



**HAL**  
open science

## Agent virtuel signeur - Aide à la communication des personnes sourdes

Alexis Héloir

► **To cite this version:**

Alexis Héloir. Agent virtuel signeur - Aide à la communication des personnes sourdes. Interface homme-machine [cs.HC]. Université de Bretagne Sud; Université Européenne de Bretagne, 2008. Français. NNT: . tel-00516280

**HAL Id: tel-00516280**

**<https://theses.hal.science/tel-00516280>**

Submitted on 9 Sep 2010

**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.

N° d'ordre: 109

# THÈSE

présentée

**devant l'Université de Bretagne Sud**

pour obtenir

le grade de : DOCTEUR EN SCIENCE ET TECHNOLOGIE DE  
L'INFORMATION ET DE LA COMMUNICATION DE L'UNIVERSITÉ  
EUROPÉENNE DE BRETAGNE  
Mention INFORMATIQUE

par

Alexis HELOIR

Équipe d'accueil : Interaction et Intelligence, axe SAMSARA

Laboratoire d'accueil : VALORIA

École Doctorale : UBS

Titre de la thèse :

*Système de communication par agent virtuel  
aide à la communication des personnes sourdes*

À soutenir le 11 janvier 2008 devant la commission d'examen

MM. :	Ronan	BOULIC	Rapporteurs
	Patrice	DALLE	
MM. :	Catherine	PÉLACHAUD	Examineur
	Sylvie	GIBET	Directeur de thèse
	Franck	MULTON	Co-directeur de thèse
	Nicolas	COURTY	Encadrant



*La vie nous donne beaucoup.*

*Nous ne comprenons pas parce que nous avons obstinément appris à nous contenter de peu.*

# Remerciements

Je remercie tout d'abord Catherine Pélachaud qui m'a fait l'honneur de présider mon jury de thèse. Je remercie également Patrice Dalle et Ronan Boulic pour avoir accepté d'être mes rapporteurs ainsi que pour leurs commentaires sur ce travail.

Je remercie Sylvie Gibet, Franck Multon et Nicolas Courty pour m'avoir encadré tout le long de ma thèse et pour leur soutien dans les moments difficiles.

Merci à l'ensemble des personnes qui, par leurs conseils précieux et le temps qu'il m'ont consacré ont contribué à l'aboutissement de ce travail. Je remercie entre autres Annelies Braffort du Limsi, Loïc Kervajan, Philippe Angèle de Visuel LSF de Rennes, les membres du laboratoire Valoria : Jean-François, Pierre-François, Salah, Régis, François, Sylvianne dont le zèle et l'efficacité m'ont épargné de nombreuses tâches administratives (ainsi qu'un week-end forcé à Toulouse!), Jo, qui apporte avec bonne humeur et entrain les nourritures terrestres nécessaires à la bonne marche du Labo. Je n'oublie pas les collègues thésards qui m'ont accompagné durant ces trois années : Ahmed, Alexandre, Bart, Chouki, Didier, Franck, Mohammed, Nadeem, Nicolas, Roméo, Sébastien, je leur souhaite à tous de concrétiser leurs rêves. Merci à Mohammed Daoudi, qui m'a ouvert les portes du monde de la recherche en septembre 2003.

Merci à ma famille, à mes parents, à qui je dois tout et à mon Frère Olivier. En rédigeant ce mémoire, j'ai souvent pensé aux membre de ma grande famille et à mes amis qui malgré l'éloignement restent très présents dans ma vie.

Enfin, je remercie la personne qui mérite d'avoir son nom au même titre que moi sur cette thèse, pour son infinie patience qui n'a d'égal que l'amour qu'elle m'a témoigné. Merci Silvia.

*A mes parents.*  
*A Silvia.*





# Table des matières

<b>Remerciements</b>	<b>1</b>
<b>Table des matières</b>	<b>0</b>
<b>Introduction</b>	<b>7</b>
<b>1 Motivation de la thèse</b>	<b>9</b>
1.1 La surdité : histoire et contexte . . . . .	9
1.1.1 Une histoire de la langue des signes française . . . . .	9
1.1.2 Contexte actuel . . . . .	10
1.1.3 Les aides technologiques, la culture des sourds et la société	11
1.2 Problèmes posés par les formes écrites . . . . .	12
1.2.1 Le français écrit n'est pas la solution . . . . .	13
1.2.2 La LSF ne possède pas (encore) de forme écrite large- ment adoptée . . . . .	13
1.2.3 De la nécessité des avatars signeurs . . . . .	14
<b>2 Organisation générale de la thèse</b>	<b>15</b>
2.1 Problématiques liées . . . . .	15
2.1.1 Problématiques liées à la linguistique . . . . .	15
2.1.1.1 La visée iconicisatrice . . . . .	17
2.1.2 Problématiques liées aux agents conversationnels . . . . .	17
2.1.3 Problématiques liées à l'animation . . . . .	18
2.2 Contributions . . . . .	18
2.2.1 Modèle d'énonciation de la langue des signes . . . . .	18
2.2.2 Modèle de spécification des gestes . . . . .	19
2.2.3 Identification des modificateurs . . . . .	19
2.2.4 Méthodes de génération du mouvement . . . . .	20
<b>3 État de l'art</b>	<b>21</b>
3.1 Études linguistiques et représentation des langues signées . . . . .	21
3.1.1 Notions de base en linguistique . . . . .	22

3.1.1.1	La notion d'arbitraire . . . . .	23
3.1.1.2	La double articulation . . . . .	24
3.1.2	Modèles phonologiques des langues signées . . . . .	25
3.1.3	Remise en cause de la phonologie par la visée illustrative . . . . .	25
3.1.4	Prise en compte de la structure spatiale de la langue des signes . . . . .	26
3.1.4.1	L'ordre des éléments . . . . .	26
3.1.4.2	Les verbes directionnels . . . . .	26
3.1.4.3	Marqueurs non manuels et flexions . . . . .	26
3.1.5	L'iconicité en langue des signes . . . . .	26
3.1.6	Conclusion . . . . .	28
3.1.7	Descriptions des langues signées . . . . .	30
3.1.7.1	Modélisation et génération d'énoncés de la LSF . . . . .	31
3.1.8	Importance de la prise de rôle . . . . .	34
3.2	Apport des agents conversationnels . . . . .	35
3.2.1	Prise en compte de la personnalité de l'agent, notion de style . . . . .	35
3.2.2	Geste co-verbal et applications aux agents conversationnels . . . . .	36
3.2.2.1	Prise en compte du style . . . . .	37
3.2.2.2	Paramètres qualitatifs du mouvement . . . . .	37
3.2.3	Gestion du style pour les agents conversationnels, "l'art et la manière" . . . . .	39
3.3	La génération de mouvements naturels et la notion de style en animation par ordinateur . . . . .	40
3.3.1	Modèles générateurs . . . . .	40
3.3.2	Modèles phénoménologiques . . . . .	42
3.3.3	Analyse-synthèse . . . . .	43
3.3.3.1	Modification de mouvements . . . . .	43
3.3.3.2	Mélange de mouvements . . . . .	44
3.3.3.3	Extraction des composants stylistiques et modification de mouvements . . . . .	45
3.4	Conclusion . . . . .	46
<b>4</b>	<b>Modèle du discours</b> . . . . .	<b>49</b>
4.1	La notion d'entité et sa représentation . . . . .	49
4.1.1	Hors visée illustrative . . . . .	51
4.1.2	En visée illustrative . . . . .	51
4.1.2.1	En transfert de taille et de forme . . . . .	52
4.1.2.2	En transfert de situation . . . . .	53

4.1.2.3	En transfert personnel . . . . .	55
4.1.3	Le contexte . . . . .	57
4.2	Modèle global du discours . . . . .	58
4.3	Influence du modèle de discours sur les paramètres expressifs .	58
<b>5</b>	<b>Modèle de Spécification des Gestes de la Langue des Signes</b>	<b>63</b>
5.1	Typologie des gestes . . . . .	65
5.1.1	Phases gestuelles non porteuses de sens . . . . .	65
5.1.1.1	Phases de préparation . . . . .	66
5.1.1.2	Phases de rétraction . . . . .	66
5.1.1.3	Phases de tenue . . . . .	67
5.1.2	Phases gestuelles porteuses de sens . . . . .	68
5.1.2.1	Phases de strokes . . . . .	68
5.1.2.2	Phases de tenue . . . . .	68
5.1.3	Règles d'organisation entre phases gestuelles . . . . .	68
5.2	Spécification du geste . . . . .	69
5.2.1	Apports d'un modèle gestétique . . . . .	70
5.2.2	Séparation entre paramètres statiques et dynamiques . .	70
5.3	Paramètres statiques . . . . .	71
5.3.1	Paramètres non-manuels . . . . .	72
5.3.2	Paramètres manuels . . . . .	72
5.3.2.1	Configuration manuelle . . . . .	72
5.3.3	Emplacement du geste . . . . .	74
5.3.3.1	Plans . . . . .	74
5.3.3.2	Zones de contact . . . . .	74
5.3.4	Modélisation de l'orientation de la main . . . . .	77
5.3.4.1	Contraintes géométriques . . . . .	78
5.4	Paramètres dynamiques . . . . .	80
5.4.1	Organisation des paramètres . . . . .	80
5.4.2	Gestes tracés . . . . .	80
5.4.3	Gestes de pointage . . . . .	81
5.4.3.1	Illustration . . . . .	81
5.5	Pistes de réflexion pour la gestion de la simultanéité en langue des signes . . . . .	82
5.6	Vers un modèle de composition des gestes de la langue des signes	84
<b>6</b>	<b>Analyse et synthèse du mouvement</b>	<b>89</b>
6.1	Matériel et méthodes . . . . .	89
6.1.1	Introduction . . . . .	89
6.1.2	Acquisition de données . . . . .	90
6.1.3	Format des données . . . . .	91

6.1.4	Séquences capturées . . . . .	91
6.2	Analyse/synthèse de séquences gestuelles . . . . .	94
6.2.1	Analyse temporelle de séquences gestuelles . . . . .	94
6.2.2	Alignement temporel adaptatif . . . . .	96
6.2.2.1	DTW à pente contrainte . . . . .	97
6.2.2.2	DTW multiniveaux . . . . .	98
6.2.2.3	DTW adaptative . . . . .	98
6.2.3	Résultats . . . . .	99
6.2.4	Discussion . . . . .	103
6.2.5	Application à la synthèse . . . . .	103
6.2.6	Conclusion . . . . .	103
6.3	Identification des paramètres d'expressivité . . . . .	104
6.3.1	Introduction . . . . .	104
6.3.2	Mise en évidence des phases du mouvement . . . . .	105
6.3.3	Segmentation automatique . . . . .	106
6.3.4	Segmentation manuelle assistée par la segmentation automatique . . . . .	107
6.3.5	Paramètres d'expressivité en langue des signes . . . . .	108
6.3.5.1	Variations temporelles et répartition des phases gestuelles . . . . .	108
6.3.5.2	Caractéristiques spatiales et structurelles de l'expressivité . . . . .	111
6.3.6	Conclusion . . . . .	115
6.4	Synthèse de gestes de la LSF . . . . .	116
6.4.1	De la spécification à la génération . . . . .	116
6.4.1.1	Méthodes de génération . . . . .	117
6.4.2	Cinématique inverse . . . . .	117
6.4.2.1	Modèle sensori-moteur . . . . .	118
6.4.2.2	Résolution de l'inversion . . . . .	118
6.4.2.3	Validation du modèle . . . . .	119
6.4.3	Prototype de composition des gestes . . . . .	121
6.4.4	Conclusion . . . . .	123
	<b>Conclusion</b>	<b>125</b>
	<b>A Corpus météo</b>	<b>129</b>
	<b>B Hiérarchie BVH du squelette</b>	<b>131</b>
	<b>C Représentation de l'agent signeur</b>	<b>135</b>
C.1	Squelette d'animation . . . . .	135
C.2	Rendu . . . . .	138

<i>Table des matières</i>	5
<b>Glossaire</b>	<b>143</b>
<b>Bibliographie</b>	<b>155</b>
<b>Table des figures</b>	<b>157</b>



# Introduction

Faciliter l'accès à l'information pour chacun requiert de nouvelles méthodes permettant un meilleur dialogue entre l'homme et la machine. Le développement d'humanoïdes virtuels capables de s'exprimer par le langage des signes est une réponse à cette préoccupation. Les technologies de l'information et de la communication investissent notre espace quotidien en suivant une progression galopante. Le foisonnement de nouveaux services qui découlent de ces technologies facilite la consultation et la production d'information pour leurs utilisateurs. Cependant, de telles avancées techniques ne sauraient être perçues comme un réel progrès pour une société si elles continuaient à exclure une partie de sa population. Un individu peut ressentir des difficultés d'accès à l'information pour diverses raisons, à savoir économiques, culturelles et linguistiques, ou physiques. Notre laboratoire tente de répondre à une barrière d'ordre linguistique : permettre à la machine d'émettre des informations grâce à la langue des signes.

Si la langue des signes est aujourd'hui reconnue comme une langue à part entière, elle a souffert en France d'un bannissement de près d'un siècle auquel les sourds paient encore un lourd tribut. Aujourd'hui, on voit apparaître de plus en plus de logiciels à même de favoriser l'accès des sourds à l'information en la présentant dans leur langue naturelle et au moyen d'outils de visualisation pertinents. Les applications qui en découlent concernent notamment la communication à distance en temps réel (visiophonie, vidéoconférence), l'enseignement assisté par ordinateur (dictionnaires informatisés, tutoriaux de langues des signes) et la diffusion d'émissions bilingues, signées et orales. Par ailleurs, les nouvelles modalités de communication, exploitées dans les systèmes de télécommunication récents, intègrent des agents artificiels capables d'améliorer les capacités d'interaction avec des utilisateurs. Ces agents peuvent jouer notamment le rôle d'assistants dans des applications interactives multimédia.

Cette thèse se concentre sur la conception et l'animation d'entités virtuelles capables de générer des gestes de la Langue des Signes. L'accent est mis sur la qualité expressive des gestes, ainsi que sur le lien entre ses représentations. Il s'agit de développer une chaîne complète de spécification et de contrôle inter-

actif d'humanoïde virtuel, en y intégrant des caractéristiques propres à la dynamique des mouvements réels. L'identification d'invariants et de paramètres biomécaniques propres au mouvement humain est effectuée à partir de l'analyse de mouvements acquis et devrait permettre de synthétiser des gestes paramétrés conduisant à des exécutions variées, voire de créer de toutes pièces de nouveaux gestes.



# Chapitre 1

## Motivation de la thèse

### 1.1 La surdité : histoire et contexte

Cette section présente la situation des sourds en France. En premier lieu, elle retrace leur histoire et met en lumière les conséquences regrettables qu'ont dû avoir les incompréhensions successives entre le monde des sourds et le monde des entendants. Un panorama de la situation actuelle est ensuite dressé. Ce panorama met l'accent sur les mesures qui ont permis la reconnaissance tardive de la langue des signes comme langue officielle et sur les conséquences en terme d'accessibilité à l'information et de reconnaissance de la communauté des sourds. Une dernière partie s'appuie sur les avancées technologiques récentes en faveur de l'intégration des sourds dans la société et les interrogations, voire les craintes qu'elles suscitent. Notre travail se place dans un tel contexte et s'inscrit à l'encontre de certaines idées reçues à propos de la culture des sourds et des moyens mis en oeuvre en faveur de l'accès à l'information. Enfin, nous positionnons notre travail par rapport à trois domaines liés à l'analyse et à la production de gestes : les études linguistiques dédiées aux langues des signes, les études en IHM dédiées aux agents conversationnels humanoïdes <sup>1</sup> et les études dans le domaine de l'animation par ordinateur.

#### 1.1.1 Une histoire de la langue des signes française

La Langue des Signes Française (désormais : LSF) est la langue historique pratiquée par les sourds, certains malentendants et une partie de la population entendante en France. La LSF est une langue à part entière et constitue l'un des piliers de l'identité de la culture sourde. Contrairement à une idée en-

---

<sup>1</sup>Les agents conversationnels humanoïdes sont couramment cités par l'acronyme agent conversationnel animé (ACA).

core répandue, Il existe de nombreuses langues signées (LS) de part le monde. Dès le début du XIX<sup>e</sup> siècle, la méthode française d'éducation des sourds, qui impliquait un enseignement reposant sur l'emploi des signes a été diffusée très largement en Europe et en Amérique par des professeurs sourds et entendants. La LSF, elle même, a donc été mise en contact avec les langues des signes locales et il s'est ainsi créé selon divers chercheurs la "famille des LS françaises", un regroupement de langues ayant des liens historiques. C'est le cas notamment de l'ASL (LS américaines), mais aussi de nombreuses autres langues telles que la NGT (LS des Pays Bas) ou la LSQ (LS québécoise). Malgré ce rôle historique important, la LSF est peut être l'une des LS qui a le plus souffert des conséquences de l'interdiction, à partir de 1880 [Moo98], de l'utilisation des LS dans l'éducation des sourds. Il faudra ensuite attendre les années 60-70 pour constater les premières initiatives visant à réhabiliter la langue des signes grâce aux mouvements de revendication des sourds, d'associations et d'institutions sensibilisées et de certains chercheurs.

### 1.1.2 Contexte actuel

En France, il faudra attendre le premier mars 2004 pour que le Sénat ne reconnaisse la langue des signes comme langue à part entière<sup>2</sup>. Cette reconnaissance a été inscrite dans la loi en février 2005<sup>3</sup>. Les dernières lois en faveur de la reconnaissance de la surdité sont en réalité très récentes et découlent d'une série de mesures prises dans le cadre de la lutte pour l'égalité des chances des citoyens en situation de handicap. L'esprit de ces lois vise à réduire les difficultés éprouvées par les personnes handicapées dans l'exercice de leurs droits et devoirs de citoyens. Elles se concentrent donc sur la mise en place de moyens (techniques ou personnels) visant à faciliter les différents aspects de l'exercice et de la jouissance de la citoyenneté (accès à l'information, démarches administratives, déplacements et transports, accès à l'emploi).

Si l'on peut se réjouir de la prise en considération des besoins spécifiques des personnes en situation de handicap, l'on peut regretter que la situation des sourds ne soit traitée *que* selon une visée palliative et que les aspects liés à l'histoire et à la culture soient occultés dans cette démarche. A côté de ces mesures officielles, de nombreux indices témoignent de la reconnaissance de la LSF, tant sur le plan universitaire (ouvertures de filières en LS dans certaines universités, recherches en linguistique, informatique, psychologie, anthropologie et sociologie dédiées aux LS) que pour le grand public (reconnaissance de

<sup>2</sup>Source : <http://www.senat.fr/leg/tas03-064.html>

<sup>3</sup>Loi n° 2005-102 du 11 février 2005 pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées. Source : <http://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=SANX0300217L>

la LS comme seconde langue pour les enfants entendants, nombre croissant d'adultes effectuant la démarche d'apprendre la LS dans le cadre associatif, par curiosité ou par plaisir).

Cependant, l'accès à la LSF et à la culture sourde reste une source de difficultés pour les enfants sourds. Ces derniers, en général isolés dans un monde d'entendants, ont des difficultés à appréhender leurs différences, car isolés de la culture des sourds. De plus, l'accompagnement médical entourant le jeune sourd dans ses premières années ne favorise pas la réalisation de son appartenance à la communauté des sourds. En effet, bien que le corps médical ait pris note du fait que les LS soient des langues riches et complètes, il privilégie la mise en oeuvre de solutions destinées à corriger le déficit sensoriel causé par la surdité. La correction de ce déficit fait appel à des techniques d'implantation de prothèses auditives. L'évocation de solutions techniques nous amène naturellement à aborder l'influence de la technologie sur la culture des sourds et les enjeux qu'elle soulève. Aujourd'hui en France, la surdité touche 6,6 % de la population. Parmi ces personnes, sont recensées 11 600 personnes atteintes de déficience auditive profonde (perte supérieure à 90 décibels) et 372 000 personnes atteintes de déficience auditive sévère (perte comprise entre 70 et 90 db). La langue des signes française est pratiquée par 100 000 personnes dont 80 000 souffrent d'un déficit auditif<sup>4</sup>.

### 1.1.3 Les aides technologiques, la culture des sourds et la société

"[...]La surdité ne doit plus exister ! [...]"<sup>5</sup>. Ce manifeste résume la position d'une partie du monde médical vis à vis de la surdité. La surdité n'est vue que comme une affection sensorielle entraînant un handicap. La compensation de ce handicap passe par le recours à des techniques d'implantation de prothèses auditives appelées implants cochléaires. L'implant cochléaire est un appareillage qui vise à restaurer un certain niveau d'audition pour certaines personnes sourdes, en stimulant directement les terminaisons nerveuses de l'audition situées dans la cochlée, au moyen d'électrodes implantés chirurgicalement. Cette solution, d'abord réservée aux surdités acquises, a pu être mise en oeuvre pour les surdités pré-linguales ou congénitales à partir des années 80. Aujourd'hui, le dépistage précoce de la surdité et la prise en charge médicale des jeunes sourds dans le cadre de la loi sur l'égalité des chances de 2005 ont largement généralisé l'implantation cochléaire des enfants sourds, leur permettant, dans une certaine mesure, d'accéder à une forme d'audition.

---

<sup>4</sup>Source : Unapeda ([http://www.unapeda.asso.fr/article.php3?id\\_article=83](http://www.unapeda.asso.fr/article.php3?id_article=83))

<sup>5</sup>Source : <http://recorlsa.club.fr/surdite/index.html>

Les sensations délivrées par l'implant cochléaire étant différentes des sensations de l'audition normale, l'implantation, pour être bien intégrée, doit être accompagnée d'une éducation auditive spécifique dispensée par des orthophonistes. L'éducation orthophonique des jeunes sourds implantés prend une place importante au détriment de la découverte du monde des sourds et de sa langue naturelle, la LSF.

Actuellement, des chercheurs, des associations de sourds et de parents d'enfants sourds s'inquiètent des conséquences que pourrait présenter une assimilation systématique des jeunes sourds et le déni de leur identité<sup>6</sup>.

Ces faits nous amènent à nous interroger sur la place des aides technologiques dans l'intégration des sourds dans la société et sur leurs conséquences identitaires. Le travail présenté dans cette thèse est ainsi motivé par les problématiques suivantes : En parallèle de la compensation sensorielle, comment peut-on proposer des aides technologiques facilitant l'accès des sourds à l'information en prenant en compte leurs besoins et leurs spécificités ? Une meilleure compréhension de la langue naturelle des sourds ne permettrait-elle pas de mieux comprendre les schémas cognitifs à l'origine des langues orales ? Les solutions techniques mises en oeuvre pour présenter l'information aux sourds par l'intermédiaire du canal visuo-gestuel ne permettraient-elle pas d'enrichir les métaphores d'interaction entre l'homme et la machine pour tous ?

Présenter des informations de manière naturelle pour la communauté sourde soulève des problèmes nombreux et variés. La suite de cette introduction est consacrée à la présentation des problèmes liés à la représentation de l'information sous une forme respectant les caractéristiques de la langue naturelle des sourds. Nous présenterons les difficultés posées, ferons le point sur quelques idées reçues et présenterons brièvement les recherches dans chacun des domaines. Nous verrons également comment ces contraintes expriment naturellement le besoin de la spécification et de la mise en oeuvre d'un humanoïde expressif capable de transcrire un discours en langue signée.

## 1.2 Problèmes posés par les formes écrites

Il est coutume de penser que le sous-titrage de messages véhiculés d'ordinaire par le canal audio-verbal est une solution au problème de l'accès à l'information pour les sourds. Cette section démontre que le sous-titrage ne fournit pas une solution universelle. Une fois fait le point sur cette idée reçue, nous nous attaquons à d'autres idées préconçues et soulevons les problématiques induites par une prise en compte pertinente des spécificités de la LSF.

---

<sup>6</sup>Source : Pour une reconnaissance effective de la langue des signes française dans l'espace éducatif des enfants sourds <http://appel.lsf.free.fr/>

### **1.2.1 Le français écrit n'est pas la solution**

L'impossibilité pour les jeunes sourds d'avoir accès à une éducation bilingue dans leurs premières années a posé des problèmes et continue de peser lourdement sur la maîtrise du français écrit par la communauté sourde. Aujourd'hui, les chiffres les plus optimistes relatifs à la proportion de sourds rencontrant des problèmes d'illettrisme en France se situent entre 60% et 80% de la population sourde<sup>7</sup>. Cet illettrisme revêt des formes variées, allant de l'absence totale de capacité à lire à une difficulté à saisir les nuances d'un texte. La maîtrise insuffisante de la langue majoritaire écrite est incontestablement un frein à l'intégration dans une société où la majorité des supports de communication fait l'hypothèse que les utilisateurs maîtrisent au minimum le français écrit.

Nous pouvons nous interroger sur la pertinence du sous-titrage des contenus informatifs reposant sur le canal acoustique dans les médias de diffusion comme réponse à l'accès à l'information pour les sourds. D'un point de vue technologique, la transmission d'informations textuelles possède nombreux avantages : elle est économique en terme de quantité d'information (on admet par exemple que le débit moyen de la parole dactylographiée est de 50bit/s, ce qui est négligeable par rapport aux 13 Kbit/s<sup>8</sup> nécessaires au transport du signal vocal acoustique compressé). L'utilisation du canal textuel dans le but de représenter un codage symbolique de la langue des signes, capable d'être interprété par le terminal de l'utilisateur sourd et rejoué sous la forme d'une animation, peut apparaître comme une solution économique et pertinente.

### **1.2.2 La LSF ne possède pas (encore) de forme écrite largement adoptée**

Les langues des signes n'ont pas développé de forme écrite historique. Les caractéristiques de la langue des signes, l'histoire propre à la culture des sourds et l'éclatement géographique de la communauté n'ont pas favorisé l'apparition d'une forme écrite utilisée par tous. La conservation des faits historiques et du patrimoine culturel qui est une motivation forte pour la création d'un système d'écriture a été réalisée par la transmission gestuelle entre sourds, par l'écriture de textes en français, puis par l'enregistrement vidéo. Des initiatives qui s'apparentent à la création d'une forme écrite de la langue des signes se sont d'abord concentrées sur une description des signes lexicaux (également appelés signes standards) à l'aide des dessins en vue de la consti-

---

<sup>7</sup>Source : websourd : [www.websourd.org](http://www.websourd.org)

<sup>8</sup>13 Kbit/s, codage LPC (GSM), la référence "historique" de débit nécessaire au transport de la voix "qualité téléphonique" non compressé s'élevant à 64Kbit/s étant progressivement remplacée.

tution de dictionnaires. Nous vous renvoyons à [BV06] pour une présentation détaillée des dictionnaires en France. Divers systèmes de notation ont également été élaborés [Beb25, PLZ<sup>+</sup>89, New87] afin de répondre à des besoins précis de description linguistique. Dans les années 60, Valérie Sutton [Sut98] a développé la forme d'écriture *Sign Writing*. Cette forme d'écriture est encore peu répandue dans la communauté des sourds, et les évaluations de la pertinence de cette forme écrite sont rares [Flo02]. L'évaluation de formalismes de représentation des LS a récemment fait l'objet d'un projet de recherche, le projet LS-SCRIPT<sup>9</sup> [AD04, AD05, RDF04]. On peut également noter en France les travaux de Brigitte Garcia visant à définir une formalisation graphique de la LSF [BG06] ainsi que ceux de [Bru07] sur l'utilisation de la vidéo comme forme *écrite* de la LSF. Le fait qu'il n'existe pas de forme écrite des langues signées proscrit de fait l'utilisation de toute forme de transcription "textuelle" de l'information. La spécification d'un média de communication adapté à la langue signée passe par l'identification de ses spécificités.

### 1.2.3 De la nécessité des avatars signeurs

La section précédente a mis en lumière les difficultés posées par l'utilisation d'un canal écrit, que ce soit par l'intermédiaire d'une langue orale (sous-titrage en français) ou par l'intermédiaire d'un codage dédié (forme écrite de la langue des signes). Les causes à l'origine de l'impossibilité d'utiliser la langue orale majoritaire sous la forme de sous-titrage sont liées à la difficulté pour les sourds de maîtriser cette langue, pour des raisons évoquées à la section 1.1.

D'un autre côté, l'inexistence de formats de codage répandus au sein de la communauté des sourds proscrit la transmission d'énoncés de la LSF par l'intermédiaire de systèmes de codages textuels. Les solutions les plus confortables pour les sourds consistent à l'heure actuelle à faire intervenir des interprètes qui traduisent en temps réel les énoncés en langue orale. Depuis un peu moins d'une dizaine d'années, les recherches sur les avatars signeurs et sur les systèmes de traduction automatique de la langue des signes ont fait naître divers prototypes de systèmes permettant de traduire et de proposer des énoncés plus ou moins conformes à la langue des signes par le canal de communication visuo-gestuel, privilégié des sourds. Ceci à l'aide d'avatars signeurs.

---

<sup>9</sup><http://lsscript.limsi.fr>

## Chapitre 2

# Organisation générale de la thèse

La réalisation d'un système capable de générer des séquences intelligibles en langue des signes est une tâche ambitieuse et complexe. Comme nous l'avons évoqué dans le chapitre introductif, il est réducteur de considérer le problème de génération d'une séquence gestuelle de la langue des signes d'un point de vue uniquement lexical, ou comme une simple concaténation de primitives de mouvements. Une telle conception conduirait à un système de génération capable de produire au mieux, des énoncés de français signé et au pire, des séquences inintelligibles. Il est donc impossible de faire l'économie des caractéristiques fondamentales de la langue des signes. Une appréhension de ces caractéristiques requiert la prise en compte de travaux dédiés à la linguistique des langues des signes.

La vocation de ce chapitre est de présenter d'une part les différentes problématiques de recherche qui sont liés à nos travaux et d'autre part, de présenter nos contributions au sein des différents modules qui composent la chaîne de spécification et de génération de séquences gestuelles que nous proposons.

### 2.1 Problématiques liées

L'illustration ci-après récapitule de manière schématique les différentes problématiques que nous abordons dans ce mémoire. Elles sont liées à la linguistique des langues des signes, aux agents conversationnels, ainsi qu'à l'animation par ordinateur. Ces trois thèmes sont soulevés dans les parties suivantes.

#### 2.1.1 Problématiques liées à la linguistique

Il est désormais acquis que les langues des signes possèdent des structures très éloignées des structures des langues orales. C'est sans aucun doute une

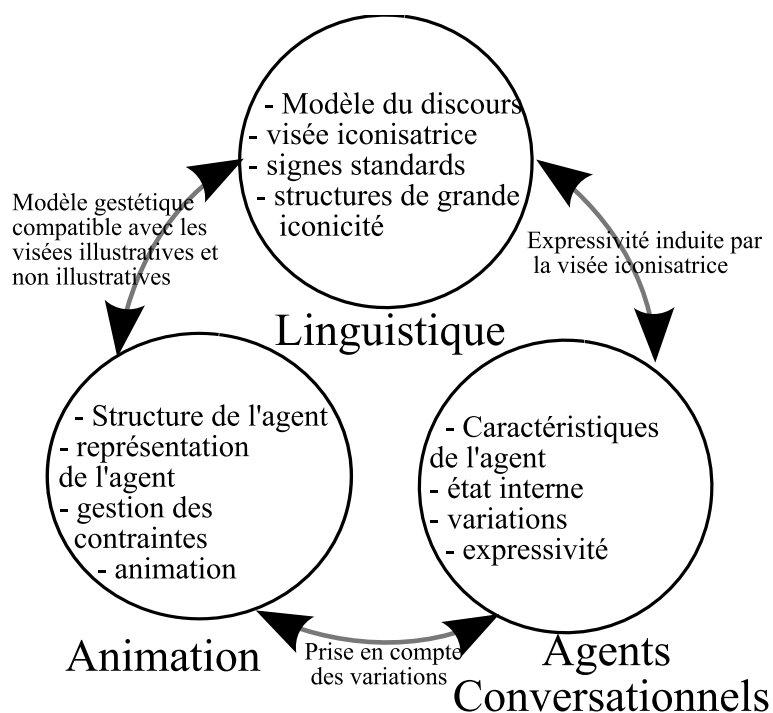


FIG. 2.1 – Schéma organisationnel de la thèse

erreur que de considérer les langues des signes selon un point de vue uniquement lexical sans prêter attention aux structures qui font la complexité et la richesse des langue des signes. Pour s'en convaincre, il suffit de regarder cette illustration tirée de l'ouvrage de Diane Brentari [Bre98] :



FIG. 2.2 – Passage polymorphémique en langue des signes américaine

On peut y voir un signeur exécuter une séquence polymorphémique<sup>1</sup> que

<sup>1</sup>Le morphème est une unité de sens du discours. Il se retrouve, par exemple, dans des



l'on pourrait traduire en français par : *“Deux créatures bipèdes et voutées se faisant face avancent en prenant garde, d'un point a à un point b”*. D'un point de vue quantitatif, en reprenant les considérations évoquées au paragraphe 1.2.1, on ne peut être qu'étonné par la richesse et la compacité de la langue des signes : en supposant que la phrase ait été signée en 1.3 secondes, le débit de l'information en LS atteindrait 700 bit/s. Une telle concision est due à la visée illustrative de la langue des signes.

### 2.1.1.1 La visée iconicisatrice

L'iconicité est la base des LS. Elle s'exprime par la ressemblance entre les signes et ce à quoi ils réfèrent. Ce mécanisme “fondé sur une mise en forme gestuelle de l'information à transmettre” est commun à toutes les langues des signes, élaborées ou primaires. Le processus d'iconicisation est le processus par lequel le locuteur va rendre iconique l'expérience. Ce processus donne accès à une visée illustrative selon laquelle le locuteur vise à reconstituer de manière imagée une expérience vécue ou imaginaire. Cette visée est due aux mécanismes cognitifs qui sélectionnent dans l'expérience ce qui peut ou doit être iconicisé et qui le restituent dans la langue, sous forme d'énoncés. “C'est le lien de ressemblance direct, plus ou moins étroit entre la chose du monde, le référent, et le signe qui s'y rapporte”[Cux00]. Nous pouvons retenir de ce modèle qu'il y a deux manières de “dire” en LS : “dire en montrant” (visée iconicisatrice) et “dire sans montrer” (lexique standard).

### 2.1.2 Problématiques liées aux agents conversationnels

La génération d'énoncés en langue des signes fait intervenir des humanoïdes de synthèse capables de reproduire des séquences gestuelles intelligibles. D'un point de vue IHM, les recherches sur les agents conversationnels animés reposent sur l'hypothèse que les utilisateurs humains sont capables d'interagir de manière plus efficace et plus agréable avec la machine lorsque le mode d'interaction est proche d'un mode de communication naturel en situation de dialogue humain-humain. Les agents virtuels signeurs peuvent donc être vus comme des Agents Conversationnels Animés (ACA) et leur mise en œuvre recoupe des problématiques communes.

Dans le cadre du discours, les ACAs déploient un large éventail de stratégies afin d'offrir une interaction crédible avec un utilisateur humain. La qualité de l'interaction est alors en partie sanctionnée par le jeu expressif de l'agent. L'expressivité d'un ACA se traduit d'une part par le choix de la stratégie

---

formations lexicales telle que “aphone”, où le préfixe “a-” constitue une unité morphémique privative qui s'ajoute au morphème traduisant la capacité à oraliser “phone” .

employée lors de l'acte de communication et d'autre part par la manière (ou la qualité expressive) avec laquelle l'agent communique. Les agents virtuels signeurs sont une catégorie d'ACA dont la particularité est le recours quasi exclusif au canal visuo-gestuel pour tout acte de communication. Nous verrons comment les modèles d'expressivité des ACAs se déclinent dans le cadre de la génération de gestes de la langue des signes, notamment en visée iconisatrice.

### **2.1.3 Problématiques liées à l'animation**

La génération de séquences de gestes expressifs requiert des algorithmes de génération de mouvements capables de produire des animations réalistes et capables de prendre en compte les variations liées au style que nous avons évoquées dans la sous-partie précédente. Cette thèse décrit comment des algorithmes de synthèse de mouvements sensori-moteurs vont pouvoir être appliqués à des méthodes d'animations basées données pour produire des animations réalistes capable de prendre en compte les aspects stylistiques des avatars pour produire des animations convaincantes, réalistes et pertinentes dans un contexte de générations de séquences gestuelles de la langue des signes française.

## **2.2 Contributions**

La section précédente a permis de présenter les problématiques s'articulant autour de la création d'un agent signeur. Dans cette section, nous allons brièvement introduire et situer nos différentes contributions. Ces dernières sont résumées dans le schéma de la figure 2.3. Ces contributions se répartissent selon trois grands domaines, situés à des niveaux d'abstraction échelonnés. A un haut niveau d'abstraction, nous proposons un modèle du discours en langue des signes. Ensuite, dans un niveau intermédiaire, nous décrivons un modèle de spécification des gestes ainsi qu'une description des modificateurs de certains gestes. Enfin, au niveau le plus proche du signal de mouvement, nous développons des méthodes de génération de mouvement, par rejeu ou cinématique inverse inspirée des modèles sensori-moteurs.

### **2.2.1 Modèle d'énonciation de la langue des signes**

Cette partie propose une description des différents modèles intervenants dans la construction des phrases de la langue des signes. La vocation de cette

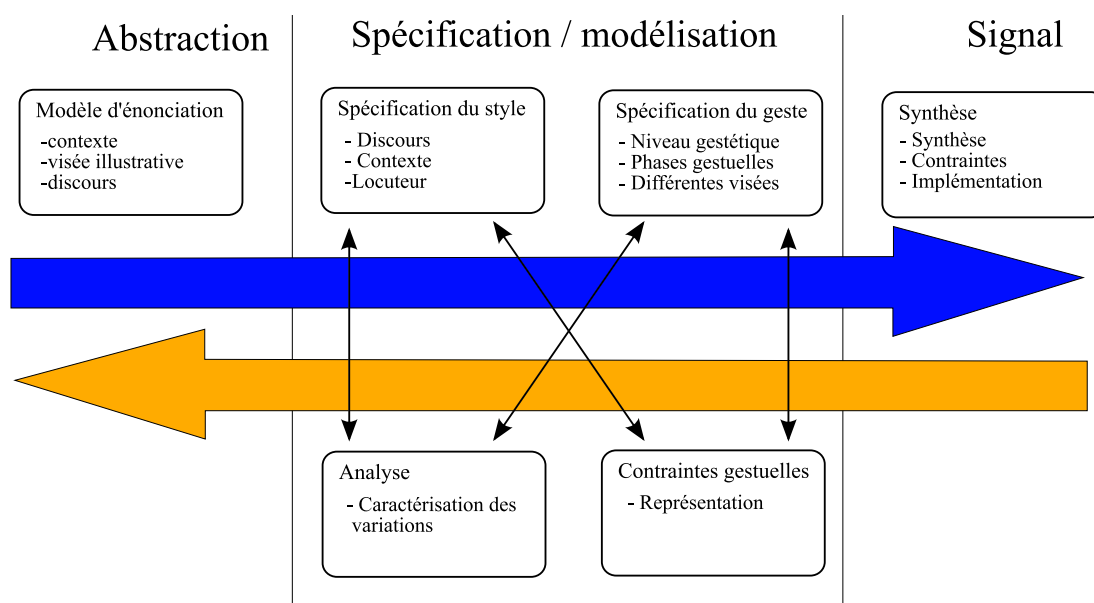


FIG. 2.3 – Contributions et domaines adressés dans ce mémoire.

partie est de présenter, de manière concise et quelque peu simplifiée, l'organisation des différentes caractéristiques propres à la langue des signes selon un formalisme communément employé en modélisation informatique. Cette partie a pour vocation de structurer, et de mettre en lumière les relations entre les différents aspects de notre recherche, en particulier, la modélisation des gestes et la modélisation du style dans le contexte du discours en langue des signes.

## 2.2.2 Modèle de spécification des gestes

Dans le contexte de génération du geste expressif, il est nécessaire de s'appuyer sur un modèle de description et de spécification du geste. Un tel modèle permet, d'une part, de générer un geste "ex-nihilo" et d'autre part, de procéder à certaines opérations de modifications du geste (transformations qui peuvent être le résultat de variations liées au style ou de contraintes physiologiques). Cette partie décrit le modèle de spécification du geste de la langue des signes que nous avons mis en oeuvre dans cette thèse.

## 2.2.3 Identification des modificateurs

Cette partie représente le travail d'analyse qui a été réalisé sur des données de mouvement de langue des signes capturées en vue d'extraire des invariants liés à certaines variations dans la gestuelle. Les résultats de l'analyse de don-

nées sont ensuite confrontées aux hypothèses émises par les études liées à la prosodie de la langue des signes et aux modèles existants de spécification et de gestion du style pour les agents conversationnels. Une classification des causes des différents variations est alors proposée puis organisée selon la portée temporelle des différentes causes de variation du geste dans le discours. La réflexion ainsi menée conduit à un modèle de spécification et de gestion des variations des gestes de la langue des signes inspiré des études linguistiques en LSF et des études liées à la modélisation des variations pour les agents conversationnels.

#### **2.2.4 Méthodes de génération du mouvement**

Cette ultime partie traite de la génération des gestes. Il s'agit de proposer une méthode qui, à partir d'une description structurée d'une phrase gestuelle et d'une description structurée de l'évolution de paramètres stylistiques au cours du temps soit capable de générer les mouvements adéquats de la langue des signes.

# Chapitre 3

## État de l'art

### 3.1 Études linguistiques et représentation des langues signées

“Lorsque l’usage de la parole est rendu impossible pour quelque raison que ce soit, les gestes manuels peuvent s’organiser en un système gestuel qui peut à son tour évoluer en une véritable langue signée” [Cux00]. Cette assertion renferme deux idées fondamentales de la langue des sourds : d’une part, la spécificité du canal utilisé et, d’autre part, l’utilisation de celui-ci comme une solution palliative à un déficit sensoriel.

Mises à part les communautés de sourds, il n’en existe aucune qui ait développé un système de communication visuo-gestuel aussi sophistiqué que ne le sont les langues orales.

Cette section présente un panorama général des études linguistiques dédiées à la langue des signes. Son organisation est chronologique et débute avec une présentation de quelques notions de linguistique qui permettent de définir le statut de langue naturelle.

Nous présentons ensuite les travaux qui ont permis aux langues signées d’accéder au statut de langue. Cette reconnaissance s’est faite au prix de la mise en place de parallèles forts entre langues signées et orales.

Dans un second temps nous abordons les travaux consacrés à la syntaxe de la langue des signes et à sa reconnaissance en tant que langue spatiale, entraînant la prise en compte de l’espace de signation et des relations qui s’y créent.

Enfin, nous abordons les travaux à portée plus globale qui sont apparues majoritairement en France dans les années 80, études visant à prendre en compte la visée illustrative comme système organisateur des langues signées.

### 3.1.1 Notions de base en linguistique

Cette sous-section pose les bases générales et les grands principes à l'origine de la linguistique moderne. Elle ne se veut pas exhaustive, mais propose un cadre simplifié et des définitions aux termes linguistiques que nous utiliserons dans la suite de ce mémoire. Cette partie s'appuie sur les ouvrages de référence suivants : [Jak73a, Jak73b, Dub, dS16]

Le langage est né d'un besoin de communiquer. La communication implique la transmission d'une information et met en jeu au moins deux acteurs : un émetteur et un récepteur. Ces acteurs de la communication véhiculent des messages à travers un canal selon le chemin suivant :

Source → émetteur → canal → message → récepteur → destinataire

La plupart des langues employées par les hommes reposent sur le canal *audio-oral* : le message est produit par l'appareil phonatoire de l'émetteur, transite par l'interface air et est reçu par l'appareil auditif du récepteur pour être interprété.

