers le référent qui possède un objet ou vers l'objet (ton, ta...)
PT:LBUOY	Pointage vers l'autre main pour lister des éléments (premièrement, en deuxième...)
PT:FBUOY	Pointage vers une main en configuration de flotteur isolé
PTBUOY	Pointage vers un élément de l'espace qui est maintenu pendant une activité de l'autre main relative à ce qui est pointé
PT:BUOY	Pointage vers un signe iconique

Tableau 5: Les pointages selon Johnston (2011)

Typologie des pointages en Langue des Signes Française (LSF) et problématiques de leur annotation

Brigitte Garcia, Marie-Anne Sallandre, Camille Schoder, Marie-Thérèse L'Huillier

Structures Formelles du Langage, UMR 7023 CNRS-Université Paris 8, 59/61 rue Pouchet, 75017 Paris
brigitte.garcia@univ-paris8.fr, marie-anne.sallandre@univ-paris8.fr,
camille.schoder@sfl.cnrs.fr, marie-therese.lhuillier@sfl.cnrs.fr

Résumé Cet article rend compte de l'annotation semi-automatique d'une séquence vidéo de dialogue en Langue des Signes Française (LSF) entre deux adultes sourds, réalisée sous ELAN (4.0.1) en vue d'une étude fonctionnelle des types de pointage dans cette langue. Après un bref descriptif du cadre théorique spécifique qui sous-tend l'analyse (Cuxac 2000), on présente en détail les choix méthodologiques réalisés et le schéma d'annotation original élaboré à cette occasion (choix du médium graphique, des pistes, de leurs relations de dépendance, des listes d'items). On s'attache à restituer les questionnements et difficultés rencontrés en cours d'annotation et les solutions successivement apportées. On montre ainsi en quoi le schéma d'annotation finalisé a résulté d'un va-et-vient heuristique entre analyse, saisie des annotations et révision progressive des caractéristiques initialement choisies (pistes, types, vocabulaire contrôlé). On conclut par un bilan des apports et limites des choix opérés et du logiciel exploité, en pointant les perspectives d'optimisation.

Abstract This article discusses the semi-automatic annotation of a dialogue recorded on video in French Sign Language (LSF) between two deaf adults, through ELAN (4.0.1), conducted for a functional study of pointing methods. The article begins with a brief description of the theoretical framework underlying our analysis (Cuxac 2000), followed by a detailed presentation of our methodological choices and the original annotation template developed for this study (including the choice of graphic medium, tiers, internal relations and lists of items). We intend to reproduce the questions and difficulties encountered during the annotation process and the solutions reached. Finally, it will be shown how the completed annotation template was obtained through a heuristic feedback between analysis, annotation and progressive revision of the initially selected properties (tiers, types, controlled vocabulary). We conclude with a summary of the advantages and limitations of the choices made and of the software used, pointing issues for future improvement.

Mots-clés : annotation, transcription, Langue des Signes Française (LSF), schéma d'annotation, ELAN, corpus, pointages, dialogue, modèle sémiologique.

Keywords: annotation, transcription, French Sign Language (LSF), schéma d'annotation, ELAN, corpus, pointings, dialogue, semiological model.

1 Introduction

L'annotation de données discursives de langue des signes (LS) pose des problèmes spécifiques qui tiennent en premier lieu au fait que ces langues ne disposent d'aucune forme graphique propre institutionnalisée et, à ce titre, d'aucun support consensuellement admis pour leur transcription (*vs*, pour les langues vocales (LV), l'API et/ou les orthographes standard plus ou moins aménagées). Dans cet arrière-plan, le présent article rend compte de la méthodologie mise en œuvre et des types de questions rencontrées pour l'annotation, centrée particulièrement sur la problématique de l'analyse des pointages, d'un court dialogue en Langue des Signes Française (LSF) entre deux adultes sourds¹.

La question de la description formelle et fonctionnelle des pointages en LS et, en amont, celle de leur statut (linguistique / non linguistique) sont parmi les plus débattues dans la littérature linguistique relative à ces langues (*e.g.* Meier, 1990, Engberg-Pedersen, 1993, Cuxac, 2000, Liddell, 2003, Johnston et Schembri, 2007, Antinoro Pizzuto et Capobianco, 2008, Meurant, 2008). Nous appréhendons quant à nous cette catégorie d'unités des LS dans le cadre spécifique du modèle dit « sémiologique » (Cuxac 2000) et dans la perspective large de contribuer à une typologie de ces unités.

Après un bref descriptif des grandes caractéristiques du modèle théorique qui sous-tend notre analyse, nous rendons compte de nos options technologiques —quel logiciel d'annotation et pourquoi?— et des principales étapes de notre démarche (section 2.). Nous décrivons ensuite le détail du schéma d'annotation élaboré à cette occasion en précisant la nature et les raisons des modifications successives apportées pour l'optimiser ; nous justifions de même les choix relatifs au médium graphique des annotations et à la détermination de listes fermées d'items (section 3.). Nous indiquons finalement quelques pistes pour l'optimisation du logiciel utilisé et, au-delà, celle de l'annotation du discours en LS (section 4.).

2 Options théoriques et méthodologiques

2.1 Cadre théorique de description de la LSF

Notre cadre théorique est celui du *modèle sémiologique* (Cuxac, 2000, Sallandre, 2003, Fusellier-Souza, 2004, Cuxac et Sallandre, 2007, Cuxac et Antinoro Pizzuto, 2010, Garcia et Derycke, 2010, Garcia, 2010). Dans ce modèle, ancré sur une approche énonciative et fonctionnelle, le regard est décrit comme jouant un rôle linguistique majeur. Condition et recteur de l'interaction, marqueur de la distinction entre plans de l'énonciation et de l'énoncé et des passages de l'un à l'autre, il est également le vecteur (et l'indice ultime) du choix, toujours possible dans les LS, entre les deux grands types de visée sémiologique que sont la visée « illustrative » (dire en montrant) et la visée « non illustrative » (dire sans montrer).