Un autre canal de transmission est immédiatement disponible : le canal visuo-gestuel. Cette fois, le message est produit par le système musculo-squelettique de l'émetteur, transite dans le vide, ou toute interface laissant passer la lumière avec un taux de distorsion acceptable et est reçu par l'appareil visuel du récepteur.

Les utilisateurs de langues orales emploient au quotidien le canal visuo-gestuel, pour appuyer leurs discours, ils ont recours à des "gestes de communication" appelés *gestes coverbaux*. La plupart du temps, nous utilisons le canal visuo-gestuel pour appuyer, préciser et colorer nos paroles. Il arrive que les informations véhiculées à travers le canal visuo-gestuel contiennent un élément capital à la compréhension du discours. Ainsi, lorsque nous demandons notre route, il arrive que le quidam qui se propose de nous venir en aide pointe une direction de l'index en disant « c'est par là ! ». Lorsque plusieurs canaux de communication sont utilisés, on parle de *communication multimodale*.

Parfois, nous nous trouvons dans une situation de communication qui ne permet pas l'utilisation du canal *audio-oral*. Les plongeurs sous-marins se trouvent dans cette situation. Pour se faire comprendre, ils ont recours à un système de communication rudimentaire, qui leur permet de transmettre de l'information à l'aide de gestes. On appelle un tel système un *code*. Bien que le *code* de communication adopté par les plongeurs dispose d'un ensemble de *signes*, il n'a pas le statut de langue, il en est de même pour les gestes qui accompagnent nos paroles, les *gestes coverbaux*.

Nous pensons qu'il est nécessaire de faire le point sur les propriétés qui permettent à un système de communication d'accéder au statut de langue. La compréhension des notions linguistiques de base qui ont conduit au consensus

définissant une langue permet, en effet, de comprendre, d'une part, l'orientation et l'évolution des recherches dédiées aux langues des signes et, d'autre part, les discussions et les interrogations suscitées par ces recherches.

### 3.1.1.1 La notion d'arbitraire

La linguistique se focalise sur l'étude des langues humaines, dites aussi langues naturelles. Elle étudie les mécanismes qui interviennent dans l'élaboration et la production du langage. La linguistique est une branche de la *sémiologie*, la discipline qui étudie les signes. La sémiologie est elle-même une branche des *sciences cognitives*, la science qui étudie les mécanismes de la pensée. Le *signe* est un élément constitutif d'une langue.

Dans la théorie linguistique moderne que nous devons pour partie à Saussure [dS95], il est admis qu'un signe, élément constitutif du discours possède un *signifiant*, représentation artificielle d'un concept (*le signifié*) et fait allusion à un référent concret ou abstrait, présent ou non lors de l'émission du message. La figure 3.1 illustre la relation entre le référent, le signifié et le signifiant à l'aide du signe [ARBRE]. Ce signe [ARBRE] fait partie d'un ensemble vaste de signe *standards*, signes stabilisés et employés par tous les sourds de France pour désigner le concept d'arbre. On dit aussi que le signe arbre est *stabilisé* :

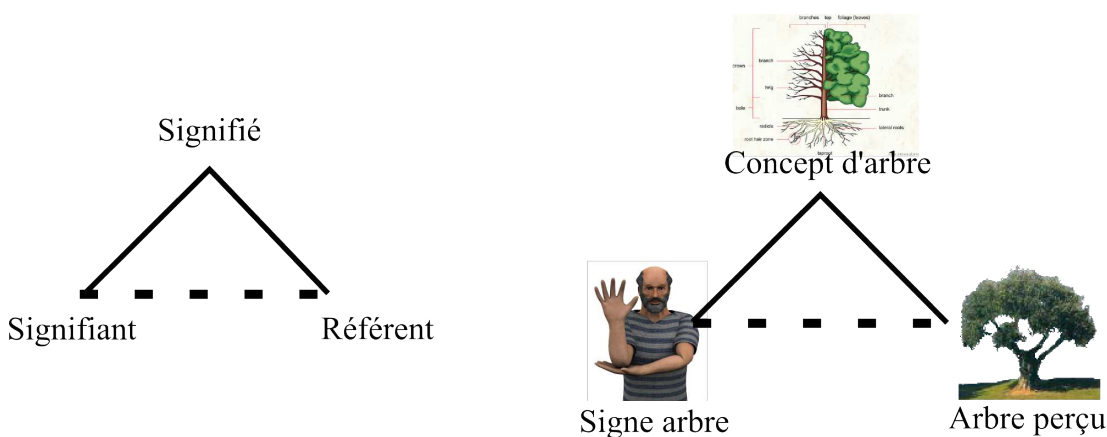


FIG. 3.1 – Relations entre signe, signifiant et signifié

D'après Saussure, le rapport entre le signifiant et le signifié est *arbitraire* (rien ne doit justifier l'emploi du signe bimanuel [arbre] pour désigner un arbre) et *immotivé* (il n'y a pas de lien empirique entre le signe et son signifiant). Cette constatation semble aller à l'encontre de l'exemple illustré ci-dessus. En effet, la configuration de la main droite (main dominante) évoque incontestablement la silhouette d'un arbre, où le tronc serait l'avant bras et les

branches les doigts du signeur. On dit que le signe [ARBRE], présente un aspect fortement *iconique*. C'est justement cet aspect non arbitraire des signifiants de la langue des signes qui a poussé certains linguistes à refuser aux langues signées le statut de langue.

Cuxac [Cux00] démontre que le caractère iconique des signifiants de la LSF n'est pas incompatible avec la définition saussurienne du signe linguistique en précisant que le concept d'arbitraire saussurien renferme en fait deux notions distinctes :

- Un premier sens, qui renvoie à la propriété de la face significative des unités linguistiques des langues orales [...]de ne pas représenter de lien de ressemblance avec la tranche de réalité qu'elles étiquettent[...];
- Un second sens, plus fondamental, qui renvoie au "caractère différentiel et négatif des signifiants de la langue". Ce qui revient à dire que la valeur d'un terme d'une langue résulte de la présence simultanée des autres termes à valeur opposable.

Autrement dit, l'aspect arbitraire que l'on observe dans les langues orales est une conséquence de la différentiation des unités de la langue à travers le canal audio-oral. L'aspect iconique des signifiants de la langue de signes ne constitue donc pas une entrave à leur fonctionnement différentiel dans le cadre de la linguistique structurale de Saussure et ne bloque pas la possibilité d'avoir accès à l'ensemble des jeux de langage.

### 3.1.1.2 La double articulation

La double articulation est une notion de base de la linguistique structurale. Elle stipule qu'une langue naturelle doit présenter un mécanisme dit de double articulation. La double articulation permet, à partir d'un petit nombre d'unités non porteuses de sens appelées *phonèmes*, de générer, en respectant un certain nombre de règles combinatoires, un nombre quasi illimité de phrases.

Les phonèmes se combinent en *morphèmes*, unités minimales porteuses de signification, eux-mêmes se combinant pour former des lexèmes (mots ou signes) qui à leur tour constituent, en accord avec la syntaxe, les phrases. Ce mécanisme a longtemps été un des critères majeurs de l'attribution du statut de "vraie" langue. Ce fait a donc poussé de nombreux chercheurs à proposer des décompositions de la langue des signes en unités non porteuses de sens. Ces travaux ont donné naissance aux études phonologiques de la langue des signes que nous présentons dans la sous-section suivante.



### 3.1.2 Modèles phonologiques des langues signées

Les premières études dédiées à la langue des signes ont d'abord été descriptives. On retrouve l'apparition des premiers dictionnaires avec Bebian et Remi-Vallade [Beb25] qui ont été les premiers à parler des langues signées comme étant "décomposables en unités structurales combinées simultanément".

Stokoe [Sto60], dans les années 60, a ensuite décrit les langues de signes à partir d'unités minimales établissant ainsi une phonologie de la langue des signes. Stokoe a démontré que la langue des signes pouvait être décomposée en phonèmes gestuels. Il les a appelés *chérèmes* et leur étude *chérologie*. Citons d'ores et déjà les trois chérèmes identifiés par Stokoe : l'endroit d'exécution du signes ou *tabulateur* (TAB), la configuration manuelle ou *désignateur* (DEZ) et le *mouvement* (SIG).

Battison [Bat78] ajoute un quatrième paramètre : l'orientation de la main ou des mains, et distingue, en outre, au niveau le plus général, quatre zones principales d'articulation du signe : le visage, le tronc, le bras et la main. Klima et Bellugi [KB79] définissent l'orientation de la paume de la main comme un paramètre mineur mais en valident l'importance en citant des paires de signes pour lesquelles ils constituent le seul critère de distinction.

L'identification et la description de la langue des signes à partir de la description de ses unités constitutives minimales non porteuses de sens, d'une part, a permis la reconnaissance des langues signées comme langue à part entière et, d'autre part, a posé les bases de la phonologie de la langue des signes. Le modèle phonologique a cependant été remis en cause au milieu des années 1980 par les travaux de Cuxac.

### 3.1.3 Remise en cause de la phonologie par la visée illustrative

Selon Cuxac, les signes standards de la langue des signes sont composés d'un agencement d'atomes porteurs de sens. Ainsi, les unités minimales qui entrent dans la composition des signes standards sont proches des unités morphémiques des langues orales. La décomposition de la langue des signes en unités atomiques porteuses de sens est une conséquence de l'iconicité vue comme mécanisme fondamental de la langue des signes. Nous abordons un peu plus loin l'iconicité des langues des signes et ses conséquences sur sa représentation.

### 3.1.4 Prise en compte de la structure spatiale de la langue des signes

Malgré l'évidence des spécificités de la langue des signes en terme d'organisation spatiale et de simultanéité des articulateurs, les études menées à la fin des années 60 se sont concentrées sur les mécanismes syntaxiques communs aux langues des signes et aux langues orales.

#### 3.1.4.1 L'ordre des éléments

Les langues des signes conservent certaines structures syntaxiques et des principes d'organisation proches de ceux des langues orales. Parmi ces concepts partagés, on trouve l'ordre des éléments majeurs tels que le sujet, le verbe et l'objet constitutifs d'une phrase. [Cux83, Mil98, RdLGH04]. Cette approche de la syntaxe est bien connue de travaux sur les langues orales et ne présente pas une approche spécifique aux langues de signes.

#### 3.1.4.2 Les verbes directionnels

Les verbes de mouvement (qui tracent un chemin dans l'espace lors de leur exécution) ou les verbes à orientations variables (signes où la direction et/ou l'orientation de la main change) constituent la grande majorité des verbes en LSF. L'identification du rôle des arguments de ces verbes utilise aussi l'espace : le point de départ du signe DONNER, par exemple, identifie cette place comme l'agent du processus et le point d'arrivée de DONNER comme la cible [Mot76].

#### 3.1.4.3 Marqueurs non manuels et flexions

Les marqueurs non manuels qui sont introduits de façon simultanée avec des signes manuels sont au cœur de la syntaxe de la LSF. Il a été montré que les marqueurs non manuels peuvent revêtir un rôle de qualificateur d'aspect, et de quantificateur, comme par exemple la marque du pluriel [Jou95].

### 3.1.5 L'iconicité en langue des signes

En linguistique, une *icône* est un signe qui a pour propriété d'imiter perceptuellement ce à quoi il réfère. En langue des signes, le processus d'iconisation est le processus par lequel le locuteur va rendre iconique l'expérience. Ainsi, le locuteur reconstitue de manière imagée une expérience vécue ou imaginée. Ou, en reprenant les termes de Cuxac, "[...] L'iconicité est le lien de ressemblance plus ou moins étroit entre la chose du monde, le référent et

le signe qui s’y rapporte [...]”. Toujours selon Cuxac, la langue des signes offre deux manières de “dire”. On peut “dire en montrant”, auquel cas il est fait appel à la visée iconique, ou “dire sans montrer”, auquel cas il est fait appel à l’agencement de signes standards.

Les usagers des langues orales ont parfois recours à une forme d’iconicité dans le discours, par l’intermédiaire de gestes coverbaux (« hier, j’ai pêché un poisson grand comme ça ! ») ou par l’intermédiaire d’onomatopées, (« lorsqu’il s’est jeté à l’eau, cela a fait “plouf !” »). Si les langues orales peuvent exprimer quasiment toute idée sans avoir recours à l’iconicité, elle est nécessaire aux langues signées. Par exemple, la syntaxe des langues des signes ne permet pas d’exprimer une assertion telle que « le chocolat mange le garçon » en se basant uniquement sur la visée non-illustrative. En revanche, il est tout à fait possible de traduire ce concept en ayant recours à la visée illustrative. Le locuteur pose le contexte de sa narration : le chocolat d’un côté, le garçon de l’autre, puis, il prend le rôle du chocolat et dévore le garçon. La bifurcation entre les visées est illustrée à la figure 3.2. En visée iconique, un locuteur donne à voir, par l’usage de structures iconiques telles que les transferts. A l’inverse, en visée non iconique, le locuteur ne donne pas à voir. Il a uniquement recours à un agencement de signes standards.

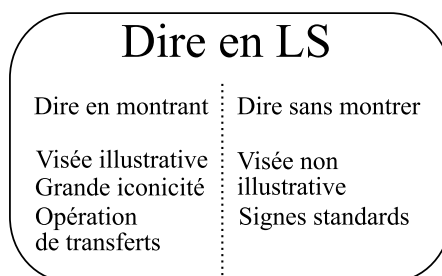


FIG. 3.2 – Bifurcation entre la visée illustrative et conceptuelle

La visée iconique a été décrite et formalisée par Cuxac puis enrichie par Sallandre [Cux00, Sal03]. Les structures de grande iconicité se distinguent selon leur visée. C’est alors que l’on parle de *transferts*. Le terme transfert renvoie à un ensemble d’opérations soumises à la visée iconisatrice du locuteur. La spécificité de chacune de ces opérations de transfert correspond à un type particulier de structure. Les structures de base sont identifiées comme transferts situationnels, transferts de taille et/ou de formes et transferts personnels. Nous présentons ci-après les trois structures minimales de grande iconicité identifiées par Cuxac :

- Le transfert de taille et de forme est utilisé pour montrer la taille ou la forme d’un élément du discours (rond, carré, taille de l’objet “telle qu’elle

- est"). C'est l'ensemble du corps (mains, visage, épaules) qui est sélectionnée dans ce type de transfert où l'on ne parle plus d'ailleurs de configuration, mais de proforme, car il s'agit d'une forme complètement iconique ;
- Le transfert situationnel (TS) sert à montrer le déplacement d'un élément par rapport à un locatif stable. La scène est vue "à la troisième personne". Le transfert de situation peut servir à mettre en place le décor, le cadre de l'énonciation ;
  - Le transfert personnel (TP) a pour finalité de reproduire l'expérience de l'un des actants de l'énoncé. Le locuteur est en prise de rôle et tout son corps est sélectionné. C'est à dire que tout son corps est mobilisé pour exprimer la situation de prise de rôle.

Reprenons l'exemple du chocolat et du garçon à la lumière des structures de grande iconicité. Pour énoncer l'assertion « le chocolat mange le garçon », le locuteur commence par présenter les acteurs du discours, en l'occurrence, le chocolat et le garçon. Ces deux éléments sont identifiés par leurs signes standards puis positionnés dans l'espace de signation à l'aide de *proformes* : il s'agit d'un transfert de situation (TS). Le locuteur pointe ensuite son regard sur le chocolat puis prend le rôle du chocolat (léger changement d'orientation des épaules et du buste puis changement d'attitude pour signifier que le locuteur prend le rôle du chocolat). En "devenant le chocolat", pour ainsi dire, le locuteur passe en transfert personnel. Son regard quitte l'interlocuteur : il incarne un rôle qui se situe dans un contexte qui exclut l'interlocuteur. C'est enfin en double transfert que le locuteur poursuit la narration : la main droite du locuteur figure la "bouche" du chocolat qui vient dévorer le garçon figuré par la main gauche du locuteur en proforme qui le représente. L'organisation de la langue des signes autour des structures iconiques peut se résumer à l'aide du schéma suivant 3.3 qui reprend, de manière simplifiée, l'organisation proposée par Sallandre [Sal03].

Ce schéma représente les deux visées : illustrative et non illustrative. et les composants qui les constituent. La visée non illustrative est ainsi exprimée à l'aide de l'agencement de signes standards, de gestes de pointage ou de séquences réalisées à l'aide de la dactylogogie. La visée illustrative se manifeste par des structures iconiques de transferts que nous avons définies au paragraphe 3.1.5.

### 3.1.6 Conclusion

Cette première sous partie nous a permis de découvrir, à travers l'évolution des recherches linguistiques dédiées à la langue des signes, les caractéristiques et les spécificités qui lui sont propres. Nous pouvons résumer cette sous-partie à l'aide du schéma suivant 3.4 :

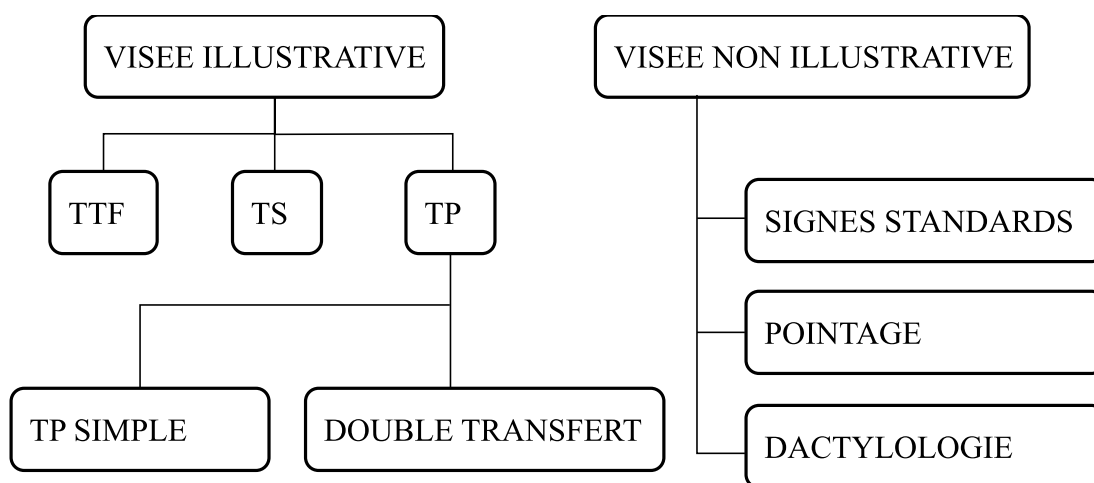


FIG. 3.3 – Modèle iconique, schéma récapitulatif

Le schéma 3.4 représente les deux manières de décrire la langue des signes que nous avons présentées dans les sections précédentes. Le premier modèle, phonologique découle directement des modèles linguistiques classiques et se cale sur les structures des langues orales sans tenir compte de la visée illustrative. Au contraire, le second modèle se base sur la visée illustrative de la langue des signes et prend en compte les structures de grande iconicité intervenant dans le discours.

Ce schéma permet de discerner différents sous-systèmes intervenants dans la langue des signes. Ces différents sous-systèmes nous offrent une base pour classer les travaux de recherche qui se sont attachés à la production d'énoncés en langue des signes :

1. La description gestétique représente les trajectoires des articulations mises en jeu par les gestes réalisées lors de leur productions ;
2. La description phonologique se situe au niveau de la description des paramètres formationnels de la langue des signes ;
3. Les unités minimales de sens et les mécanismes qui régissent leur agencements appartiennent au niveau morphologique ;
4. Certaines opérations syntaxiques qui s'appliquent aux langues orales peuvent être appliquées aux langues des signes. Par exemple, l'ordre des mots dans la phrase ;
5. Récemment, des recherches ont pris en compte des relations spatiales et géométriques entre les entités intervenant dans le discours en langue des signes. Cette prise en compte de deux aspects de la visée illustrative en langue des signes que sont les transferts de situation et les transferts de

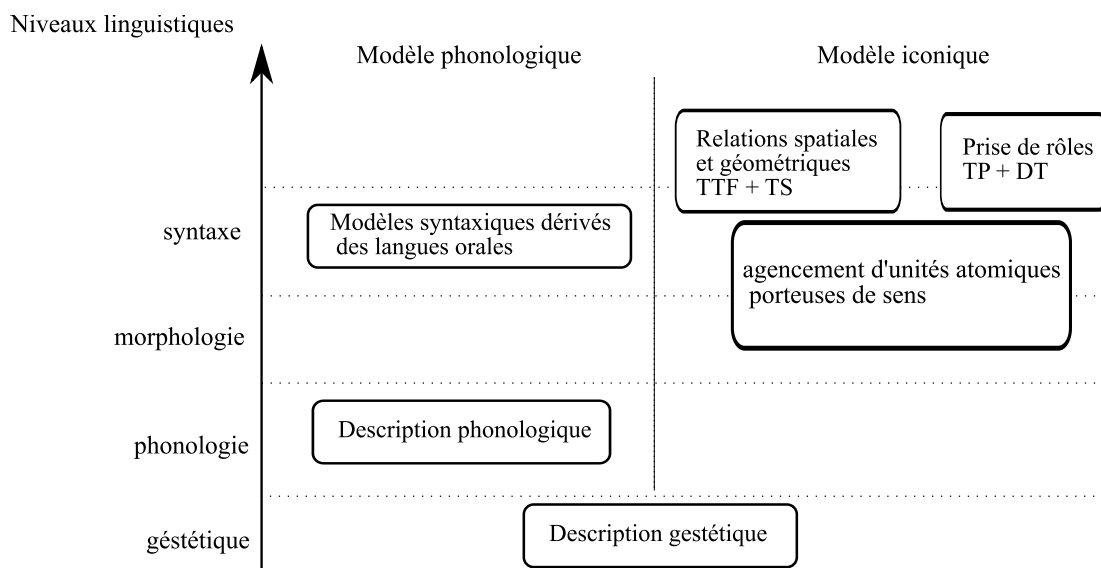


FIG. 3.4 – Organisation des différents champs d'études des LS.

forme constitue un premier pas dans la prise en compte de l'iconicité de la langue des signes ;

- Enfin, un troisième mécanisme de la visée illustrative est représenté par les prises de rôles. Les prises de rôles, aussi appelés *transferts personnels*, sont un aspect fondamental du discours, notamment dans le cadre du récit et ont été peu adressées dans le cadre de la modélisation de la langue des signes.

### 3.1.7 Descriptions des langues signées

Les premiers travaux dédiés à une description formelle et informatisée de la langue des signes sont sans doute ceux de Teuber, Battison et leurs collaborateurs [Bat78]. La description proposée repose sur les travaux de Stokoe [Sto60] et la finalité est la constitution d'un dictionnaire linguistique.

Un système de notation informatisé est "HamNoSys" [PLZ<sup>+</sup>89], acronyme d'Hamburg Notation System. Ce système a pour vocation de décrire les langues signées par leurs paramètres formationnels, c'est à dire, au niveau d'unités gésétiques non porteuses de sens.

De par son objectif de description, "HamNoSys" se veut donc être un outil international, c'est à dire qu'il peut être a priori utilisé pour décrire toute langue des signes, indépendamment de son origine. C'est d'ailleurs ce qui a été entrepris au sein du projet "E-Sign" [ZVRvdS04], où le système de notation "HamNoSys" a été utilisé pour décrire les langues des signes anglaises, néerlandaises

et allemandes. Nous rappelons que ces trois langues des signes ont des fondements différents et qu'elles présentent donc des dissemblances importantes.

Le fait que le système soit dédié à une description de bas niveau implique une certaine complexité et une certaine exhaustivité dans l'étendue des possibilités de notation et de description des signes par hamnosys. Ainsi, le répertoire de signes d' "HamNoSys" comporte environ 200 signes, uniquement pour les configurations manuelles, position, orientation et mouvement.

Les premières interfaces informatiques dédiées aux langues des signes offraient des interfaces de consultation de dictionnaires proposant des séquences vidéo pré-enregistrées de lexèmes [FCD<sup>+</sup>93]. La généralisation de la vidéo a permis de réaliser des applications destinées à la fois à l'apprentissage de la langue orale par les sourds et à celui de la langue des signes par les entendants.

L'accès direct à la séquence filmée se fait par une liste alphabétique. C'est le principe adopté par le centre de la langue des signes allemande, à Hambourg, éditeur de plusieurs dictionnaires dédiés à diverses branches professionnelles dont certains sont disponibles en ligne <sup>1</sup>. De telles réalisations sont les pendants des ouvrages lexicographiques papier.

Parmi les ouvrages lexicographique papiers, nous ne pouvons faire l'impasse sur le dictionnaire publié par l'I.V.T. [Moo98]. Cet ouvrage présente la particularité d'offrir un ordre de recherche inverse à celui jusque là considéré et constitue, en ce sens, le premier dictionnaire bilingue-français-LSF. Il est en effet possible de retrouver un signe à l'aide des paramètres formationnels principaux du signe : configuration de la main, lieu d'articulation et nombre d'articulateurs impliqués.

### **3.1.7.1 Modélisation et génération d'énoncés de la LSF**

En 1998, les travaux de Thierry Lebourque ont ouvert la voie aux systèmes de spécification et de génération du geste de la langue des signes par le système intitulé "Gessyca" [LG99]. Ce système repose sur une description formelle des paramètres formationnels de la langue des signes, couplée à des méthodes de génération du geste basées sur les paradigmes sensori-moteurs tel que proposé par Sylvie Gibet [GM94].

En 2001, Olivier Losson expose à son tour un système de spécification du geste de la langue des signes [Los00]. Ce système repose sur une étude approfondie de la phonologie de la langue des signes, en particulier les travaux de Battison [Bat78] et ceux de Scott Liddel [Lid89]. La description des signes se cantonne à une formalisation des éléments lexicaux constitutifs de la langue.

---

<sup>1</sup>Source : <http://www.sign-lang.uni-hambourg.de/projects/SLDicts.html>

En parallèle, le projet européen "E-Sign" , porté par plusieurs laboratoires européens, notamment le centre de linguistique de Hambourg. [Ken01] propose un système complet de spécification et de génération de langue des signes. Ce projet est également porté par des organismes de spécification de normes d'animation pour la spécification MPEG-4 [PPR02].

Un peu plus tard, apparaissent les travaux de Huenerfauth, dédiés à la création de personnages signeurs capables de prendre en compte les aspects directionnels des verbes ainsi que les relations spatiales entre les entités de la langue des signes [Hue06]. Huenerfauth présente également une méthode de représentation et de synchronisation des gestes de la langue des signes.

En 2006, Lenseigne et Dalle ont proposé une modélisation de l'espace de signation et des relations sémantiques entre les éléments du discours qui s'expriment de façon géométrique dans l'espace de signation du locuteur [LD06]. Ce modèle a été appliqué à la description d'énoncés narratifs en LSF et de situations de dialogues.

Un avatar signant répondant au nom de "Elsi" a vu récemment le jour au Limsi<sup>2</sup>. [FB06]. Cet avatar est animé par un infographiste à partir de séquences vidéos par la méthode de *rotoscopie*<sup>3</sup> bien connue en animation traditionnelle. Ces travaux ont pour objectif la création d'un modèle de spécification de gestes de la langue des signes basé sur une grammaire spatiale prenant en compte les travaux précédemment cités [Cux00, Lid03, LD06].

Progressivement, le monde industriel et le monde des médias se sont intéressés aux applications de la génération de gestes de la langue des signes pour un public sourd. Outre les résultats du projet "E-Sing" , on peut noter les initiatives française de Nexiad et de Websourd, qui ont proposé des interfaces homme machines robustes et originales destinées à la présentation de messages en langue des signes dans des lieux publics tels que les gares de la SNCF [snc06, snc07]. Beaucoup plus loin de nos frontières, un projet national chinois basé à Beijing a monté une chaîne automatique de transcription de texte en langue des signes chinoises [XG00].

La figure 3.5 recense sur une frise chronologique des illustrations des travaux cités dans cette section :

Afin d'avoir une vision globale des différentes études et d'apprécier les problématiques dressées par ces travaux, nous proposons une classification synthétique représentée sous la forme d'un tableau. Ce tableau met en parallèle les travaux présentés dans le cadre de la représentation, la génération de séquences de gestes de la langue des signes en les confrontant aux domaines que nous avons décrits dans la section précédente 3.1.5. Cette présentation

---

<sup>2</sup><http://www.limsi.fr>

<sup>3</sup>La rotoscopie est à l'origine une technique d'animation consistant à transformer une scène filmée en un dessin animé. Elle a été par la suite adaptée à l'infographie 3D.



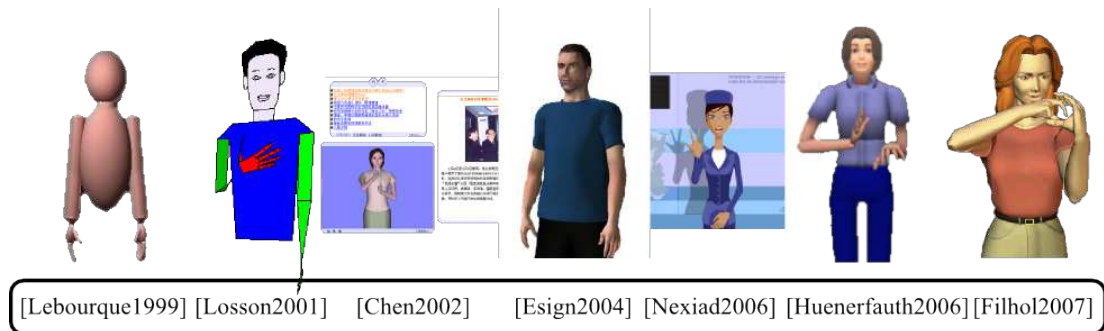


FIG. 3.5 – Systèmes de génération de gestes de la langue des signes par ordre chronologique

donne une vision globale des différents travaux, de leurs objectifs et des problématiques adressées.

Travaux	Gest.	Phon.	Lex.	stx. LO	TTF+TS	TP+DT	objectifs
Hamnosys2001	X						Notation
Brentari1998		X					Modèle ling.
Lebourque1999	X		X				Spec./Gén.
Losson2001	X	X	X				Génération
Kennaway2001	X		X	X			Trad./Gén.
Kennaway2003	X		X	X			Spec./Gén.
Nexiad2006			X				Génération
Chen2004	X	X	X	X			Trad./Gén.
Speers2001		X	X				Description
Kervajan2007		X	X	X			Traduction
Lenseigne2006					X		Description
Filhol2007		X	X		X		Spéc./Gén.
Huenerfauth2006			X	X	X		Trad./Gén.

TAB. 3.1 – Récapitulatif des travaux dédiés à la représentation et la génération de la langue des signes. **Gest.** = représentation gestétique, **Phon.** = représentation phonologique, **Lex.** = utilisation du lexique standard, **stx. LO** = utilisation des structures syntaxiques communes aux langues orales, **TTF + TS** = transfert de taille et de forme et transfert de situation **TP + DT** = transfert personnel et double transfert.

Il est intéressant de constater que l'évolution des travaux dédiés à la modélisation informatique et la production d'énoncés en langue des signes a suivi une évolution comparable à celle qu'ont connus les travaux linguistiques dédiés aux langues signées.

Les premiers travaux se sont concentrés sur les aspects gestétiques et phonologiques de la langue des signes et la façon dont les paramètres formation-

nels s'agencent pour former des éléments lexicaux appartenant à l'ensemble des signes standards. Dans un deuxième temps, une partie de la syntaxe de la langue des signes a été prise en compte, pour la modélisation de verbes directionnels, des phénomènes flexionnels et la prise en compte de l'ordre des éléments. Enfin, des structures iconiques ont été prises en compte, Parmi ces structures iconiques, on trouve les transferts de situations (gestion des classificateurs selon Liddel) et les transferts de taille et de forme. A notre connaissance, les travaux dédiés à la génération des gestes de la langue des signes n'ont pas encore pris en compte les transferts personnels. Or, ces transferts, décrits au paragraphe 3.1.5 sont couramment utilisés en langue des signes française.

### 3.1.8 Importance de la prise de rôle

Une étude menée en 2005 par Martin [Mar05] dans le cadre de son mémoire de master, nous donne quelques chiffres à propos du taux de prise de rôle dans un récit. Cette étude a été faite à partir de deux extraits du corpus vidéo réalisé dans le cadre du projet "LS-COLIN" [Ca01]. Ces deux extraits représentent un enregistrement vidéo d'un récit narratif mettant en scène des animaux dans une prairie. Ce récit, que nous appellerons "L'histoire du cheval", a été créé par Hickmann en 1982 afin d'étudier l'organisation des référents spatiaux et temporels dans l'acquisition du langage [Hic03]. "L'histoire du cheval" est un récit en images : il est décrit par une succession de 5 vignettes. Martin a répertorié pour les deux locuteurs la répartition des visées lors de leur interprétation de l'historiette du cheval. Il en ressort les résultats suivants :

- Le premier locuteur a produit 35, 5% de signes dans le cadre d'une prise de rôle ;
- Le second locuteur a produit 76, 8% de signes dans le cadre d'une prise de rôle.

Ces chiffres nous éclairent sur la part que prend la prise de rôle dans le cadre du discours. Cette importance de la prise de rôle dans le discours et le fait qu'à notre connaissance, elle n'ait pas été prise en compte dans les systèmes de modélisation de la langue des signes nous a amené à nous intéresser à la modélisation de la prise de rôle dans un contexte de génération du geste signé. Aussi proposons nous, dans le chapitre quatre, une modélisation de la prise de rôle, et plus précisément, les transferts personnels [Cux00] dans un système de modélisation du discours de la langue des signes.

Dans le contexte de la génération du mouvement, il est nécessaire de se demander comment il est possible d'exprimer l'idée de la prise de rôle, et, d'une manière plus générale, comment l'on peut traduire la visée illustrative fondamentale en langue des signes. Cette interrogation nous a mené à considérer les travaux de recherches dédiés à la description et à modélisation des paramètres

“qualitatifs” du mouvement tels que l’extension spatiale ou la tension du geste. Nous allons voir, dans la suite de cet état de l’art, les travaux dédiés à la caractérisation des paramètres qualitatifs du mouvement humain et comment ces travaux permettent de traduire la valeur illustrative de la langue des signes.

## 3.2 Apport des agents conversationnels

Les agents conversationnels sont un type d’interface visant à transférer les propriétés de l’interaction *homme-homme*, avec toute leur richesse, à l’interaction *homme-machine*. Elle font donc intervenir des représentations virtuels d’agents humanoïdes appelés *agents* ou *avatars*. Ces agents, peuvent être doués de gestes expressifs et de parole. Il s’agit alors de personnages virtuels multimodaux. Les agents conversationnels sont donc capables de communiquer avec l’utilisateur via à la fois plusieurs éléments dont les expressions faciales, la direction du regard, les gestes des mains, les postures corporelles, etc. [CSPC01].

L’efficacité d’un agent est en partie liée à sa crédibilité au cours de l’interaction. Les recherches dédiées aux agents conversationnels stipulent que le taux d’acceptation des agents par les utilisateurs est largement influencé par des paramètres non verbaux tels que l’attitude générale de l’humanoïde, sa posture, sa tenue vestimentaire, sa façon de bouger [LBAM04]. Ces caractéristiques sont le reflet de la personnalité de l’agent et peuvent être expliquées par son imprégnation culturelle et professionnelle, son origine sociale, ses caractéristiques physiologiques et son état émotionnel [Ken93, Nei92a]. Nous nous intéressons plus particulièrement à la modélisation des attributs propres à l’agent que nous pourrions exploiter dans le cadre de transferts de rôle en langue des signes. Nous nous concentrons sur la modélisation des paramètres expressifs des agents qui sont portés par le canal visuo-gestuel et le lien qui unit ces paramètres avec les attributs et les intentions d’un agent.

### 3.2.1 Prise en compte de la personnalité de l’agent, notion de style

L’une des préoccupations liées à la spécification d’agents conversationnels animés est la prise en compte d’attributs qui sont en relation avec la personnalité de l’agent. En effet, des études ont montré que les utilisateurs humains montraient des modes d’interactions *hommes-machines* comparables avec le mode d’interaction *homme-homme* [NST94]. Ce fait a été illustré par le paradigme CASA, acronyme correspondant à “Computers are social agents” [CSPC01]. La nécessité d’insuffler une “personnalité” propre à un agent pour augmenter son réalisme et améliorer son acceptation vis-à-vis d’utilisateurs humains a

fait l'objet de recherches dans le cadre de l'animation d'agents conversationnels. Ruttkay et Pelachaud [RPPN04] proposent un cadre théorique pour gérer les aspects propres à la personnalité d'un agent. Dans leur travaux, la manifestation de la personnalité d'un agent est regroupé sous un terme appelé *style*. Ce terme regroupe les attributs suivants :

- Le style dépend de la personnalité de l'agent ;
- Le style dépend de la culture de l'agent ;
- Le style dépend de l'état interne de l'agent.

Parmi les éléments qui ont une influence sur le style répertoriés par Ruttkay et al., on trouve la nationalité, le genre, l'âge, l'état physique, et l'humeur de l'agent. Pour Ruttkay et coll., le style a une influence sur deux plans dans la conception d'agents conversationnels. D'une part, il influence le choix de la modalité et du type de geste utilisé pour exprimer une intention. D'une autre part, il agit sur la manière dont est exécuté le geste.

Ces deux préoccupations distinctes sont appréhendées par l'utilisation des dictionnaires de gestes et de répertoires [RN05]. Ce paradigme permet de définir un répertoire de gestes pour chaque agent, étant composé ce répertoire d'un ensemble organisé élaboré à partir de gestes atomiques. L'agencement des gestes atomiques est réalisé par une grammaire mettant en jeu des contraintes spatiales et temporelles, des opérateurs de composition parallèles (&) ou séquentiels (+)[RN01].

### 3.2.2 Geste co-verbal et applications aux agents conversationnels

Les études dédiées à l'expressivité du geste ont tenté d'apporter une réponse à l'identification de paramètres pertinents pour la description des mouvements réalisés par les humains en situation de communication. Les travaux initiateurs ont été menés par McNeill [Nei92a]. L'approche de Mc Neill catégorise les gestes en cinq classes :

- Gestes iconiques : gestes qui représentent des caractéristiques du sujet évoqué (cf 3.1.5) ;
- Gestes déictiques : gestes qui font référence à un point dans l'espace faisant référence à un sujet évoqué («C'est par là !») :
- Gestes rythmiques : mouvements de la main réalisés lorsque le locuteur souhaite renforcer son discours ("Poing sur la table"),
- Emblèmes : les emblèmes sont des codes admis par une communauté ("Salut de la main, OK, pouce vers le haut").

Le modèle de McNeill a été mis en œuvre par différents travaux dédiés à la conception d'agents conversationnels dans des études menées par Cassel [CPB+94, CSPC01], Badler [BCC99] et Thalmann [BT96, CPMTT99]. Ces mo-

dèles ont permis la réalisation d'agents conversationnels capables de produire des gestes coverbaux pertinents en situation de dialogue ou d'interaction.

Les travaux de McNeill ont donc ouvert la voie à la conception d'avatars capables de produire les "bons" gestes coverbaux en situation de dialogue. Cependant, bien que les mouvements générés par ces agents soient "corrects", ils continuent à manquer de "naturel" et sont perçus comme robotiques par les observateurs humains. Pour pallier ce manque de naturel, des études se sont consacrées à l'analyse des nuances observées entre différentes réalisations de gestes expressifs.

### 3.2.2.1 Prise en compte du style

Les précédentes recherches dédiées aux agents conversationnels et aux interactions *homme-machine* ont mis en évidence le fait que les utilisateurs ont tendance à observer avec la machine des comportements qu'ils pourraient mettre en place avec d'autres humains. Comme nous avons déjà cité précédemment, ce constat a été à l'origine de la naissance du paradigme CASA [NST94]. Le réalisme des agents conversationnels dépend largement de caractéristiques non verbales, telle que la richesse des modalités utilisées en situation de dialogue, la cohérence et la pertinence des gestes co-verbaux exécutés.

Des personnes différentes vont exploiter de diverse façon les moyens de communication non verbales à leur disposition. Ces personnes vont également manifester une perception subjective différente d'un agent conversationnel selon la stratégie expressive que ce dernier déploie. Il a ainsi été montré que la réponse d'un utilisateur à un agent conversationnel est liée aux caractéristiques de l'agent telles que son statut social supposé, sa personnalité ou sa tenue vestimentaire [WSS94]. Il a par exemple été montré que les utilisateurs appréciaient plus les agents qui leur ressemblaient.

### 3.2.2.2 Paramètres qualitatifs du mouvement

La problématique du naturel du mouvement n'est pas circonscrite à l'animation par ordinateur : des artistes ont tâché de créer des animations convaincantes depuis l'invention du dessin animé. Ces techniques d'animation image par image ont d'ailleurs servi de guides à d'autres techniques procédurales [Las87, TJ95].

L'analyse qualitative du mouvement et son application à la danse a motivé la démarche de Laban (1879-1958) [Lab88] qui a mis en place un système de notation des mouvements corporels connu comme "LMA", acronyme pour Laban Movement Analysis. Ce système permet de représenter un mouvement à partir de paramètres décrivant la posture d'un sujet et la cinématique des mou-

vements en rapport avec l'environnement du sujet. Sa dimension temporelle, représentée à la manière d'une partition musicale permet de décrire précisément la séquence et la coordination.

Le mouvement est décomposé en cinq paramètres : corporel (*body*), de forme (*shape*), d'espace (*space*) et d'effort (*effort*) ainsi qu'un dernier, *relation*, qui caractérise les interactions avec l'environnement du sujet. Le paramètre corporel identifie les différentes parties du corps impliquées lors de la production du mouvement. Celui de forme décrit l'aspect du mouvement exécuté par le sujet. Le paramètre spatial décrit le chemin suivi lors de l'exécution du mouvement et précise l'occupation du corps en fonction de l'environnement. Le paramètre d'effort donne des informations d'ordre qualitatives à propos du mouvement, il peut être mis en parallèle avec les annotations de type { *legato, forte...* } que l'on retrouve sur les partitions de musique.

La "LMA" a inspiré les travaux de Chi et Badler pour le modèle d'expressivité du mouvement, baptisé "EMOTE" (Expressive Motion Engine) [CCZB00]. Celui-ci a pour vocation de proposer un modèle qualitatif du mouvement indépendant des gestes de communications spécifiés. On peut voir "EMOTE" comme un filtre qui ajoute un caractère naturel à un mouvement obtenu par des méthodes d'animation traditionnelles, comme l'illustre la figure 3.6 :

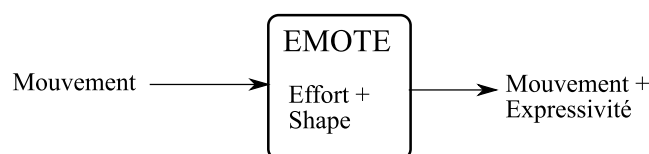


FIG. 3.6 – EMOTE modifie un mouvement existant

Quelques années plus tard, des travaux dirigés par Pelachaud [HMP06, MBP05] ont adopté une approche comparable en proposant un système permettant une sélection des types de gestes réalisés ; en ce sens, la démarche rejoint les travaux de Badler/Cassel mais elle permet également de moduler la manière dont les gestes sont réalisés. Le modèle d'expressivité est inspiré, d'une part, de résultats présentés dans des travaux en psychologie sociale [Wal98, Gal92] et, d'autre part, d'une étude approfondie d'un corpus vidéo [MHO<sup>+</sup>03].