La mise en œuvre systématisée de l'une et l'autre de ces deux visées permet de rendre compte de l'existence et des structures de deux grands types d'unités gestuelles en LSF (dans les LS) :

- les unités lexématiques (UL)² —caractéristiques des productions hors visée illustrative ;

¹ Cette expérience d'annotation (comme le présent article) s'inscrit dans le cadre de l'atelier *DEGELS* (*Défi Geste Langue des signes*) organisé par *LPL* (CNRS-Université de Provence) et le *LIMSI* (CNRS-Orsay) [<http://degels.limsi.fr/>]. Le corpus à annoter, d'une durée de 54 secondes, est mis à disposition par les organisateurs [CRDO d'Aix-en-Provence, réf. oai:crdo.fr:crdo000767]. Les métadonnées relatives aux locuteurs ne sont pas précisées. La focalisation de l'annotation sur les phénomènes de pointage répond à la consigne générale commune donnée par les organisateurs du *Défi*.

² La désignation antérieure de « signe standard » (*e.g.* Cuxac, 2000) pour les unités lexicales conventionnalisées semblait exclure du « standard » de la LSF les unités par définition non conventionnalisées que sont les unités de transfert. Ceci allant à l'encontre du modèle sémiologique pour lequel ces unités non conventionnelles sont justement au cœur de la norme des LS, le terme préférentiellement adopté depuis peu est celui d'« unité lexématique » (Cuxac, 2009, Cuxac et Antinoro Pizzuto, 2010).

- les unités de transfert (UT), résultantes structurales et expression de la mise en œuvre d'une visée illustrative (également dites « structures de grande iconicité »).

Deux autres types d'unités existent dans le discours en LS : les unités dactylogiques (qui, moyen interlingue d'intégration des mots de la LV et par là étroitement associées au lexique, relèvent clairement d'une visée non illustrative) et les unités de pointage. Ces dernières, très fréquentes dans le discours³ constituent un objet particulièrement intéressant pour la description des LS et la compréhension de leurs spécificités :

- a) les pointages occupent en effet une place à part, probablement à mi-chemin des deux grands types de structures décrits ci-dessus, dans la mesure où, relevant par nature d'un « montrer » (de type indexical), ils peuvent se combiner aussi bien aux UL qu'aux UT (Sallandre, 2003 : 154, pour une description des pointages apparaissant dans des unités en discours rapporté) ;
- b) en eux-mêmes formellement identiques aux pointages de la gestualité dite coverbale, ils remplissent cependant des fonctions linguistiques clés en LS, notamment par la dynamique de leur couplage avec le regard.

La littérature présente diverses typologies des phénomènes de pointage dans différentes LS (*e.g.* Engberg-Pedersen, 1993 pour la DSL ; Emmorey et Reilly, 1995, Liddell, 2003 pour l'ASL ; Johnston et Schembri, 2007 pour l'Auslan⁴). Ces typologies ont cependant été le plus souvent proposées dans des perspectives théoriques qui situent hors champ (ou, au mieux, considèrent comme périphériques) les unités de type « UT » (Garcia, 2010) et envisagent les phénomènes de pointage indépendamment des comportements du regard. Dans le cadre du modèle sémiologique, pour lequel ce sont au contraire les UT qui constituent le centre des LS et qui octroie un rôle essentiel au regard et à la dynamique de son couplage avec les pointages, un essai de typologie fonctionnelle des pointages a été proposé par Cuxac pour la LSF (Cuxac, 2000), repris par Fusellier-Souza pour trois LS émergentes (Fusellier-Souza, 2004). Garcia, Sallandre et Fusellier-Souza (2009) ont par ailleurs amorcé, en relation avec le travail systématique mené par Rinfret (2008) pour la LSQ (LS québécoise), une analyse du rôle du pointage dans l'instanciation d'une entité référentielle et, entre autres, dans l'expression de la définitude (en LSF mais également en LS Mauricienne). Ces travaux constituent les références de l'analyse présentée ici et de son annotation.

2.2 Choix méthodologiques et aspects techniques du travail d'annotation

Nous avons opté pour le logiciel d'annotation semi-automatique ELAN, dans sa version 4.0.1 pour Windows, 4.0.0 pour Mac OS X. Les raisons de ce choix sont diverses.

L'annotation de discours en LS implique de recourir à un éditeur de partition en raison, d'une part, du caractère multiparamétrique de ses unités —elles mobilisent parallèlement paramètres manuels et corporels—, d'autre part, de l'exploitation linguistique que font ces langues de l'espace et du temps (pertinence des relations spatiales et temporelles entre les divers composants paramétriques et, notamment, entre unités manuelles, regard et pointage).

Parmi les logiciels d'annotation, la préférence octroyée à ELAN par rapport notamment à ANVIL (Kipp, to appear) tient à une raison essentielle. Engagées dans la réalisation d'un grand corpus de LSF (projet ANR *Creagest*, 2007-2012, Balvet *et al* 2010) nous sommes tenues par un impératif d'interopérabilité avec les grands corpus de LS déjà élaborés ou en cours d'élaboration (Johnston, 2008a pour le corpus Auslan ; Schembri, 2008 pour le corpus BSL ; Crasborn, Zwitserlood, Ros, 2008 pour le corpus NGT ; Hanke, König, Wagner, Matthes, 2010 pour le corpus DGS). Tous, de fait, recourent à ELAN. Opter ici pour ELAN participe de l'ensemble de nos expérimentations des fonctionnalités de ce logiciel. Pour autant, nous sommes conscientes que les deux logiciels présentent chacun leurs avantages propres (Boutet et Braffort, 2011).

L'annotation du corpus de dialogue étudié a mis à contribution trois annotatrices différentes. Notons toutefois que, celles-ci ayant travaillé de manière successive, il n'a pas été nécessaire de recourir à la fonction « fusion des annotations », disponible dans ELAN. Au total, l'élaboration finalisée de l'annotation

³ Cuxac (2000 : 283) indique 950 pointages à valeur de reprise d'un signe lexématique sur un corpus d'une heure.

⁴ La DSL est la LS danoise, l'ASL la LS américaine, l'Auslan la LS australienne.