Les paramètres d'expressivité identifiés dans [HMP06] sont au nombre de six :

- *Activation globale* : quantité de mouvement investie lors d'un tour de conversation ;

- *Portée spatiale* : amplitude des mouvements ;
- *Portée temporelle* : durée des mouvements ;
- *Fluidité* : degré de continuité des mouvements ;
- *Puissance* : propriétés dynamiques des mouvements ;
- *Répétition* : battements ou répétitions.

Ce modèle a été confronté à une évaluation poussée [BHMP06] qui a permis de confronter la pertinence des paramètres proposés pour générer trois comportements expressifs différents :  *Brusque, tonique et mou*. Les résultats expérimentaux montrent que les six paramètres d'expressivités ont une influence statistiquement significative pour différencier les qualités expressives des gestes.

### 3.2.3 Gestion du style pour les agents conversationnels, "l'art et la manière"

En définitive, les études dédiées à la gestion et à la modélisation de l'expressivité pour les agents conversationnels s'accordent pour considérer deux problématiques liées au style. D'une part, il s'agit de gérer un modèle de décision permettant de choisir une stratégie d'expression et un geste parmi un répertoire disponible. Cette gestion se situe au niveau comportemental. D'autre part, il s'agit de générer des gestes répondant à des critères permettant de nuancer les mouvements tels que le temps d'exécution ou l'amplitude des mouvements. Cette organisation est reprise par le schéma 3.7 qui fournit un cadre de référence dans lequel situer les études dédiées au style et à l'expressivité du mouvement humain.

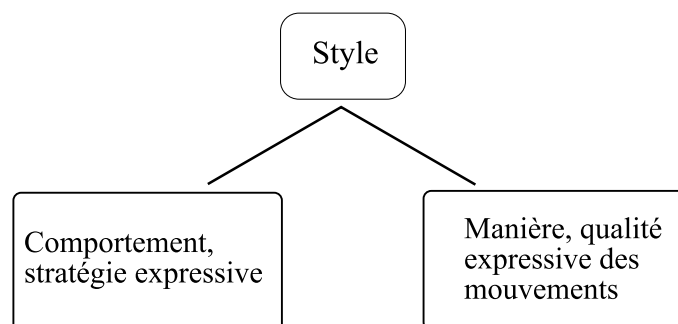


FIG. 3.7 – Le style regroupe deux problématiques

### 3.3 La génération de mouvements naturels et la notion de style en animation par ordinateur

Les recherches dédiées à l'animation d'entités virtuelles ou robotiques se sont très tôt focalisées sur la vraisemblance de leur apparence, de leurs mouvements et de leurs comportements. En 1970, un article de M. Mori [Mor70] établit grâce au concept d'*Uncanny Valley* une relation entre la vraisemblance d'une entité artificielle et la réponse émotionnelle d'un être humain à son égard. Le schéma de l'illustration reprend cette relation.

Ce schéma montre que plus une entité virtuelle nous ressemble, plus notre réponse émotionnelle est positive. Cependant, arrivé à un certain point de réalisme, les détails trahissant sa non-humanité, bien qu'infimes, provoquent un sentiment opposé à celui recherché, c'est à dire, une réaction de rejet vis à vis de l'entité. Ainsi donner une apparence humaine risque d'aboutir à l'effet opposé de celui poursuivi dans la mesure où l'entité artificielle trouble, et rendre plus difficile son acceptation par des utilisateurs humains.

Comme dans beaucoup de domaines liés à l'imitation des performances humaines, la recherche dédiée au mouvement humain a connu trois grandes phases :

- Les premières études, descriptives, furent réalisées afin de comprendre la cinématique [Mar94] du mouvement humains et les mécanismes permettant le mouvement [Mar83] ;
- Des systèmes de génération du mouvement se sont ensuite basés sur ces descriptions. Les modèles qui s'attachent à reproduire les causes à l'origine du mouvement peuvent être classés dans la catégorie des modèles générateurs, les modèles qui représentent la cinématique des structures au cours du temps peuvent être classés dans la catégorie des phénoménologiques ;
- L'avènement des outils de capture de mouvement et l'augmentation des capacités de stockage des ordinateurs grand public dans les années 1990 ont conduit les chercheurs à développer des approches basées sur des données de mouvement capturés. Il semble que nous n'ayons pas aujourd'hui quitté la troisième phase [BRRP97].

#### 3.3.1 Modèles générateurs

La caractéristique principale d'un modèle générateur réside dans le fait qu'il s'attache essentiellement à la cause du mouvement. Parmi les méthodes visant à produire des mouvements naturels, celles basées sur des modèles physiques consistent à définir des contrôleurs spécifiques dont l'action est d'ap-



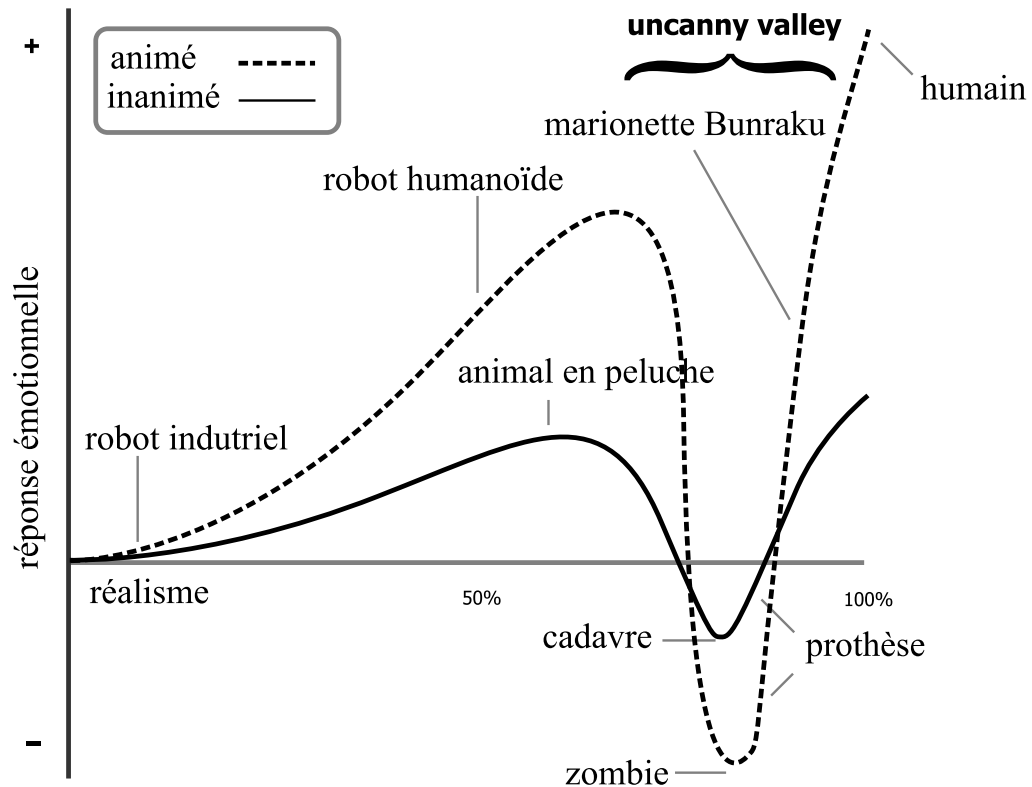


FIG. 3.8 – Shéma illustrant le concept d’uncanny valley, d’après Masahiro Mori et Karl MacDorman.

plier des forces et des couples au système mécanique représentant le corps articulé du personnage de manière à ce que l’évolution du système corresponde au mouvement désiré. Ainsi, à partir de la connaissance des forces appliquées à chaque articulation du squelette, on peut calculer l’état du système articulé en intégrant les équations du mouvement au cours du temps. Ces équations assurent intrinsèquement le réalisme des mouvements générés.

Il reste néanmoins à spécifier les forces et les couples à appliquer sur l’humain virtuel afin que son évolution dans l’environnement corresponde au mouvement désiré. La détermination automatique de ces efforts n’ayant pas de solution analytique dans le cas général, le problème de l’animation dynamique d’un humain virtuel devient un problème de contrôle inverse moins intuitif que celui du contrôle cinématique inverse. Plusieurs abords visent à répondre à ce problème de contrôle : une approche basée sur la gestion des contraintes [Arn89, Dum90, WFB87] ; une autre consistant à simplifier le modèle dynamique ou à en simplifier le contrôle [Popovic 99], [Fang 03], et enfin une

approche basée sur la conception de contrôleurs spécifiques [HWBO95, HP97, FvdPT01]. Pour finir, des méthodes hybrides mixent des aspects propres au contrôle cinématique et dynamique. Ces procédés combinent des modèles basés sur les données capturées et des modèles dynamiques, tirant de cette façon parti du réalisme des mouvements capturés ainsi que du réalisme physique de l'approche dynamique [ZH99, ZH02, ZMCF05].

### 3.3.2 Modèles phénoménologiques

La notion de style est difficile à définir. Cependant, la séparation entre la stratégie expressive et les aspects qualitatifs du mouvement nous offre un cadre permettant de représenter les travaux dédiés à la prise en compte d'une certaine notion de style dans les mouvements humains. Les techniques d'animations dédiées aux humanoïdes virtuels se sont employées, à partir du milieu des années 90, à proposer des modèles prenant en compte certaines facettes des aspects naturels du mouvement humain. Bien que la littérature regorge de travaux à ce sujet, les problématiques abordées par chaque travail et les objectifs dénommés par des termes récurrents (*naturel, style, expression, texture...*) ont souvent peu de choses en commun. Par exemple, la notion de style décrite dans la littérature dédiée à l'animation revêt de nombreuses significations : état interne du personnage animé ([ABC96]), différence de réalisation d'une classe de mouvements (danse) entre différents sujets ([BH00]), paramètres qualitatifs et quantitatifs de différents modes de déplacements ([SF06, UGB<sup>+</sup>04, Mar06, Tro02]), application à la cinématique inverse ([GMHP04]), gestes expressifs ([NF05, MZF06, SDO<sup>+</sup>04]). Nous allons décrire ces travaux en les organisant selon les techniques employées.

Dans le but d'insuffler cette coloration naturelle au mouvement, la majorité des méthodes d'animation prenant en compte le style reposent sur des séquences capturées. Les premières techniques se proposent de modifier des mouvements existants à l'aide de méthodes d'animation classiques, de techniques de filtrage au niveau signal ou par une combinaison des deux. Une seconde classe de méthodes reposent sur l'hypothèse que le mélange de mouvements capturés (par essence naturel) doit produire un mouvement naturel. Ces méthodes de mélanges de mouvements ont recours à des techniques d'alignement non linéaires, ainsi que des méthodes d'évaluation de distances entre séquences et postures de mouvements. Enfin, une troisième catégorie de méthodes s'emploie à modéliser des composants du mouvement plus ou moins indépendant, et comportant une sémantique plus ou moins bien définie.

### **3.3.3 Analyse-synthèse**

#### **3.3.3.1 Modification de mouvements**

Nous avons vu dans la section consacrée aux agents conversationnels comment s'architecturaient les solutions dédiées à la prise en compte du style. Nous allons voir dans cette sous-section comment est gérée la prise en compte des paramètres qualitatifs de l'expressivité. La stratégie de modification de mouvements consiste à perturber un mouvement initial, obtenu par capture, pour respecter de nouvelles contraintes ou s'adapter à une morphologie différente d'agent. Le problème est alors de respecter ces nouvelles contraintes tout en conservant les caractéristiques du mouvement initial.

Généralement, la morphologie de l'acteur humain qui réalise le mouvement et celle du personnage de synthèse qui le reproduit ne sont pas identiques. Or, calquer le mouvement d'un humain sur un personnage aux proportions différentes peut introduire certains artefacts comme le glissement du pied d'appui ou sa pénétration dans le sol. Cela est d'autant plus problématique pour les gestes expressifs, où le respect des contacts est indispensable à la bonne compréhension des mouvements. Des collisions des membres avec le corps peuvent aussi apparaître dans le cas où l'on modifie la taille des segments du squelette sans changer les valeurs des angles de la séquence de mouvement capturé par exemple. Il devient alors nécessaire de réaliser certaines adaptations sur les trajectoires articulaires.

Afin de remédier à ce problème, la méthode de Gleicher [Gle98] s'appuie sur une stratégie d'optimisation de contraintes spatio-temporelles pour calquer un mouvement capturé sur un squelette dont la structure articulaire est la même mais qui possède des segments de longueurs différentes. On arrive ainsi à assurer les contraintes géométriques tout en conservant un mouvement continu. Notons que cette adaptation conserve le style du mouvement, en guise d'exemple on peut citer l'adaptation d'un geste d'un personnage de grande taille (adulte) sur un personnage de taille plus modeste (enfant) a pour résultat un enfant marchant comme un adulte. Aucune redistribution de la dynamique du mouvement n'est effectuée : le réalisme physique apporté par cette technique est limité aux contraintes géométriques imposées.

Pour tenir compte de la physiologie, l'approche proposée par Komura et coll. [KSK00] enrichit le modèle de Gleicher en calculant et modifiant par dynamique inverse les forces liées à une structure musculaire basée sur celle de l'humain. Ces techniques étant coûteuses en terme de temps de calcul, des méthodes ont été proposées pour contrôler uniquement les extrémités des membres [CK00], et plus particulièrement des pieds d'appui pour éviter le phénomène de glissement [Mé03]. Ces méthodes se basent sur de simples contraintes cinématiques et permettent de conserver les conditions nécessaires à

l'interactivité de l'animation. Elles se révèlent être performantes, mais ne prennent pas en compte les aspects sensori-moteurs liés au contrôle du mouvement. [CB06, MBBT00]

### 3.3.3.2 Mélange de mouvements

En partant du principe qu'une posture peut s'exprimer comme une combinaison linéaire de plusieurs autres, de nombreuses méthodes de mélange de mouvements ont été développées. Un exemple simple permet d'illustrer leur motivation : une séquence où un personnage fait un geste de salutation avec le bras en marchant peut être décomposée en deux séquences. La première est une séquence de marche, la seconde, un geste de salutation d'un personnage en position statique d'autre part. Un certain nombre de contraintes doivent être vérifiées pour assurer la cohérence de l'animation finale : par exemple, on ne peut pas mélanger simultanément deux mouvements antagonistes, c'est à dire qu'on ne peut pas mélanger, par exemple, un mouvement mettant en jeu une flexion au niveau du coude gauche avec un mouvement mettant en jeu une extension au niveau du coude gauche.

Une manière d'assurer la synchronisation entre deux séquences compatibles est de normaliser leurs durées respectives. Cette opération, appelée *déformation dynamique du temps* (Dynamic Time Warping), est issue de la recherche en traitement de la parole [MR81] et a été introduite dans le domaine de l'animation par Bruderlin et coll. [BW95]. Le DTW permet d'aligner dans le temps des mouvements qui présentent des distorsions non-linéaires. Le terme dynamique de DTW est dû au recours à la programmation dynamique pour résoudre le problème d'alignement temporel [MR81]. Une fois alignés, le mélange entre les mouvements devient pertinent. Ces alignements ont toujours été réalisés sur des mouvements courts (mouvements d'arts martiaux) ou des séquences quasi-périodiques (mouvements de marche).

La multiplication de séquences de mouvements disponibles a permis aux méthodes reposant sur de grandes quantités de mouvement de s'imposer. Des solutions ont été imaginées pour extraire la structure sous-jacente d'un ensemble de mouvements et permettre d'effectuer en temps interactif les transitions les plus probables entre les différentes séquences d'une même base de mouvements. La solution la plus courante est de considérer un graphe de mouvements dont les nœuds sont des postures et les arcs des transitions possibles entre deux configurations [KGP02]. Ainsi, un nouveau mouvement est introduit dans le graphe sous la forme d'une suite de nœuds-postures et leurs transitions. Pour calculer la distance entre les nœuds du graphe, plusieurs métriques existent. Ces métriques utilisent la similarité entre les informations angulaires des postures [LCP<sup>+</sup>02] ou des coordonnées cartésiennes des articulations des

postures [KGP02].

Pour des applications particulières, des informations peuvent enrichir les nœuds pour influencer le choix des transitions et offrir un contrôle de haut niveau. Ainsi, Lee et Hoon [LL04] ont proposé d'associer aux nœuds-postures une grille de positions clés qui guident automatiquement la recherche de chemin et offre par exemple la possibilité de modéliser le combat entre deux boxeurs autonomes. D'autres méthodes d'interpolation haut niveau permettent de contrôler l'interpolation à l'aide de contraintes géométriques simples spécifiées par l'utilisateur[LCP<sup>+</sup>02]. Si ces modèles de génération peuvent être contrôlés par un jeu réduits de paramètres (direction et vitesse d'une marche), ils sont incompatibles avec la multitude de paramètres décrivant les gestes expressifs.

La faiblesse de ces techniques découle du fait suivant : si deux séquences de mouvements capturés respectent systématiquement les lois de la dynamique par nature, leur interpolation ne garantit en rien de conserver cette propriété. Par exemple, les contraintes liées au non glissement des pieds au sol ainsi que les contraintes d'équilibre ne sont pas garanties après une interpolation. Il peut donc s'avérer nécessaire d'effectuer une phase de correction du mouvement après la génération du mouvement [KG03].

### **3.3.3.3 Extraction des composants stylistiques et modification de mouvements**

Nous devons les premières études dédiées à la modélisation du style à partir de données à Unuma et coll. [UAT95] et à Amaya [ABC96]. La méthode proposée par Unuma repose sur une modélisation multi-échelle du mouvement dans le domaine fréquentiel. Les différents niveaux sont extraits par analyse de Fourier sur les trajectoires des angles aux articulations des squelettes à modifier. Un agencement pertinent de différentes classes de composants permet de générer des mouvements doués d'une expressivité particulière.

La modification d'expressivité est réalisée par la transposition des données de mouvements exécutées selon des styles différents dans un espace caractérisé par la vitesse d'exécution du mouvement ainsi que son extension spatiale. Une fois qu'un modèle d'émotion est inféré à partir de données d'apprentissage, le modèle peut être appliqué à des mouvements neutres pour modifier (filtrer) les paramètres d'expressivité qu'ils véhiculent. La nature fréquentielle de l'analyse proposée par Unuma requiert des mouvements pseudo-périodiques comme des cycles de marche. De plus, la combinaison des différents composants du mouvement nécessite un alignement préalable type DTW. La méthode proposée par Amaya est dédiée à la traduction stylistique, et pas à la génération de mouvement.

En étiquetant manuellement une base de mouvements décrivant des ac-

tions simples exécutées selon différents styles, Rose, Cohen et Bodenheimer ont introduit la notion de verbes et d'adverbes [RCB98]. Selon ce paradigme, les actions élémentaires (verbes) sont modifiées à l'aide de paramètres (adverbes) permettant de modifier la qualité expressive du mouvement. Un espace d'interpolation entre les actions de base est construit à l'aide de fonctions à bases radiales. La cinématique inverse est utilisée pour résoudre des contraintes telles que les appuis au sol. Avant la construction de l'espace d'interpolation, les fonctions sont alignées par des méthodes type DTW. La faiblesse de cette technique est le jeu restreint d'adverbes qu'il est possible de mettre en oeuvre.

Dans un but similaire, Brand & Hertzmann [BH00] proposent les dits *automates stylistiques*, "style machines". Ces automates sont conçus dans le but de modéliser une classe de mouvements relativement complexe (mouvements de danse) réalisés selon des styles différents à l'aide d'un système d'états discret capable d'une part de modéliser partie du mouvement commune à toutes les réalisations et d'autre part, d'être capable de passer d'un style à l'autre par des excursions plus ou moins grande autour du modèle commun. Les réalisations différentes du mouvement peuvent revêtir des propriétés telles que l'émotion, la personnalité, mais également la technicité ou la compétence dans la réalisation de mouvements complexes (danse).

Un avantage de cette technique est qu'elle ne nécessite pas de segmenter, d'annoter ou d'aligner les données de mouvements, la construction du modèle se faisant par apprentissage. Cela implique que l'utilisateur n'a pas la possibilité de fixer les paramètres contrôlant le modèle inféré, retournant l'avantage en défaut, si l'utilisateur a besoin que la structure du modèle inféré ait une sémantique significative.

### 3.4 Conclusion

Cet état de l'art nous a permis, dans un premier temps, de faire le point sur le statut de la langue des signes en tant que langue à part entière, en faisant des parallèles avec des caractéristiques propres aux langues naturelles telles que la double articulation et la dissociation entre le signifiant et le signifié pour les signes linguistiques qui la composent. Une fois ce constat établi, nous avons passé en revue les caractéristiques propres aux langues signées, en tant que langage hautement spatialisé.

L'analyse des études linguistiques initiées par Cuxac nous ont permis de prendre en compte l'importance de la visée illustrative de la langue des signes et de son importance fondamentale dans la structure des langues signées. La visée illustrative s'exprime dans le modèle de l'iconicité qui classe les méca-

nismes liés à la visée illustrative selon les mécanismes de transfert. Les trois grandes classes de transferts que sont les transferts de situation, de taille, de forme et personnels ont été présentés. Nous avons ensuite montré l'importance des prises de rôles exprimés lors des transferts personnels. Une revue des travaux dédiés à la modélisation de la langue des signes nous a permis de recenser les mécanismes liés à la visée illustrative qui ont été adressés par les chercheurs depuis les années 1970. Cette revue de la littérature nous a permis de constater que la visée illustrative n'avait été que en partie adressée, puisque aucun des travaux répertoriés n'a adressé les mécanismes de transfert personnels.

La réflexion que nous avons engagée à propos des besoins nécessaires à la modélisation des transferts personnels nous a conduit à considérer la littérature dédiée aux agents conversationnels. Ces travaux nous ont en effet semblé pertinents dans la mesure où ils se sont intéressés à la modélisation de l'état interne de l'agent et sur l'influence de cet état sur les stratégies communicatives déployées par l'agent et sur la manière d'exécuter ces stratégies. Il nous paraît en effet pertinent de mettre en parallèle le choix de la stratégie communicative employée par un agent avec l'élection de la visée employée par un signeur en situation de discours ou de dialogue.

Les paramètres d'expressivité décrits et mis en œuvre par les agents conversationnels, au delà d'augmenter la naturalité des expressions produites renseignent sur des aspects propres à l'état interne de l'agent et sur ses intentions pendant le discours (état émotionnel, emphase, origine culturelle). Appliqués à la langue des signes, les paramètres d'expressivité vont pouvoir, selon la visée employée, enrichir et qualifier le discours, soit à la manière d'adjectifs ou d'adverbes, à la fois en visée non-illustrative ou en visée illustrative.

La littérature dédiée aux agents conversationnels propose des modèles d'expressivité inspirés d'études dédiées à l'évaluation qualitative de mouvement humains ou d'études psycho-cognitives relative à la communication non-verbale. Ces modèles proposent une paramétrisation sémantiquement satisfaisante, mais leur mise en œuvre conduit à des gestes manquant de réalisme dans la mesure où ces modèles que l'on pourrait qualifier d'*experts* ne parviennent pas à prendre en compte l'infinité de nuances contenue dans le mouvement humain.

La problématique liée à la génération de mouvements naturels a mobilisé de nombreux animateurs puis de chercheurs en animation par ordinateur depuis l'invention de l'animation. C'est donc cette problématique qui a motivé la troisième partie de l'état de l'art dédiée à la génération de mouvements naturels en animation par ordinateur. Les solutions proposées se sont dans un premier temps inspirées du savoir faire des experts en animation et en robotique pour s'orienter progressivement vers une démarche d'analyse-synthèse. La démarche d'analyse-synthèse, facilitée par l'apparition d'outils

de captures et de méthodes de stockage de plus en plus performants et de moins en moins coûteux a engendré trois courants de méthodes de générations de mouvements que sont le mélange/concaténation de mouvement, l'adaptation/modification de mouvements et l'inférence de modèles de mouvements *enfouis* dans les données par apprentissage, identification, traitement statistique ou traitement du signal.



# Chapitre 4

## Modèle du discours

Le chapitre consacré à l'état de l'art nous a permis de considérer l'importance de la visée illustrative en langue des signes. Les structures de grande iconicité forment en effet le noyau sur lequel reposent l'organisation des LS.

Par conséquent, un système informatique prétendant prendre en compte la langue des signes se doit de prendre en considération les structures de grande iconicité. Nous avons vu dans la littérature que les systèmes de modélisation et de génération de la langue des signes se limitent à la prise en compte des signes standards stabilisés et que, très récemment, des travaux adressent une partie seulement des structures de grande iconicité, en l'occurrence les transferts de situation [LD06, Hue06, BL06] ainsi que les transferts de taille et de forme [BL06]. Dans ce chapitre, nous proposons un modèle de discours capable de prendre en compte certains mécanismes mis en jeu dans le cadre de la visée iconique de la langue des signes tels que les transferts personnels.

Le constituant de base de notre modèle de discours repose sur la notion d'entité. Une entité représente un concept qui est appréhendé par l'agent signeur. Une entité peut représenter une personne, un objet ou une notion plus abstraite dont il est question dans le cadre du discours. Nous allons voir comment la visée employée par l'agent signeur va influencer le mode de représentation d'une entité et les attributs qu'elle est censée contenir.

### 4.1 La notion d'entité et sa représentation

Une entité peut représenter une personne, un objet ou une notion plus abstraite dont il est question dans le cadre du discours. Une entité dans le discours est liée à un référent (cf chapitre 3, paragraphe 3.1). Nous nous appuyons sur la notion de référent proposée par Lenseigne et coll. [LD06] où il représente une unité sémantique du discours pouvant être spécialisée en *date*,

*lieu, action ou objet*. Un référent revêt une *fonction* dans le discours. La représentation gestuelle d'une entité peut se faire hors visée iconique à l'aide d'un signe standard qui la représente (par exemple, un signeur utilisera le signe standard "arbre" pour représenter un arbre). Ce signe standard ou *stabilisé* représente, par convention, l'entité. En visée iconique, la représentation d'une entité ne passe plus par les signes standards mais par une représentation gestuelle des propriétés pertinentes de l'entité selon sa fonction dans le discours :



FIG. 4.1 – Locuteur représentant un boxer en visée iconisatrice.

Par exemple, lorsque un locuteur représente un chien en visée iconisatrice, il va s'attacher à représenter le chien évoqué "tel qu'il est", en faisant référence à ses caractéristiques marquantes. Si le chien dont il est question est un boxer à la démarche pataude, le locuteur s'appliquera à représenter la physionomie et la démarche du boxer. Un locuteur représentant un chien en visée iconisatrice est représenté à la figure 4.1. Autrement dit, en visée illustrative, le locuteur s'attache à représenter le référent tel qu'il est. En visée non illustrative, un locuteur fait référence au signifié, "le chien" à l'aide d'un signe standard.

Le diagramme de la figure 4.2 offre une représentation à l'aide du formalisme UML des relations entre l'entité, son référent et des fonctions qu'il revêt.

La fonction peut revêtir trois types : agent, localisation spatiale ou temporelle. Ces fonctions permettent de construire des spécialisations d'actions,

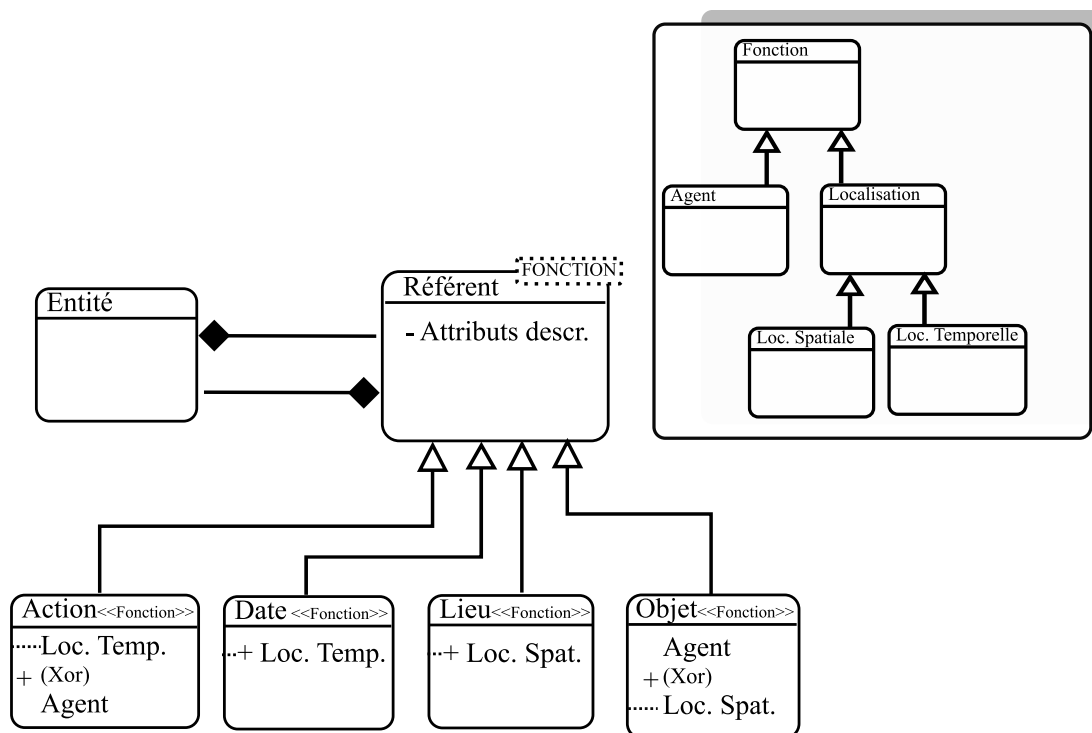


FIG. 4.2 – Diagramme de classe représentant l'entité et son référent

de date, de lieu ou d'objet, héritant de la classe *réfèrent*. Celui-ci référence une entité et la qualifie à l'aide d'attributs descriptifs.

#### 4.1.1 Hors visée illustrative

Hors visée illustrative, une entité peut être représentée à l'aide d'un signe standard ou signe stabilisé. Lorsqu'il n'existe pas de signe standard pour représenter une entité (par exemple, pour introduire une personne entendante ne disposant de nom signé), celle-ci est épelée à l'aide de la dactylologie.

Le diagramme de la figure 4.3 représente le lien entre une entité et les signes standards qui la réfèrent.

#### 4.1.2 En visée illustrative

En visée illustrative, le signeur s'attache à représenter les aspect importants du référent de l'entité dans le contexte du discours. Ces aspects importants peuvent être des caractéristiques marquantes et représentatives du référent (caractéristiques physiques, forme globale, taille), de relations spatiales et/ou topologiques entre les référents, ou encore des traits relatif au comportement,

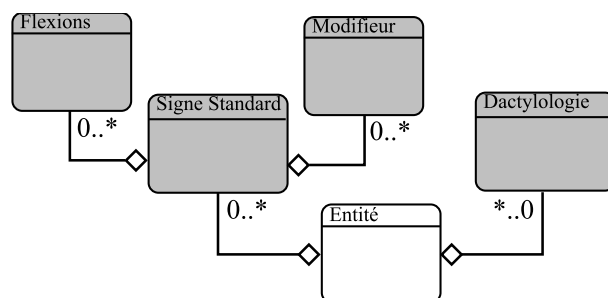


FIG. 4.3 – Représentation d’une entité hors visée iconisatrice

au caractère, à la culture ou à l’état mental d’un actant de l’énoncé (cf. l’exemple du boxer au paragraphe 4.1). Selon le cas, ces aspects seront représentées à l’aide de l’une des trois structures de grande iconicité identifiées par Cuxac : transferts de taille et/ou de forme, transfert de situation et transfert personnels. Ces trois structures ont été sous-catégorisées par Cuxac [Cux00] puis Sallandre [Sal03].

#### 4.1.2.1 En transfert de taille et de forme

La structure de taille et de forme permet de représenter la taille et/ou la forme, partielle ou globale, de lieux, d’objets ou de personnages. Les gestes permettant de figurer ces tailles/ou formes dans l’espace de signation se composent :

- D’une configuration de la main (ou des mains) indiquant une forme de base : une oreille de chien est représentée par un cône, esquissé par le pouce, l’index et le majeur ;
- D’un mouvement et d’une orientation de la main (ou des mains) signifiant le déploiement de la forme dans l’espace : selon la physionomie du chien, la forme du cône représentant l’oreille sera différente ;
- D’un emplacement de départ du déploiement : les oreilles du chien sont représentées sur la tête au dessus du front de l’agent.

Un inventaire de configurations manuelles décrivant les formes illustrées a été élaboré par Cuxac [Cux00]. Les formes décrites par les configurations manuelles peuvent représenter, entre autres, l’épaisseur (plus ou moins), la forme (sphérique, carré, rectangulaire) ou un agencement (rayonnant, aligné, grillagé, en éventail). Lorsque plusieurs transferts de formes sont réalisés pour caractériser un objet complexe ou un personnage, on parle de *descripteurs*. Les descripteurs peuvent être préférés au lexique standard dans la mesure où ils offrent une représentation plus pertinente de l’objet dans son contexte. Par exemple, dans un récit tiré du corpus de Cuxac [Cux00] mettant en scène deux

chiens de morphologies très différentes (un boxer et un chien loup), les deux protagonistes du récit sont décrits à l'aide de transferts de forme sans que le signe standard [CHIEN] ne soit utilisé.

Selon la visée illustrative, les transferts de taille et/ou de forme décrivent des attributs significatifs d'un référent faisant partie de l'énoncé. Par exemple, les attributs des chiens caractérisent le museau, les oreilles, les bajoues du boxer. La figure 4.4 illustre à l'aide d'un diagramme d'objet les attributs des deux référents correspondants aux deux chiens protagonistes du récit tiré du corpus de Cuxac :

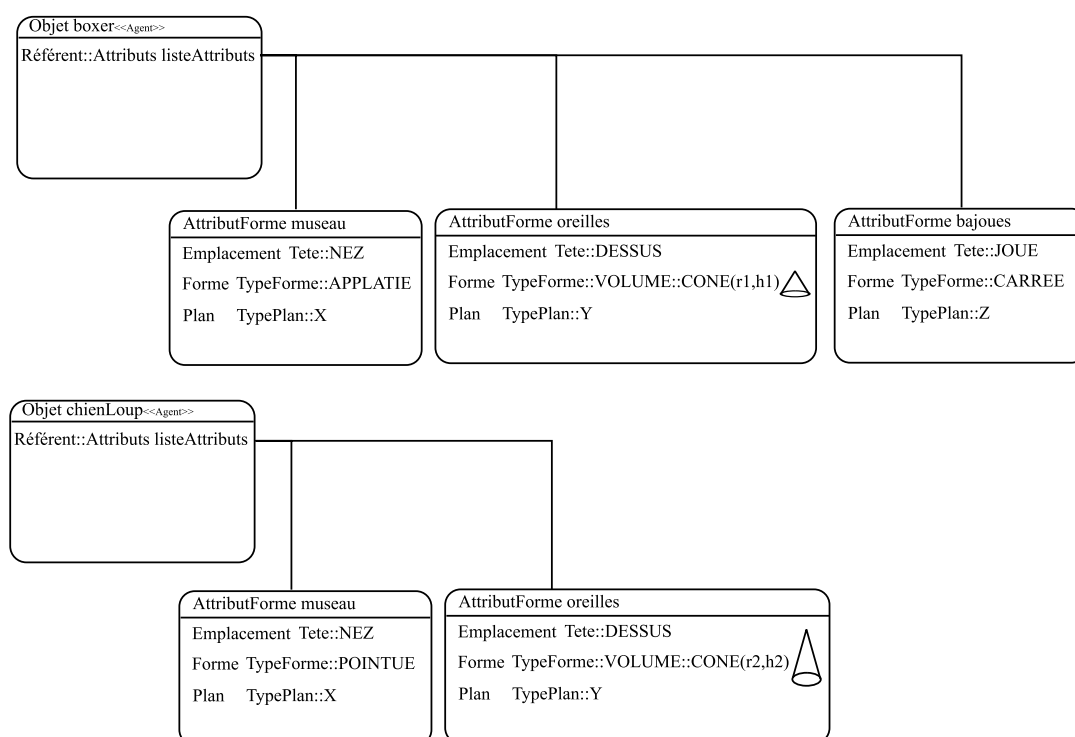


FIG. 4.4 – Spécification d'attributs de taille ou de forme pour les référents

#### 4.1.2.2 En transfert de situation

Par un transfert de situation, le locuteur vise à reproduire dans son espace de signation les scènes qui figurent le déplacement spatial d'un actant par rapport à un locatif stable (réfèrent spécialisé en localisation spatiale, cf. paragraphe 4.2). L'action effectuée par la main dominante s'inscrit dans un paradigme restreint de configurations que l'on appelle *proformes*. Les énoncés de transferts situationnels se réalisent de la façon suivante :

- Configuration du locatif par la main dominée, par exemple une piste de décollage : main dominée à plat, bras horizontal ;
- Configuration de l’agent par la main domante, comme pour représenter un avion : pouce et auriculaire de la main dominante écartés suggérant les ailes) ;
- Mouvement effectué par la main dominante, représentant la trajectoire de l’avion relativement à la piste ;
- Emplacement d’arrivée, où la main dominante est “en l’air”, bras tendu.

Lorsqu’un signeur représente, par exemple, un avion en train de décoller, la main dominée est animée d’un mouvement de va et vient sous la main dominante pour figurer le sol qui défile sous les ailes de l’avion. La relation qui lie l’actant et le locatif est de type trajectoire. Cette relation s’applique sur l’actant et sa localisation et est paramétré par une trajectoire et un profil de vitesse. L’illustration 4.5 représente l’élaboration d’un transfert de situation selon notre modèle du discours. L’exemple employé est le transfert de situation représentant un avion qui décolle :

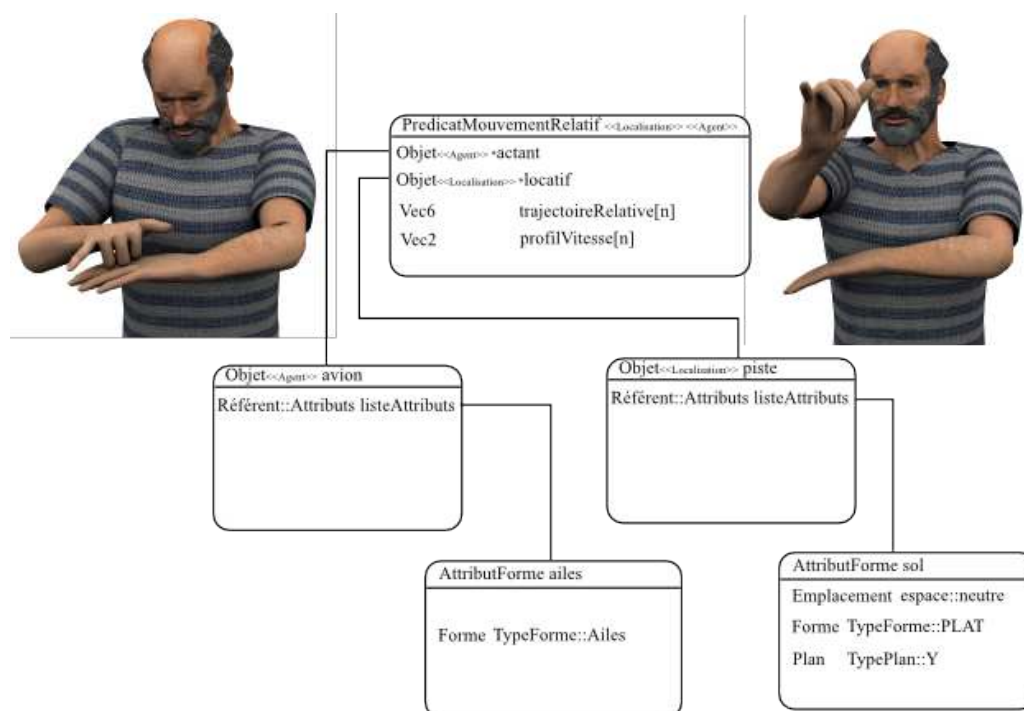


FIG. 4.5 – Spécification des relations spatiales en transfert de situation

En transfert de situation, la réalisation de la trajectoire a une fonction hautement illustrative. La manière dont l’agent réalise le mouvement offre, en effet,

une description objective du déplacement réalisé par l'actant (forme de la trajectoire, profil de vitesse) et des relations spatiales qui le lient avec l'objet référé (relations topologiques telles que décrites dans [BL06]).

#### 4.1.2.3 En transfert personnel

Les structures de transfert personnel reproduisent, en mettant en jeu tout le corps du locuteur, une ou plusieurs actions effectuées ou subies par un actant impliqué dans l'énoncé, humain ou animal le plus fréquemment, mais ce peut-être aussi un objet non-animé qui devient donc animé.



*Je suis une théière, voici mon anse et mon bec verseur.*

FIG. 4.6 – Transfert personnel appliqué à un objet non animé : la théière est “personnifiée”

A la différence des transferts situationnels, spécialisés dans des déplacements et des rapports de localisation, les transferts personnels sont à même d'exprimer l'action, dans le cours de son accomplissement. En transfert personnel, le locuteur “incarne” un sujet personnifié dont il est question dans le discours. En incarnant ce sujet, le locuteur rend compte, dans une certaine mesure, des caractéristiques physiques, comportementales et culturelles ainsi que de l'état interne (états d'esprit, affectif et mental) et physique de l'objet personnifié.

Le répertoire gestuel déployé par les transferts personnels est sans aucun doute illimité. De plus, son fondement illustratif rend le choix d'une gestuelle fortement dépendante du contexte du discours et de l'intention illustrative du locuteur. Il semble donc a priori impossible de modéliser de manière formelle

l'ensemble des mécanismes intervenant dans les situations liées au transferts personnels.

Si, par exemple, la représentation des traits de comportements de personnages incarnés semble impossible à décrire de façon systématique, d'autres caractéristiques, telles que l'état mental et physique peuvent être prises en compte et rendues lors de la génération du mouvement. Le fait est que la recherche dédiée aux agents conversationnels s'est employée à proposer des modélisations de l'état interne des agents et de ses conséquences sur leur expressivité. Une revue de ces recherches a été dressée en seconde partie de l'état de l'art 3.2. La figure 4.7 illustre l'organisation des classes d'objets impliqués lors d'un transfert personnel.