(du corpus vidéo au format .avi jusqu'au fichier .eaf final) a requis une trentaine d'heures, la seule réalisation concrète du schéma d'annotation sous ELAN ayant représenté une part conséquente de ce temps. Les principales étapes du processus d'ensemble ont été les suivantes :

- premier séquençage en « unités de sens » (hors ELAN) ; validation de cette analyse par une locutrice sourde dont la LSF est la langue principale ; catégorisation linguistique des unités ainsi identifiées et, parmi elles, repérage des pointages ; ajustement corollaire de la traduction ;
- sur ces bases, élaboration d'une première version du schéma d'annotation (types, pistes, vocabulaires contrôlés) ;
- travail d'annotation à proprement parler :
 - annotation par la première annotatrice des pistes 'unité de sens', 'regard', 'traduction' et 'catégorie' pour chacun des deux locuteurs (cf. *infra* Figure 1.)
 - annotation (par les deux autres annotatrices travaillant de conserve sur un même poste) des autres pistes, liées directement à l'analyse des pointages : va-et-vient entre saisie des valeurs et modifications du schéma d'annotation et du vocabulaire contrôlé. La section suivante revient plus en détail sur les motivations et l'enseignement de ce nécessaire va-et-vient entre annotation et révision du schéma d'annotation.

3 Le schéma d'annotation

3.1 Description générale

Cette annotation concernant une situation de dialogue, nous avons opté pour un ordre correspondant à celui des prises de parole : le premier locuteur est ainsi désigné par « L1 », son interlocuteur par « L2 ». Ce choix est un pis-aller dans la mesure où c'est le discours du second intervenant qui constitue l'essentiel du propos.

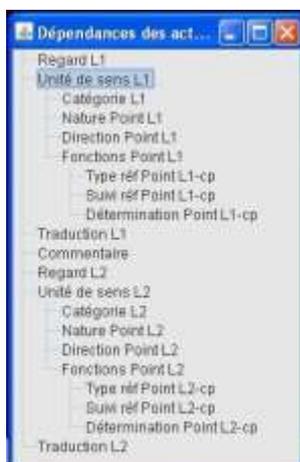


Figure 1 : Représentation du schéma d'annotation

Nous avons représenté le discours de chaque interlocuteur par deux groupes de pistes distincts. Le schéma d'annotation comporte vingt-et-une pistes organisées comme suit : dix pistes pour L1, les mêmes dix pistes pour L2 et une piste 'commentaire' valable pour les deux locuteurs. Les pistes pour chaque locuteur sont : 'regard', 'unité de sens', 'catégorie', 'nature du pointage', 'direction du pointage', 'fonctions du pointage', 'type de référence du pointage', 'suivi référentiel du pointage', 'détermination du pointage', 'traduction'.

Comme le montre la figure 1, les pistes indépendantes sont le regard L1 et le regard L2, la traduction L1 et la traduction L2 et le commentaire. Celles-ci ont un type associé mais sans spécification, elles sont seulement alignées sur le temps. En revanche, la piste 'unité de sens' est le « parent » de toutes les autres

pistes auxquelles nous avons associé une relation de dépendance du type ‘association symbolique’. C’est en effet à partir de la segmentation en unités de sens que nous définissons la catégorie attachée à chaque unité, en suivant la typologie décrite à la section 2.1.

Les six pistes restantes sont dédiées à la description fine des pointages, avec deux niveaux de hiérarchie :

- les pistes ‘nature du pointage’, ‘direction du pointage’ et ‘fonctions du pointage’ sont seulement dépendantes de ‘unité de sens’ et une valeur doit être affectée à chaque pointage ;
- les pistes ‘type de référence du pointage’, ‘suivi référentiel du pointage’ et ‘détermination du pointage’ sont dépendantes de la piste ‘fonctions du pointage’ et leurs valeurs ne sont remplies que si la valeur « référentielle » a été sélectionnée dans la piste ‘fonctions du pointage’.

Notre choix de nous focaliser sur les grands types fonctionnels de pointage a eu pour conséquence que certains aspects formels du pointage ne sont pas mentionnés dans cette annotation. Par exemple, nous n’avons pas jugé indispensable de préciser quelle main effectuait le pointage, dans quel point exact de l’espace de signation se situait le pointage, etc. Tous ces éléments relèvent de choix des chercheurs et seules des annotations complémentaires sur un échantillon plus grand de corpus pourront permettre de dire si ces choix se révèlent pertinents ou non.

3.2 Evolution du schéma d’annotation

On peut considérer, *a posteriori*, que le schéma d’annotation a subi trois étapes majeures d’évolution avant d’apparaître dans sa version actuelle. Les changements ont concerné avant tout les pistes relatives aux pointages.



Figure 2 : Représentations du schéma d’annotation intermédiaire (partie gauche) et du schéma d’annotation final (partie droite)

Au départ, nous avons défini deux groupes de pistes pour les pointages : l’un concernait leurs formes, l’autre leurs fonctions. Mais nous avons une difficulté avec la hiérarchie de ces pistes et les types à associer. Nous avons donc tenté un deuxième schéma d’annotation, intermédiaire (partie gauche de la figure 2) dans lequel figurait une série de pistes où l’indication de la forme ou de la fonction du pointage n’apparaissait pas (il n’y avait pas d’hyperonyme visible). Cet « aplatissement » de la hiérarchie des niveaux d’information ne nous convenant pas, nous avons rétabli l’hyperonyme ‘fonctions du pointage’ en y ajoutant un vocabulaire contrôlé plus précis (voir section 3.4) et en lui associant trois pistes « enfants ».

Ce dernier changement a accru la lisibilité et la précision de l’ensemble. Il reste cependant deux problèmes d’ordre général :

- la hiérarchie conceptuelle ne peut être rendue visible que si des valeurs sont annotables pour l'hyperonyme ; c'est pour cette raison que l'hyperonyme « forme » n'apparaît pas.
- les pistes dépendant de 'fonctions du pointage' ne sont pas des hyponymes mais des spécificateurs pour seulement une partie des fonctions envisageables.

Ces diverses manipulations dans le schéma d'annotation nous ont néanmoins aidées à clarifier notre typologie linguistique. Ainsi, l'élaboration même du schéma d'annotation a nettement eu une fonction heuristique.