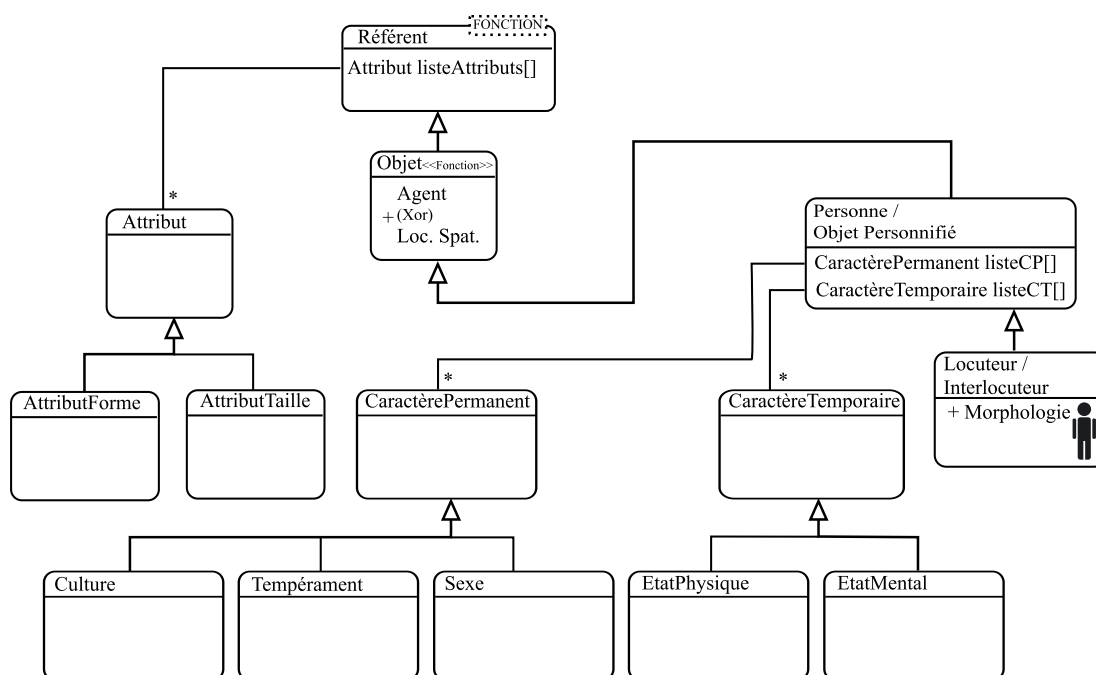


FIG. 4.7 – Diagramme de classe représentant les référents impliqués lors d'un transfert personnel et leurs attributs

Dans ce diagramme, nous avons choisi de considérer que le *réfèrent* de type *personne* est une extension du type *objet*. La spécificité de la *personne* est qu'elle possède des attributs ses caractéristiques de long terme (âge, propriétés physiques, tempérament, sexe, culture) et, d'autre part, des caractéristiques de moyen terme (état de fatigue, émotionnel, comportement) en plus des attributs de tailles et de formes présentes à un niveau de la classe *réfèrent*. Ces caractéristiques sont prises en compte lorsqu'un locuteur réalise un transfert personnel, c'est à dire lorsqu'il incarne une personne ou un objet. Le fait de considérer



le locuteur comme une spécialisation de la classe *personne* permet de l'intégrer comme une entité à part entière du discours. Cette généralisation de la notion de personne s'avère particulièrement utile dans un contexte de dialogue, où plusieurs interlocuteurs vont partager le même contexte d'énonciation.

Le transfert personnel "en état pur" décrit précédemment connaît quelques variantes dans le discours en langue signée. Par exemple, lorsqu'un locuteur en transfert personnel a recours à un transfert de situation, le locuteur se trouve en situation de double transfert personnel. Par contre, quand un locuteur a recours, brièvement, à un transfert personnel pour évoquer un objet ne participant pas à l'action (procès) de l'énoncé, on parle de *pseudo transfert* (la première image de la théière peut représenter un pseudo-transfert personnel). Quand un locuteur a recours à des stéréotypes empruntés à la culture environnante (entendante), le plus souvent pour représenter finement l'état d'esprit du personnage incarné (se gratter la tête pour exprimer la perplexité), il s'agit d'un *pseudo-transfert personnel*. Cependant, lorsque le locuteur a recours à des mimiques et une gestuelle qualifiant une action réalisée ou une phrase composée de signes standards, nous sommes en présence d'un *semi-transfert personnel*.

L'exemple suivant illustre un semi-transfert personnel, les traits expressifs de CONTENT se retrouvent dans la reprise du substantif HOMME qui devient HOMME<sub>[CONTENT]</sub> mais aussi dans la suite des actions qu'il accomplit.

HOMME CONTENT HOMME<sub>[CONTENT]</sub>  
SE-PROMENER<sub>[CONTENT]</sub> BAVARDER<sub>[CONTENT]</sub> REGARDER<sub>[CONTENT]</sub>

*l'homme, content se promène, bavarde, regarde*

Le pseudo-transfert personnel a un rôle qualifiant qui se superpose à un énoncé initialement exécuté hors visée illustrative. Tel un complément de manière, il enrichit et qualifie le procès réalisé hors visée illustrative. Ces deux dernières sous-classes de transferts personnel nous intéressent particulièrement dans la mesure où elles permettent de caractériser l'état interne du personnage incarné.

### 4.1.3 Le contexte

*L'entité* est liée à un *référent*. Le référent est une classe abstraite qui représente une unité sémantique du discours et qui peut être spécialisé en *date*, *lieu*, *action* et *objet*. Chacune de ces spécialisations revêt une *fonction* dans le discours. Certaines fonctions sont capables de référencer un contexte énonciatif. Il s'agit des

fonctions de *localisation* soit spatiale, soit temporelle. Le référencement entre les *fonctions* de localisation et le *contexte énonciatif* permet de gérer les commutations de contextes qui sont susceptibles d'intervenir dans un récit narratif. Imaginons par exemple qu'un locuteur, après avoir instancié une localisation temporelle (le 11 septembre 2001), décide de narrer ce qu'il faisait ce jour-là. Il va être capable de se projeter dans le contexte pointé par le *référent locatif temporel* de l'entité [11 septembre], de réaliser une commutation de contexte et de narrer son expérience personnelle en ce jour (ce jour-là, sa voiture était en panne, etc.) (cf. corpus LS-COLIN [Ca01]).

Ainsi, chaque référent dont la fonction permet de référencer un nouveau contexte peut potentiellement être une voie vers un nouveau contexte énonciatif. Un tel mécanisme, permet un empilement de contextes gigognes. Lors du discours rapporté, catégorie de transfert personnel illustré par Sallandre [Sal03], on assiste également à une permutation de contexte.

## 4.2 Modèle global du discours

Le schéma 4.8 illustre une représentation générale du modèle de discours. Ce schéma représente les entités dont il est question en discours et les référents qu'elles représentent. Les référents occupent une fonction dans le discours. Ces fonctions peuvent avoir une valeur locative (spatiale ou temporelle) ou une valeur d'actant. Les différents attributs portés par les référents permettent de renseigner les mécanismes participants de la visée illustrative. Les attributs renseignent en effet des propriétés pertinentes des référents qui sont décrites dans le cadre de la visée illustrative. Les référents de localisation peuvent référencer un contexte particulier. La permutation de contexte est susceptible d'apparaître en visée illustrative lors d'un récit ou du discours rapporté [Sal03]. Le contexte de l'énonciation possède lui-même des attributs susceptibles d'influencer le discours, à savoir le genre discursif, l'environnement, le contexte social.

## 4.3 Influence du modèle de discours sur les paramètres expressifs

Dans cette section, nous nous intéressons à la manière de réaliser les gestes, autrement dit, quelle va être l'influence des attributs des référents sur la réalisation au niveau articulatoire et perceptif des gestes. La section précédente nous a permis d'identifier des mécanismes à même de participer à de telles variations. Ces mécanismes et les attributs qui sont à leur origine varient selon la

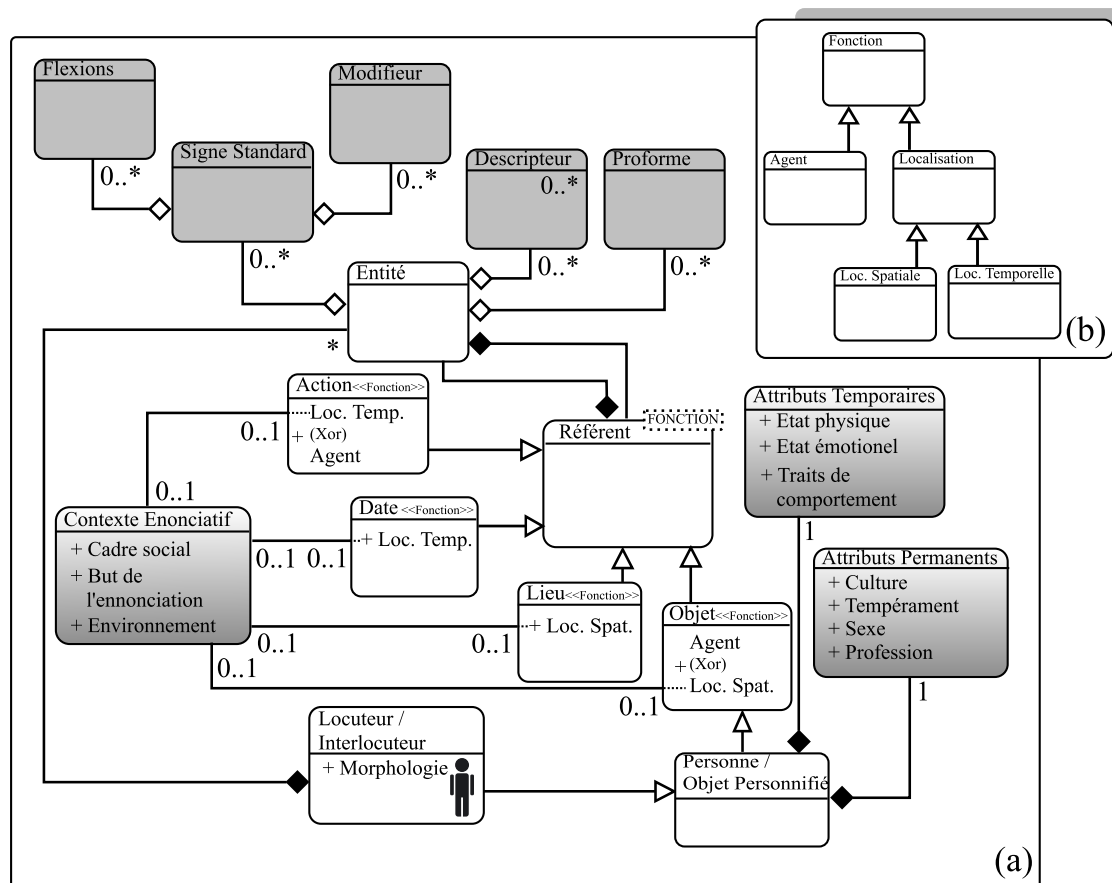


FIG. 4.8 – Diagramme de classes reprenant le formalisme de UML et représentant le modèle du discours en LS

catégorie employée. Nous avons vu que les variations sont portées par des traits. La difficulté vient de ce que l'expression de ces traits – donc des variations – est présente sur plusieurs niveaux linguistiques. Les variations ont donc une expression possible sur plusieurs niveaux. De plus, chaque niveau, indépendamment des autres, peut présenter une variation sur un même trait. Par exemple, l'amplitude peut être modifiée sur les niveaux d'attributs permanents, temporaires, de contexte énonciatif, prosodique et/ou syntaxique. Mais les niveaux sont organisés de manière hiérarchique de telle sorte que le niveau supérieur est prioritaire et impose une sorte d'étalon de la mesure de variation pour le niveau inférieur, l'ensemble fonctionnant de manière incrémentale. De plus, la longueur du segment affecté par une modification est plus longue pour un niveau donné que pour celui d'en dessous.

Le tableau 4.1 propose une première approche de cette articulation entre

variations et niveaux hiérarchiques, les variations portant sur deux phénomènes opposables : l'augmentation et la diminution de l'intensité de la présence d'un trait. Ce tableau est le fruit d'une réflexion menée conjointement par l'auteur et par Loïc Kervajan, doctorant en sciences du langage [HK07].

Cause	Exemple	Trait Expressif	Effet
Caractères permanents	Culture introvertie	Amplitude des gestes Rythme Regard	Moindre Moins marqué Moins investi
	Tempérament expansif	Amplitude des gestes Rythme	Plus dynamique Plus marqué
	Locuteur gaucher	Δ Structurelles	Mains inversées :
État Mental	Enthousiasme	Amplitude des gestes Profil temporel	Plus grande Plus rapide
	Étonnement	Sourcils Rythme	Relevés Geste suspendu
	Énervement	Profil temporel Rythme	Plus dynamique Plus marqué
État physique	Fatigue	Amplitude des gestes Rythme Profil temporel	Moindre Moins marqué Moins dynamique
Contexte énonciatif	Urgence	Amplitude des gestes Répétitions Profil temporel Rythme	plus grande Plus fréquentes Plus dynamique Plus marqué
	Cadre informel	Profil temporel Rythme	Moins dynamique Moins marqué
MouvementRelatif	Décollage Avion	Forme mouvement Profil vitesse	Selon trajectoire
Attribut taille	Transfert de taille	Amplitude du geste	Selon taille
Attribut forme	Transfert de forme	Forme du geste	Selon forme
Morpho-Syntaxe	Pluriel	Répétitions Amplitude des gestes Δ Structure articulaire	Dépend du signe

TAB. 4.1 – Tableau synthétique des causes et modifications

Ce tableau demeure une approximation, dans le sens où les éléments relevés restent à quantifier précisément les uns par rapport aux autres. Sur la ligne syntaxique, nous avons évoqué les modificateurs. Ils peuvent être réducteurs ou augmenteurs, affecter différents traits (spatiaux et temporels) et différents paramètres de formation des gestes (manuels et non manuels). Les modificateurs de taille et de volume affectent l'amplitude en jouant différemment sur les paramètres, les modificateurs de manière ou de quantité affectent l'amplitude et la répétition.

Ce tableau est donc le premier pas vers une typologie fine à réaliser croisant précisément les traits affectés et le type d'affectation (plus précis que "augmenté" vs "diminué") avec les niveaux concernés. Pour la syntaxe, cela demande, d'une part, une étude relevant les comportements induits par tous les modificateurs possibles et, d'autre part, une étude sur la catégorisation des effets de

ces modificateurs (infiniment peu > un petit peu > un peu > normalement > un peu plus > beaucoup > infiniment).



## Chapitre 5

# Modèle de Spécification des Gestes de la Langue des Signes

Le chapitre précédent nous a permis de représenter un modèle du discours à haut niveau d'abstraction dans lequel s'agencent des entités constitutives du discours. La sélection et l'agencement des signifiants gestuels relatifs aux entités mises en jeu dans le discours est réalisée par un processus de détermination sémantico-cognitif dont le cadre dépasse la portée de cette thèse [BL06]. Cependant, nous avons jugé pertinent de prendre en compte les différentes spécialisations des référents que nous rappelons dans l'illustration de la figure 5.1.

Notre problématique est la suivante : il s'agit de proposer un modèle de spécification des gestes de la langue des signes qui soit capable de décrire de manière concise et non-ambigüe les mouvements participant à la construction des référents des entités intervenant dans un énoncé. Le point de départ de notre modèle de spécification est donc une suite de référents, accompagnés de leurs flexions éventuelles et des relations qui lient les entités auxquels les référents se rapportent. Notre modèle doit également être capable de spécifier des gestes de transitions ne participant pas directement à la sémantique du message produit. Ces gestes sont identifiés comme des phases gestuelles non porteuses de sens par la littérature dédiée à la langue des signes et à l'étude des gestes co-verbaux [Nei92a, KvGvdH98].

L'apport d'un modèle de spécification du geste de la langue des signes est de deux ordres. D'une part, il fournit une description non ambigüe des gestes. D'autre part, la représentation formelle de leurs paramètres permet de définir des opérateurs susceptibles de modifier des gestes préalablement spécifiés. La possibilité d'appliquer ces modificateurs offre la possibilité de rendre compte de la grande variabilité des gestes de la langue des signes. Deux grands principes guident notre système de spécification des gestes constitutifs

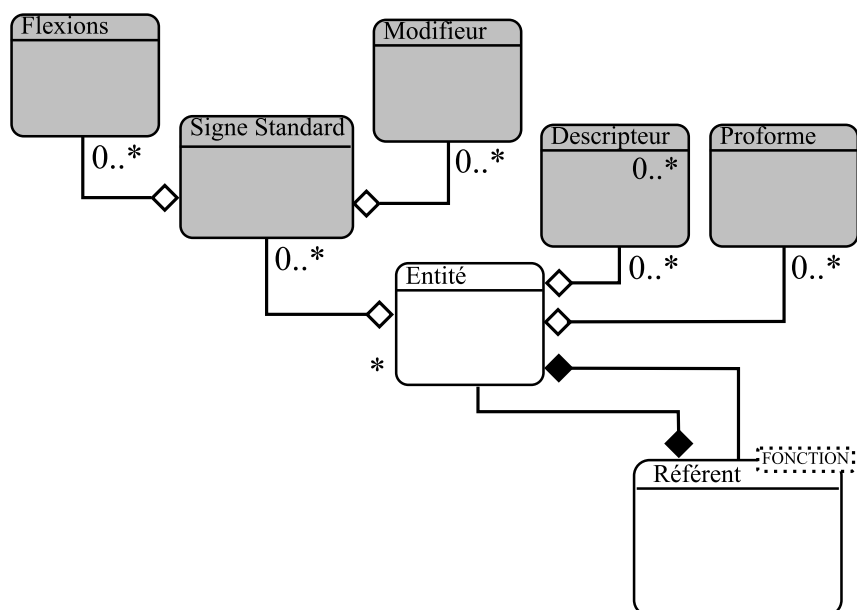


FIG. 5.1 – Relation entre l'entité, son référent et les signes associés

des énoncés en langue des signes. Premièrement, notre modèle de description doit répondre aux objectifs de concision et de compacité : il doit comporter le jeu de descripteurs le plus réduit possible tout en garantissant des descriptions précises. Deuxièmement, il doit attacher de l'importance à la prise en compte de l'aspect variationnel des gestes dans le cadre de la langue des signes évoqué au chapitre précédent. C'est la prise en compte, en amont, de la spécification de l'aspect variationnel de la langue des signes sur le plan temporel et spatial qui constitue l'originalité du modèle que nous proposons.

Une première section dresse une typologie de l'ensemble des gestes gérés par le système de spécification. L'identification des différents types de gestes se base sur des études dédiées à l'étude du geste expressif [Ken93] et des différentes phases intervenant dans la production de la langue signée [KvGvdH98, BP94]. Une seconde section est dédiée à une présentation générale du modèle de spécification des gestes. Cette section présente le concept central de notre modèle, c'est à dire le découplage entre les paramètres statiques et les paramètres dynamiques des gestes mis en jeu dans la production de séquences en langue des signes, ainsi que ses fondements théoriques. Une troisième section est consacrée à la spécification des paramètres statiques constitutifs d'un geste. Nous verrons que ces paramètres statiques peuvent s'exprimer à l'aide de contraintes que nous expliciterons. La quatrième section est dédiée à la spécification des paramètres dynamiques des gestes. Ces paramètres sont présen-



tés de manière hiérarchique en fonction de l'influence des paramètres sur le déplacement final. Dans une cinquième section, nous présenterons quelques opérateurs de modification du mouvement ainsi que la façon dont ils modifient le signe de base à l'aide de quelques exemples représentatifs. Dans cette section, nous soulèverons des problèmes liés à la synthèse qui seront traités dans le chapitre 6.

## 5.1 Typologie des gestes

Cette première section a vocation à identifier le segment minimal décrit par notre modèle de spécification. Pour ce faire, nous nous appuyons sur les travaux réalisés dans le domaine de l'étude des gestes de communication [Nei92a, Ken93]. Ces travaux se sont basés sur l'étude des séquences gestuelles dans leur globalité, à l'opposé des études dédiées aux langues de signes qui se sont concentrées sur le caractère lexical des signes en se concentrant sur l'opposition entre les parties dynamiques et les parties statiques intervenant lors de l'exécution des signes standards [Lid90]. En considérant la production de signes comme un flux continu d'informations, les chercheurs étudiant le geste ont identifié des phases de mouvement qualitativement différentes [Ken93]. Kita [KvGvdH98] en propose une organisation illustrée par la figure 5.4. Quatre phases principales ont ainsi été distinguées : les phases de préparation, de rétraction, de tenue et de *stroke*<sup>1</sup>. D'un point de vue fonctionnel, les phases gestuelles se distinguent par leur portée signifiante. C'est cette distinction qui guide la description des phases dans les sous-sections suivantes. Les phases gestuelles s'agencent pour former des syntagmes gestuels, c'est à dire, un ensemble de gestes formant une seule unité fonctionnelle. Une succession de syntagmes permet de construire une phrase. Cette organisation est reprise dans l'illustration 5.4.

### 5.1.1 Phases gestuelles non porteuses de sens

A la fois dans la littérature dédiée à l'étude des gestes co-verbaux et celle consacrée à la langue des signes, il existe une distinction forte entre deux types de gestes : les gestes porteurs de sens et ceux non porteurs de sens. Cette sous-section traite des gestes non porteurs de sens qui interviennent dans la construction d'une phrase gestuelle. Pour les phases de préparation et de rétraction, l'objectif se concentre sur une configuration à atteindre : soit la configu-

---

<sup>1</sup>*Stroke* est un terme anglais directement issu de la typologie de Mc Neill pour lequel nous n'avons pas trouvé de traduction satisfaisante. Le terme Stroke est défini au paragraphe 5.1.2.1

ration de départ de la phase suivante, soit une configuration neutre ou une configuration de repos.

### 5.1.1.1 Phases de préparation

Les phases de préparation permettent de placer dans la position initiale d'un geste les articulateurs mis en jeu dans la production d'un geste. Bien que ces phases aient été identifiées par des travaux dédiés à l'étude des gestes conversationnels, elles n'ont, à notre connaissance, été que peu décrites par la littérature consacrée à la langue des signes [KvGvdH98, VM04, Jou95]. L'introduction d'une phase spécifique liée à la préparation avant l'exécution d'un signe limite l'influence de la coarticulation sur le signe à réaliser et facilite sa compréhension. La figure 5.2 illustre une phase de préparation entre deux signes standards :

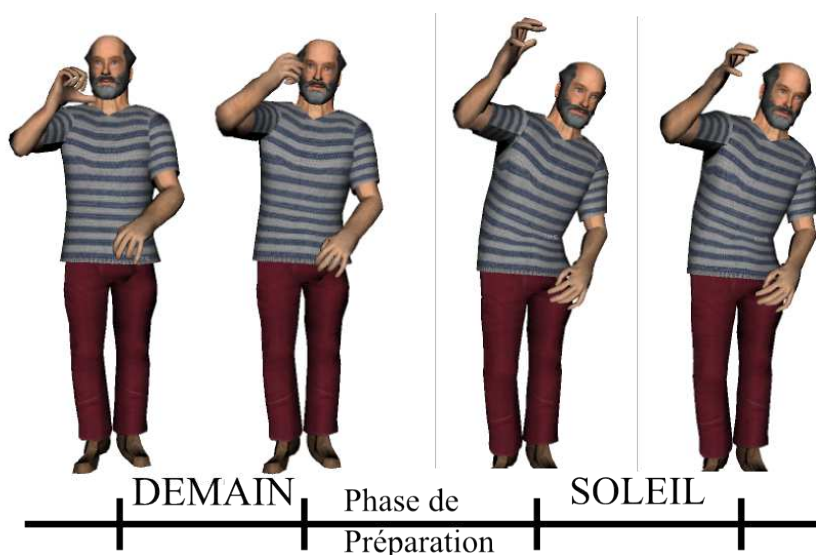


FIG. 5.2 – Illustration d'une phase de préparation : «Demain, soleil (il y aura)»

La spécification des phases de préparation est directe et s'appuie sur la dernière posture de la phase précédent la préparation et la première posture de la phase suivant la préparation.

### 5.1.1.2 Phases de rétraction

Les phases de rétraction permettent de placer les articulateurs mis en jeu dans la production d'un geste dans une position neutre, ou de repos. Au cours de l'énonciation, entre deux signes, la phase de rétraction peut être fortement

tronquée. On parle alors de *rétraction partielle*. En langue des signes, les phrases sont en général conclues par des phases de rétraction. Au sein d'une unité gestuelle, entre deux phrases, les rétractions sont en général réalisées de façon partielle. La figure 5.3 illustre une rétraction partielle intervenant entre deux signes au sein d'une phrase et une rétraction complète à la fin d'une phrase :

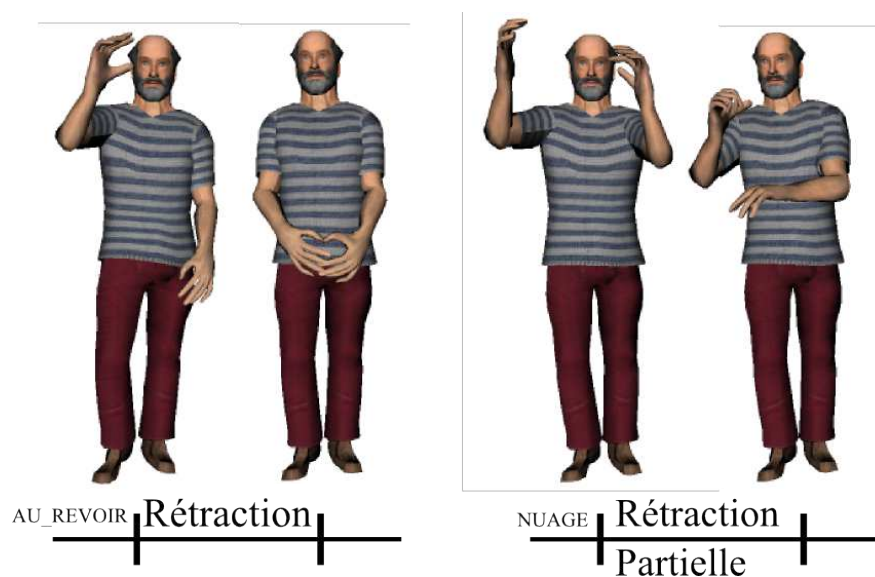


FIG. 5.3 – Illustration d'une phase de rétraction partielle (NUAGES) et d'une phase de rétraction totale en fin de séquence (AU\_REVOIR)

La spécification des phases de rétraction s'appuie sur la dernière posture de la phrase précédente et sur une posture neutre, ou de repos qui est atteinte si la phase de rétraction est exécutée entièrement.

### 5.1.1.3 Phases de tenue

Les phases de tenue encadrent les phases de stroke. Les phases de tenue ont été identifiées dans le cadre des recherches dédiées aux gestes co-verbaux [Kit90] et celles consacrées aux gestes de la langue des signes, à la fois d'un point de vue syntaxique [Gro79], phonologique [Lid93] et plus global [Cux00]. D'un point de vue fonctionnel, ces phases de tenue, très courtes (de l'ordre d'un dixième de seconde), facilitent la perception des phases expressives et améliorent la compréhension des gestes. La spécification des phases de tenue est également directe : il s'agit de spécifier un intervalle temporel pendant lequel la configuration articulaire courante est maintenue.

## 5.1.2 Phases gestuelles porteuses de sens

Les phases gestuelles porteuses de sens permettent l'identification des signes au cours de leur exécution. Elles sont caractérisées, au niveau qualitatif, par un effort plus prononcé, au sens de la notation de Laban [Lab88, KvGvdH98].

### 5.1.2.1 Phases de strokes

Les phases de strokes contiennent des gestes qui comportent une composante dynamique agissant sur les articulateurs. Les gestes dynamiques se distinguent des phases de préparation et de rétraction par leur intention. En effet, lors des strokes, la tâche est spécifiée par un élément dynamique, une trajectoire à suivre dans l'espace, un mouvement de supination/pronation ou un changement de configuration manuelle à respecter. Parmi les gestes dynamiques, on distingue les gestes simples, qui ne présentent qu'un changement de configuration unique, et les gestes complexes, qui présentent plusieurs changements successifs.

### 5.1.2.2 Phases de tenue

Les tenues peuvent également être porteuses de sens : en effet, certains signes ne présentent aucun mouvement. Ce fait a d'abord été identifié par des travaux consacrés à la représentation et la reconnaissance de la langue des signes [VG04, Lid89], puis au sein d'études dédiées aux gestes co-verbaux [Kit90].

## 5.1.3 Règles d'organisation entre phases gestuelles

Nous proposons la spécification formelle de l'organisation des gestes élémentaires au sein des phrases puis des séquences gestuelles à l'aide de la forme étendue de Backus-Naur [ML86]. Cette description se base sur la proposition de [KvGvdH98]. Le schéma de la figure 5.4 propose une description visuelle de l'organisation de ces éléments.

$$\langle \textit{Phrase Gestuelle} \rangle ::= \langle \textit{Syntagme Gestuel} \rangle \{ \langle \textit{Syntagme Gestuel} \rangle \}$$

$$\langle \textit{Syntagme Gestuel} \rangle ::= (\langle \textit{Préparation} \rangle) \langle \textit{Phase Expressive} \rangle (\langle \textit{Rétraction} \rangle)$$

$$\langle \textit{Phase Expressive} \rangle ::= (\langle \textit{tenue} \rangle) \langle \textit{Geste Dynamique} \rangle (\langle \textit{tenue} \rangle)$$

$$\langle \textit{Phase Expressive} \rangle ::= \langle \textit{tenue} \rangle$$

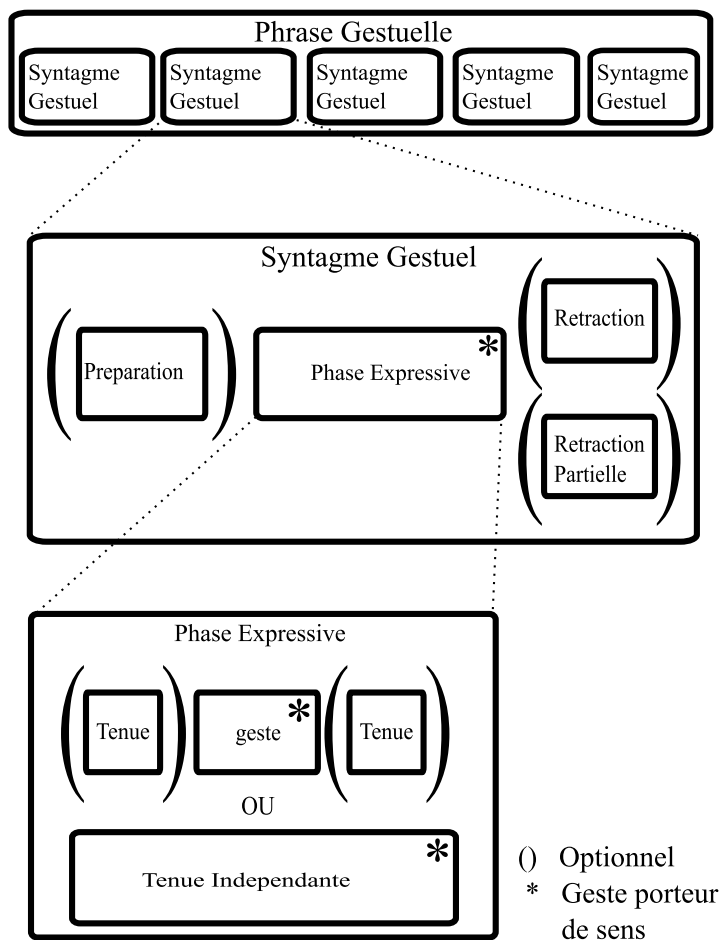


FIG. 5.4 – Organisation de la phrase gestuelle et décomposition en phases

## 5.2 Spécification du geste

La section précédente a présenté une organisation de la structure d'une séquence gestuelle dérivée des travaux dédiés aux gestes expressifs [Nei92b, KvGvdH98]. La plus petite unité constitutive de la séquence gestuelle a été identifiée comme la phase gestuelle. Cette section présente le modèle de spécification des phases du geste. Ce modèle est destiné à proposer une description concise et précise des gestes, tout en respectant leur variabilité, à la fois spatiale et temporelle.

### 5.2.1 Apports d'un modèle gestétique

Les modèles phonologiques, tels qu'ils ont été décrits [Lid89] permettent de prendre en compte les paramètres formationnels du geste et permettent sa resynthèse dans des modèles d'animation par ordinateur [LG99, Los00, Ken01]. Les modèles de spécification existants reposent sur une description phonologique des gestes de la langue des signes, qui convient, dans une certaine mesure, à la génération d'assemblages lexicaux. Ces modèles se concentrent sur la description du signe, et, à l'exception de Losson, [Los00] qui spécifie des opérateurs syntaxiques modifiant des signes standards. Ils ne prennent pas en compte la variabilité susceptible d'intervenir dans la production des gestes.

Notre objectif est donc de proposer un langage de spécification du geste qui soit capable de rendre compte des opérations de bas niveau (niveau articulatoire) qui interviennent dans la génération des gestes expressifs et qui soit capable de restituer les variations intervenant à des niveaux élevés, notamment en ce qui concerne les structures de grande iconicité. Le modèle que nous proposons s'inspire des travaux de Diane Brentari [Bre98]. Le modèle théorique sur lequel nous nous basons [Bre98] propose une description articulatoire des signes architecturée autour des travaux de Liddel [Lid89]. Nous pensons donc que la facette dédiée à la description articulatoire du geste peut nous aider à établir une description complète et précise des phases gestuelles mises en jeu en langue des signes quelque soit la visée adoptée par le locuteur.

### 5.2.2 Séparation entre paramètres statiques et dynamiques

L'idée maîtresse de notre modèle de spécification des gestes est la séparation des caractéristiques de la description en deux branches : la première branche rassemble les paramètres stables lors de la production d'un geste et la seconde branche rassemble les paramètres dynamiques impliqués lors de la production d'un signe. Par exemple, le signe standard illustré à la figure 5.5 présente un paramètre dynamique (flexion au niveau du coude). Les autres paramètres (configuration et orientation de la main dominante) sont statiques : ils ne changent pas au cours de l'exécution du mouvement.

Suivant cette philosophie, nous pouvons décrire un signe en séparant, d'une part, les attributs statiques au cours de la production et, d'autre part, les attributs dynamiques au cours de la production. Une vision schématique de notre modèle est proposée à l'illustration 5.6. Nous allons maintenant nous attacher à la description de la branche regroupant les caractéristiques intrinsèques des signes.



FIG. 5.5 – Le signe standard chewing-Gum

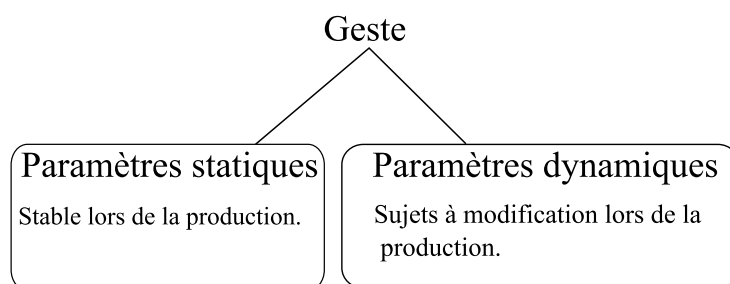


FIG. 5.6 – Vision générale du modèle prosodique de la langue des signes

### 5.3 Paramètres statiques

La branche dédiée à la description des paramètres statiques peut se résumer à l'aide du schéma 5.7.

Le premier niveau hiérarchique au sein de la branche des caractéristiques intrinsèques regroupe les caractéristiques de localisation de l'exécution du geste et sa configuration de base. La branche dédiée aux articulateurs regroupe les caractéristiques des bras, des mains impliqués dans la production d'un signe. La main est certainement le plus complexe des articulateurs. De ce fait, elle occupe une structure complexe dans notre modèle de spécification.

Les sous-sections suivantes décrivent chaque composant de la branche des paramètres stables.

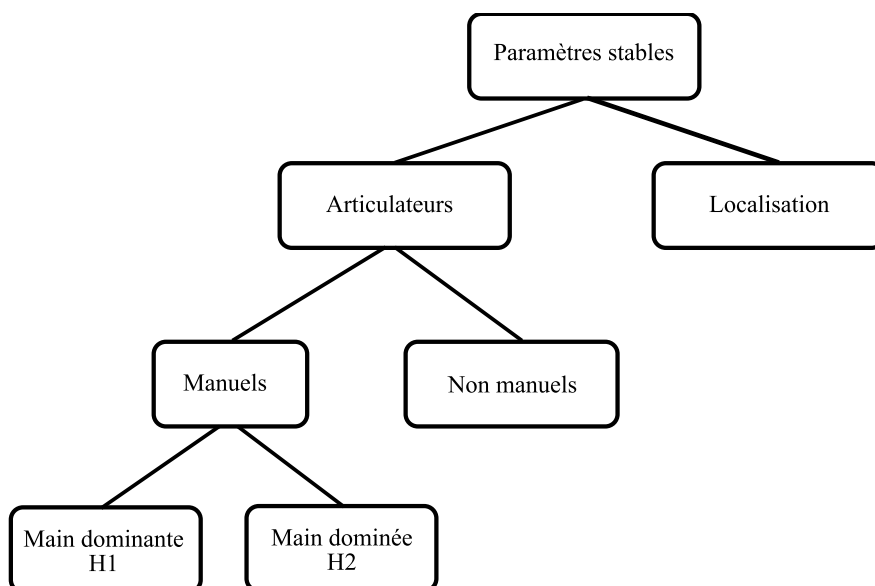


FIG. 5.7 – Organisation de la branche des paramètres statiques au cours de la production d'un geste

### 5.3.1 Paramètres non-manuels

Les articulateurs non manuels regroupent les mimiques faciales, la direction du regard, les mouvements du cou, les haussements d'épaules et l'orientation du buste.

### 5.3.2 Paramètres manuels

La branche des articulateurs manuels regroupe la main dominante et la main dominée, respectivement nommées H1 et H2. Nous verrons par la suite que pour la majorité des signes, il existe une dépendance forte entre la main dominée et la main dominante et que cette dépendance peut être modélisée par notre description à l'aide d'un jeu réduit de contraintes.

Lorsqu'un signe est bi-manuel, il fait apparaître la structure H2. La branche H2 peut regrouper deux caractéristiques : la localisation place et la configuration configuration. Selon le mouvement, H2 a deux alternatives : copier le mouvement de H1 (avec un déphasage éventuel) ou ne pas bouger.

#### 5.3.2.1 Configuration manuelle

Leur nombre et leur nature varient suivant les auteurs. Stokoe en a répertorié 19, Klima et Bellugi 20. Liddel et Johnson en font état de 150. En LSF,



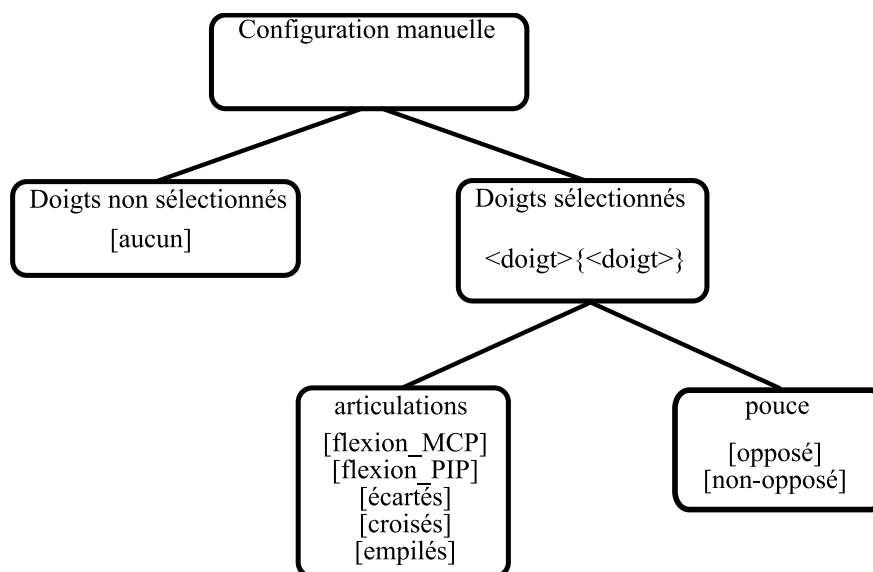


FIG. 5.8 – Organisation des paramètres stables pour la configuration manuelle

Braffort [ACT94] pose 139 configurations dont 55 comme les principales configurations statiques. Un large éventail de configurations manuelles peuvent être générées par un processus de sélection (de doigts et pouces) + modificateur. L'introduction des doigts sélectionnés permet, d'une part, de distinguer les parties dynamiques de la main lors de la production du geste des parties statiques. Les doigts sélectionnés sont en effet les doigts sujets au mouvement lors de la production du signe. D'autre part, cette spécification des gestes permet une description concise et non ambiguë des configurations manuelles.

**Doigts non sélectionnés** Les doigts non sélectionnés sont ceux qui ne subissent pas de modification lors du mouvement. Les doigts non sélectionnés peuvent ainsi être fléchis ou tendus. Par défaut, la configuration des doigts non sélectionnés est fléchie.

**Doigts sélectionnés** Le groupe de doigts sélectionnés rassemble deux types d'informations : la spécification de leurs valeurs de flexion selon leurs articulations et une liste de doigts effectivement sélectionnés.

**Articulations** Le groupe des articulations des doigts sélectionnés est destiné à spécifier la configuration des articulations des doigts appartenant au groupe des doigts sélectionnés. Trois articulations sont mises en jeu pour chaque doigt : l'articulation métacarpo-phalangiennes, régie par le paramètre <MCP> ; les ar-

ticulations, interphalangiennes proximales et interphalangiennes distales, spécifiées à l'aide du paramètre <PIP>, sachant qu'une relation linéaire lie les valeurs angulaires de flexions interphalangiennes distales et proximales :  $\phi_{DIP} = \frac{2}{3}\phi_{PIP}$ . Sans spécification, les valeurs aux articulations sont tendues. Une description détaillée des articulation de la main est présentée en annexe.

**Pouce** Il est nécessaire de dissocier le pouce des autres doigts sélectionnés au vu des propriétés articulaires du pouce par rapport aux autres doigts. Ce groupe permet de spécifier un paramètre de contact du pouce par rapport aux autres doigts.

La figure 5.9 illustre quelques configurations manuelles décrites par notre modèle.

### 5.3.3 Emplacement du geste

La branche dédiée à la localisation permet de spécifier la position de la main par rapport à l'agent. Dans notre modèle, la localisation est spécifiée par les paramètres suivants : un point dans l'espace de signation, un vecteur normal à un plan ou un couple point/ normale. La figure 5.11 illustre la structure spécifiant l'emplacement d'un geste.

#### 5.3.3.1 Plans

L'introduction d'un paramètre de type plan, caractérisé par sa normale dans le modèle de description, permet d'introduire un lieu dans lequel le geste est exécuté. Selon les attributs dynamiques du mouvement associé au signes, le mouvement peut être inscrit dans le plan ou perpendiculaire au plan spécifié. Nous verrons dans la sous-section suivante que le plan entre en jeu dans la spécification implicite de l'orientation de la main. Le plan n'étant défini que par la direction de sa normale, il est nécessaire de l'ancrer dans l'espace de signation du signeur. Pour ce faire, nous spécifions un point d'attache, défini dans l'espace entourant le signeur à l'aide de coordonnées cartésiennes.