3.3 Attributs de l'annotation

Une fois déterminées les pistes et leurs relations de dépendance, se pose la question des modalités de l'étiquetage, tant en ce qui concerne le choix d'un médium graphique et des conventions régissant son utilisation, que pour ce qui regarde le choix d'établir ou non des listes fermées de valeurs (vocabulaires contrôlés).

1/ Médium graphique d'annotation et conventions

Le seul médium de notre annotation est le français écrit, c'est-à-dire la forme écrite standard de la LV cohabitant avec la LSF. Il s'agit pour nous d'un choix par défaut, dont nous mesurons très largement les limites et, au-delà, les incidences potentiellement préjudiciables à divers niveaux. Outre le caractère hétérogène et en partie arbitraire de l'usage qui en est fait (cf. *infra*), le tagging *via* les mots de la LV surimpose nécessairement à la langue étudiée les structures et catégories issues d'une autre langue ou, *a minima*, risque de masquer ou de faire oublier la spécificité des phénomènes linguistiques ainsi annotés (e.g Cuxac, 2000, Pizzuto et Pietrandrea, 2001, Di Renzo *et al*, 2009, Garcia, 2001, 2010).

Nous y recourons néanmoins, faute de solution alternative satisfaisante. En effet :

- L'adoption d'« Idgloses » n'est envisageable que s'il existe pour la LS annotée une base conséquente d'UL glosées de manière « consistante » c'est-à-dire cohérente et fondée sur l'analyse de corpus en quantité suffisante (cf. Johnston & Schembri, 1999, Johnston, 2001, 2008b). Ce n'est pas le cas actuellement pour la LSF. Notons en outre qu'une telle base, que nous appelons de nos vœux, ne résoudrait pas pour autant la question de l'annotation des UT, sauf à y intégrer les unités de sens de niveau paramétrique qui composent les deux grands types d'unités gestuelles, UL et UT (sur ce point précis, v. Cuxac 2004, Boutet et Garcia 2007).
- Les systèmes de notation spécifiquement élaborés pour rendre compte de la forme signifiante des unités gestuelles des LS présentent d'autres types de limites. Le plus répandu et sans doute le plus complet des systèmes de type monolinéaire, *HamNoSys*, conjugue deux difficultés : a/ du fait du principe de représentation monolinéaire, identifier une même unité lexématique sous les formes diverses qu'elle peut prendre en discours (forte variation morphologique interne) est le plus souvent impossible⁵, b/ le système est peu opérationnel pour rendre compte des UT (Garcia, 2001, 2010, Boutora, 2008). A la fois global et multilinéaire, combinant analogique et digital, apte à représenter de manière également 'lisible' UL et UT tout en rendant compte de leur isomorphie structurelle, le système *Sign Writing* pourrait, lui, constituer une alternative intéressante (Di Renzo *et al* 2009). Il pose néanmoins deux problèmes non encore résolus : l'absence de standard pour la restitution graphique de la structure interne des UL (une même UL peut faire l'objet de plusieurs restitutions graphiques concurrentes) et, surtout, l'absence de compatibilité actuelle avec les logiciels d'annotation automatique comme ELAN ou ANVIL (Di Renzo *et al*, 2009, Bianchini, *et al*, 2010).

⁵ Raison pour laquelle Johnston (2001) par exemple opte pour une représentation en *HamNoSys* de la seule forme non fléchie (« stem ») des signes conventionnels, forme d'étiquetage somme toute guère plus informative quant à la forme signifiante effective du signe en discours que ne l'est la glose en LV. Ceci d'ailleurs conduit l'auteur à la considérer comme dorénavant non nécessaire, la vidéo intégrée suffisant selon lui à fournir cette information sur la forme signifiante.

Reste que le recours au français écrit est pour nous un pis-aller et qu'une alternative doit être trouvée. Les types d'usage que nous en faisons sont par ailleurs hétérogènes.

a) Pour trois des pistes, le français écrit joue (mal et de manière seulement fragmentaire) le rôle qui est dévolu aux systèmes de transcription dans le cas de l'annotation de corpus de LV (API ou orthographe standard).

Il en est ainsi, pour la piste 'unité de sens', où nous adoptons les conventions suivantes :

- Pour les UL, nous recourons à la « glose » classique, où le mot français en lettres capitales (par exemple [PLAGE]) prétend référer non pas directement au sens mais d'abord à la forme signifiante (tel signe de LSF qui se réalise de telle façon précise, et qui a tel sens), suivant Cuxac (2000 : 326). Faute d'un recours possible à des IdGloses, le choix des gloses est ici fondé sur les associations interlingues traditionnelles et, dans le cas précis de la séquence analysée, aux labialisations concomitantes lorsqu'elles existent (par exemple 'David' pour [LE DAVID] labialisé par le locuteur L1, 00:41.500).
- Pour les UT, nous recourons aux minuscules, ce choix graphique signalant simplement que contrairement aux UL, les UT ne sont pas des unités conventionnalisées. Seule l'existence d'une base de données intégrant de manière conséquente les composants forme-sens (morphèmes) constituant (de manière régulière) les UT pourrait permettre de pallier le côté actuellement artisanal et non satisfaisant de ces étiquettes.
- Pour les pointages, catégorie à part mais dont les valeurs ne sont précisément pas encore établies, nous avons également opté pour les minuscules.

De même, pour les deux pistes 'reg' et 'direction', les items qui constituent le vocabulaire contrôlé décrivent un aspect de la forme signifiante (cf. *infra* 2/).

b) Concernant les autres pistes, le français écrit est en revanche le médium de l'annotation à proprement parler (« tagging »). Il rend en effet compte des catégories métalinguistiques de l'analyse (métalangage), ceci concernant aussi bien les items servant à désigner les pistes (« Type référence pointage », par exemple) que, pour l'annotation proprement dite, les items composant les vocabulaires contrôlés, comme « défini », « indéfini », « démonstratif », etc.