#### 5.3.3.2 Zones de contact

La spécification d'un point d'attache de manière absolue, à l'aide de coordonnées de  $\mathbb{R}^3$ , n'est pas adaptée à la spécification des points appartenant au corps ou au visage du signeur. La position de ces points dépend en effet de la configuration de la structure articulaire du signeur. Nous reprenons les zones de contact définies dans [Bre98], chaque zone appartenant à un segment du corps du signeur. Quatre segments sont ainsi définis :

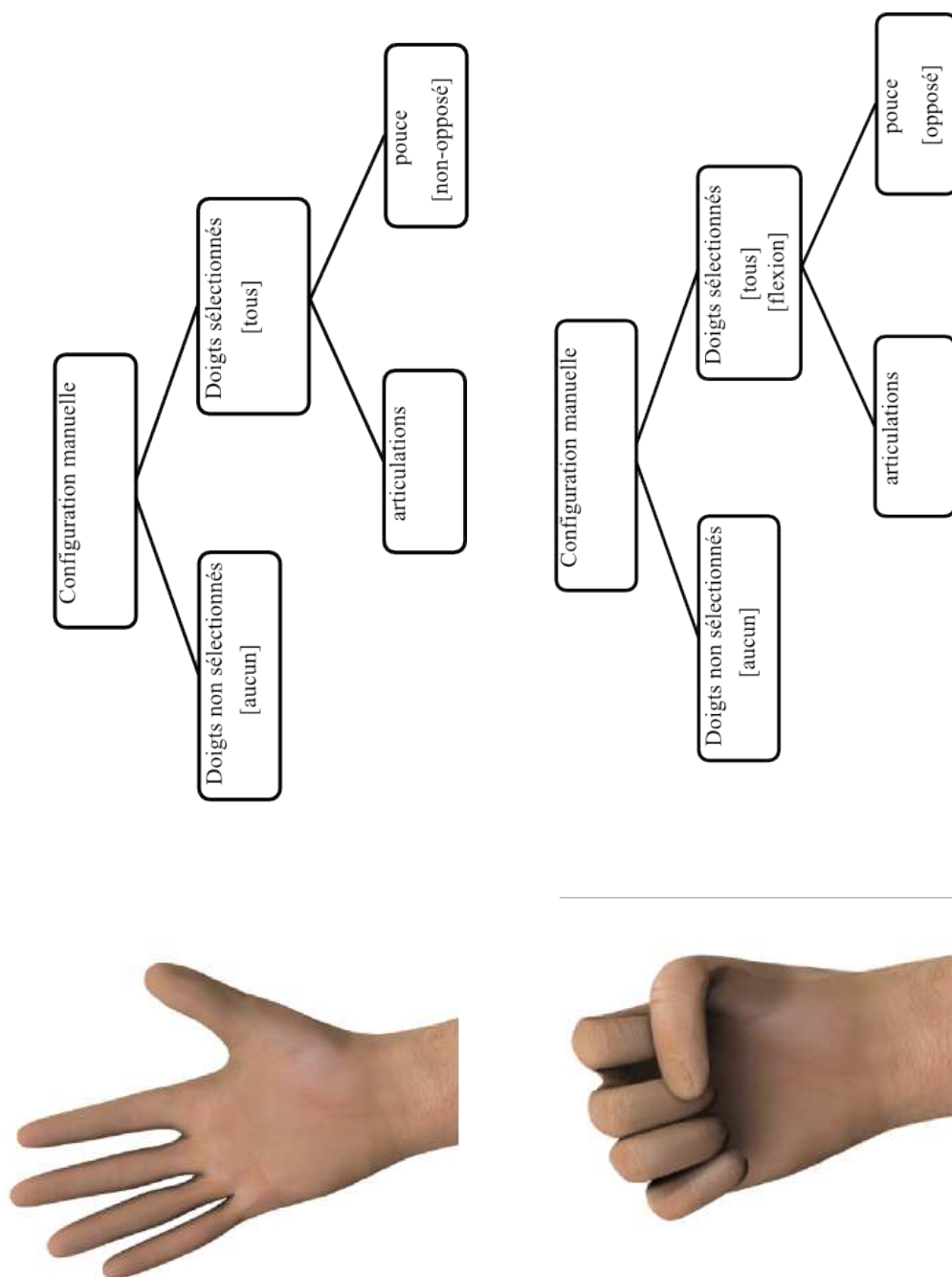


FIG. 5.9 – Quelques configurations décrites par le modèle

1. Tête;
2. Corps;

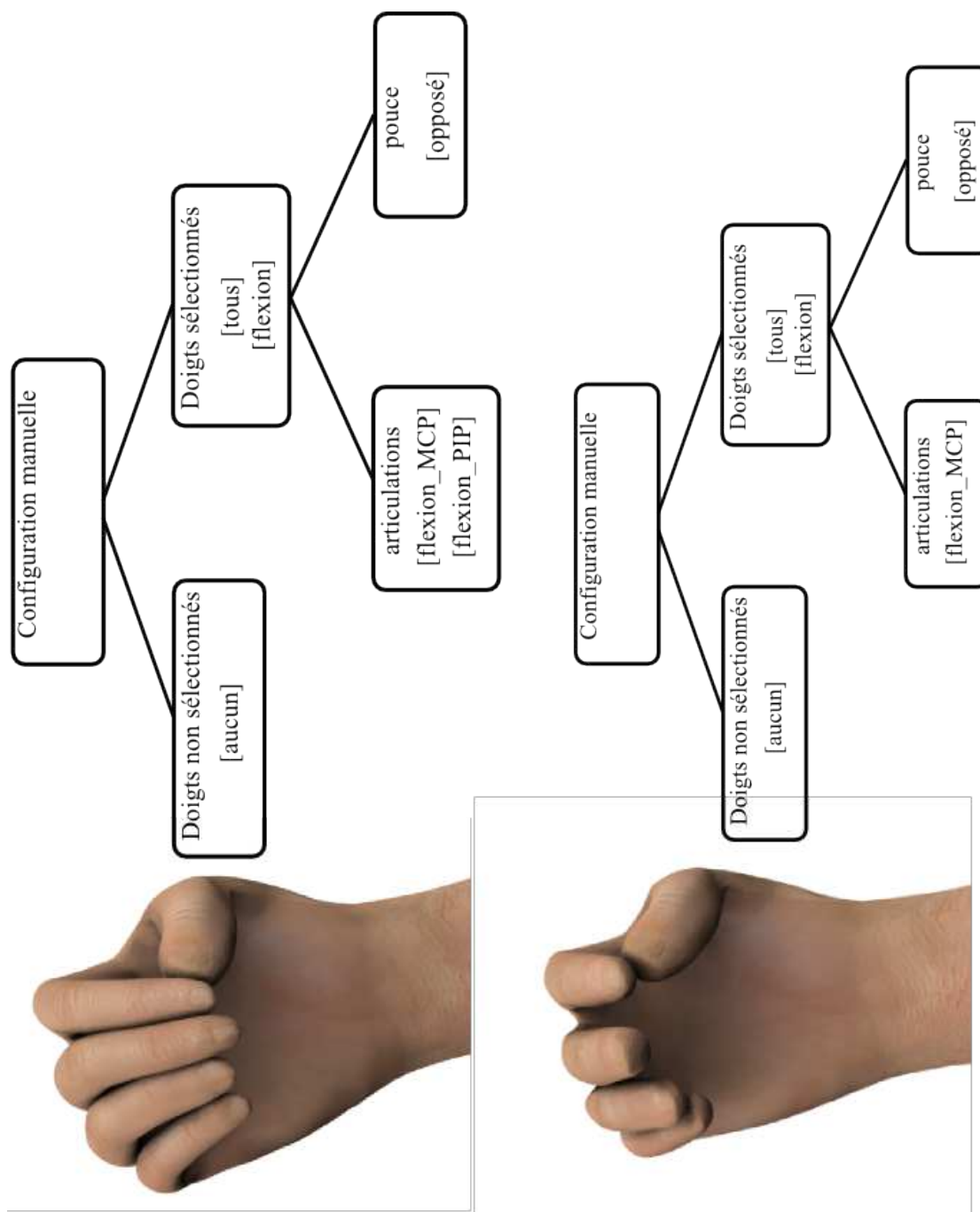


FIG. 5.10 – Quelques configurations décrites par le modèle

3. Bras;
4. Main.

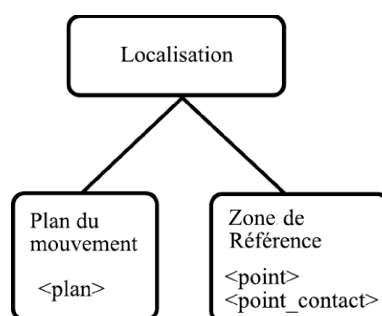


FIG. 5.11 – Modélisation de la localisation

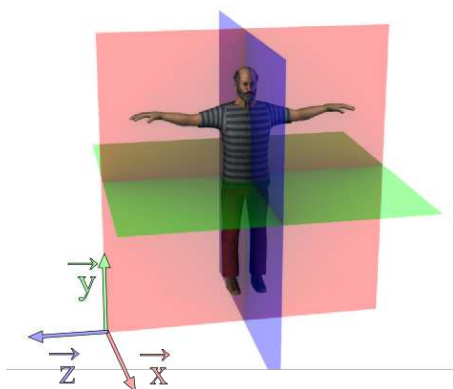


FIG. 5.12 – Plans de l'espace de signation

Chaque segment est lui-même découpé en huit zones de contact, ce qui donne quarante zones de contact pour le corps du signeur. D'après Cuxac, la LSF comporte une soixantaine d'emplacements pertinents [Cux83], ce qui nous inciterait à l'avenir à raffiner la décomposition des zones. Les segments et les zones de contact associées sont repris en annexes. A titre illustratif, nous reprenons, à la figure 5.13 les zones de contact définies pour la main.

### 5.3.4 Modélisation de l'orientation de la main

L'orientation porte sur un aspect légèrement différent des articulateurs classiques en langue des signes. Si Stokoe n'a pas choisi de reconnaître l'orientation comme l'un des paramètres formationnels des signes, il a été reconnu par Battison [Bat78]. Nous pensons que, conformément au modèle de Brentari [Bre98], l'orientation est en mesure de produire une différenciation lexicale entre deux signes, celle-ci n'ayant guère besoin d'être exprimée par un paramètre dédié. En d'autres termes, l'orientation doit cesser d'être considérée comme un paramètre phonologique, mais plutôt comme la conséquence d'un agence-



FIG. 5.13 – Les huit emplacements définis pour la main

ment de contraintes articulaires. Il est alors possible d'exprimer une orientation manuelle comme la conséquence d'une relation entre des emplacements sur la main et des contraintes de position localisés sur le corps ou dans des plans de l'espace de signation.

La figure 5.14 illustre comment l'interaction entre le plan et la localisation font émerger une configuration unique pour la main. Les deux spécifications présentées respectivement pour les signes [VIEUX] et [GLACE] ne diffèrent en effet que par le plan par rapport auquel s'exécute le signe. A la différence des précédents modèles [Los00, LG99], l'orientation des mains n'est donc pas considéré comme un paramètre formationnel à part entière mais comme la conséquence d'un jeu de contraintes. Nous verrons au chapitre 7 que cette gestion implicite de l'orientation permet d'obtenir des poses plus naturelles respectant les contraintes physiologiques du signeur.

#### 5.3.4.1 Contraintes géométriques

Beaucoup de signes présentent des coordinations fortes entre la main dominante et la main dominée. Par exemple, l'attribut [Alternating] permet de spécifier le déphasage du mouvement du bras de la main dominée par rapport

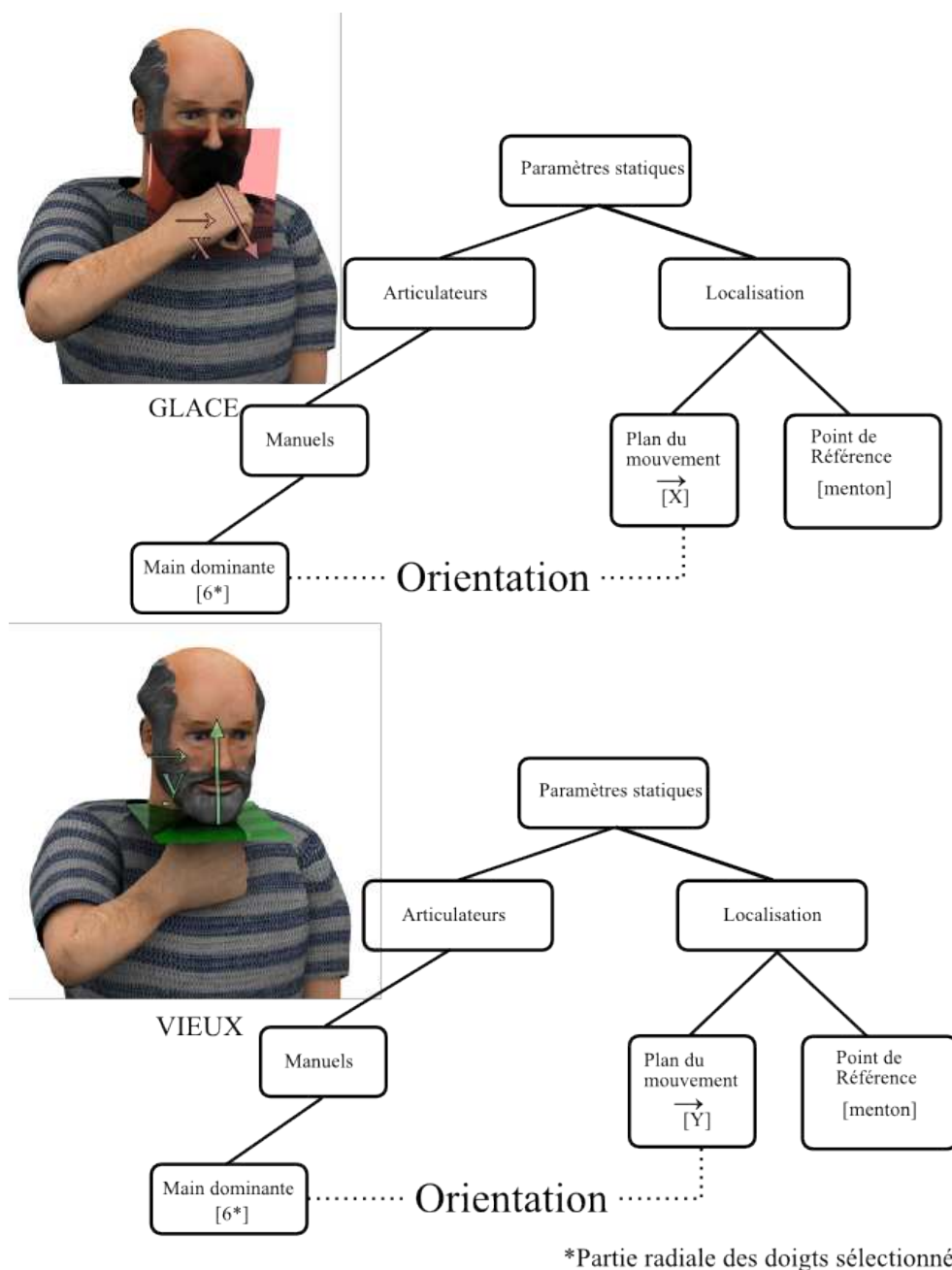


FIG. 5.14 – L’orientation définie comme une relation entre le vecteur normal à une zone de contact et une localisation ou un plan d’articulation

à celui de la main dominante dans un signe bimanuel. De même, l’attribut de symétrie [Symetrical] rend compte de l’aspect symétrique du mouvement des bras lors de l’exécution d’un signe bimanuel.

## 5.4 Paramètres dynamiques

La branche des paramètres dynamiques est la branche regroupant les propriétés d'un geste qui varient au cours de sa production. La séparation entre les paramètres statiques et dynamiques permet de mettre en relief l'importance du mouvement dans les gestes expressifs et permet de gérer de manière pertinente les variations susceptibles d'apparaître au cours du discours.

### 5.4.1 Organisation des paramètres

La branche qui caractérise les paramètres dynamiques de notre modèle ordonne les paramètres de manière hiérarchique, selon le nombre de degrés de libertés influencés. Ainsi, un paramètre ayant une influence en amont de la structure articulaire de l'avatar sera placé en amont dans la branche consacrée aux paramètres dynamiques. La figure 5.15 reprend cette organisation.

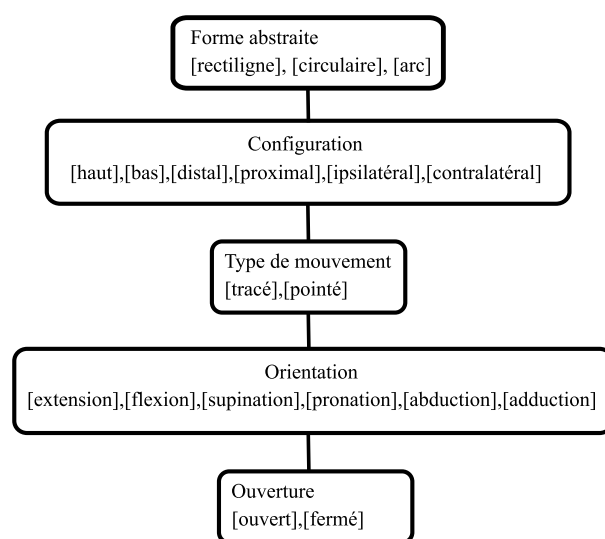


FIG. 5.15 – Hiérarchie des caractéristiques prosodiques du modèle de description

### 5.4.2 Gestes tracés

Les gestes de tracé décrivent une trajectoire inscrite dans un plan idéalisé de l'espace de signation. Notre modèle spécifie les gestes de tracé selon une trajectoire adoptée par les gestes. Cette trajectoire peut revêtir différentes formes :

1. Trajectoire rectiligne ;



2. Trajectoire curviligne ;
3. Trajectoire circulaire.

Les gestes de tracé ont la particularité d'être inscrits dans le plan du mouvement spécifié par les paramètres statiques comme spécifié au paragraphe 5.3.3.1.

### 5.4.3 Gestes de pointage

Les gestes de pointage décrivent une trajectoire orientée selon la normale à un plan idéalisé dans l'espace de signation. Cela souligne l'aspect directif des gestes de pointage. Par conséquent, seules deux trajectoires s'appliquent aux gestes de pointage :

1. Trajectoire rectiligne ;
2. Trajectoire curviligne.

Les gestes de pointage sont par conséquent orientés selon la normale du plan du mouvement spécifié dans les paramètres statiques comme spécifié au paragraphe 5.3.3.1.

#### 5.4.3.1 Illustration

Reprenons les deux gestes que nous avons utilisés à titre d'exemple lors de la spécification des paramètres statiques dans le paragraphe 5.14. Ces gestes correspondent aux signes standards [VIEUX] et [GLACE]. Nous avons vu que la seule différence au niveau de la spécification des paramètres statiques se situe au niveau du plan d'exécution du mouvement du geste. Attardons-nous maintenant aux paramètres dynamiques de [VIEUX] et [GLACE]. Ces paramètres dynamiques se situent en amont de la branche des paramètres dynamiques. En effet, le mouvement décrit lors de la réalisation de ces deux signes met en jeu les articulations du coude et de l'épaule. Les mouvements impliqués dans la production des signes [VIEUX] et [GLACE] ont tous deux une trajectoire rectiligne. En revanche, le signe [GLACE] est tracé : le mouvement décrit par la main devant le visage du signeur s'inscrit dans le plan défini par la normale  $\vec{[X]}$ , à l'inverse, [VIEUX] est pointé : le mouvement décrit par la main lors de sa production est orienté selon la normale  $\vec{[Y]}$  du plan spécifié dans ses paramètres statiques.

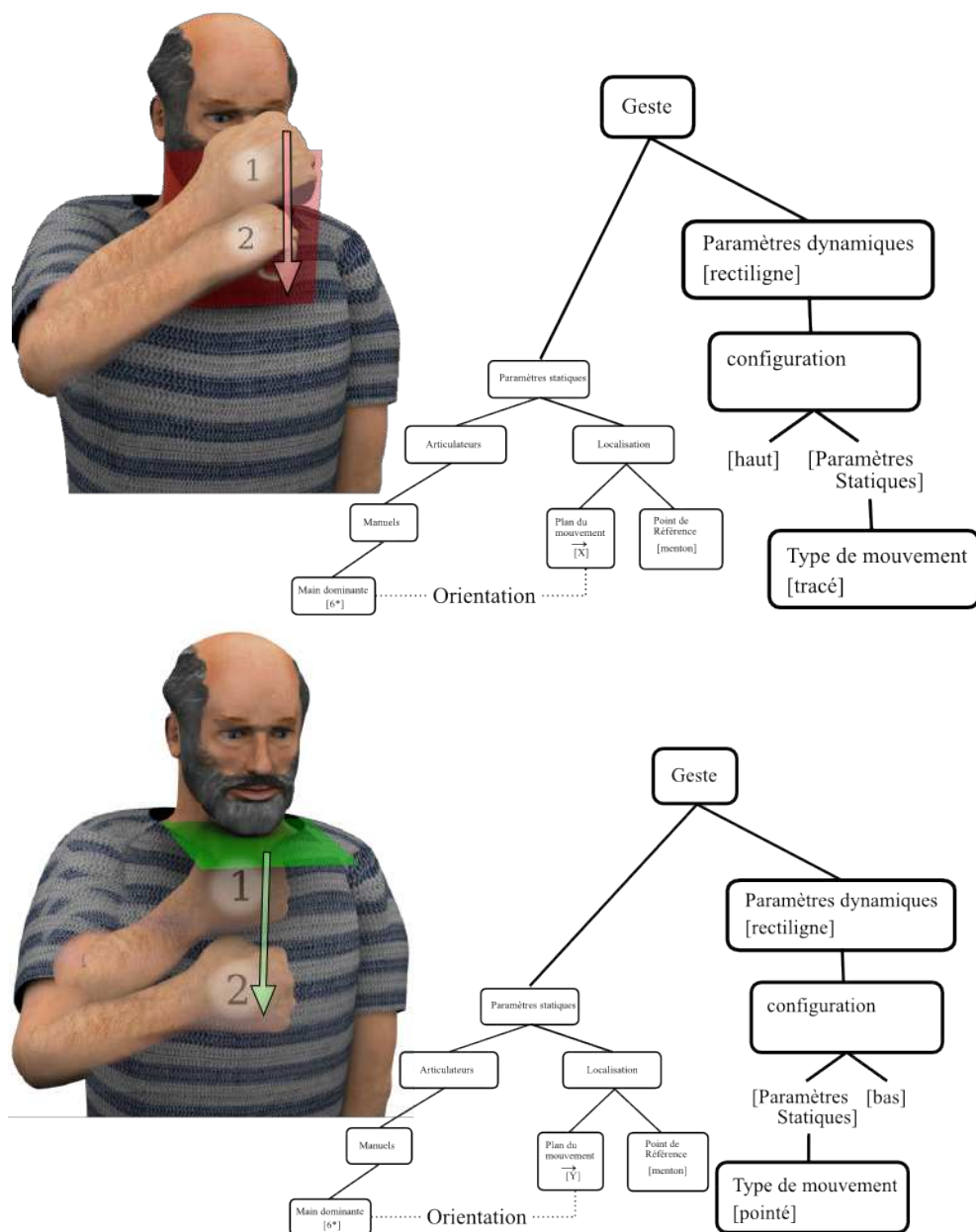


FIG. 5.16 – Deux spécifications de mouvement différentes : tracé et pointé.

## 5.5 Pistes de réflexion pour la gestion de la simultanéité en langue des signes

Un aspect fondamental de la langue des signes est la simultanéité des informations, ou le fait que lors de la production des gestes, de nombreux paramètres

soient réalisés de concert. Cette co-occurrence des paramètres des mouvements et des gestes de la langue de signes a été identifiée comme caractère multi-linéaire des canaux de la langue des signes par Cuxac [Cux00]. Un système de génération de la langue des signes se doit donc de prendre en compte cet aspect simultané et multi-linéaire.

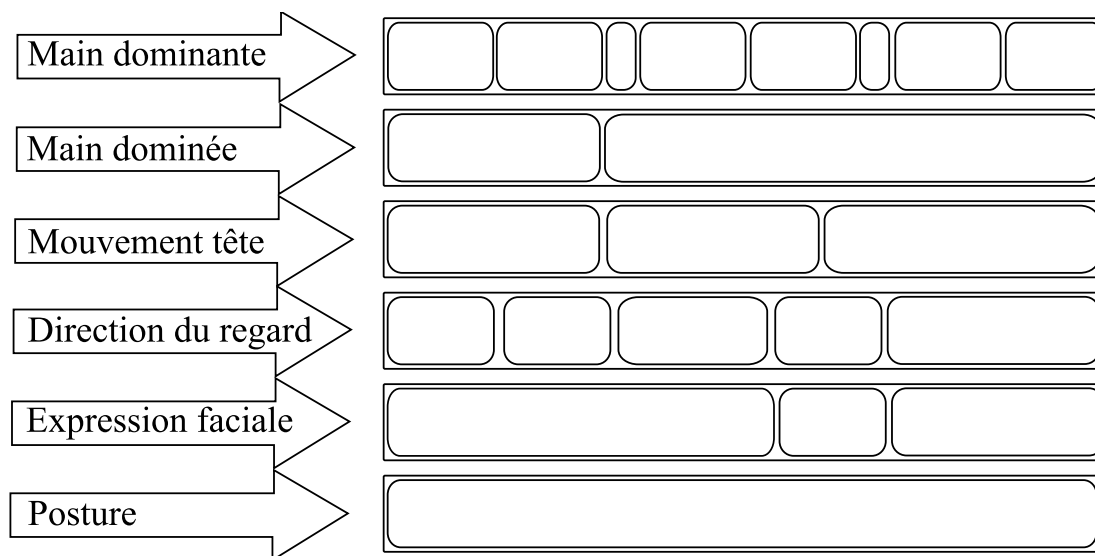


FIG. 5.17 – Représentation des différents canaux intervenants en langue des signes, aspect multi-linéaire

Le geste regroupe de nombreuses modalités. La figure 5.17 illustre l’organisation d’un geste complexe selon un ensemble de canaux participants à sa formation. Cet exemple révèle les différents niveaux de coordination entre les canaux (mouvement de tête et main dominée) et l’indépendance temporelle d’autres canaux (expression faciale et main dominante). Ces schémas de coordination peuvent changer au cours de l’exécution (expression faciale et direction du regard).

L’agencement multi-linéaire des différents canaux peut être illustré à l’aide d’un exemple représentant une séquence exécutée en situation de grande iconicité et proposé par Cuxac. Dans cet exemple, la rencontre de deux chiens est narrée par le locuteur. Un chien ressemble à un boxer, l’autre à un chien loup. A chaque protagoniste est associé un ensemble de descripteurs le caractérisant (ces descripteurs sont illustrés au chapitre 4 4.1.2.1). Chacun des chiens se voit également attribuer une démarche particulière. Ainsi, pour illustrer la démarche du boxer, les coudes sont écartés du corps et les poings sont serrés. En situation de transfert personnel, et à l’aide de plusieurs traits simultanés représentés dans la figure 5.18, le locuteur va pouvoir “donner à voir” un chien

ressemblant à un boxer, marchant, regardant (un autre chien), avec envie et en bavant beaucoup. Cette scène a recours à différents traits expressifs simultanés. Ces traits s'expriment sur des articulateurs différents et se combinent pour former la scène finale.

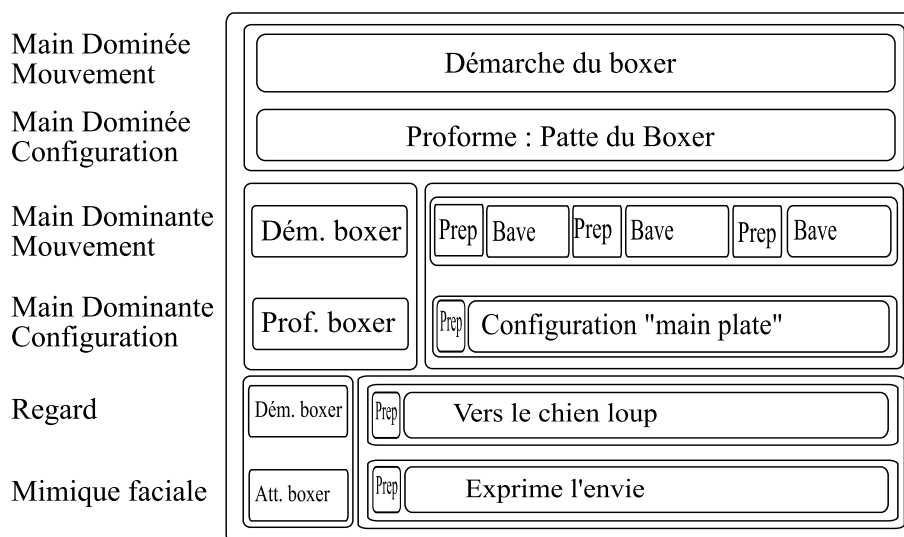


FIG. 5.18 – Partition du mouvement décrivant l'exemple tiré de [Cux00] : “Tout en marchant, le boxer regarde le chien loup [...] et bave beaucoup”

## 5.6 Vers un modèle de composition des gestes de la langue des signes

Le parallélisme de la langue des signes se manifeste par l'aspect simultané des signes dans l'exécution du mouvement. Cet aspect simultané, s'illustre à la fois en situation de grande iconicité (voir l'exemple du chien), mais également hors visée illustrative. En effet, et comme le présente Cuxac, la majorité des signes standards du lexique de la LSF peuvent être décomposés en unités minimales à valeur de morphèmes. La valeur morphémique des unités de base des gestes étant liée à la valeur iconique illustrant le sens référé.

Selon cet ordre, la combinaison de gestes ne peut être réalisée en mélangeant directement les gestes par mélange de mouvement, ou interpolation. En effet, certains gestes sont spécifiés à l'aide de contraintes exprimées relativement à des parties du corps du signeur susceptibles d'être modifiées par la participation d'autres signes intervenant dans la relation de composition inter-signes. Cet exemple est illustré dans la figure 5.19.

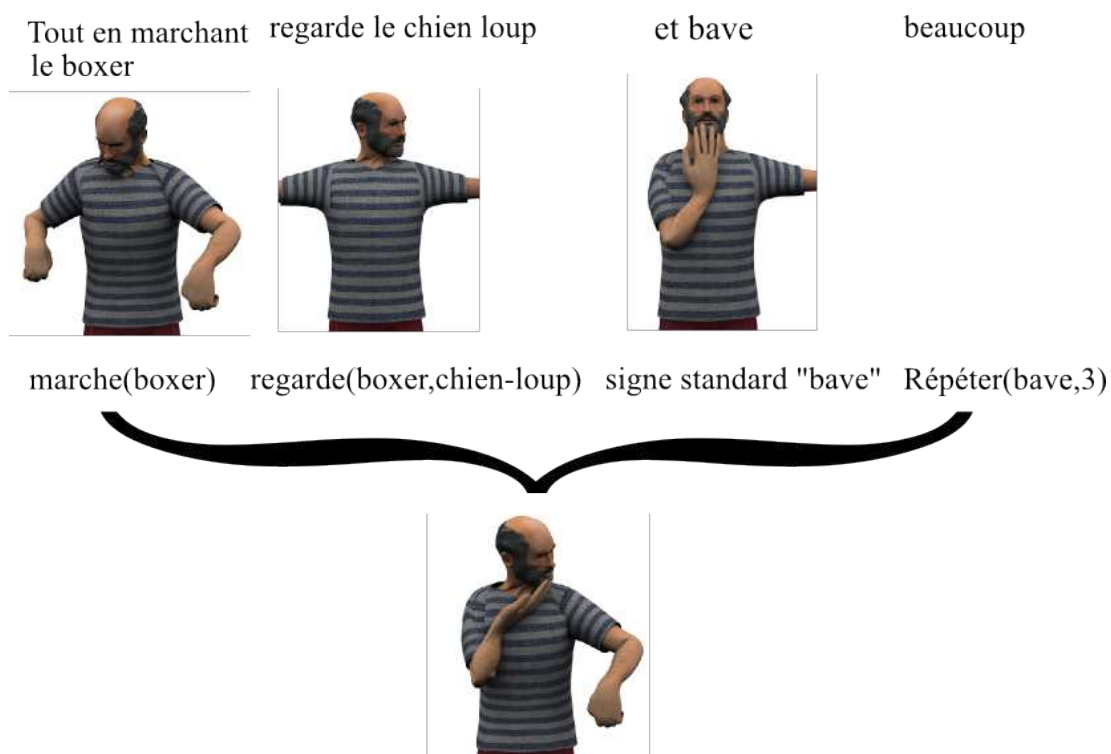


FIG. 5.19 – Compositionnalité morphémique en visée iconisatrice : l'exemple du boxer [Cux00]

Le mécanisme de composition entre signes atomiques porteur d'information au niveau morphémique nous amène à nous interroger sur les modèles de compositionnalité de ces gestes élémentaires et le niveau auquel ils doivent être traités. Intuitivement, on éliminera la possibilité de traiter la composition des gestes au niveau du signal de mouvement brut. Comme l'illustre la figure 5.20, un mélange de mouvement brut réalisé à partir des prototypes de signes de base ne conserve pas les caractéristiques nécessaires à la bonne compréhension du signe telles que les contraintes de contact spécifiées sur l'agent.

La compositionnalité des gestes de la langue des signes pourrait être réalisée en agrégeant les spécifications énoncées dans la seconde section de ce chapitre. En d'autres termes, la compositionnalité d'un geste de la langue des signes revient à spécifier une algèbre capable de gérer des opérations élémentaires entre gestes spécifiés selon le formalisme introduit en section 2 de ce chapitre. La compositionnalité des gestes de la langue des signes n'a pas été adressée dans cette thèse. Elle soulève néanmoins des défis très stimulants, telles que l'identification et la gestion des contraintes incompatibles, la spécification de niveaux de priorités entre contraintes. Dans le cadre de l'anima-

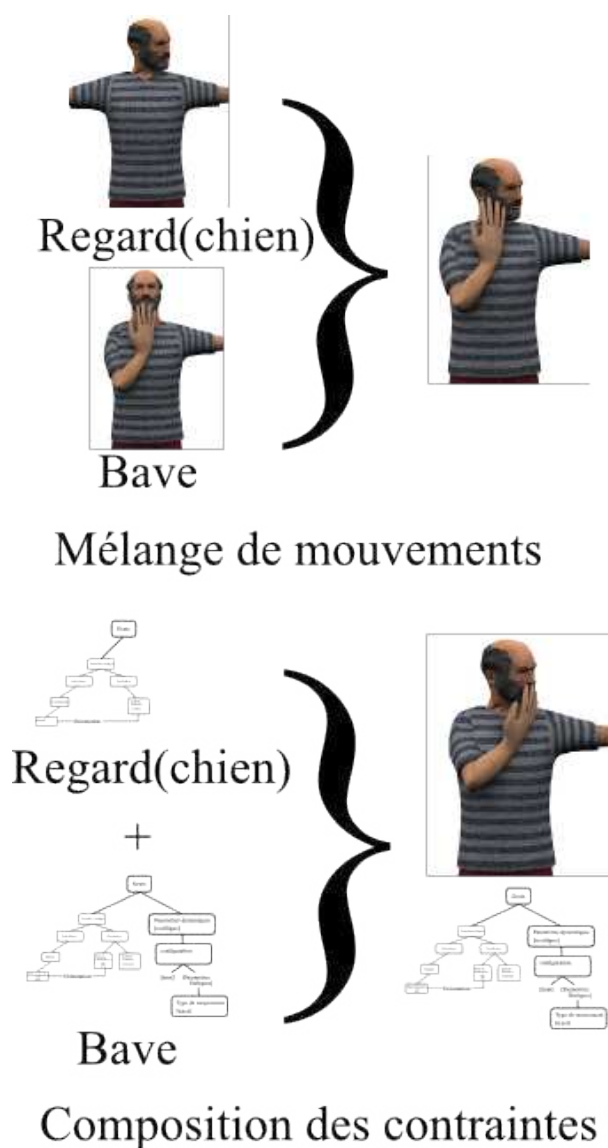


FIG. 5.20 – Le mélange de mouvements ne permet pas de respecter les contraintes nécessaires à la restitution intelligible d'un signe composé. Le signe représenté dans l'image du haut, obtenu par mélange de mouvements ne conserve pas la contrainte de contact entre la partie intérieure des doigts et le menton. Au contraire, le signe représenté dans l'image du bas, obtenu par mélange de contraintes la respecte.

tion par ordinateur, nous pouvons citer les travaux menés par Multon et coll. [KM05], Boulic et coll. [BB98] ainsi que Gleicher et coll. [GLS97]. Dans un cadre plus général, Ruttkay [Rut94] a proposé des méthodes de satisfaction de con-

traintes non rigides.





# Chapitre 6

## Analyse et synthèse du mouvement

Dans ce chapitre, nous présentons les travaux et les résultats qui nous ont permis de proposer une caractérisation de certains facteurs d'expressivité en langue des signes. Notre approche est basée sur une étape d'analyse de mouvements capturés. La première section est dédiée à la description de notre protocole de capture de mouvements, au format des données ainsi qu'à la nature des mouvements capturés. Une seconde section est consacrée à l'identification des aspects temporels de l'expressivité. Nous abordons ensuite les aspects spatiaux puis les aspects structurels touchant à l'organisation interne des signes. Une dernière section présente les méthodes de génération du mouvement que nous avons implantées en vue de produire les gestes de la langue des signes.

### 6.1 Matériel et méthodes

#### 6.1.1 Introduction

La modélisation et la synthèse de gestes expressifs supposent un aller-retour entre une analyse de données expérimentales et des simulations qui fournissent une vue simplifiée de notre compréhension d'un modèle. L'analyse permet de proposer des hypothèses sur le contrôle ou la coordination des mouvements mais permet également de valider des modèles théoriques tels que le modèle de description des gestes proposé dans le chapitre 5.

La constitution d'une base de mouvements pertinente est donc un prérequis indispensable à notre démarche reposant sur l'analyse et la synthèse. Les chercheurs se sont déjà appuyés sur des données de mouvement pour valider des hypothèses linguistiques [Sal03] ou relatives au contrôle moteur [NMB04] ou aux interactions communicatives réalisées par les agents conversationnels [Kip04]. Les données à partir desquelles nous avons construit notre étude se situent au confluent des trois disciplines sus-mentionnées et répondent à des

interrogations réparties sur un spectre qui va de la synchronisation des mouvements à la qualification de modificateurs d'ordre linguistique.

Aussi nous décrivons et représentons des aspects très différents de l'information véhiculée dans nos données de gestes. Nous restons proches du signal et étudions des trajectoires articulaires pour mettre en évidence les caractéristiques de bas niveau tels que les alignements temporels. Nous adoptons une approche plus globale en ce qui concerne les caractéristiques de plus haut niveau en nous appuyant sur une segmentation manuelle à partir d'enregistrements vidéo.

Dans les sous-sections suivantes, nous décrivons les données qui nous ont permis de réaliser nos études. Nous présentons, dans un premier temps, les protocoles de capture de données pour les décrire, ensuite, de manière plus globale et qualitative. Une dernière sous-section est consacrée à une discussion critique de notre corpus.

### 6.1.2 Acquisition de données

Notre protocole d'acquisition de mouvements a été conçu pour capturer l'ensemble des modalités entrant en jeu lors de la production de gestes de communication. Ce protocole repose sur l'utilisation de deux techniques complémentaires : la capture optique, pour les expressions faciales et le mouvement du corps du signeur, et la capture par gants de données, capable de restituer le mouvement des mains et des doigts ???. L'utilisation de ces gants de données pour la capture des mouvements des doigts du signeur a l'avantage d'éviter les problèmes d'occultations visuelles entre les doigts lorsque les mains se croisent, par exemple. Les données de mouvement capturées sont ensuite rééchantillonnées et fusionnées après la phase d'acquisition.

Cette méthode a bénéficié de l'expertise de plusieurs équipes françaises : les laboratoires M2S<sup>1</sup> de Rennes II et le LESP<sup>2</sup> pour leur compétence dans le domaine biomécanique, le laboratoire LINC<sup>3</sup> pour l'animation faciale. Le laboratoire M2S est intervenu autour de la capture optique par l'apport d'un réseau de caméras *Vicon-MX*<sup>4</sup>. Le LEST nous a fait bénéficier de son expérience en capture de mouvements de mains par l'intermédiaire de gants de données *Cybergloves*<sup>5</sup>. Le LINC, quant à lui, nous a apporté son aide autour de la capture de mouvements du visage. La figure 6.2 illustre notre protocole d'acquisition des gestes.

---

<sup>1</sup>Mouvement, Sport, Santé

<sup>2</sup>Laboratoire Energétique de Systèmes et Procédés

<sup>3</sup>Laboratoire de recherche en Informatique, Génie Industriel et Communication

<sup>4</sup>Source : <http://www.vicon.com/products/viconmx.html>

<sup>5</sup>Source : <http://www.immersion.com/3d/products/cyberglove.php>

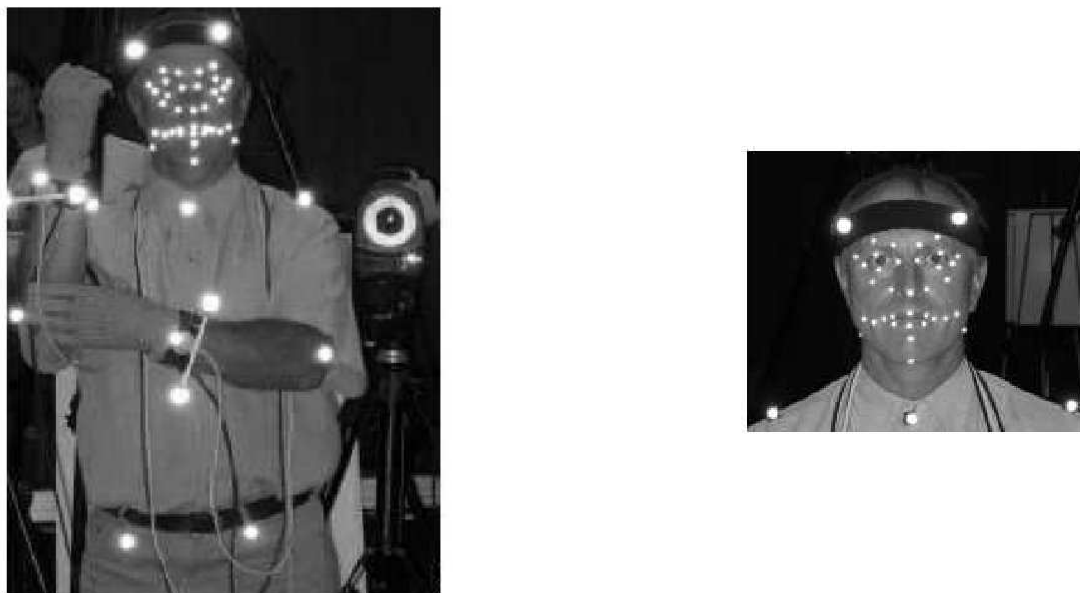


FIG. 6.1 – Systèmes d’acquisition des mouvements, répartition des réflecteurs acquis par le système de capture optique

### 6.1.3 Format des données

Une fois acquises, les données sont fusionnées et enregistrées dans le format de capture de mouvement Biovision hiérarchique (BVH). En comptant le corps (19 articulations) et les mains (22 articulation), la structure articulaire est composée de 63 articulations. La structure articulaire de notre squelette d’animation est présentée en annexe B.

### 6.1.4 Séquences capturées

Nous avons demandé à un signeur sourd, formateur de profession<sup>6</sup>, de signer à plusieurs reprises un bulletin météorologique dont la transcription est présentée en annexe A. A chaque reprise, nous avons demandé au signeur de modifier son style d’énonciation. La durée moyenne d’une séquence atteint 60 secondes, la fréquence d’acquisition est de 120 Hertz. Quatre séquences ont été utilisées dans nos expérimentations.