Autrement dit, le recours au français écrit correspond tantôt à une fonction de « transcription » (piste 'unités de sens'), tantôt à une fonction d'annotation (et, en fait, ni l'une ni l'autre, l'étape de 'transcription' à proprement parler faisant spécifiquement défaut en ce qui concerne les LS par rapport aux LV : cf. Pizzuto et Pietrandrea, 2001, Garcia, 2010).

2/ Listes de vocabulaires contrôlés

Seules les pistes 'unité de sens' (L1 et L2), 'traduction' (L1 et L2) et 'commentaire' n'ont pas de vocabulaire contrôlé car leur valeur est nécessairement libre en ce qui concerne la traduction, et encore en partie indéfinie pour ce qui est des unités de sens du fait de l'absence, pour le moment, d'une base de données qui permettrait de fixer des étiquetages stables (par exemple du type « forme plate ondulant »).

Pour les autres pistes, huit vocabulaires contrôlés ont été créés et associés à un type⁶. La liste ainsi que les valeurs des entrées de chaque vocabulaire contrôlé sont présentées ci-dessous :

- 'reg' (pour « regard ») : 'interlocuteur', 'forme de la main', 'vague', 'espace de signation', 'pointeur', 'fermeture des yeux'. Ces valeurs indiquent les caractéristiques formelles du comportement du regard, celles-ci correspondant aux principales fonctions de ce paramètre qui ont été répertoriées par Cuxac (2000). Ainsi notamment de la distinction entre les visées, selon que le regard est sur l'interlocuteur (visée non illustrative) ou non (visée illustrative) ou encore de la fonction d'activation d'un point ou d'une zone de l'espace ('espace de signation').

⁶ Le vocabulaire contrôlé et le type portent des noms très proches, par commodité. Par exemple : le vocabulaire contrôlé « détermination » est associé au type « T Détermination Point ».

- ‘cat’ (pour « catégorie ») : ‘UL’, ‘TTF’, ‘TS’, ‘TP’, ‘TP divers’, ‘Pointage’, ‘Dactylogogie’, ‘Autre’. Ici encore, nous reprenons les principales catégories proposées par Cuxac (2000) en ayant fait le choix de ne pas mentionner le panel de toutes les UT car cela n’était pas une priorité pour l’étude des pointages et le corpus proposé ne présentait pas ce panel (à une exception près, en 00:38.748, l’unité « Le David » que nous avons considérée comme un pseudo-TP, donc étiquetée ‘TP divers’). Pour la valeur ‘Autre’, notons enfin que nous ne l’avons finalement pas utilisée car nous avons mentionné les éléments intéressants le cas échéant dans la rubrique ‘commentaire’. Cette dernière valeur serait donc sans doute à supprimer pour des annotations ultérieures.
- ‘nature’ (pour « nature du pointage », c’est-à-dire le type de forme, manuelle ou non, que peut prendre un pointage) : ‘index’, ‘main plate’, ‘regard’, ‘autre’. En fait, les deux valeurs les plus utilisées ont été les deux premières, ce qui est conforme à la littérature sur le sujet. Nous n’avons pas relevé de pointage uniquement exprimé par le regard, bien que cette forme particulière soit attestée (Antinoro Pizzuto & Capobianco, 2008).
- ‘direction’ (pour « direction du pointage ») : ‘locuteur’, ‘interlocuteur’, ‘énoncé de l’interlocuteur’, ‘énonciation’, ‘main dominée’, ‘main dominante’, ‘point de l’espace’, ‘zone de l’espace’. La direction d’un pointage est une information importante tant en informatique qu’en linguistique et les approches sont nombreuses dans ce domaine (cf. section 2.1). Nous avons choisi une description assez neutre mais qui intègre une approche énonciative et interactionnelle du langage. Comme indiqué dans la section précédente, la valeur des items est de type formel, indiquant la nature de ce qui est pointé. Deux valeurs doivent toutefois être explicitées à cet égard, celles d’‘énonciation’ et d’‘énoncé de l’interlocuteur’. L’item ‘énonciation’ réfère à une dimension qui n’a pas d’actualisation matérielle (le plan de l’énonciation tel que le définit Jakobson (1963)) : il s’agit de qualifier ces types particuliers de pointage qui sont caractérisés par une moindre tension musculaire et une faible valeur indexicale, dont la fonction est de type « modalo-énonciatif » (gestion de l’interaction). L’item ‘énoncé de l’interlocuteur’ renvoie aux cas où l’index pointeur est dirigé non pas vers l’espace de signation du locuteur mais bien plutôt vers l’espace dans lequel l’interlocuteur a produit son énoncé et, au sein de cet espace (désormais partagé), placé telle entité discursive.
- ‘fonct point’ (pour « fonction du pointage ») : ‘gestion de l’interaction’, ‘référentielle’, ‘métalinguistique’, ‘poétique’. Ce vocabulaire est celui qui a connu le plus de modifications au cours de ce travail. L’idée de départ était de reprendre les six fonctions du langage de Jakobson (1963) et d’associer l’une ou l’autre d’entre elles à chaque pointage. Nous avons finalement conservé telles quelles les fonctions ‘référentielle’, ‘métalinguistique’ et ‘poétique’ et nous avons regroupé les trois fonctions en lien avec l’interaction (fonctions conative, phatique et expressive) en une seule fonction que nous avons appelée ‘gestion de l’interaction’. Ce choix a simplifié notre annotation tout en nous permettant de conserver une cohérence théorique.
- ‘construct réf’ (pour « construction de référence ») : ‘actancielle’, ‘spatiale’, ‘temporelle’. Les trois domaines référentiels attestés dans la littérature (e.g. Hendriks et Watorek, 2008) ont été simplement repris.
- ‘suivi réf’ (pour « suivi référentiel ») : ‘introduction’, ‘reprise’. La littérature dans ce domaine (e.g. Trevisiol, Watorek, Lenart, 2010) propose en général trois formes de suivi référentiel que sont l’introduction, le maintien et la reprise d’une entité dans le discours, les deux dernières formes étant parfois regroupées. C’est ce dernier choix que nous avons fait pour plus de simplicité.
- ‘détermination’ : ‘défini’, ‘indéfini’, ‘démonstratif’, ‘pluriel’, ‘quantifieur’, ‘autre’. Le champ de la détermination est un champ d’études important pour les LV mais seulement en cours d’exploration pour les LS (e.g. Rinfret, 2008 ; Garcia, Sallandre, Fusellier-Souza, 2009), c’est pourquoi nous lui avons attribué des valeurs classiques tout en ménageant la possibilité de trouver une valeur nouvelle, non encore identifiée, notée ‘autre’ (un exemple se trouve en 00:42.132).