Les différentes séquences énoncent un même bulletin de prévision météorologique. La différence entre les séquences réside dans la qualité expressive que nous avons demandée au signeur de traduire. Trois séquences gestuelles ont été prises en compte. Dans les deux premières séquences, il a été demandé

<sup>6</sup>Il s’agit d’Alain Cahu de la société Polycom, Rennes

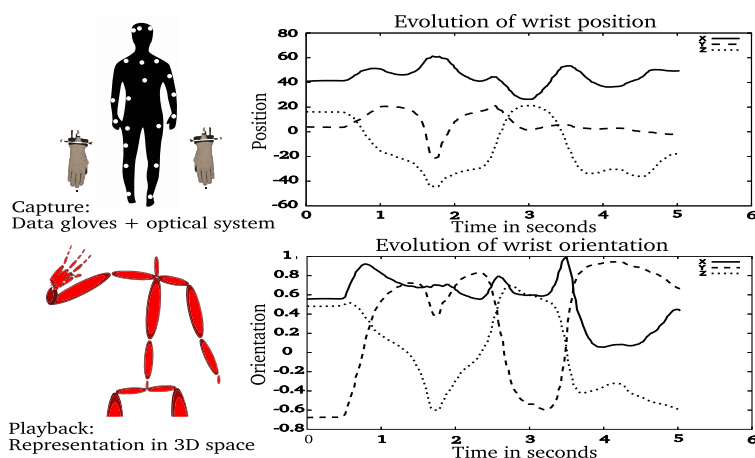


FIG. 6.2 – Systèmes d’acquisition des mouvements, ce système fait intervenir un outil de capture optique et des gants de données. Une fois acquises, les trajectoires de mouvements (courbes) sont rééchantillonnées puis fusionnées.

au signeur de produire des gestes avec une qualité expressive neutre. Une de ces deux séquences a été retenue. Dans la troisième séquence, il a été demandé de produire des gestes avec une qualité expressive "brusque". Enfin, dans la quatrième séquence, il a été demandé au signeur de produire des gestes suivant une qualité expressive "détendue" ou "molle".

La séquence gestuelle produite peut être décrite à l’aide de la séquence lexicale suivante : { BONJOUR, S’IL VOUS PLAÎT, REGARDEZ MOI, AUJOURD’HUI, SIX, JUILLET, LE TEMPS, BRETAGNE, MATIN, NUAGES TRAVERSEMENT, APRES MIDI, PLUIE, DEMAIN, SOLEIL, NAGER, ALLER, IL FAUT, VENDREDI, NUAGES, MATIN, BROUILLARD, A BIENTOT, AU REVOIR }. Le texte présenté au signeur se trouve en annexe A

Le tableau suivant reprend les caractéristiques de chaque séquence :

Séquence	Expressivité	Durée (s)	Nombre d’éléments lexicaux
1	Neutre	45	25
2	Brusque	33	25
3	Désinvesti, mou	52	25

TAB. 6.1 – Trois réalisations de la séquence météorologique

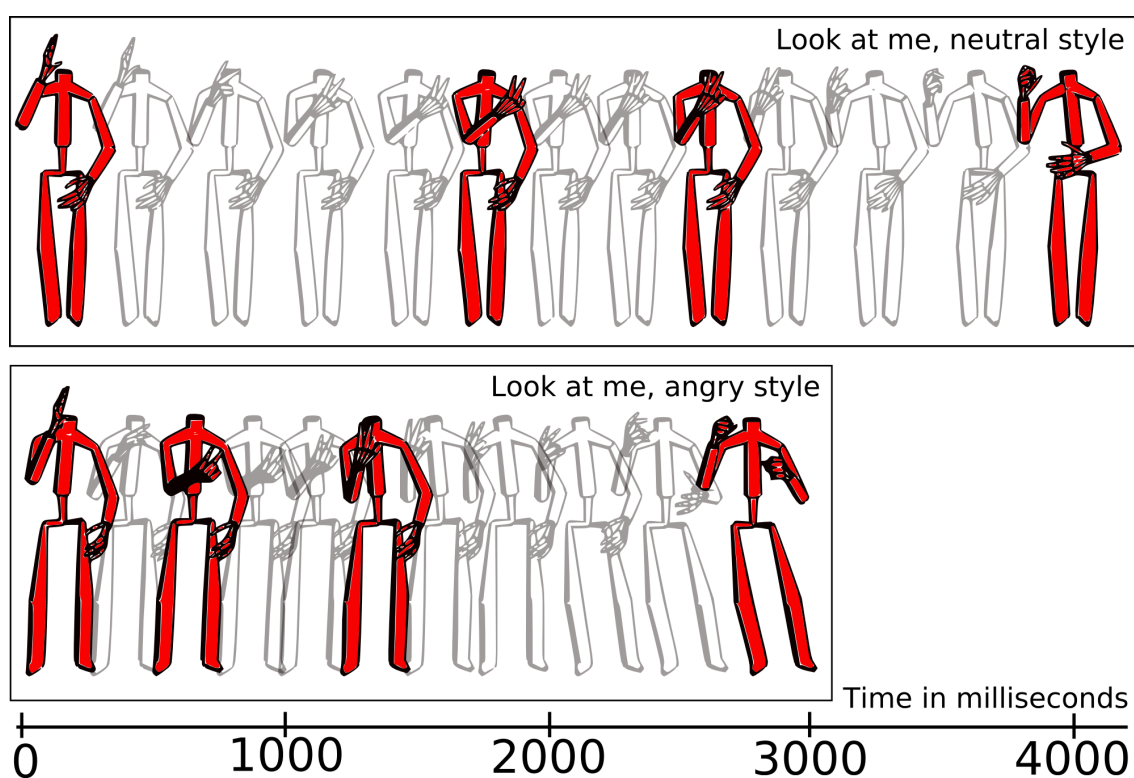


FIG. 6.3 – Le signe REGARDEZ\_MOI exécuté par le signeur. La séquence du dessus correspond à une expressivité neutre tandis que la séquence du dessous correspond à une qualité expressive “brusque”.

## 6.2 Analyse/synthèse de séquences gestuelles

Une fois les données acquises, la première partie de notre démarche s'est focalisée sur les aspects signal du mouvement. Plus particulièrement, nous nous sommes proposés de caractériser les variations d'ordre temporel observables entre différentes réalisations d'une même séquence de mouvement selon des qualités expressives différentes. Pour cette étude, nous nous sommes donnés une définition de la qualité expressive la plus proche possible du signal, c'est à dire, la variabilité observée entre deux réalisations d'une même séquence gestuelle.

D'un point de vue signal, l'analyse des données de mouvement soulève deux difficultés. En premier lieu, ces données sont, par nature, mutidimensionnelles ; elle traduisent, en effet, l'évolution simultanée de plusieurs articulateurs plus ou moins indépendants. En deuxième lieu, les séquences de gestes ne sont pas prédictibles et ne peuvent pas être assimilées à des phénomènes pseudo-périodiques, comme peut l'être la marche, par exemple. Par conséquent, l'analyse de séquences gestuelles est hors de portée de la plupart des méthodes ayant été proposées jusqu'à aujourd'hui dans le cadre de l'animations par ordinateur [HPP05, SF06].

Notre travail se concentre sur l'alignement temporel de différentes réalisations de gestes expressifs de communication. Pour ce faire, nous émettons l'hypothèse qu'il existe une "composante commune" aux différentes réalisations d'une séquence gestuelle donnée. Nous montrons ensuite qu'un espace de représentation s'appuyant sur une analyse en "composante principales pondérées" est compatible avec l'hypothèse d'existence de "composante commune". L'alignement temporel est réalisé à partir de cet espace de représentation et optimisé par une approche multi-niveaux que nous appelons *DTW adaptative*.

### 6.2.1 Analyse temporelle de séquences gestuelles

L'alignement de séquences multidimensionnelles complexes telles que des séquences de gestes nécessite des techniques de réduction d'espace de représentation. L'analyse en composantes principales (ACP) fait partie de ces techniques d'analyse de données. Le principe de cette méthode est d'effectuer une transformation linéaire qui permet d'exprimer un ensemble de donnée multidimensionnelles dans une nouvelle base orthonormale, dont les axes ordonnés - les composantes principales - supportent les variations des données par ordre décroissant d'importance. Les premiers axes factoriels extraits par ACP sont alignés sur les plus grandes variations des données de départ, alors que les derniers, qui ne supportent que les variations infimes, peuvent être négligés

ou modélisés par une loi de variation.

Les coordonnées angulaires des matrices de séquences gestuelles sont préalablement converties afin d'être représentées par des cartes exponentielles. Cette représentation offre une grande robustesse aux combinaisons linéaires induites par l'analyse en composantes principales. Pour plus de détails sur les représentations des orientations utilisées au cours de ce travail de thèse, nous vous invitons à consulter l'annexe C.1.

L'espace de représentation que nous avons choisi pour caractériser les données de mouvements dans un espace de dimension réduite est inspiré de l'ACP pondérée proposée par Forbes et col. [FF05] à des fins d'indexation de mouvements. Cette méthode permet d'optimiser les recherches en projetant l'ensemble des mouvements à indexer dans un sous-espace décrit par un nombre réduit de composantes principales extraites d'un mouvement de référence  $M_{ref}$ . Le mouvement de référence choisi par Forbes et col. est le mouvement qui a servi à calibrer l'appareil de capture de données. Ce mouvement, que l'on nomme en anglais *range of motion* est en quelque sorte un inventaire du débattement possible de chaque degré de liberté du signeur. Ce choix est pertinent lorsque l'on ne dispose d'aucun a priori sur le type de mouvements à indexer.

Or, les mouvements dont nous disposons sont très proches les uns des autres : il s'agit de différentes réalisations d'une même séquence gestuelle. Il est donc pertinent de choisir comme mouvement de référence l'une des séquences de notre jeu d'étude et de construire le sous-espace de référence sur cette séquence. Ces postures propres sont les vecteurs propres de la matrice de covariance de la matrice de mouvement de référence préalablement centrée. La suite des postures propres  $U_{ref}$  est obtenue par la décomposition en valeurs singulières (SVD) de la matrice de référence  $M_{ref}$ . La décomposition SVD de  $M_{ref}$  s'écrit alors :

$$M_{ref} = U_{ref} \Sigma_{ref} V_{ref}^T$$

Avec  $\Sigma_{ref}$  une matrice diagonale composée de réels en ordre décroissant.  $U_{ref}$  est une base orthonormale de postures propres de  $M_{ref}$  et  $V_{ref}^T$  exprime les coordonnées de  $M_{ref}$  dans la base  $U_{ref} \Sigma_{ref}$ .

Les matrices de mouvement  $M_i$  correspondant aux réalisations suivantes sont ensuite projetées sur la base extraite du mouvement de référence  $M_{ref}$  :

$$M_i = U_{ref} \Sigma_{ref} V_i^T$$

d'où

$$V_i^T = \Sigma_{ref}^+ U_{ref}^T M_i$$

$\Sigma_{ref}^+$  est la transposée de  $\Sigma_{ref}$  dont les valeurs de la diagonale non nulles ont été inversées.  $V_{ref}^T$  exprime les coordonnées de  $M_{ref}$  dans la base  $U_{ref}\Sigma_{ref}$ .

Comme l'illustre la figure 6.4, la base de posture propre de la séquence de référence permet de décrire de manière économique les trois séquences étudiées (neutre, molle et brusque). Les quatre premières postures propres extraites de la matrice de référence permettent d'expliciter entre 70% et 80% de la variance totale des différents mouvements.

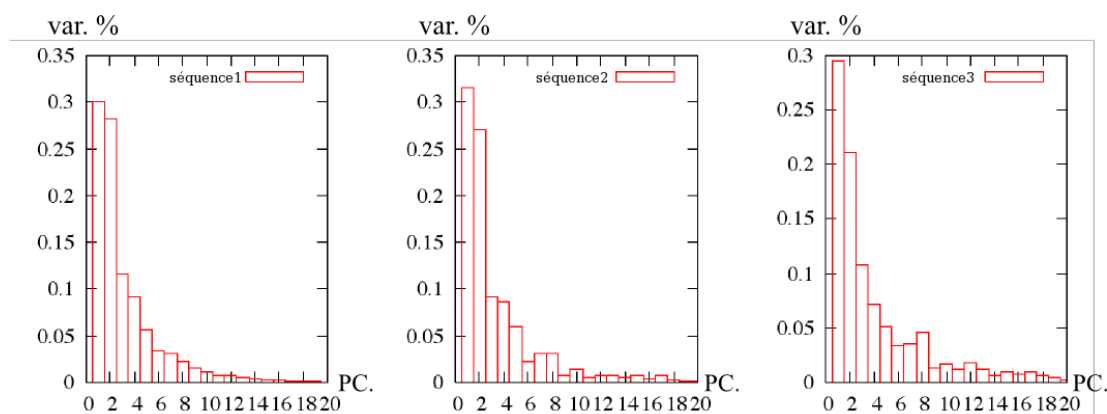


FIG. 6.4 – Proportion de variance explicitée par les composantes des matrices de mouvements projetées sur la base des postures propres du mouvement neutre (mouvement 1)

La représentation de chaque matrice de mouvement dans une base commune de postures propres permet de définir une fonction de distance cohérente entre les postures de matrices de mouvement. En effet, quelle que soit la séquence de mouvement projetée, celle-ci s'exprime à l'aide d'une combinaison linéaire de  $n$  postures propres extraites du mouvement de référence.

La distance  $\delta(p_i, q_j)$  entre deux postures  $p_i$  et  $q_j$  et exprimée dans  $U_{ref}\Sigma_{ref}$  est alors calculée à partir des 4 premières composantes de  $p_i$  et  $q_j$ .

$$\delta(p_i, q_j) = (p_i(1) - q_j(1))^2 + \dots + (p_i(4) - q_j(4))^2$$

La fonction de distance ainsi déterminée permet la construction de matrices de distances sur lesquelles appliquer notre algorithme d'alignement temporel adaptatif (aDTW).

## 6.2.2 Alignement temporel adaptatif

La problématique de comparaison de séries temporelles est classiquement adressée par l'algorithme de *Dynamic Time Warping* (DTW) [MR81]. Cet al-



gorithme peut être défini comme une mesure de distance entre deux séries présentant des variations temporelles (dilatations ou contractions) non homogènes. La distance donnée par l'algorithme DTW a été utilisée dans de nombreux champs d'applications telles que la reconnaissance de la parole[Rab93], la reconnaissance de gestes[Cor01] ou l'animation par ordinateur [KG03].

L'algorithme de DTW procède en calculant une distance entre chaque paire possible de représentants des séquences à comparer. Cette étape permet de construire une matrice de distances cumulatives à partir de laquelle est extrait un chemin de moindre coût la traversant. Dans sa version de base, l'algorithme DTW pose deux problèmes : d'une part, la phase de construction de la matrice de distances cumulatives implique une complexité quadratique selon la longueur des séquences à comparer ; d'autre part, le comportement local de ce type d'algorithme tend à générer des chemins d'appariements comportant de fortes discontinuités. Afin de répondre à ces limitations, nous nous appuyons sur une extension multi-niveaux contrainte de l'algorithme DTW. L'approche multi-niveaux permet de prendre en compte les variations temporelles globales et locales intervenant entre différentes réalisations d'une séquence gestuelle tout en garantissant une continuité dans les chemins d'appariement en contraignant la pente du chemin d'appariement obtenu.

Nous présentons dans un premier temps les réponses apportées dans le cadre de la synthèse du mouvement en vue d'améliorer la continuité des chemins d'appariement de l'algorithme de DTW standard. Dans un second temps, nous montrons l'algorithme de DTW multiniveaux proposé par Salvador et col. [SPC04]. Enfin, nous exposons notre proposition, le DTW adaptatif, se basant à la fois sur la contrainte de pente et le DTW multiniveaux.

### 6.2.2.1 DTW à pente contrainte

Il arrive que les chemins d'appariement obtenus par DTW présentent des discontinuités fortes qui se manifestent sous la forme de chemins comportant d'importantes zones de pentes horizontales ou verticales. La DTW à pente contrainte a été conçue afin de limiter ces discontinuités. Si  $Q$  et  $P$  sont deux séries temporelles de longueurs respectives  $m$  et  $n$ ,  $\delta$  une fonction de distance entre deux éléments  $q_i$  de  $Q$  et  $p_j$  de  $P$ , et  $k$  un réel positif, la construction de la matrice de distances cumulatives de la DTW à pente contrainte peut être définie à l'aide de la fonction  $\gamma\{q_i, p_j\}$  définie de la manière suivante :

$$\gamma\{q_i, p_j\} = \delta\{q_i, p_j\} + \min(\gamma\{q_{(i-1)}, p_{(j-1)}\}, k\gamma\{q_{(i-1)}, p_j\}, k\gamma\{q_i, p_{(j-1)}\})$$

$$DTW(Q, C) = \gamma\{q_m, p_n\}$$

$k$  permet de contraindre le chemin d'appariement le long de la diagonale de la matrice de distance entre  $Q$  et  $C$ , permettant ainsi une meilleure continui-

té dans les chemins extraits. Cette méthode a été notamment utilisée par Kovar et Gleicher [KG03]. L'inconvénient de cette méthode est de ne pas prendre en compte les différents niveaux de variation temporelle du mouvement. Elle peut en particulier quitter le chemin optimal d'alignement si les deux séquences présentent de fortes variations temporelles locales.

### 6.2.2.2 DTW multiniveaux

La complexité quadratique de l'algorithme de DTW rend peu performantes les comparaisons de séries temporelles longues. Une optimisation multi-niveaux a été proposée par Salvador et col. [SPC04] sous le nom de FastDTW. Cette technique consiste à réduire la complexité de la forme standard du DTW par un sous-échantillonnage successif des séries à comparer. Une fois la résolution la plus basse atteinte, un chemin d'alignement est calculé sur les séries sous-échantillonnées par la forme classique de l'algorithme de DTW. Le chemin obtenu sert ensuite de guide pour contraindre l'espace la zone traitée au niveau de résolution supérieur. La complexité de FastDTW est linéaire et permet d'extraire des chemins d'alignement quasi optimaux entre des séries temporelles. Cependant, l'aspect multi-niveaux de l'algorithme FastDTW a vocation à optimiser la forme standard de DTW, il ne permet pas de contrôler la continuité des chemins d'alignement obtenus.

### 6.2.2.3 DTW adaptative

Afin de proposer une contrainte adaptative de la pente du chemin d'appariement calculé par le DTW, nous proposons d'adapter niveau par niveau la contrainte de pente de la courbe d'appariement calculée par le DTW multiniveaux. L'algorithme récursif proposé est illustré par le pseudo-code de la figure 6.5. Il commence par créer deux sous-séries de résolution inférieure en sous-échantillonnant la matrice de mouvement passée en paramètre ( fig. 6.5, l. 5-6 ). La contrainte de pente de chemin d'appariement est ensuite mise à jour ( l. 7-9 ). Un chemin d'appariement  $[p, q]$  est ensuite extrait des matrices à basse résolution ( l. 10 ) puis projeté à la résolution double ( l. 11 ). Ce chemin est ensuite affiné à l'itération suivante de l'algorithme ( l. 5-9 ). Le cas de base est atteint lorsque la longueur d'une des matrices de mouvement dépasse le seuil fixé en paramètre. Dans ce cas, un chemin d'appariement est retourné par la version standard de l'algorithme DTW.

**Algorithm 1** adaptative DTW algorithmAdaptativeDTW( $X, Y, \text{radius}, K$ )**Require:** $X$  - a projected motion $Y$  - a projected motion

radius - distance to search out of the projected warp path from the previous resolution

 $K$  - slope constraint at the finest resolution**Ensure:** $[p, q]$  a warp path along the motions  $X$  and  $Y$ 

```

1: minTSSize = radius + 2 \\ The min size of the coarsest resolution
2: if length( $X$ )  $\leq$  minTSSize OR length( $Y$ )  $\leq$  minTSSize then
3:   RETURN DTW( $X, Y, K$ )
4: else
5:   shrunkX = downsample( $X$ ) \\ length(shrunkX) = 1/2  $\times$  length( $X$ )
6:   shrunkY = downsample( $Y$ )
7:   if  $K > 1$  then
8:      $K = K - 1$  \\ Adjust slope constraint for coarser resolution
9:   end if
10:  ( $p, q$ ) = AdaptativeDTW(shrunkX, shrunkY, radius,  $K$ )
11:  window = ExpandResWindow( $p, q, X, Y, \text{radius}$ )
12:  RETURN DTW( $X, Y, \text{window}, K$ )
13: end if

```

FIG. 6.5 – Algorithme de DTW adaptatif

### 6.2.3 Résultats

Les résultats d'appariement de notre algorithme ont été comparés aux appariements obtenus par les approches DTW standard et par une version de DTW à pente contrainte telle qu'elle a été décrite dans [KG03]. Nos résultats sont illustrés à la figure 6.7. La figure 6.7 met en comparaison un alignement temporel obtenu par DTW classique, DTW contraint et DTW adaptatif. Les courbes représentent l'évolution de l'influence du premier vecteur propre de chaque mouvement au cours du temps. Les deux mouvements comparés sont deux réalisations de la séquence météo exécutées selon un style "neutre". La figure 6.8 représente l'alignement temporel obtenu entre la réalisation "neutre" et la réalisation "brusque" de la séquence. Cette figure montre la capacité du DTW adaptatif à aligner deux séquences présentant des variations spatiales

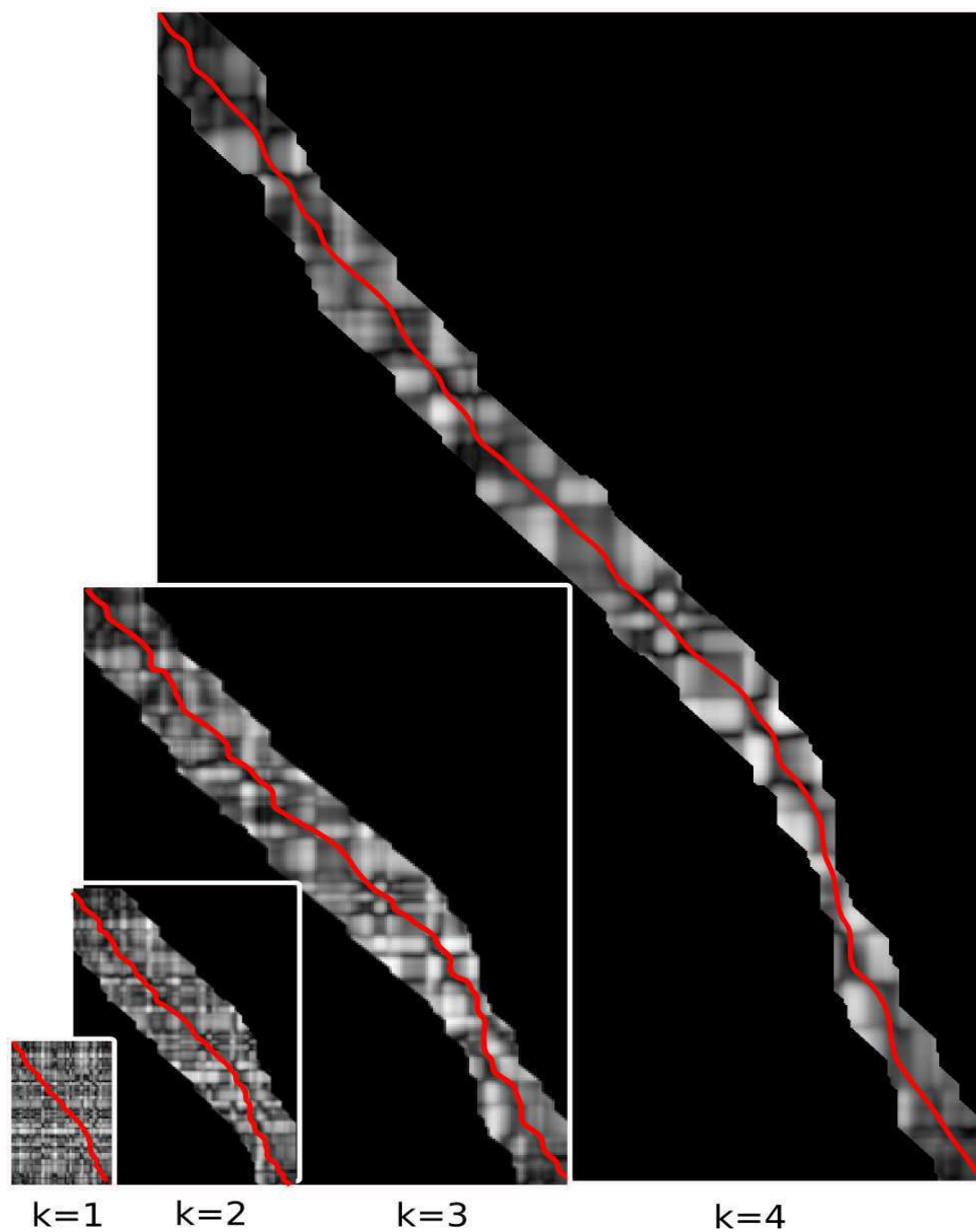


FIG. 6.6 – A chaque itération, le chemin d'appariement obtenu est projeté sur le niveau supérieur, il fournit ainsi un guide plus contraint lors de l'itération suivante

relativement importantes.

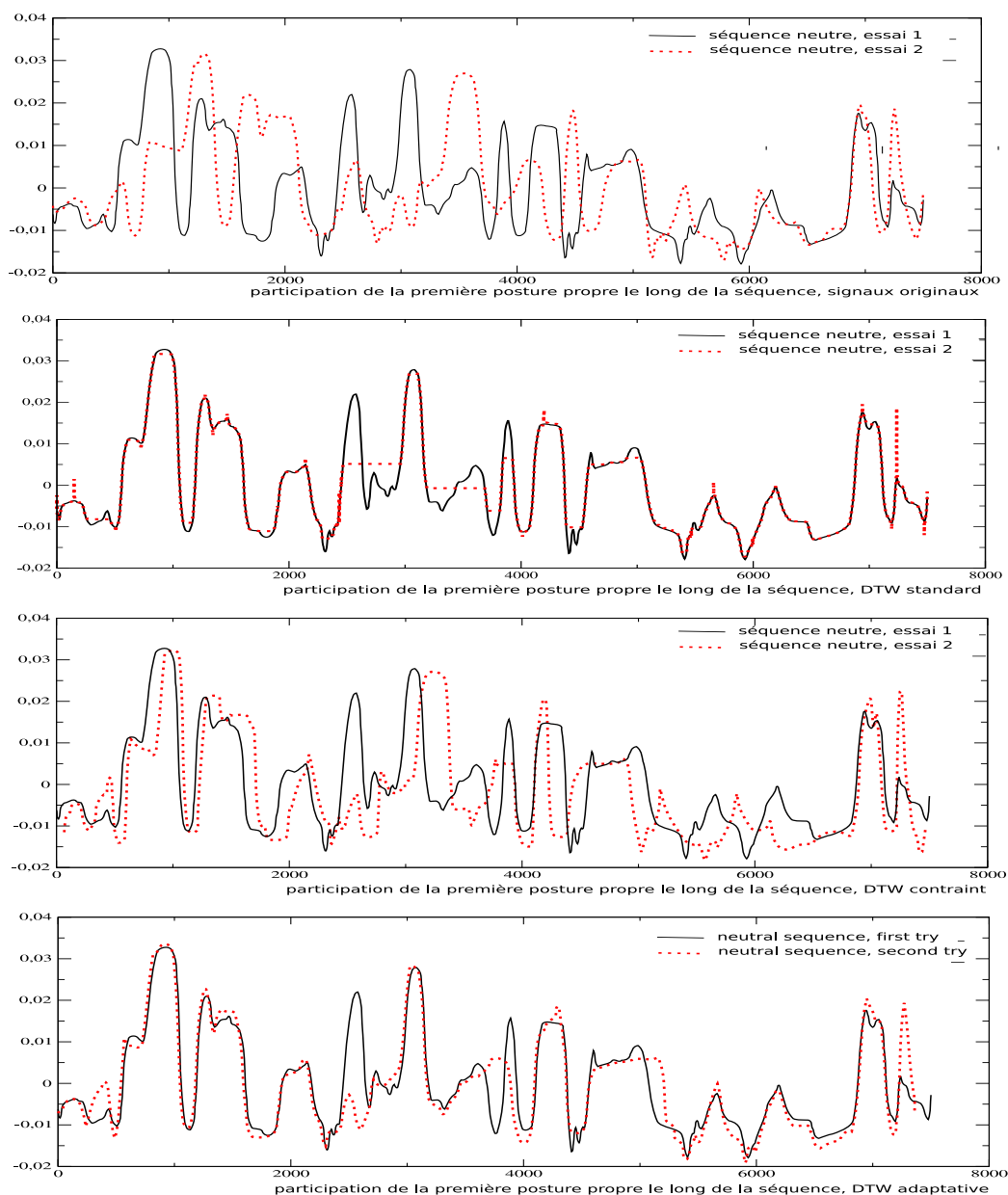


FIG. 6.7 – De haut en bas : signaux originaux, alignement par DTW standard, DTW à pente contrainte et DTW adaptative. L’algorithme de DTW adaptative propose un alignement limitant les discontinuités et capable de s’adapter à la fois aux perturbations locales et globales entre séries temporelles

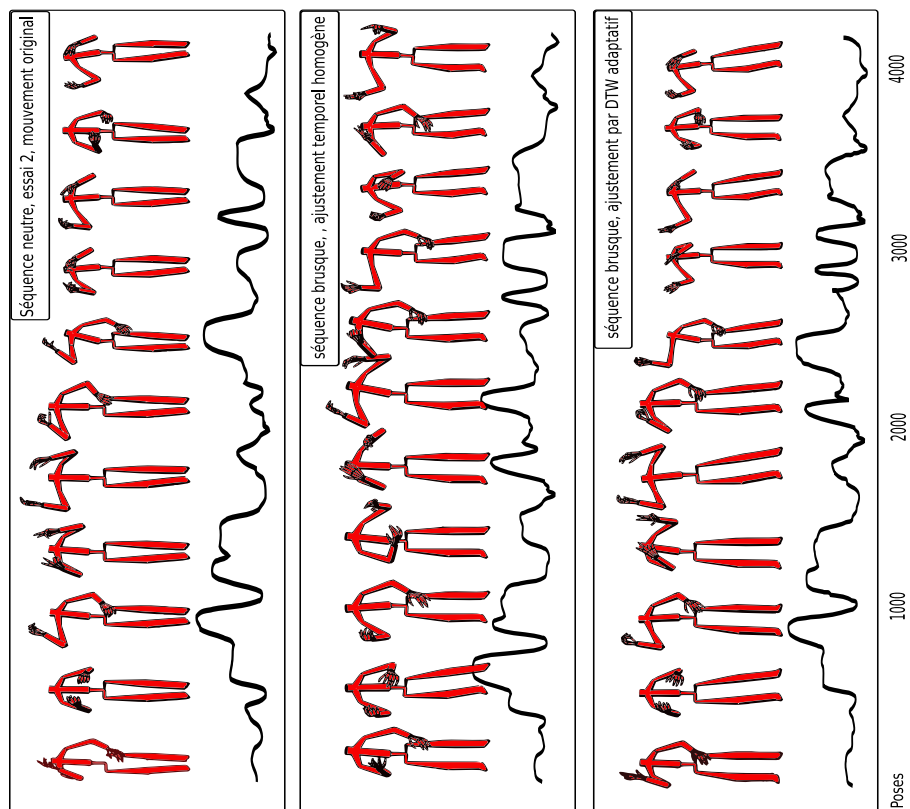


FIG. 6.8 – Résultats d’alignements temporels illustrés à l’aide de postures. Les deux premières séquences représentent deux réalisations réalisées respectivement de manière neutre et brusque. La troisième séquence illustre l’alignement de la séquence brusque sur la séquence neutre. Les courbes présentées sous les poses représentent l’évolution de l’influence de la première posture principale de la base de représentation du mouvement ( $\Sigma_{ref}^+ U_{ref}^T$ )

## 6.2.4 Discussion

La fonction de distance entre poses que nous avons proposée permet de comparer des postures issues de séquences réalisées selon des styles différents. La méthode d'alignement proposée ensuite permet d'extraire des chemins d'appariement fluides. Les alignements obtenus peuvent être exploités dans différents contextes : mélange de mouvement, facilitation de la segmentation manuelle des mouvements ou traduction de style. L'une des limitations de notre méthode est de ne pas prendre en compte certaines variations spatio-temporelles induites par le style. Notamment, les répétitions rythmées de certains gestes qui interviennent dans les séquences "brusques".

## 6.2.5 Application à la synthèse

Une application directe de l'alignement de mouvements pour la synthèse de mouvement basée données est le rejeu par mélange de mouvement. En exploitant le chemin d'alignement entre plusieurs réalisations d'une même séquence gestuelle selon différentes qualités expressives, il est possible de modifier de façon interactive la qualité expressive du geste rejoué. Cette application est illustrée à la figure 6.9. On y trouve deux réalisations d'une même séquence (regardez-moi en LSF). L'extraction préalable d'un chemin d'appariement a permis de mettre en correspondance chaque poses deux à deux. Un utilisateur est en mesure de choisir interactivement la proportion de caractère brusque dans le mouvement lors du rejeu. Le mélange de mouvement est réalisé par interpolation.

## 6.2.6 Conclusion

D'un point de vue signal, l'étude de la variabilité au niveau des séries temporelles multidimensionnelles nous a posé le problème de l'alignement pertinent de séquences gestuelles communicatives longues de plusieurs milliers de postures. Ce problème nous a amené à proposer un algorithme d'alignement temporel multi-niveaux et à pente adaptative. Nous avons baptisé cet algorithme *Adaptive Dynamic Time Warping*, (ADTW). Il permet d'aligner des séquences gestuelles communicatives longues de manière pertinente tout en garantissant une continuité satisfaisante dans les chemins d'appariement. Ces alignements peuvent être directement utilisés dans le cadre de la synthèse de gestes expressifs en permettant une synthèse par mélange de postures.

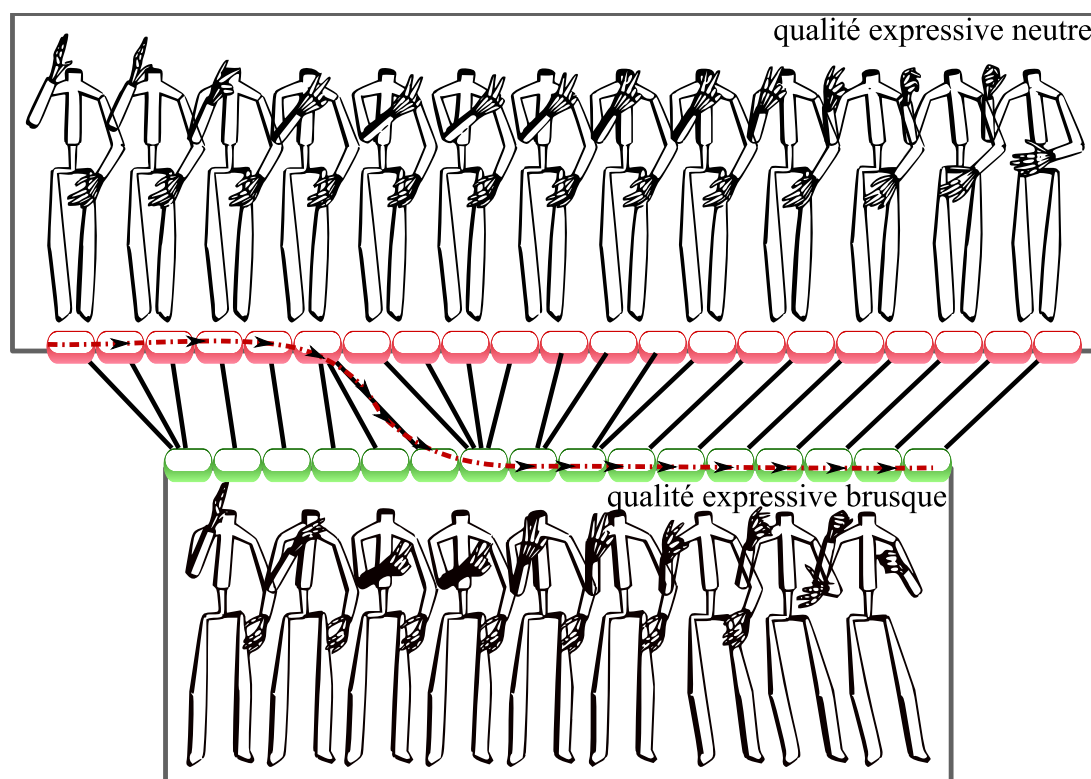


FIG. 6.9 – Application de l’alignement à la synthèse : une fois alignés, il est aisé de modifier la qualité expressive d’un mouvement de manière continue et fluide par simple interpolation

## 6.3 Identification des paramètres d’expressivité

### 6.3.1 Introduction

La section précédente a fait l’objet d’une analyse temporelle des facteurs d’expressivité entre différentes réalisations d’une même séquence gestuelle selon des qualités expressives différentes. Cette caractérisation est basée sur une prise en compte des mouvements proches du signal et présente un aspect systématique inhérent aux propriétés automatiques de la méthode d’alignement. Une fois l’alignement réalisé, nous avons vu qu’il est possible, par rejeu, de moduler la manière dont se déroule le geste dans le cadre de la synthèse. L’aspect temporel lors de l’exécution des gestes constitue un paramètre entrant en jeu dans l’expressivité.

Dans cette section, notre démarche repose sur une analyse qualitative des mouvements afin d’y identifier les paramètres d’expressivité intervenant en langue des signes. L’identification de ces paramètres nécessite au préalable



l'isolement de phases gestuelles conformes à celles présentées en 5.1. Il a donc été nécessaire de recourir à des méthodes d'extraction pertinentes. Nous présentons en première section nos travaux vis à vis de la segmentation des gestes et la méthode hybride mettant en jeu des algorithmes de segmentation automatique assistant une segmentation manuelle. Une fois isolées, les séquences de gestes sont comparées de manière à mettre en évidence les paramètres d'expressivité, au niveau spatial, temporel et au niveau de la structure des signes.

### 6.3.2 Mise en évidence des phases du mouvement

Le modèle de spécification des gestes que nous avons proposé au chapitre 5 se base sur une segmentation du flux gestuel selon trois niveaux : une phrase gestuelle est décomposée en syntagmes, eux-même découpés en phases gestuelles. Les phases gestuelles se distinguent par leur valeur sémantique et par leurs caractéristiques cinématiques. Nous rappelons ici les différentes phases que nous avons identifiées :

- Stroke : il s'agit d'une phase porteuse de sens et contenant un élément dynamique ;
- Préparation : phases non porteuses de sens permettant d'amener les articulateurs dans une configuration correspondant à la configuration de départ de la phase suivante ;
- Rétraction : phase non porteuse de sens permettant d'amener les articulateurs dans une configuration neutre ou de repos.
- Tenue : phase porteuse de sens pendant laquelle les articulateurs mis en jeu conservent leur configuration.

Afin de confronter notre modèle théorique aux données expérimentales que nous avons capturées, il est nécessaire de repérer les différentes phases de mouvement se succédant dans les mouvements capturés. Autrement dit, il est nécessaire de les segmenter. La segmentation de mouvements a fait l'objet de travaux dédiés à la langue des signes [Ca01] ainsi qu'à l'étude de gestes expressifs [KvGvdH98, Kip04]. Les méthodes de segmentation alors présentées se sont basées sur une détermination manuelle des phases de mouvements, prenant en compte de multiples canaux [Ca01] ou non ([KvGvdH98]).

S'il est indiscutable que les méthodes automatiques ne sont pas en mesure d'égaliser un annotateur humain bénéficiant d'une expertise dans le champ d'application de la segmentation, il est également vrai que la plupart des opérations mises en jeu lors du processus d'annotation sont répétitives, d'autant plus si le processus doit être réalisé selon différents canaux d'annotation (mains droite et gauche, bras droit et gauche, corps). A partir des données de mouvements acquises lors du processus décrit au paragraphe 6.1.2, nous avons opté pour une méthode de segmentation hybride mettant en jeu des algorithmes de dé-

tection de phases et un outil interactif d'annotation se basant sur les vidéos enregistrées lors de la capture.

### 6.3.3 Segmentation automatique

Bien que certains travaux [FMJ02, KG03] aient montré qu'une segmentation automatique peut être réalisée de manière directe, en comparant la dérivée seconde des trajectoires angulaires ou cartésiennes des différents degrés de liberté des squelettes d'animation, Hodgins et col.[BSP<sup>+</sup>04] ont montré qu'une segmentation plus pertinente peut être réalisée grâce à une approche utilisant l'analyse en composante principale (ACP). L'algorithme de segmentation est basé sur l'hypothèse que le sous-espace permettant de représenter avec précision un mouvement présentant un comportement unique doit être plus faible que la dimension nécessaire à la représentation exacte d'un mouvement présentant deux comportements distincts. Ainsi, lorsqu'un mouvement présente une transition entre deux comportements différents, l'erreur de reconstruction des postures projetées sur le sous-espace des postures propres de dimension  $r$  augmente rapidement. La figure 6.10 illustre l'augmentation de cette erreur entre deux postures de configuration manuelle.

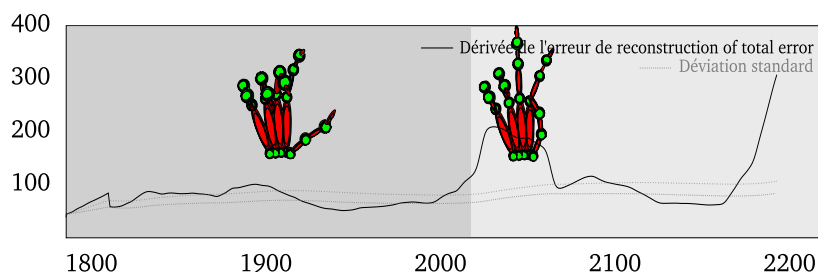


FIG. 6.10 – Segmentation automatique réalisée par l'algorithme proposé par [BSP<sup>+</sup>04]. Une marque de segmentation est posée lorsque la dérivée de l'erreur de reconstruction dépasse trois déviations standard de la moyenne de l'erreur

Nous avons appliqué l'algorithme de segmentation automatique à une séquence capturée représentant une succession des postures de main droite signifiant l'alphabet dactylogique en langue des signes. La séquence a été réalisée par l'auteur. La séquence contient 7200 postures et est échantillonnée à 120Hz. Nous avons fixé à trois le nombre de composantes principales nécessaires à la reconstruction du mouvement, celui-ci garantissant un signal reconstruit dont la variance se situe entre 75% et 95% de la variance portée par le signal original.

En parallèle de la segmentation automatique, nous avons demandé à un observateur humain non signeur de segmenter manuellement la séquence dacty-

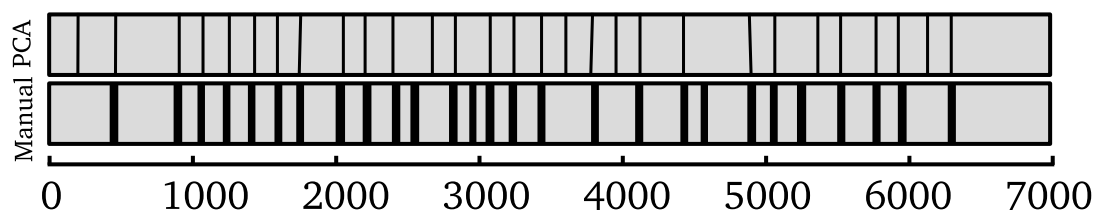


FIG. 6.11 – Comparaison entre les résultats de segmentation automatique et manuelle

lographique enregistrée. La figure 6.11 présente une comparaison entre la segmentation manuelle et la segmentation automatique. L'observateur humain a identifié 27 zones tandis que la segmentation par ACP a identifié 29 zones. Parmi ces zones, 22 se recouvrent.

### 6.3.4 Segmentation manuelle assistée par la segmentation automatique

Afin d'étudier la succession de phases gestuelles constituant une phrase gestuelle, nous nous sommes appuyés sur la méthode de segmentation proposée par Kita [KvGvdH98]. Cette méthode a initialement été appliquée pour l'annotation manuelle de gestes enregistrés sous forme de séquences vidéo. Bien qu'il existe au moins un travail qui ait appliqué cette spécification à la segmentation automatique des gestes dans un but de resynchronisation entre des mouvements de main et le reste du corps [MZF06], cette méthode s'est avérée insuffisante pour segmenter de manière non ambiguë les différentes phases du mouvement. Aussi avons nous eu recours à une segmentation manuelle assistée par une segmentation automatique. Selon Kita [KvGvdH98], les phases de mouvement sont caractérisées soit par un changement brutal de l'orientation de la main, soit par une discontinuité dans le profil de vitesse du mouvement décrit par le poignet. Comme l'identification des changements brutaux du poignet sont difficiles à repérer par la lecture manuelle d'enregistrements vidéos, nous avons enrichi le processus de segmentation décrit par Kita par l'introduction de deux canaux supplémentaires extraits de la représentation tridimensionnelle du mouvement.

Ces canaux supplémentaires sont dédiés à l'évaluation du changement de configuration des mains au cours du temps. Un canal correspond à une main, à chaque posture  $i$  dans la séquence capturée, une distance  $D_i$  est calculée selon deux postures manuelles consécutives. Ces deux postures sont préalablement centrées et alignées dans l'espace cartésien. La somme des distances au carré entre les positions des articulations correspondantes des deux postures à com-

parer donne la distance entre ces deux postures :  $D(A_i, A_{i+1}) = \sum_j \|(\vec{P}_j - \vec{P}'_j)\|$ .

Nous fixons un seuil pour décider d'une zone de changement de configuration lorsque la distance entre deux postures dépasse deux déviations standard au-delà de la moyenne des distances calculées sur la séquence totale. Les zones de changements ainsi détectées sont ensuite importées sous la forme de fichiers XML dans le logiciel d'annotation interactif Anvil [Kip04].

Le reste du processus de segmentation a été réalisé à la main à partir des enregistrements vidéos des séances de captures et grâce à l'outil ANVIL[Kip04]. La segmentation manuelle a été réalisée le long des trois séquences de gestes capturées et a conduit à l'identification des phases que nous avons présentées dans le chapitre 5. L'illustration 6.12 illustre l'interface d'annotation d'Anvil, selon les canaux que nous avons définis. La section suivante présente les résultats d'ordre qualitatifs et quantitatifs que nous avons obtenus sur les gestes de la langue des signes.

### 6.3.5 Paramètres d'expressivité en langue des signes

#### 6.3.5.1 Variations temporelles et répartition des phases gestuelles

Dans cette section, nous présentons une caractérisation des variations temporelles induites par l'expressivité observées sur les séquences gestuelles capturées. Les résultats se concentrent sur l'organisation des phases gestuelles et sur les variations de profils temporels entre les séquences gestuelles et entre des éléments lexicaux.

**Répartition entre phases** Le tableau 6.2 reprend la répartition des phases de mouvement annotées au long de la séquence gestuelle, selon la qualité expressive. Ce tableau montre que quasiment chaque phase de stroke est précédée d'une préparation. Nous observons en revanche une influence plus sensible de la qualité expressive sur la fréquence des phases de rétraction. Les phases de rétraction permettent de marquer des coupures dans la séquence gestuelle. Par conséquent, nous pouvons supposer qu'une qualité expressive brusque tend à appuyer l'aspect saccadé du mouvement en introduisant des phases de retractions qui renforcent l'idée de rythme. Les phases de tenue sont plus fréquentes lors de la séquence molle. De la même manière que les retractions renforcent l'impression de gestuelle rythmée, les phases de tenue ont tendance à prolonger le mouvement, à le fluidifier en quelque sorte. Ces observations peuvent ainsi être prises en compte lors de la synthèse de phrases gestuelles expressives. La probabilité d'apparition de ces phases étant modélisée à l'aide d'un processus stochastique dont les paramètres seraient inférés à partir d'observations plus larges.

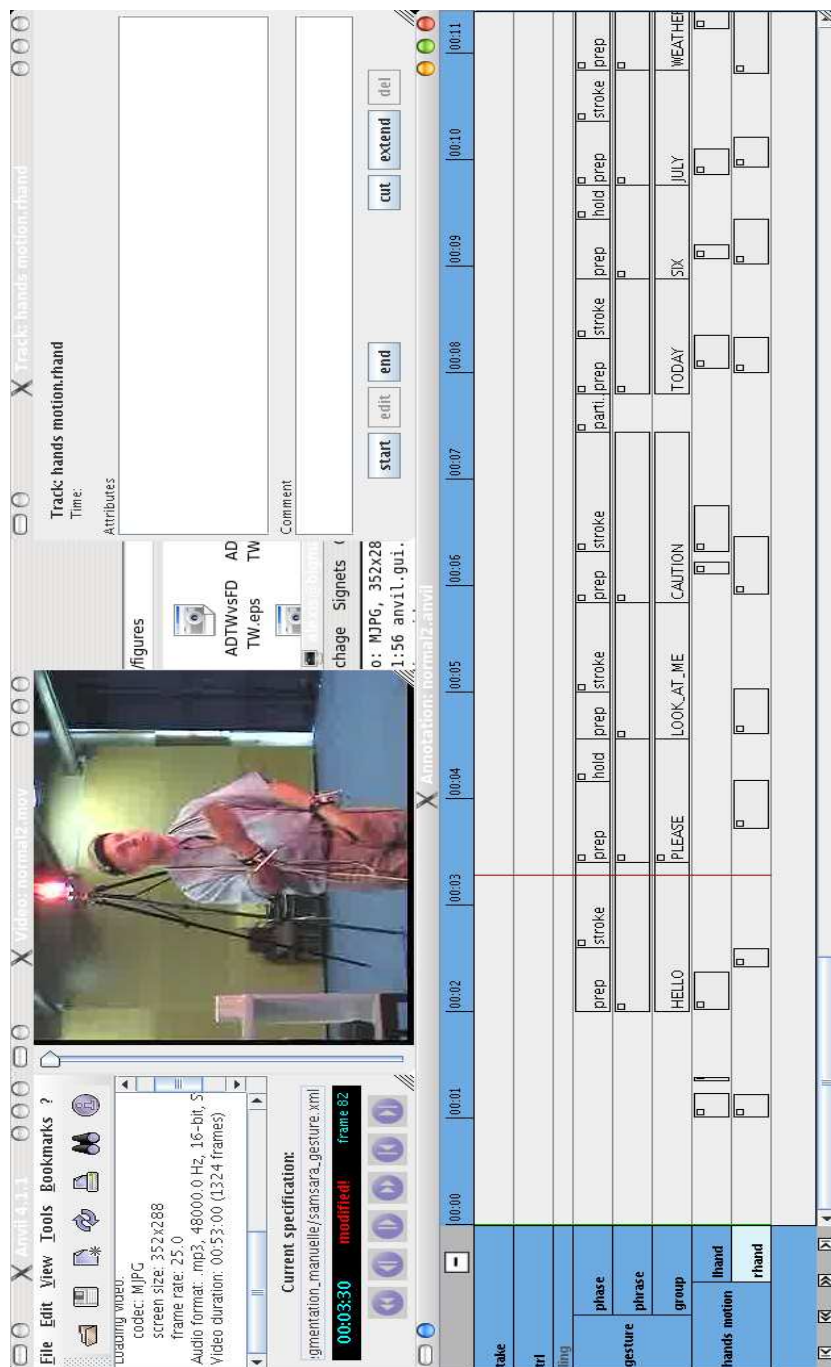


FIG. 6.12 – Session d’annotation manuelle à l’aide de l’outil Anvil. Les canaux “lhand” et “rhand” sont extraits de manière automatique et assistent la segmentation

séquence	préparation	stroke	rétraction	rétraction partielle	tenue	total
neutre	27 (40%)	27 (40%)	2 (3%)	6 (9%)	5 (7%)	67
brusque	26 (39%)	27 (40%)	8 (12%)	2 (3%)	4 (6%)	67
mou	26 (37%)	27 (38%)	9 (13%)	2 (3%)	7 (10%)	71

TAB. 6.2 – Répartition des phases selon la qualité expressive

	Préparation (moyenne, écart-type)	Stroke (moyenne, écart-type)
neutre	38%, 19	93%, 5
brusque	43%, 16	95%, 8
mou	37%, 12	86%, 8

TAB. 6.3 – Ratio  $HS\Delta/Mov$  relevé sur les trois séquences de mouvement selon la phase du mouvement

**Coordination inter-canaux** Afin de comparer les coordinations entre les canaux mis en jeu dans les séquences gestuelles, une analyse temporelle inter-canaux a été réalisée en comparant les phases où s'opère un changement de configuration manuelle combinées à un mouvement réalisé par le complexe bras/épaule.

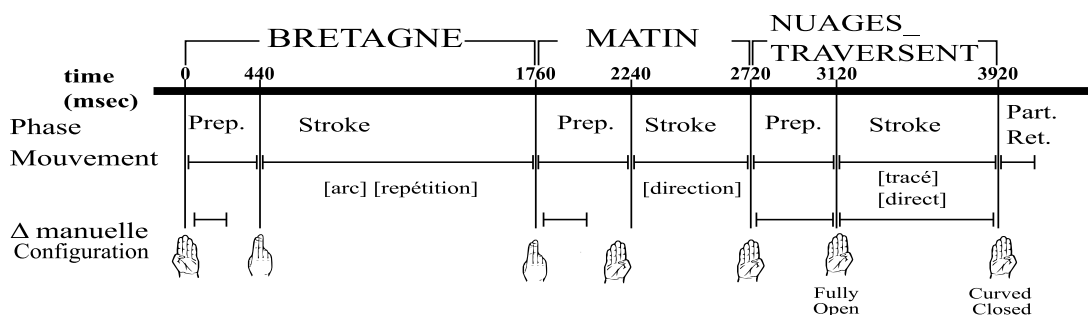


FIG. 6.13 – Identification des phases de mouvement et relations temporelles observées entre les mouvements du bras et de la main

En établissant des comparaisons entre la durée des phases bras/mains, nous avons pu établir des schémas de coordination systématiques. Le ratio entre la durée de changement d'une configuration manuelle et la durée de déplacement de la main donne une information sur la coordination entre les canaux dédiés au mouvement de la main et à la configuration manuelle. Les résultats sont exprimés pour chaque qualité d'expressivité dans le tableau 6.3.

Le tableau 6.3 nous renseigne, d'une part, sur les similarités entre la configuration manuelle et le mouvement de la main en fonction du style et, d'autre part, sur les différences observées entre le couplage des changements de configuration manuelle et des mouvements de la main entre les phases gestuelles.

Il se trouve que les changements de configuration/mouvement apparaissant au sein d'un signe standard présentent un fort couplage tandis que des phases non porteuses de sens, telles que les phases de préparation ou de rétraction, présentent un faible couplage. Ces observations coïncident avec les hypothèses émises par les études dédiées à la phonologie de la langue des signes qui stipulent que les synchronisations fortes entre les canaux lors de mouvements porteurs de sens sont un indice vers la reconnaissance de "syllabes" [Bre98], ou d'unités de segments minimaux en langue des signes. A l'inverse, des gestes non porteurs de sens ne présentent pas une telle coordination entre articulatoires. Lors de la synthèse, ces observations constituent un indice important sanctionnant la coordination ou la non-coordination des canaux intervenant en langue des signes.

**Profils d'appariement temporel entre signes** Nous nous intéressons ici aux variations temporelles qui apparaissent entre plusieurs réalisations d'un même mouvement simple exécuté selon des qualités expressives différentes. Afin de mettre en évidence les déformations temporelles induites par les différents traits d'expressivité entre des gestes simples, des alignements temporels ont été réalisés pour chaque unité lexicale contenant un mouvement simple. La méthode utilisée pour réaliser ces alignements a été présentée dans la partie 6.2.2.3.

Les chemins d'appariement temporels présentés dans la figure 6.14 ont été calculés en prenant en compte la phase de préparation précédant le geste. Nous avons choisi d'inclure la phase de préparation dans nos calculs de chemins d'appariement pour la raison qu'il est difficile de déterminer une limite franche entre la fin d'une phase de préparation et le début du stroke suivant.

La figure 6.14 illustre les déformations temporelles induites par les traits expressifs employés lors de l'exécution des gestes. Les courbes représentées en traits clairs représentent les chemins d'appariement temporels obtenus grâce à l'algorithme aDTW entre un mouvement simple réalisé selon une qualité expressive et son équivalent neutre. La courbe la plus épaisse représente une spline obtenue par régression cubique sur l'ensemble des courbes calculées. L'intérêt de telles courbes en synthèse est de pouvoir moduler les aspects temporels de l'expressivité lors de la production des gestes, tel que le montre l'illustration 6.15.

### 6.3.5.2 Caractéristiques spatiales et structurelles de l'expressivité

Cette section décrit les variations spatiales et structurelles qui ont été observées entre différentes réalisations de gestes simples selon des qualités expressives différentes.

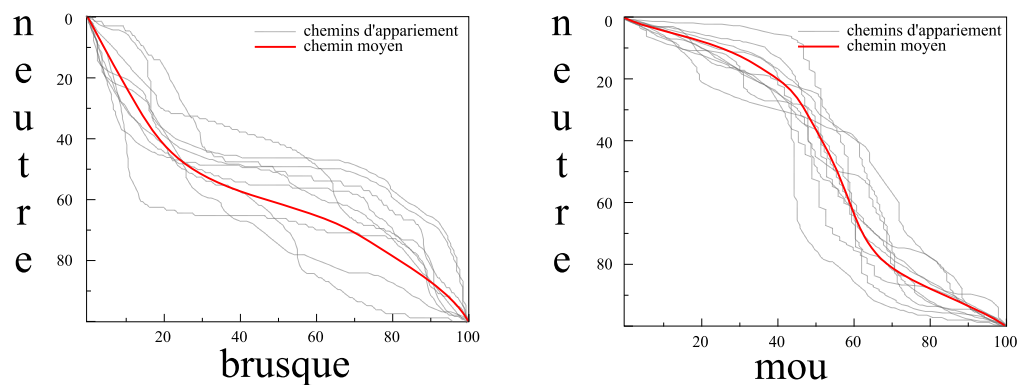


FIG. 6.14 – Alignements temporels extraits de gestes simples réalisés selon des qualités expressives différentes

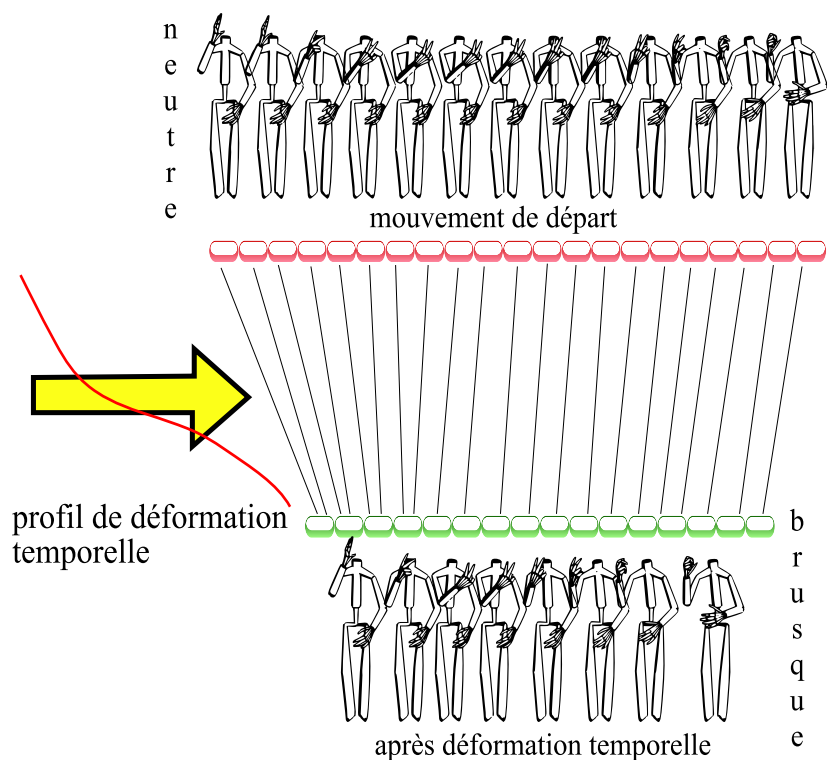


FIG. 6.15 – Application des profils d'alignements inter-styles à la synthèse de gestes expressif



**Proximalisation/Distalisation** Les différentes qualités expressives sont susceptibles de modifier l'amplitude d'un geste. En d'autres termes, les facteurs d'expressivité sont susceptibles de dilater ou de contracter la trajectoire accomplie par les poignets dans l'espace de signation du signeur. Ce phénomène a été décrit à la fois dans la littérature dédiée aux agents conversationnels [Zha01, PML04] et dans la littérature consacrée à la phonologie des langues des signes [Bre98] où le phénomène est décrit sous le terme de *proximalisation/distalisation* et se traduit par une migration des articulations mises en œuvre dans la réalisation d'un mouvement. Un exemple de proximalisation/distalisation tiré de notre corpus d'étude est représenté à la figure 6.16a. Un geste simple de type signe standard (VENT) initialement réalisé de manière neutre (image du centre) peut être distalisé (image de gauche), ou proximalisé (image de droite). Lorsque qu'un geste est distalisé, les articulations mises en jeu dans la production du geste ont tendance à s'éloigner du corps du signeur (distales). Cependant, lorsqu'un geste est proximalisé, les articulations mises en jeu dans la production du geste se rapprochent du corps du signeur (contribution des épaules et du corps).

Un autre exemple est donné à la figure 6.16b avec le geste simple tiré du signe IL\_FAUT, réalisé dans la forme neutre par les articulations des coudes (image du centre). Ce mouvement, dans sa forme distalisée, est réalisé en partie par les articulations des poignets, tandis que dans sa forme proximalisée, la contribution des épaules devient prépondérante.

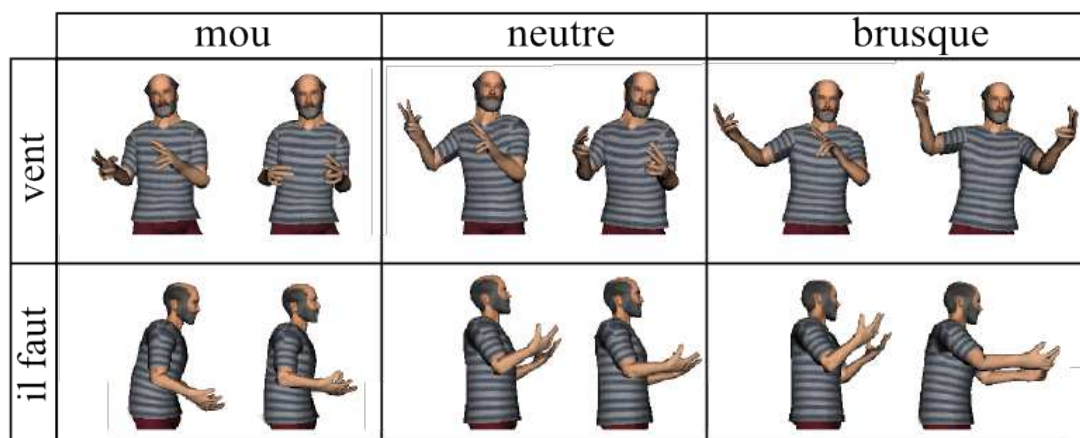


FIG. 6.16 – Exemple de signes subissant la proximalisation / distalisation

Le mécanisme de proximalisation/distalisation peut être modélisé à partir du formalisme introduit au chapitre 5 en migrant les articulations normalement mises en jeu lors de l'exécution du mouvement. La détermination de l'amplitude et de la direction (proximale ou distale) de cette migration est

sanctionnée par une valeur scalaire [prox/dist] comprise entre 0 (mouvement réalisé par la tête et le torse) et 5 (mouvement réalisé par les phalanges).

**Chute de la main dominée** La chute de la main dominée est observée lorsqu'un signe initialement bi-manuel est réalisé par une seule main, en l'occurrence la main dominante. Parmi les réalisations de signes que nous avons capturées, ce phénomène se produit avec le signe PLUIE dans la séquence exécutée selon une qualité expressive molle. La figure 6.17 illustre ce mécanisme. Le mécanisme de *weak drop* a notamment été décrit par Battison [Bat78] qui a identifié un certain nombre de contraintes favorisant ou interdisant la chute de la main dominée. Par exemple, la chute de la main dominée ne peut apparaître sur les signes à deux mains mobiles subissant un mouvement alterné.

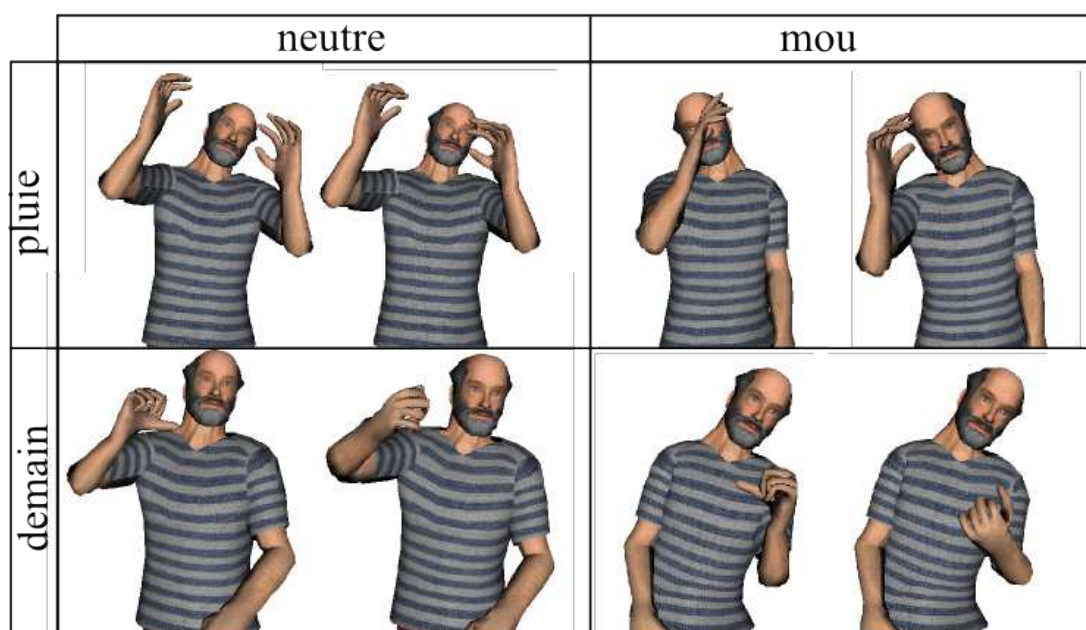


FIG. 6.17 – Exemples de signe subissant une chute de main dominée

Le signe PLUIE est un signe bi-manuel à deux mains mobiles. Lors du mouvement, les deux mains sont en phase et aucun contact n'a lieu. Le signe PLUIE est donc un candidat idéal pour subir l'abandon de la main dominée. L'abandon de main dominée est révélateur d'un état de décontraction, d'un niveau de langue dit relâché. L'état interne de l'agent et le contexte de l'énonciation ont donc une influence sur la probabilité d'observer un abandon de main dominée. Cela correspond bien à l'expressivité molle du mouvement où l'on voit apparaître le signe PLUIE avec chute de main dominée.

Le modèle de description présenté au chapitre 5 permet de modéliser directement le phénomène de chute de la main dominée. Le phénomène est ainsi représenté par l'abandon de l'attribut [bi-manuel]. La description formelle des gestes proposée permet également de décider si un geste est potentiellement éligible à la chute de main dominée : un geste bi-manuel présentant un contact ne peut être sujet à la chute de main dominante, il en est de même pour un signe présentant un mouvement alterné (ex. bicyclette).

### **6.3.6 Conclusion**

Cette section nous a permis de présenter une étude qualitative et quantitative de gestes de la langue des signes, et de l'influence de paramètres d'expressivité s'exprimant sur les plans spatiaux, temporels et structurels. Les causes des facteurs d'expressivité peuvent être liées aux caractéristiques internes de l'agent signeur ou au contexte du discours. Les différentes analyses des mouvements capturés nous ont permis d'identifier certains facteurs de la variabilité des gestes de la LSF, à la fois d'un point de vue sémantique et structurel et d'un point de vue du signal. L'étude de la variabilité à un niveau de description macroscopique se ramène à l'étude de la variabilité inter-phases et inter-canaux : elle met en évidence des schémas de coordination motrice, et la façon dont ces schémas varient d'une exécution gestuelle à une autre. Ces résultats nous renseignent sur les schémas de coordination à prendre en compte lors de la synthèse. L'étude de la variabilité spatiale et structurelle s'est cantonnée à une analyse qualitative, basée sur les résultats de segmentation. Les différents facteurs d'expressivité ont été associés au modèle de description du geste présenté au chapitre 5 afin de permettre leur prise en compte lors de la synthèse. A l'avenir, il est nécessaire de réaliser une analyse quantitative de la variabilité spatiale des gestes en fonction de leur qualité expressive. Une telle évaluation nécessite la mise en place de métriques fiables permettant d'évaluer l'extension spatiale du geste.

## 6.4 Synthèse de gestes de la LSF

Les sections précédentes ont proposé une analyse des mouvements en termes de paramètres d'expressivité en langue des signes. Ces paramètres d'expressivités ont ensuite été placés dans le contexte de notre modèle de spécification du mouvement présenté au chapitre 5. Cette section présente les méthodes et algorithmes de génération de mouvements qui s'appuient sur les modèles de spécification.

### 6.4.1 De la spécification à la génération

L'illustration de la figure 6.18 présente une représentation en couches de notre architecture de génération de mouvement. Cette architecture s'appuie en entrée sur une phrase gestuelle découpée en syntagmes. Ces syntagmes sont ensuite découpés en phases, elle-mêmes spécifiées selon le modèle proposé au chapitre 5. Les spécifications choisies sont ensuite modulées selon le style, d'un point de vue spatial, structurel et temporel.

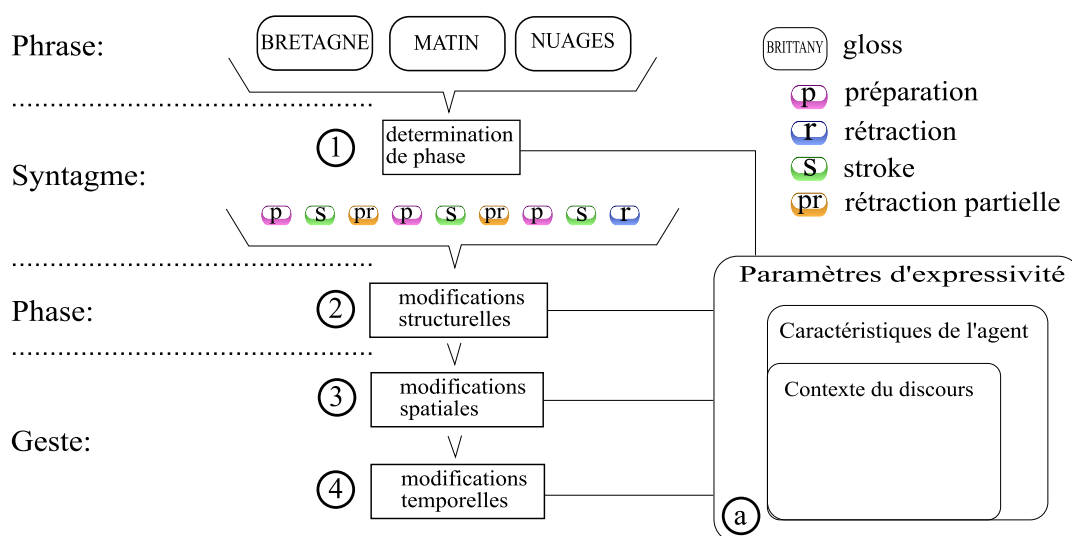


FIG. 6.18 – Architecture générale du processus de génération

A chaque niveau, les paramètres d'expressivité déterminés au préalable influencent les traitements. Au niveau de la détermination des phases, un mouvement traduisant une qualité expressive brusque comporte par exemple des phases de rétraction plus nombreuses. Au niveau du traitement d'une phase, et de la spécification du geste élémentaire, des modifications relatives à la

structure du geste, à sa vitesse d'exécution ou à son extension spatiale permettent également de moduler la qualité expressive des gestes générés.

D'une part, le besoin de gérer de concert des groupes d'effecteurs différents et, d'autre part, le besoin d'une méthode de génération de mouvements à la fois flexible et précise nous impose la mise en œuvre d'algorithmes de génération de mouvements complémentaires. En effet, la spécification du geste peut tantôt être délivrée dans l'espace cartésien, sous la forme d'une direction à suivre, d'une cible à atteindre ou d'une trajectoire à décrire, tantôt délivrée dans l'espace des articulation sous la forme d'une séquence à rejouer. Nous verrons dans la suite de cette section comment est traitée la cinématique inverse pour des groupes d'effecteurs simultanément mis en jeu dans la production des gestes, et comment est spécifiée la tâche pour chaque groupe d'effecteur.

#### 6.4.1.1 Méthodes de génération

Selon le groupe d'articulation choisie, une méthode de contrôle d'animation est préférable à une autre. Par exemple, les mouvements "ambient" traduisant la posture et attitude générale du signeur peuvent être synthétisés à l'aide de boucle de mouvements pré-capturés sur le corps du signeur [EMT05]. Notre méthode de génération de gestes comprend trois modes de génération : par jeu pur, par interpolation ou par cinématique inverse. Selon la spécification du geste, selon la phase de mouvement considérée et selon le groupe d'articulateurs mis en jeu, une méthode est préférable. Par exemple, les phases de préparation peuvent être avantageusement générées par interpolation. Dans ce cas, il suffit de spécifier une posture de départ, une posture de fin et une durée pendant laquelle opérer la transition. S'ils sont disponibles, certains signes standards peuvent être directement rejoués à l'aide de leurs trajectoires articulaires. Ces trajectoires peuvent éventuellement être déformées temporellement selon les profils de déformations déterminés en 6.3.5.1. Les phases de gestes produites en visée iconicisatrices requièrent en revanche des méthodes de génération par cinématique inverse. En effet, l'iconicité des gestes de la LSF se traduit essentiellement par un ensemble de relations spatiales entre entités participant au mouvement produit (construction d'un espace 3D de signation, positionnement des référents dans cet espace).

#### 6.4.2 Cinématique inverse

La cinématique inverse regroupe les méthodes permettant d'obtenir le positionnement d'une partie d'un squelette articulaire dans l'espace cartésien en cherchant une posture valide dans l'espace articulaire.

### 6.4.2.1 Modèle sensori-moteur

L'approche proposée pour le modèle de génération relève à la fois de l'automatique et des théories du contrôle moteur. Il s'agit d'un *système bouclé* qui exploite un *retour sensoriel* pour accomplir une *tâche donnée*. C'est en ce sens que nous le qualifions de *système sensori-moteur* de génération de mouvements [GM94].

### 6.4.2.2 Résolution de l'inversion

Le modèle de contrôle sensori-moteur repose sur une stratégie de minimisation d'une fonction de potentiel appelée *fonction d'erreur*  $E(x_t, x)$ . La minimisation est obtenue par une méthode de descente de gradient itérée jusqu'à ce que  $E$  ait convergé vers un minimum local. On considère que  $E(x_t, x)$  mesure l'écart entre la position d'observation en cours et la position de la cible à atteindre. Les coordonnées d'état  $\vec{q}$  sont réajustées à l'aide du vecteur  $\vec{\Delta q}$  obtenu à partir du gradient de cette erreur. Ce fonctionnement sensori-moteur est décrit par le système d'équations suivant :

$$\vec{\Delta q} = -g(E(x, x_t)) \cdot \vec{\nabla}(E(x, x_t)) \quad (6.1)$$

Avec  $E(x_t, x)$ , une fonction d'erreur correspondant à la norme du vecteur  $\vec{e} = (\vec{x} - \vec{x}_t)$  au carré.  $E(x_t, x) = \langle \vec{e}, \vec{e} \rangle$ . D'autre part,  $g(E(x, x_t))$  est une fonction de gain dépendante de l'erreur  $E$  que nous précisons en fin de sous-section.

Reprenons dans l'équation (6.1) l'expression du gradient de l'erreur  $\vec{\nabla}(E(x, x_t))$

$$\vec{\nabla}(E(x, x_t)) = \begin{bmatrix} \frac{\partial E(x_t, x)}{\partial q_1} \\ \vdots \\ \frac{\partial E(x_t, x)}{\partial q_n} \end{bmatrix} \quad (6.2)$$

or  $E(x, x_t) = |\overrightarrow{M(q)} - \vec{x}_t|^2$ , l'expression (6.2) peut s'écrire sous la forme d'un produit de matrices :

$$\vec{\nabla}(E(x, x_t)) = 2 \begin{bmatrix} \frac{\partial(M(q)_1)}{\partial q_1} & \cdots & \frac{\partial(M(q)_m)}{\partial q_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial(M(q)_1)}{\partial q_n} & \cdots & \frac{\partial(M(q)_m)}{\partial q_n} \end{bmatrix} \begin{bmatrix} M(q)_1 - (x_t)_1 \\ \vdots \\ M(q)_m - (x_t)_m \end{bmatrix} \quad (6.3)$$

Nous reconnaissons dans l'expression (6.3) la transposée de la Jacobienne de  $M(q)$ ,  $J^T$  ainsi que le vecteur d'erreur  $\vec{e}$ . Par conséquent,  $\vec{\nabla}(E(x, x_t))$  se résume à :

$$\vec{\nabla}(E(x, x_t)) = 2J^T \vec{e} \quad (6.4)$$

Ce qui nous amène à une expression directe du vecteur  $\vec{\Delta q}$  en remplaçant le gradient de l'erreur dans (6.4) :

$$\vec{\Delta q} = -g(E(x, x_t)) \cdot J^T \vec{e} \quad (6.5)$$

Cette équation correspond effectivement à la méthode d'optimisation par descente de gradient et garantit la convergence vers un minimum local pour  $g$  suffisamment petit.

D'un point de vue algorithmique, le fait de ne pas avoir à calculer la pseudo-inverse de la matrice Jacobienne simplifie la complexité de chaque pas de résolution. Le fait que la boucle de résolution ne comprenne que des éléments de cinématique directe permet également de se soustraire des problèmes de singularités des matrices jacobiniennes lors de configurations pathologiques (chaîne cinématique tendue).

Afin de prendre en compte plusieurs contraintes simultanées lors d'une résolution, nous adoptons le mélange de contraintes par priorité, tel qu'il a été défini par Badler et al [BPW93].

#### 6.4.2.3 Validation du modèle

Afin de tester les performances et la robustesse de notre solution de cinématique inverse, une application interactive permettant de gérer en temps interactif plusieurs chaînes articulaires commandées en position et en orientation a été développée. Cette application met en jeu un avatar sur lequel sontinstanciées trois chaînes articulaires correspondant à la main droite, gauche et à l'orientation du visage. Le tableau suivant reprend les articulations entrant en jeu dans chaque chaîne articulaire.

Chaîne	Art1	dof	Art2	dof	Art3	dof	Art4	dof
larm	lclavicle	2	lhumerus	3	lradius	2	lhand	2
rarm	rclavicle	2	rhumerus	3	rradius	2	rhand	2
eyes	head	2	headDir					

TAB. 6.4 – Chaînes articulaires mises en jeu dans l'application

Chaque chaîne cinématique est commandée par la spécification d'une cible à atteindre par l'extrémité de la chaîne. La spécification de la cible est réalisée à la fois en position et en orientation. Dans l'interface utilisateur, les cibles sont représentées par des trièdres fournissant une représentation symbolique de la

position et de l'orientation de la cible spécifiée. Afin de donner à l'utilisateur la possibilité de spécifier les cibles avec précision, trois vues en projection orthogonales accompagnent une vue en perspective. La figure 6.19 fournit une capture d'écran de l'interface.

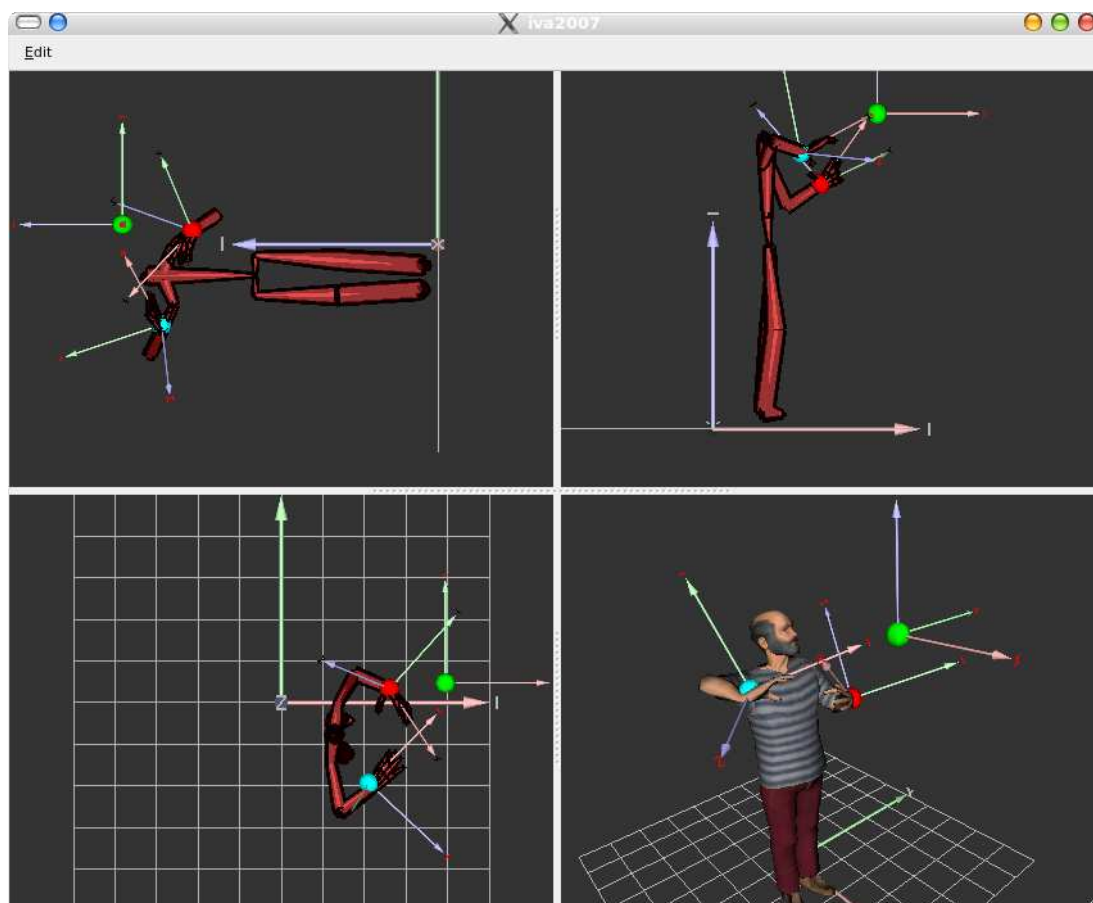


FIG. 6.19 – Interface interactive pour l'évaluation de la génération sensori-motrice

Cette interface nous a permis de valider la robustesse et la pertinence du modèle sensorimoteur de génération du geste. Elle nous a également permis d'ajuster les constantes intervenant dans les fonctions de gain afin de trouver le meilleur compromis entre stabilité et réactivité des chaînes aux changements de configuration.

Les trajectoires de cinématique inverse sont obtenues en manipulant en temps interactif deux chaînes articulées (bras gauche et droit) mues à l'aide du modèle sensori-moteur. La commande est spécifiée à l'aide de cibles, chaque chaîne étant commandée par trois cibles formant un trièdre. La prise en compte



des cibles multiples est réalisé par mélange.

Les cibles suivent une trajectoire préenregistré et sont mises à jour toutes les 40 millisecondes. Les configurations des chaînes articulées sont recalculée toutes les 10ms. La résolution d'une chaîne cinématique (9 degrés de liberté) prend 420 microsecondes. sur un ( AMD Athlon(tm) 64 Processor 3500+ ).

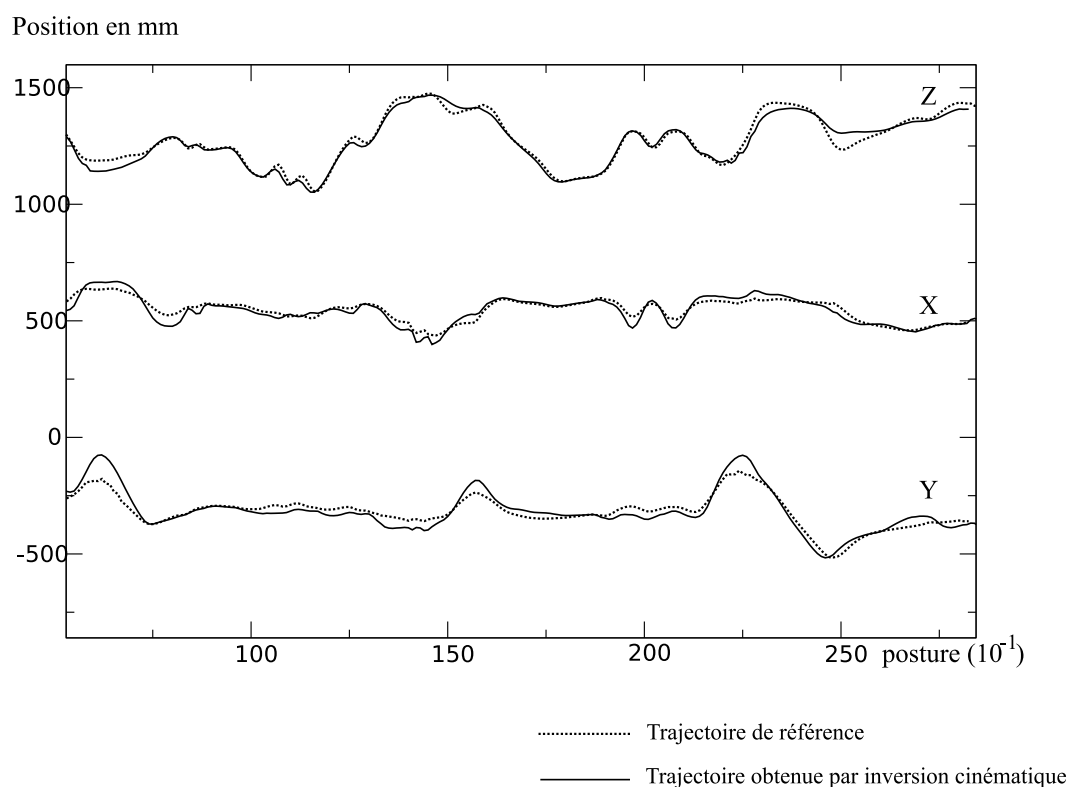


FIG. 6.20 – Modèle sensori-moteur

Le graphique de la figure 6.20 compare les différences entre les trajectoires du poignet de la main droite obtenues à l'aide de l'inversion cinématique et les trajectoires de mouvements de référence capturées. L'écart maximal entre les deux trajectoire est inférieur à 20mm.

### 6.4.3 Prototype de composition des gestes

Nous avons vu dans les chapitres 3 et 5 combien il était important de prendre en compte l'aspect multilinéaire dans la création et la spécification de gestes. Afin d'illustrer cet aspect multi-linéaire, et de mettre en évidence l'importance des synchronisations entre les différents canaux, une application interactive a été développée. L'interface utilisateur est présentée à l'illustration 6.21.