Ainsi, une question qui émerge au sujet de ces vocabulaires contrôlés est de savoir s’il est préférable de laisser des valeurs qui ne seront pas utilisées dans un corpus donné mais qui ont été attestées par ailleurs, ou si l’on doit se limiter aux seules valeurs présentes dans ce corpus donné. Nous avons opté ici pour une position intermédiaire, qui ménage une certaine « marge de liberté » dans un cadre donné et qui confère à l’annotation une dimension de découverte.

4 Remarques conclusives et perspectives

Il n'est pas question ici de présenter une analyse fine des résultats ni de détailler chaque piste ou groupe de pistes. Ainsi, le tableau 1 permet de récapituler les résultats quantitatifs issus de l'annotation des pointages.

Caractérisation des pointages :	Locuteur 1	Locuteur 2	Total
Nbre/ Nbre unité sens	5/29	15/83	20/112
% pointages	17%	18 %	17,85 %
Nature	index : 5	index : 13 main plate : 2	index : 18 main plate : 2
Fonction	gestion interaction : 2 référentielle : 3	gestion interaction : 4 référentielle : 11	gestion interaction : 6 référentielle : 14
Type de référence*	spatiale : 3	spatiale : 9 actancielle : 2	spatiale : 12 actancielle : 2
Détermination*	défini : 1 démonstratif : 2	défini : 6 démonstratif : 3 autre : 2	défini : 7 démonstratif : 5 autre : 2

*Nota : ne concerne que les 14 pointages référentiels.

Tableau 1 : Tableau récapitulatif des résultats de l'annotation des pointages

A la lumière de ce tableau, un point nous semble particulièrement saillant : le pourcentage de pointages (17,85%) largement supérieur à celui notamment relevé dans Sallandre (2003) pour le corpus LS-COLIN, qui était de 1% à 8% selon les locuteurs en transversal des genres discursifs (narration et explication). Notons toutefois que ce corpus LS-COLIN, strictement monologique, ne comportait pas de discours de type description d'une organisation spatiale qui, précisément, caractérise le corpus ici annoté. Ces caractéristiques discursives (dialogue, description de l'espace) expliquent le taux élevé des pointages et, en particulier, le nombre important de pointages de type modalo-énonciatif (gestion de l'interaction).

L'annotation reflète bien sûr l'analyse faite par le chercheur. L'emploi d'un schéma d'annotation astreint toutefois à une grande rigueur dans la catégorisation et à des choix conséquents de valeurs. Elle facilite à ce titre l'identification des cas problématiques. Nous voudrions illustrer ce point par une courte discussion concernant un exemple du corpus qui nous a posé question et qui reflète à lui seul la difficulté et l'intérêt d'analyser finement les pointages en LSF. Il s'agit de l'unité que nous avons glosée par « avant la descente » (timer : 00:18.800). Nous avons fait le choix de la coder comme un pointage dans une zone de l'espace. Elle peut tout aussi bien être analysée comme un transfert situationnel, dont elle présente toutes les caractéristiques (mouvement dynamique de la main dominante, locatif par la main dominée, regard et mimique faciale accompagnant le mouvement de la main dominante). Il est probable qu'elle soit, en fait, un mixte de ces deux catégories. Ce genre d'unité nous interroge sur les frontières de ce que l'on considère comme un pointage et fonde la nécessité de maintenir une piste « commentaire ».



Figure 3 : Arrêt sur image de l'unité « avant la descente », 00 :18.800, (locuteur 2, à droite)

Au final, l'annotation proposée mêle deux niveaux, correspondant à deux objectifs, nécessaires à l'analyse visée :

- a) Une volonté d'annoter finement tout le discours, en segmentant chaque unité de sens et en lui attribuant une catégorie sémantico-énonciative (UL, transfert, etc.), et en indiquant la nature/direction du regard (fondamental dans notre modèle). Ceci nous fournit le cadre global indispensable.
- b) Une focalisation sur les pointages, à la fois dans leur forme et dans leur fonction. Comme indiqué plus haut, nous n'avons pas souhaité cependant entrer dans le détail des aspects formels, mais plutôt conjuguer aspects formels et fonctionnels, l'enjeu étant le dégagement de grands types.

Astreignant à une catégorisation rigoureuse des critères et principes d'analyse, la conception d'un schéma d'annotation s'avère heuristique. Le temps requis pour finaliser un schéma d'annotation doit être compensé, à terme, par la puissance des requêtes possibles. La manipulation d'ELAN nous permet toutefois de relever un certain nombre d'aspects moins performants voire de manques dans les fonctionnalités dont nous voudrions suggérer l'optimisation ou l'ajout :

- pouvoir hiérarchiser le schéma d'annotation de différentes manières, par exemple en ayant la possibilité de choisir des types qui n'auraient qu'une fonction d'indicateur des hiérarchies conceptuelles (hyperonyme). Dans les faits, un seul type de hiérarchisation est actuellement disponible.
- pouvoir disposer d'une palette graphique avec insertion de couleurs et d'icônes serait d'une grande aide (pour l'instant, une couleur peut être sélectionnée seulement pour un piste entier, pas pour un segment unique).
- pouvoir générer un document pdf à partir d'un fichier eaf. serait utile pour la conservation, la lecture et le transport des annotations.

Plus largement, il nous semble indispensable d'œuvrer à pallier les limites actuelles relatives au médium graphique servant de support à l'annotation. Outre les avancées que l'on peut attendre des recherches sur *Sign Writing* notamment, ceci devrait selon nous passer par la mise en place d'une base de données permettant un inventaire cohérent (et fondé sur des analyses de discours) des UL de la LSF d'une part, des composants forme-sens des deux grands types d'unités gestuelles (UL et UT), d'autre part (sur ce point, cf. Boutet et Garcia 2007, Garcia 2010).

Remerciements

Nous remercions les organisateurs de l'atelier DEGELS 2011, Annelies Braffort et Leila Boutora. De plus, les tutoriaux⁷ qui nous ont guidées dans notre apprentissage d'ELAN ont été très utiles, que leurs auteurs Dominique Boutet et Annelies Braffort en soient remerciés. Nous tenons enfin à remercier tous les membres du projet ANR Creagest qui ont stimulé nos questionnements pour cette annotation.

Références

ANTINORO PIZZUTO, E. & CAPOBIANCO, M. (2008). Is pointing "just" pointing? Unravelling the complexity of indexes in spoken and signed discourse. *Gesture*, 8: 1, 82-103.

BALVET, A., COURTIN, C., BOUTET, D., CUXAC, C., FUSELLIER-SOUZA, I., GARCIA, B., L'HUILLIER, M-T. ET SALLANDRE, M-A, (2010). « The Creagest Project: a Digitized and Annotated Corpus for French Sign Language (LSF) and Natural Gestural Languages ». In: Dreuw, Philippe et al. (eds.): *LREC 2010*. Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), 19-21 mai 2010, Malte. Paris: ELRA. 469-475.

BIANCHINI, C.S., GIANFREDA, G., DI RENZO, A., LUCIOLI, T., PETITTA, G., PENNACCHI, B., LAMANO, L. ROSSINI, P. (2010). Ecrire une langue sans forme écrite: réflexions sur l'écriture et la transcription de la Langue des Signes Italienne (LIS). *Trav. Linguist. CerLiCO* (Ed. PUR, Rennes), 24: in press.

BOUTET, D., BRAFFORT, A. (2011), ELAN et ANVIL, logiciels d'annotation multimodale : principes et différences, Journée de formation organisée sous l'égide de l'ATALA (Association pour le Traitement Automatique des Langues des Signes), 22 janvier 2011. http://tals.limsi.fr/tuto/1.Annotation_principes.pdf

BOUTET, D., GARCIA, B., (2007), Compositionnalité morpho-phonétique de la Langue des Signes Française (LSF) et exploration des relations structurales entre paramètres», C. Cuxac, P. Dalle et C. Rétoré (dir.), *Traitement Automatique des Langues*, vol. 48, n° 3, Paris, 93-114. <http://www.atala.org/article438,438>

BOUTORA, L. (2008). *Fondements historiques et implications théoriques d'une phonologie des langues des signes. Etude de la perception catégorielle des configurations manuelles en LSF et réflexion sur la transcription des langues des signes*. Thèse de Doctorat en Sciences du Langage, Université Paris 8.

CRASBORN, O., ZWITSERLOOD, I., ROS, J. (2008). *Corpus NGT. An open access digital corpus of movies with annotations of Sign Language of the Netherlands*. Centre for Language Studies, Radboud University Nijmegen. <http://www.let.ru.nl/corpusngt/>.

CUXAC, C. (2000). *La Langue des Signes Française ; les Voies de l'Iconicité. Faits de Langues*, 15-16, Oprys, Paris.

CUXAC, C., (2004), 'Phonétique' de la LSF : une formalisation problématique. *Sillexicales, Linguistique de la LSF : recherches actuelles*, 4, 93-113.

CUXAC, C. (2009). Langue des Signes et gestuelle co-verbale : pour un programme commun de recherches. D. Boutet et C. Cuxac (dir.), *Le signifiant gestuel*, Les Cahiers de Linguistique Analogique, 5, Dijon, 181-228.

CUXAC C., SALLANDRE M-A. (2007). Iconicity and arbitrariness, in French Sign Language: Highly Iconic Structures, degenerated iconicity and diagrammatic iconicity. In Pizzuto E., P. Pietrandrea & R. Simone (eds), *Verbal and Signed Languages : Comparing Structures, Constructs and Methodologies*. Berlin: Mouton de Gruyter, 13-33.

CUXAC, C., ANTINORO-PIZZUTO, E. (2010). Emergence, norme et variation dans les langues des signes : vers une redéfinition notionnelle. In B. Garcia et M. Derycke (dir.), *Sourds et langue des signes. Norme et variations*, Langage et Société, n° 131, mars 2010, 37-53

⁷ <http://tals.limsi.fr/tuto/tuto.html>

DI RENZO, A., GIANFREDA, G., LAMANO, L., LUCIOLI, T., PENNACHI, B., ROSSINI, P., BIANCHINI, C., PETITTA, G., ANTINORO PIZZUTO, E. (2009), Representation - Analysis - Representation : novel approaches to the study of face-to-face and written narratives in Italian Sign Language (LIS). Communication présentée lors du *Colloque International sur les langues des signes* (CILS), Namur, 16-20 novembre 2009, Palais des Congrès de Namur, Belgique.

EMMOREY, K., & REILLY, J. (eds) (1995). *Language, Gesture and Space*. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, Hove, UK.

ENGBERG-PEDERSEN E. (1993). *Space in Danish Sign Language: The semantics and morphosyntax of the use of space in a visual language*. Hamburg: Signum Press.

FUSELLIER-SOUZA I. (2004). *Sémiogenèse des langues des signes : Étude de langues des signes primaires (LSP) pratiquées par des sourds brésiliens*. Thèse de doctorat en Sciences du Langage, Université Paris 8.

GARCIA, B. (2001). Conditions et enjeux de l'élaboration d'une notation pour les langues des signes, in *Oralité et gestualité. Interactions et comportements multimodaux dans la communication*, Christian CAVÉ, Isabelle GUAÏTELLA et Serge SANTI (éds), Actes du colloque *Orage 2001*, Laboratoire Parole et Langage — CNRS et Université de Provence, Association GeVoix, Paris, L'Harmattan : 329-333.

GARCIA, B. (2010). *Sourds, surdit , langue des signes et  pist mologie des sciences du langage. Probl matiques de la scripturisation et mod lisation des bas niveaux en langue de signes fran aise (LSF)*. M moire d'Habilitation   diriger les recherches, Universit  Paris 8.

GARCIA, B. et DERYCKE (2010). Introduction. B. Garcia et M. Derycke (coord.), *Sourds et langue des signes. Norme et variations*, Langage et Soci t , n  131, mars 2010, 5-17.

GARCIA, B., SALLANDRE, M-A, FUSELLIER-SOUZA, I. (2009). R le du pointage dans l'expression de la d finitude en langue des signes. Colloque international *Du geste au signe, le pointage dans les langues orales et sign es*, organis  par l'UMR 8163 "Savoirs, Textes, Langage" (CNRS et universit s Lille 3 & Lille 1), 4 et 5 juin 2009, Universit  de Lille 3, Villeneuve d'Ascq. (conf rence non publi e).

HANKE, T., K NIG, L., WAGNER, S., MATTHES, S. (2010). DGS Corpus & Dicta-Sign: The Hamburg Studio Setup. In: Dreuw, Philippe et al. (eds.): *LREC 2010*. Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), 22-23 mai 2010, Malte. Paris: ELRA. 106-109.

JAKOBSON, R. (1963). *Essais de linguistique g n rale*. Paris, Editions de Minuit.

JOHNSTON, T. (2001). The lexical database of Auslan (Australian Sign Language). In Wilbur R. ( d.), *Sign Language & Linguistics*, vol. 4 n  1/2, John Benjamins Publishing Company, Amsterdam / Philadelphia, 145-169.

JOHNSTON, T. (2008a). Corpus of grammar and discourse strategies of deaf native users of Auslan (Australian Sign Language), Endangered Languages Archive, SOAS, University of London. <http://elar.soas.ac.uk/node/3>. Public access from 2012.

JOHNSTON, T. (2008b). Corpus linguistics and signed languages: no lemmata, no corpus. In Proceedings of *The 3rd Workshop on the Representation and Processing of Sign Languages*, 1 r juin 2008, sixi me  dition de *Language Resources and Evaluation Conference* (LREC 2008), Marrakech, Maroc, 82-87.

JOHNSTON, T. & SCHEMBRI A. (1999). On defining lexeme in a sign language. *Sign Language & Linguistics*, 2 (1), 115-185.

JOHNSTON, T. & SCHEMBRI, A. (2007). *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge: Cambridge University Press.

KIPP, M. (to appear) Multimedia Annotation, Querying and Analysis in ANVIL. In: M. Maybury (ed.) *Multimedia Information Extraction*, Chapter 19, IEEE Computer Society Press.

LIDDELL S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.

MEIER, R. (1990). Person deixis in American Sign Language. In Susan Fischer and Patricia Siple (eds.), *Theoretical Issues in Sign Language Research*. Chicago and London: University of Chicago Press, 175-190.

MEURANT, L. (2008). *Le regard en langue des signes. Anaphore en langue des signes française de Belgique (LSFB) : morphologie, syntaxe, énonciation*. Rivages Linguistiques. Presses Universitaires de Namur, Presses Universitaires de Rennes.

PIZZUTO, E. & PIETRANDREA, P. (2001). The notation of signed texts: open questions and indications for further research. In Wilbur, R.B. (ed), 2001, *Sign Transcription and Databases Storage of Sign Information, Sign Language & Linguistics*, volume 4, number 1/2, John Benjamins Publishing Company, Amsterdam/Philadelphia, 29-43.

PIZZUTO, E., ROSSINI, P. & RUSSO, T. (2006). Representing signed languages in written form: questions that need to be posed. In ELRA, European Language Resources Association (eds), *LREC 2006 – 5th International Conference on Language Resources and Evaluation. Workshop Proceedings - W-15 : Second Workshop on the Representation and Processing of Sign Languages*, 1-6.

RINFRET, J. (2008). *L'association spatiale du nom en Langue des Signes Québécoise: Formes, fonctions et sens*. Thèse de Doctorat en linguistique, Université du Québec à Montréal.

SALLANDRE, M-A. (2003). *Les unités du discours en Langue des Signes Française. Tentative de catégorisation dans le cadre d'une grammaire de l'iconicité*. Thèse de Doctorat en Sciences du Langage, Université Paris 8.

SCHEMBRI, A. (2008). British Sign Language Corpus Project: Open access archives and the Observer's Paradox. In Proceedings of *The 3rd Workshop on the Representation and Processing of Sign Languages*, 1^{er} juin 2008, sixième édition de *Language Resources and Evaluation Conference (LREC 2008)*, Marrakech, Maroc.

TREVISIOL, P., WATOREK, M., LENART, E. (2010). Topique du discours /topique de l'énoncé – réflexions à partir de données en acquisition des langues. M. Chini (éd) *Topic, information structure and acquisition*. Franco Angeli

Index des auteurs

Badin, Flora	95	Ferré, Gaëlle	29
Beupoil, Pauline	15	Garcia, Brigitte	107
Ben Mlouka, Monia	77	Guardiola, Mathilde	45
Bigi, Brigitte	45	L'Huillier, Marie-Thérèse	107
Blondel, Marion	15	Lefebvre-Albaret, François	59
Boutet, Dominique	15	Morgenstern, Aliyah	15
Boutora, Leïla	3	Sallandre, Marie-Anne	107
Braffort, Annelies	3	Schoder, Camille	107
Caët, Stéphanie	15	Segouat, Jérémie	59
Chételat-Pelé, Emilie	95	Tellier, Marion	45
Dalle, Juliette	77		
Dalle, Patrice	77		
Devos, Nadège	95		

Index des mots-clés

anaphore	15	Langue des Signes Française	3, 59, 95, 107
annotation	45, 107	modèle sémiologique	107
allocentrée	15	multimodalité	3, 29, 45
automatique	77	méthode d'annotation	3
manuelle	77	méthodologie d'annotation	95
clignement	95	pointage	15, 29, 95, 107
corpus	107	manuel	77
dialogue	107	reconnaissance automatique	77
Elan	107	regard	95
espace de signation	59	schéma d'annotation	3, 107
geste	45	d'encodage	29
coverbal	59	transcription	107
gestualité	3	Transcription Orthographique Enrichie	45
gestuelle coverbale	15		
langue des signes	77		