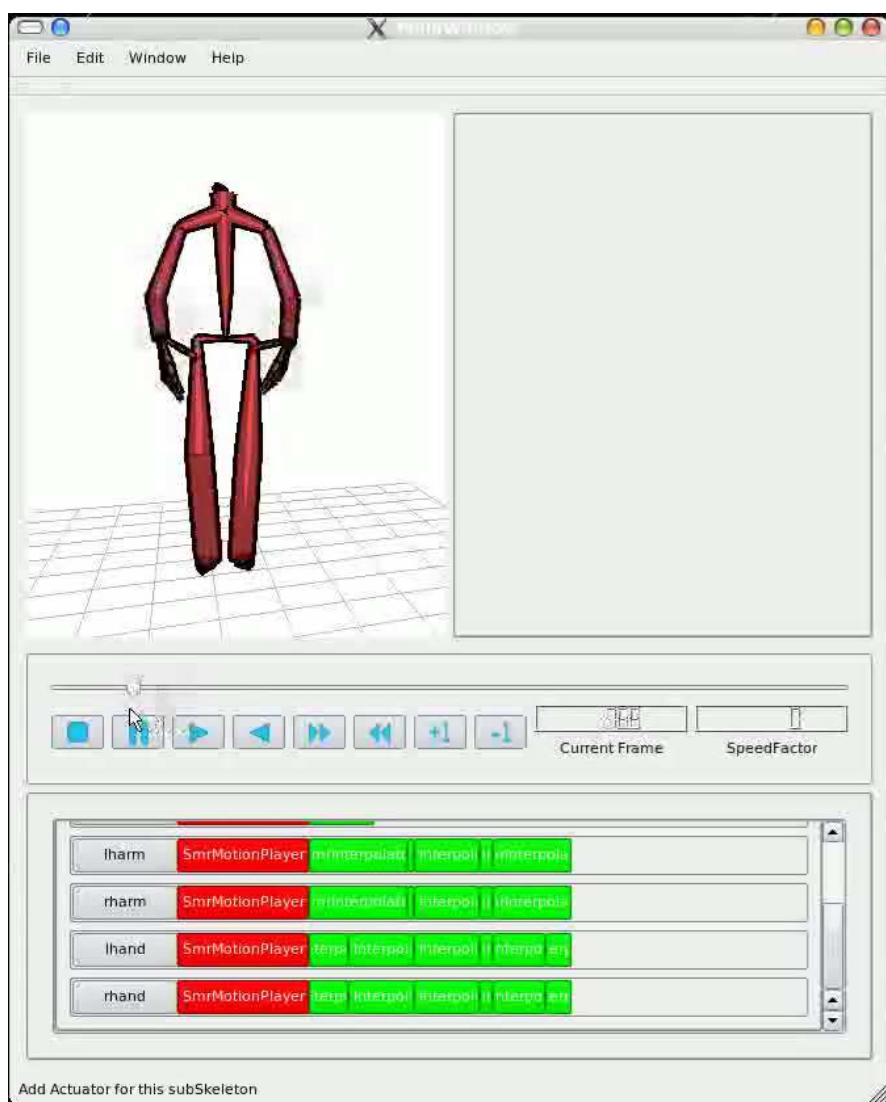


FIG. 6.21 – Interface interactive pour la compositionnalité multi-linéaire des gestes

Cette application permet de construire et d’animer un squelette articulaire à l’aide de briques dédiées au contrôle et à l’animation de parties du squelette global. Chaque sous partie du squelette peut se voir attribuer un actuateur en charge de modifier sa configuration. Trois types d’actuateurs sont pour l’instant pris en compte : interpolation entre deux postures, spécification d’une pose par cinématique inverse et rejeu de séquence préalablement capturée. Son interface utilisateur permet de générer interactivement des séquences gestuelles tout en prenant en compte les différentes phases intervenant dans la généra-

tion d'un geste. Les éventuelles synchronisations entre articulateurs sont pour l'instant laissées à la charge de l'utilisateur. Cette application ainsi que celle présentée dans la section précédente se basent sur une bibliothèque dédiée à la synthèse d'image que nous avons baptisée *Smr*. A l'heure où ce mémoire est rédigé, la bibliothèque *Smr* fait l'objet d'un dépôt APP<sup>7</sup>.

#### 6.4.4 Conclusion

Cette section a brièvement présenté les méthodes de génération du mouvement employées dans le cadre de la synthèse de gestes de la LSF. Dans un premier temps, les étapes séparant la spécification d'une séquence gestuelle à la production du mouvement ont été présentées. Parmi ces étapes, nous retrouvons la spécification du geste en phases et la prise en compte de paramètres d'expressivité à chaque niveau. L'aspect multilinéaire propre aux LS requiert la prise en compte de jeu de groupes d'effecteurs différents, pouvant évoluer de manière indépendante ou liée. Afin de mettre en évidence l'influence des différents paramètres d'expressivité, une application de génération interactive du mouvement a été réalisée. Cette application permet de construire des séquences gestuelles prenant en compte des jeux d'effecteurs différents regroupés en canaux indépendants. Une suite immédiate à donner à ce travail serait de générer suffisamment de séquences pour être en mesure de quantifier l'influence de chaque paramètre expressif.

---

<sup>7</sup>Agence pour la protection des programmes. <http://app.legalis.net/>



# Conclusion

Les travaux présentés dans cette thèse concernent la spécification et la génération de séquences gestuelles expressives en langue des signes française à partir de l'analyse extraite de données capturées. Nous nous sommes en particulier intéressés à la prise en compte de l'expressivité et de son implication dans le cadre de la visée iconicisatrice qui sous-tend la langue des signes.

Après avoir rappelé le contexte social et historique dans lequel s'inscrivent nos travaux, nous avons présenté en chapitre 3 les mécanismes linguistiques fondamentaux qui structurent la langue des signes. Ces mécanismes sont décrits dans le cadre de l'iconicité, qui a été définie par le lien de ressemblance entre ce qui est dit et la réalité décrite. Il s'agit de dire en montrant. Un locuteur qui a recours à la visée iconicisatrice déploie l'éventail des structures de grande iconicité qui sont à sa disposition. Parmi celles-ci nous distinguons les transferts de taille ou de forme, les transferts de situation et les transferts personnels. Bien que ces transferts puissent se combiner pour former des structures plus complexes, nous nous sommes limités à une représentation des trois principaux transferts à l'aide de notre modèle. Le choix de l'une de ces structures implique une prise de décision qui dépend du contexte de discours et d'attributs propres au signeur.

Les informations portées par le contexte de discours et l'état interne du signeur ayant une influence sur le mode de production des séquences de gestes en langue de signes, nous avons proposé dans le chapitre 4 un modèle du discours prenant en compte les aspects propres à l'énoncé et aux éléments constitutifs de l'énoncé regroupés sous la notion d'entité. L'apport de notre modèle est de proposer une tentative d'organisation des éléments du discours fournissant les éléments nécessaires à la prise de décision du mode d'énonciation (visée non-iconicisatrice ou structures de grande iconicité).

La génération informatisée de gestes expressifs en langue de signes nécessite un modèle de spécification. Les travaux existants portant sur ces modèles de spécification se sont cantonnés à proposer des approches directement inspirées des travaux sur sa phonologie. En conséquence, les séquences générées à partir de ces modèles peuvent difficilement dépasser la construction de séquences par agencement lexical. En particulier, ces modèles ne prenant pas en

compte les structures de grande iconicité évoquées ci-dessus ne peuvent prétendre adresser la langue des signes dans sa globalité.

Le modèle que nous proposons dans le chapitre 5 offre une spécification du geste au niveau articulatoire. Cette spécification de bas niveau permet de gérer un large éventail de mouvements. Ces mouvements peuvent appartenir à des phases expressives et porter du sens dans le cadre d'une visée iconisatrice ou non-iconisatrice. Ils peuvent également appartenir à des phases non-expressives telles que les phases de préparation ou de rétraction.

La caractérisation des paramètres d'expressivité des gestes de la langue de signes nécessite l'étude de performances réelles réalisées par des sourds signeurs. Nous avons pu bénéficier de données de mouvements capturés sur un signeur sourd qui illustrent une séquence gestuelle exécutée selon différents paramètres expressifs. L'analyse des différentes réalisations de la séquence, présentée au chapitre 6, nous a permis de mener une étude à la fois qualitative et quantitative des paramètres d'expressivité en langue de signes. Nous proposons un algorithme original d'alignement temporel adaptatif qui caractérise globalement la qualité expressive des mouvements effectués sur les séquences de signal complètes. Nous avons également identifié des variations d'ordre structurel au niveau de l'organisation gestétique des signes et des modifications de la répartition de certaines phases des gestes non-porteurs de sens, telles que les phases de préparation et de rétraction.

La génération des gestes de LSF par ordinateur nécessite la mise en œuvre d'algorithmes capables de gérer l'évolution des différents articulateurs du squelette d'animation au cours du temps, conformément à la spécification du geste passée en entrée du modèle de génération. Nous avons proposé une méthode de génération des mouvements mélangeant rejeu, interpolation et cinématique inverse permettant de gérer de manière robuste et souple plusieurs chaînes articulaires.

La représentation, la modélisation et la génération d'un signal issu d'une langue naturelle sont extrêmement complexes et nécessitent des collaborations transdisciplinaires fortement impliquées. Une connaissance superficielle des mécanismes linguistiques structurant le signal à produire peut engager tout un travail de recherche sur une voie stérile ou, pire, sur une voie qui renie la spécificité de la communauté ayant adopté la langue en question.

Notre travail de recherche a été à maintes reprises remis en question au fur et à mesure que notre compréhension des mécanismes structurant la langue des signes se précisait. Ces remises en cause nous ont permis de corriger et de raffiner progressivement notre modèle. Elles nous ont également forcés à considérer des aspects d'ordre linguistique et à nous éloigner des préoccupations consacrées exclusivement à l'animation par ordinateur. Il existe toujours un décalage entre les modèles linguistiques décrivant une langue et les modèles

informatiques les mettant en oeuvre. La langue de signes ne fait pas exception à la règle. L'apport de ce travail de recherche pourrait être de réduire le décalage existant.





# Annexe A

## Corpus météo

Texte initial Bonjour

Attention, s'il vous plaît, regardez moi.  
Aujourd'hui six juillet, voici le temps prévu.  
Le matin, des nuages traverseront la Bretagne.  
L'après midi, il va pleuvoir.  
Demain, le soleil brillera.  
Il fera chaud et sec.  
ce sera le moment d'aller nager.  
Vendredi, il y aura des orages violents.  
Le vent soufflera très fort.  
La nuit, il y aura des orages violents en mer.  
le matin, il y aura du brouillard en mer.  
A bientôt

Texte transcrit Bonjour

S'il vous plaît / regardez moi / Attention  
Aujourd'hui / 6 juillet / le temps / prévu  
La Bretagne / le matin / des nuages traversent  
L'après midi / il pleut.  
Demain / le soleil brillera  
Chaud / sec / Il fera  
Nager / aller / ce sera le moment  
Vendredi / orages violents / il y aura  
Le vent soufflera fort.  
La nuit / en mer / il y aura des orages violents  
Le matin / en mer / il y aura du brouillard.  
A bientôt



# Annexe B

## Hiérarchie BVH du squelette

```
HIERARCHY ROOT pelvis
{
  OFFSET 0 0 0
  CHANNELS 6 Xrotation Yrotation Zrotation Xposition Yposition Zposition
  JOINT lfemur
  {
    OFFSET 0 96.711 0 CHANNELS 3 Xrotation Yrotation Zrotation JOINT ltibia
    {
      OFFSET 0 0 - 400.394
      CHANNELS 3 Xrotation Yrotation Zrotation JOINT lfoot
      {
        OFFSET 0 0 - 397.564
        CHANNELS 3 Xrotation Yrotation Zrotation JOINT ltoes
        {
          OFFSET 91.659 0 0 CHANNELS 3 Xrotation Yrotation Zrotation End Site
          {
            OFFSET 0 0 0)
          }
        }
      }
    }
  }
}
JOINT rfemur
{
  OFFSET 0 - 96.711 0 CHANNELS 3 Xrotation Yrotation Zrotation JOINT rtibia
  {
    OFFSET 0 0 - 400.394
    CHANNELS 3 Xrotation Yrotation Zrotation JOINT rfoot
    {
      OFFSET 0 0 - 397.564
      CHANNELS 3 Xrotation Yrotation Zrotation JOINT rtoes
      {
        OFFSET 91.659 0 0 CHANNELS 3 Xrotation Yrotation Zrotation End Site
        {
          OFFSET 0 0 0)
        }
      }
    }
  }
}
JOINT thorax
{
  OFFSET - 0 0 150.555 CHANNELS 3 Xrotation Yrotation Zrotation JOINT head
  {
    OFFSET 0 0 364.332 CHANNELS 3 Xrotation Yrotation Zrotation End Site
    {
      OFFSET 0 0 0)
    }
  }
  JOINT lclavicle
  {
    OFFSET 0 0 298.293
    CHANNELS 3 Xrotation Yrotation Zrotation JOINT lhumerus
    {
      OFFSET 0 153.424 - 16.4259
      CHANNELS 3 Xrotation Yrotation Zrotation JOINT lradius
      {
        OFFSET 0 0 - 263.893
        CHANNELS 3 Xrotation Yrotation Zrotation JOINT lhand
        {
          OFFSET 0 0 - 204.185
        }
      }
    }
  }
}
```

```

CHANNELS 3 Xrotation Yrotation Zrotation JOINT Left - hand
{
  OFFSET 0 0 0
  CHANNELS 3 Xrotation Yrotation Zrotation JOINT LThumb
  {
    OFFSET 5 0 10
    CHANNELS 3 Xrotation Yrotation Zrotation JOINT LThumb_CARPE
    {
      OFFSET 0 0 0
      CHANNELS 3 Xrotation Yrotation Zrotation JOINT LThumb_TM
      {
        OFFSET 0 10 0
        CHANNELS 3 Xrotation Zrotation Xrotation
        JOINT LThumb_MCP
        {
          OFFSET 45 0 0
          CHANNELS 3 Xrotation Yrotation Zrotation
          JOINT LThumb_IP
          {
            OFFSET 40 0 0
            CHANNELS 3 Xrotation Yrotation Zrotation End Site
            {
              OFFSET 35 0 0)
            }
          }
        }
      }
    }
  }
}
JOINT LIndex
{
  OFFSET 5 0 10
  CHANNELS 3 Xrotation Yrotation Zrotation JOINT LIndex_MCP
  {
    OFFSET 90 0 0
    CHANNELS 3 Xrotation Yrotation Zrotation JOINT LIndex_IPP
    {
      OFFSET 45 0 0
      CHANNELS 3 Xrotation Yrotation Zrotation
      JOINT LIndex_IDP
      {
        OFFSET 28 0 0
        CHANNELS 3 Xrotation Yrotation Zrotation End Site
        {
          OFFSET 25 0 0)
        }
      }
    }
  }
}
JOINT LMiddle
{
  OFFSET 5 3 - 0
  CHANNELS 3 Xrotation Yrotation Zrotation JOINT LMiddle_MCP
  {
    OFFSET 95 0 0
    CHANNELS 3 Xrotation Yrotation Zrotation JOINT LMiddle_IPP
    {
      OFFSET 55 0 0
      CHANNELS 3 Xrotation Yrotation Zrotation
      JOINT LMiddle_IDP
      {
        OFFSET 30 0 0
        CHANNELS 3 Xrotation Yrotation Zrotation End Site
        {
          OFFSET 25 0 0)
        }
      }
    }
  }
}
JOINT LRing
{
  OFFSET 5 3 - 10
  CHANNELS 3 Xrotation Yrotation Zrotation JOINT LRing_MCP
  {
    OFFSET 90 0 0
    CHANNELS 3 Xrotation Yrotation Zrotation JOINT LRing_IPP
    {
      OFFSET 50 0 0
      CHANNELS 3 Xrotation Yrotation Zrotation JOINT LRing_IDP
      {
        OFFSET 30 0 0
        CHANNELS 3 Xrotation Yrotation Zrotation End Site
        {
          OFFSET 25 0 0)
        }
      }
    }
  }
}

```





# Annexe C

## Représentation de l'agent signeur

Cette section présente le modèle d'agent humanoïde que nous avons employé au long de nos travaux. Ce modèle repose sur une structure squelettique composée de segments de longueur fixe reliés par des d'articulations. Afin de fournir un rendu convainquant, le squelette d'animation déforme une surface continue, fermée et déformable sur laquelle sont projetés une série de textures renforçant le réalisme des postures affichées. La décomposition de notre humanoïde peut suivre la hiérarchie proposée par Chadwick [CHP89].

L'enveloppe externe habille le squelette de contrôle de l'agent. Elle permet de représenter de manière continue les éléments superficiels de l'agent tels que sa peau, son visage et ses vêtements. Une surface est continue et susceptible d'être déformée par plusieurs éléments du squelette du contrôle.

### C.1 Squelette d'animation

La plupart des méthodes d'animation de personnages virtuels reposent sur une représentation hiérarchique dont un certain nombre d'os idéalisés forment les arcs. Les noeuds reliant les os du graphe d'animation représentent les articulations. Les articulations sont vues comme des liaisons mécaniques possédant des degrés de liberté (*ddl*) en rotation. L'animation d'un humanoïde synthétique peut être réalisée en agissant sur les degrés de liberté en rotation des articulations du squelette.

Cette représentation hiérarchique est basée sur des articulations parfaites. Les segments étant réputés de longueur fixe, la position de points caractéristiques du squelette  $\vec{X}$  ne dépend que des angles appliqués à chaque degré de liberté  $\vec{q}$ . Les rotations induites par les différentes articulations peuvent être paramétrées de différentes façons.

- Les matrices homogènes permettent de représenter les repères sous forme

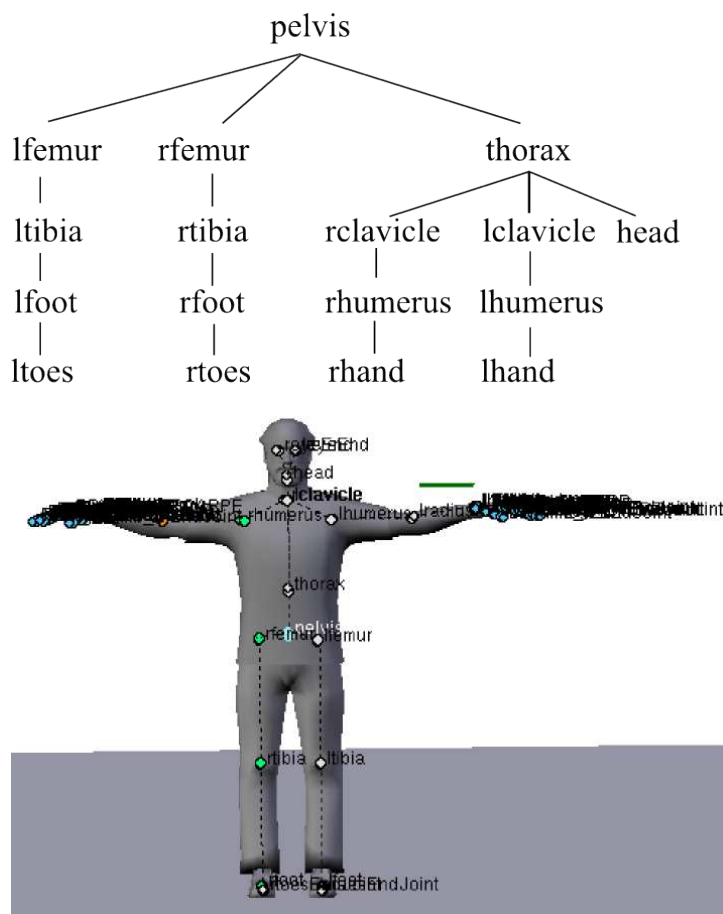


FIG. C.1 – Représentation hiérarchique du squelette

matricielle. Il est alors possible de les cumuler en utilisant une multiplication de matrices. Les matrices homogènes permettent de coupler les rotations, cependant, cette représentation pose un problème de singularité pour l'interpolation.

- Les angles d'Euler sont décrits par trois rotations successives autour des axes X, Y et Z. Cette représentation simple, compacte et intuitive pose néanmoins un problème de singularité connu sous le nom de "Gimbal Lock". Cette singularité se traduit par la perte d'un degré de liberté après une séquence de rotations d'angle  $\frac{\pi}{2}$ .
- Une articulation peut être représentée à l'aide du formalisme *swing-twist* [BB00, Kor85]. Ce formalisme consiste à représenter les trois degrés de liberté en rotation d'une articulation à l'aide d'une première rotation qui sanctionne la direction du segment rattaché à l'articulation puis d'une rotation du membre autour de son axe. Cette représentation a l'avantage



d'exprimer de façon naturelle les articulations de l'humain et offre une paramétrisation réaliste des butées articulaires à l'aide d'ellipses sphériques.

- Les quaternions, proposés par Lord Williams Hamilton [Ham44] sont une extension des nombres complexes. L'extension est réalisée en introduisant trois axes orthonormés  $i$ ,  $j$  et  $k$  tels que  $ii = jj = kk = ijk = -1$ . Un quaternion  $Q$  s'exprime donc à l'aide de quatre scalaires dans un espace de dimension quatre tel que  $Q = w + xi + yj + zk$ . Les rotations peuvent être représentées à l'aide de quaternions unitaires. L'expression des rotations sous la forme de quaternions permet des interpolations optimales, empruntant le plus court chemin (géodésique) sur la sphère unitaire dans l'espace des quaternions.
- Les cartes exponentielles [Gra98] sont une simplification des quaternions, ou, plus exactement le résultat de la projection des rotations exprimées dans l'espace des quaternions de dimension 4 sur un hyperplan de dimension 3 tangent à l'hypersphère unitaire et centré sur un quaternion représentatif de l'ensemble des rotations à représenter. Cette approximation locale permet de linéariser la représentation des quaternions, donnant ainsi accès à l'ensemble des opérations définies dans l'espace euclidien, au prix de l'approximation induite par la projection sur le plan tangent.

Ces différentes paramétrisations ont été comparées dans des articles de référence [Gra98, BB00]. La conclusion est qu'il n'y a pas de représentation "optimale", mais qu'il revient à l'utilisateur la responsabilité de choisir la paramétrisation la mieux adaptée à l'usage qu'il souhaite faire des rotations dans le système qu'il conçoit. Notre système répond à deux préoccupations : il doit être capable de rejouer des mouvements capturés par les méthodes que nous avons décrites au chapitre six et doit être capable de réaliser une adaptation de mouvements selon les spécifications que nous avons formulées au chapitre cinq. Dans un contexte de rejeu de mouvement, la représentation des rotations à l'aide de quaternions s'impose. Cette représentation permet des interpolations simples et robustes, ce qui s'avère utile dans un contexte de mélange de mouvements basé sur les valeurs d'orientation des articulations ainsi que sur les interpolations temporelles. Dans un contexte d'adaptation de mouvement, la paramétrisation des rotations à l'aide du formalisme *swing-twist* est en revanche la paramétrisation qui nous a semblé la mieux adaptée. Nous avons donc choisi d'utiliser ces deux représentations dans notre système de génération de mouvements. Plutôt que de proposer des méthodes permettant de passer d'une représentation à l'autre, nous avons choisi de considérer la représentation d'une articulation à l'aide du formalisme *swing-twist* comme une spécialisation de la représentation à l'aide d'un quaternion unitaire. Un

diagramme de classe à l'aide du formalisme UML illustre notre choix architectural pour la modélisation des articulations dans notre système.

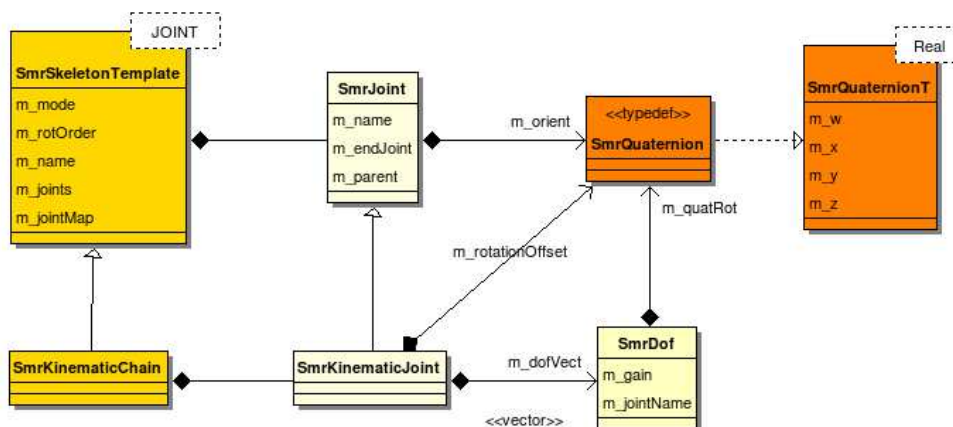


FIG. C.2 – Diagramme de classe représentant notre paramétrisation des rotations

Cette modélisation décrit la classe générique `SmrSkeletonTemplate`. Cette classe est un conteneur générique d'articulations (`JOINT`). Une articulation est une classe de type `SmrJoint`. Cette classe est composée d'une instance de quaternions (`SmrQuaternionT`). La classe `SmrKinematicJoint` étend la classe `SmrJoint`. Cette classe décrit une articulation dans le cadre de la résolution de chaînes cinématiques par cinématique inverse. la classe `SmrKinematicJoint`, en plus des attributs hérités de `SmrJoint` est composée de degrés de liberté exprimés en rotation. Chaque degré de liberté (`SmrDof`) est composé d'un axe (`SmrVector3` et d'un quaternion.

## C.2 Rendu

La manière dont les différents éléments du squelette influencent la déformation de parties de surface est sanctionnée par une pondération au niveau de chaque nœud de l'enveloppe externe. Dans un souci d'optimisation des performances, et pour bénéficier d'un rendu de l'animation en temps interactif, il est nécessaire de réduire au maximum le nombre de nœuds composant la surface externe.

L'utilisation d'une enveloppe externe, si elle permet d'affiner le rendu des poses obtenues au fil de l'animation n'est guère suffisante pour offrir une

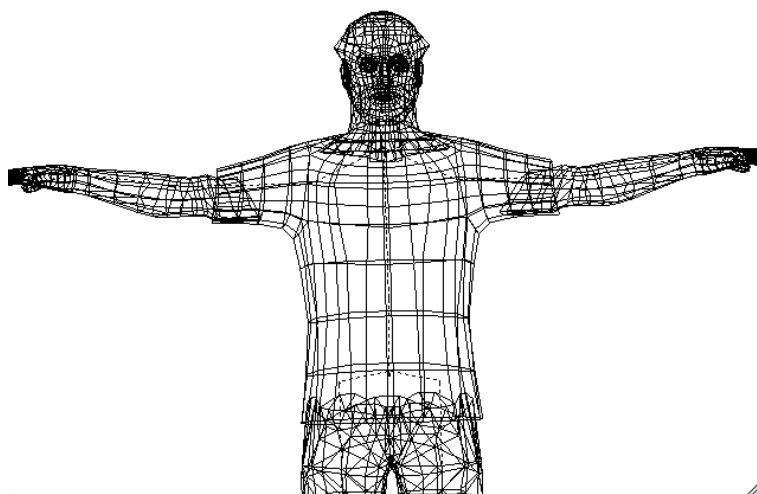


FIG. C.3 – Maillage de l'enveloppe externe

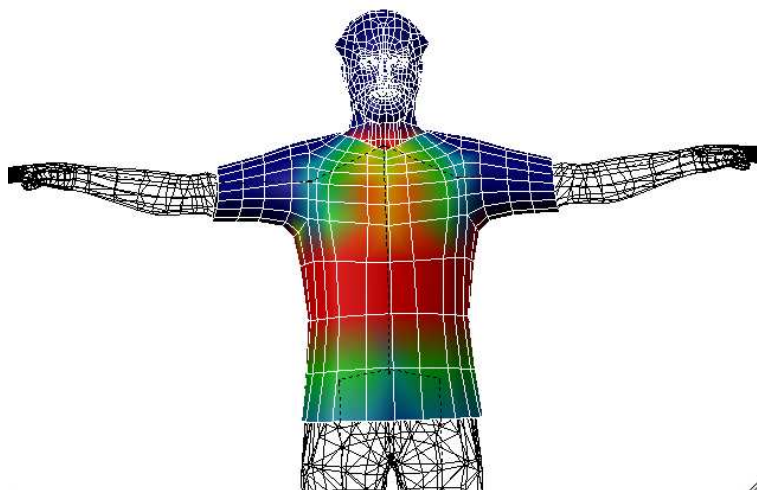


FIG. C.4 – Représentation graphique de l'influence d'une articulation sur le maillage. L'articulation considérée se nomme 'thorax'

représentation réaliste de l'agent. L'économie volontaire réalisée au niveau du nombre de nœuds composant le maillage en vue de simplifier les calculs nécessaires au rendu des différentes poses implique une géométrie simplifiée se caractérisant par de larges zones d'à-plat.

L'utilisation de méthodes d'interpolation de normales lors de la passe de rendu et un simple plaquage de texture peuvent aisément rajouter une dose de réalisme au rendu des poses de notre agent. L'association de l'information de couleur au pixel correspondant lors de la passe de rendu est réalisée dans

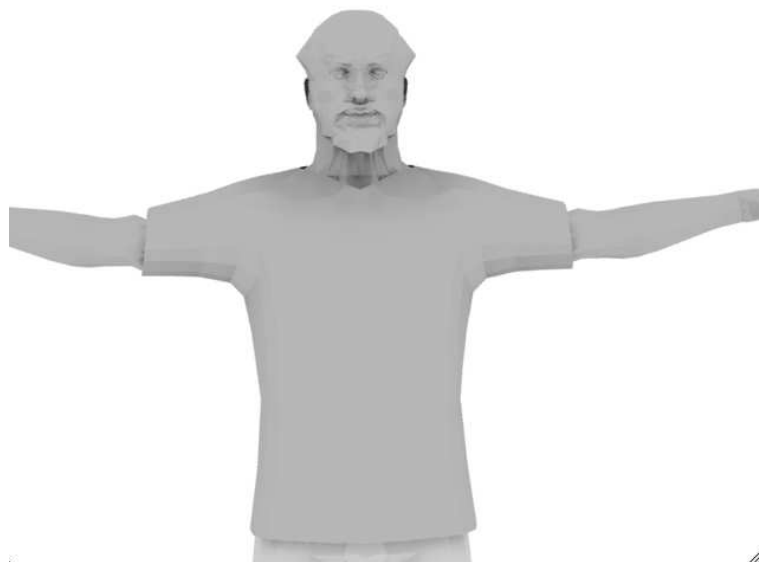


FIG. C.5 – La faible densité de points formant le maillage ne permet pas une représentation fine de l'enveloppe de l'agent

l'espace  $(u, v)$  des textures relativement aux noeuds de la surface représentant l'enveloppe de l'agent. Cette opération est réalisée en temps interactif grâce à l'utilisation d'un *fragment program* qui déporte les étapes nécessaires au placage de texture sur le processeur graphique de la carte vidéo du système.

Une méthode permet de raffiner localement la surface de l'enveloppe sans avoir à rajouter de complexité à la géométrie du maillage surfacique initial. Cette méthode emploie des cartes de normales qui sont appliquées lors de la passe de rendu afin de rajouter les détails relatifs aux détails surfaciques de l'enveloppe. A l'instar des cartes de textures, l'application des cartes de normales est déportée sur le processeur graphique.

Une fois les problèmes de réalisme inhérents au rendu traités, il s'agit d'animer la structure squelettique à l'origine de la déformation de l'enveloppe externe.



FIG. C.6 – Le placage de texture renforce le réalisme du rendu des poses de l'agent

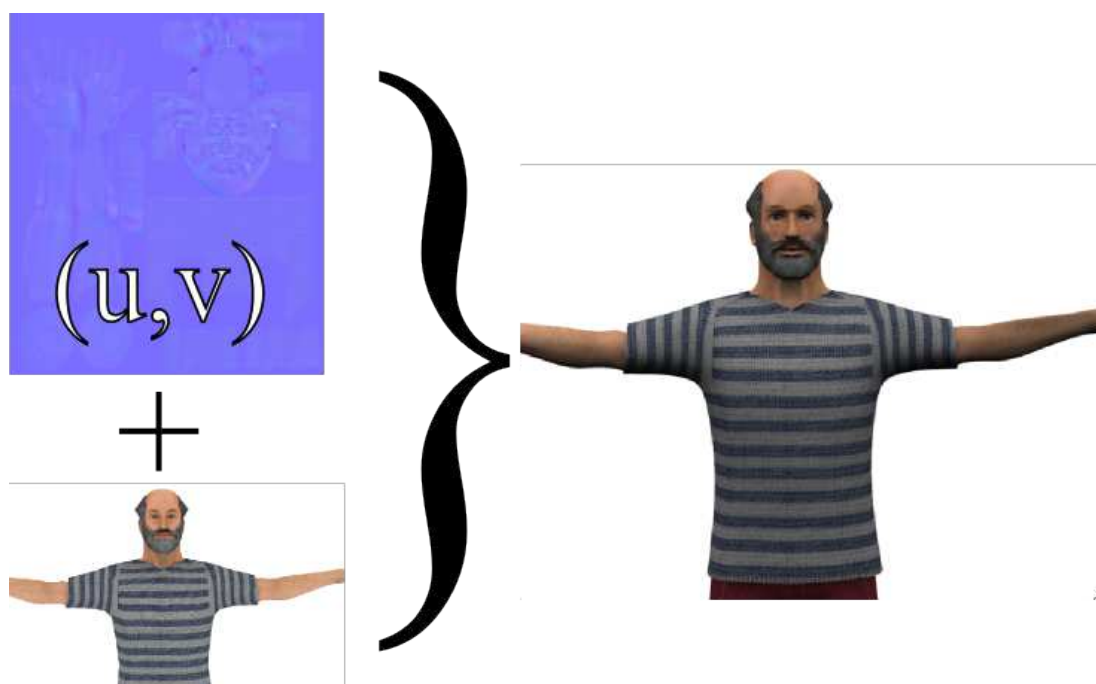


FIG. C.7 – La prise en compte de cartes de normales raffine les détails locaux de la surface de l'enveloppe



# Glossaire

**Actant** : Notion qui permet d'étendre la gamme de sujets et agents de l'action, en intégrant par exemple des entités dénuées d'intentionnalité.

**Double articulation** : Décomposition du langage humain en deux strates : la première articulation et la seconde articulation.

**Langue** : système de signes linguistiques vocaux, graphiques ou gestuels qui permet la communication entre individus.

**Langue des signes** : Langue utilisée par une communauté de sourds

**Linguistique structurale** : Courant réunissant un groupe d'écoles dans lequel la langue est étudiée comme un système doté d'une structure décomposable.

**Morphème** : Plus petite unité porteuse de sens qu'il soit possible d'isoler dans un énoncé.

**Gestétique** : Niveau linguistique qui décrit les gestes en se focalisant sur les éléments mis en jeu dans le mouvement.

**Oraliste** : Personne pratiquant une langue orale

**Paramètre qualitatif du geste** : décrit ce qui est relatif à la qualité, qui est du domaine de la qualité (et non des choses mesurables)

**Paramètre quantitatif du geste** : les données quantitatives sont des données qui peuvent être mesurées. On associe des chiffres et des unités de mesure aux données quantitatives. Ces données proviennent de calculs mathématiques ou de mesures.

**Première articulation** : Strate constituée par les unités significatives que sont les unités lexicales et grammaticales.

**Seconde articulation** : Strate constituée par les unités distinctives dépourvues de sens en elles-mêmes.

**Sémiologie** : discipline consacrée à l'étude des signes.

**Signe** : objet, geste, représentation d'objet ou représentation abstraite ayant une signification naturelle ou conventionnelle pour un groupe humain

donné. C'est également le composant de base des langues des sourds pour communiquer de manière visuelle, précise et intuitive.

**Signe linguistique** : unité d'expression du langage, communément un mot, une entité formée par la réunion d'un signifié et d'un signifiant.

**Signifiant** : Forme sonore ou visuelle.

**Signifié** : Concept

**Sciences cognitives** : Ensemble de disciplines scientifiques visant à l'étude et la compréhension des mécanismes de la pensée humaine ou artificielle.

**Visée illustrative ou iconisatrice** : visée qui tente de montrer tout en disant, de "donner à voir", de dire en montrant.

**Expression sans visée illustrative** : Expression dépourvue de contenu illustratif, constituée principalement de signes standards pour une langue des signes.



# Bibliographie

- [ABC96] K. Amaya, A. Bruderlin, and T. Calvert. Emotion from motion. In Wayne A. Davis and Richard Bartels, editors, *Graphics Interface 96*, pages 222–229. Canadian Human-Computer Communications Society, 1996.
- [ACT94] A. Braffort, C. Collet, and D. Teil. Hand configuration pre-processing tool for sign language recognition. In *RESNA Annual Conference*, 1994.
- [AD04] G. Aznar and P. Dalle. Computer support for signwriting written form of sign language. In *Proc. of Workshop RPSL, LREC*, pages 109–110, 2004.
- [AD05] G. Aznar and P. Dalle. Variations dans la représentation écrite d’un signe en signwriting. In *Proc. of TALN*, 2005.
- [Arn89] B. Arnaldi. *Conception du noyau d’un système d’animation de scènes tridimensionnelles intégrant les lois de la mécanique*. PhD thesis, Université de Rennes 1, 1989.
- [Bat78] R. Battison. *Lexical borrowings in American Sign Language*. Silver Spring, 1978.
- [BB98] P. Baerlocher and R. Boulic. Task-priority formulations for the kinematic control of highly redundant articulated structures. *IEEE IROS*, 1998.
- [BB00] P. Baerlocher and R. Boulic. Parametrization and range of motion of the ball-and-socket joint. In *Proc. of AVATARS’2000 Conference*, 2000.
- [BCC99] N. Badler, N. CHi, and S. Chopra. Virtual human animation based on movement observation and cognitive behaviors models. In *Proc. of Computer Animation conf.*, 1999.
- [Beb25] A. Bebian. *Mimographie, ou essai d’écriture mimique, propre à régulariser le langage des sourds-muets.* L. Colas, 1825.

- [BG06] D. Boutet and B. Garcia. Finalités et enjeux linguistiques d’une formalisation graphique de la langue des signes française (lsf). *Glottopol, revue de sociolinguistique en ligne.*, 7, 2006.
- [BH00] M. Brand and A. Hertzmann. Style machines. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 183–192. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.
- [BHMP06] S. Buisine, B. Hartmann, M. Mancini, and C. Pelachaud. Conception et evaluation d’un modèle d’expressivité pour les gestes des agents conversationnels. *Revue en Intelligence Artificielle RIA, Special Edition “Interaction Emotionnelle”*, 20, 2006.
- [BL06] A. Braffort and F. Lejeune. Spatialized semantic relations in french sign language : Towards a computational modelling. In *Proc of Int. Gesture Workshop*, 2006.
- [BP94] D. Brentari and H. Poizner. A phonological analysis of a death parkingsonian signer. *Languages and cognitive processes*, 9 :69–99, 1994.
- [BPW93] N. I. Badler, C. B. Phillips, and B. L. Webber. *Simulating humans : computer graphics animation and control*. Oxford University Press, Inc., New York, NY, USA, 1993.
- [Bre98] Diane Brentari. *A prosodic model of sign language phonology*. Bradford Book, 1998.
- [BRRP97] B. Bodenheimer, C. Rose, S. Rosenthal, and J. Pella. The process of motion capture : Dealing with the data. In *Computer Animation and Simulation’97, proceedings of Eurographics International Workshop.*, pages 3–18, September 1997.
- [Bru07] Jean-Louis Brugeille. Un formalisme de transcription de la lsf. In *Proc. of TALS workshop*, 2007.
- [BSP<sup>+</sup>04] J. Barbic, A. Safonova, J.-Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Pollard. Segmenting motion capture data into distinct behaviors. In *Proc. of Int. Conf. on Graphics interface*, pages 185–194, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2004. Canadian Human-Computer Communications Society.
- [BT96] P. Becheiraz and D. Thalmann. A model of nonverbal communication and interpersonal relationship between virtual actors. In *Proc. of Computer animation*, 1996.

- [BV06] F. Bonnal-Vergès. Langue des signes française : des lexiques des xviii et xixe siècles à la dictionnaire du xxie siècle. *GLOT-TOPOL, Revue de sociolinguistique en ligne*, 7, 2006.
- [BW95] A. Bruderlin and L. Williams. Motion signal processing. In *SIGGRAPH 95 : Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 97–104, New York, NY, USA, 1995. ACM Press.
- [Ca01] C. Cuxac and al. *Projet LS-COLIN. “ Quel outil de notation pour quelle analyse de la LS ? ”*. Journées “Recherches sur la langue des signes ” UTM, Le Mirail, Toulouse., 2001.
- [CB06] B. Le Callenec and R. Boulic. Interactive motion deformation with prioritized constraints. *Graphical Models*, 68 :175–193, 2006.
- [CCZB00] D. M. Chi, M. Costa, L. Zhao, and N. I. Badler. Emote. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 173–182. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.
- [CHP89] J. E. Chadwick, D. R. Haumann, and R. E. Parent. Layered construction for deformable animated characters. In *SIGGRAPH’89 : Proceedings of the 16th Annual Conference on Computer Graphics and Interactive Techniques*, pages 243–252, New York, NY, USA, 1989. ACM Press.
- [CK00] K-J. Choi and H-S. Ko. Online motion retargetting. *Journal of Visualization and Computer Animation*, 11(5) :223–235, 2000.
- [Cor01] Andrea Corradini. Dynamic time warping for off-line recognition of a small gesture vocabulary. In *Proc. of IEEE ICCV Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, page 63, 2001.
- [CPB+94] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. Animated conversation : rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *SIGGRAPH’94 : Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, pages 413–420, New York, NY, USA, 1994. ACM Press.
- [CPMTT99] T. Capin, I. Pandzic, N. Magnenant-Thalmann, and D. Thalmann. *Avatars in network environments*. Wiley Interscience, 1999.
- [CSPC01] J. Cassel, J. Sullivan, S. Prevost, and E. Churchill. *Embodied Conversational agents*. MIT Press, 2001.

- [Cux83] C. Cuxac. *Le langage des sourds*. Payot (Paris), 1983.
- [Cux00] C. Cuxac. *Faits de langue : la langue des signes française (LSF) : les voies de l'iconicité*. Ophrys(Gap), 2000.
- [dS16] F. de Saussure. *Cours de linguistique générale*. Charles Bailly and Albert Sechehaye, 1916.
- [dS95] F. de Saussure. *Essai de Linguistique Générale*. Payot (Paris), 1995.
- [Dub] J. Dubois. *Dictionnaire de linguistique*. Larousse.
- [Dum90] G. Dumont. *Animation de scènes tridimensionnelles : la mécanique des solides comme modèle de synthèse du mouvement*. PhD thesis, Université de Rennes 1, 1990.
- [EMT05] A. Egges and N. Magnenat-Thalmann. Emotional communicative body animation for multiple characters. In *V-Crowds'05*, pages 31–40, 2005.
- [FB06] M. Filhol and A. Braffort. A sequential approach to lexical sign description. In *LREC 2006 - Workshop on Sign Languages*, Genova, Italy, 2006.
- [FCD<sup>+</sup>93] N. Frishberg, S. Corazza, L. Day, S. Wilcox, and R. Schulmeister. Sign language interfaces. In *Proc. of conf. on Human Factors in Computing SystemsGorchow*, 1993.
- [FF05] K. Forbes and E. Fiume. An efficient search algorithm for motion data using weighted pca. In *SCA '05 : Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 67–76, New York, NY, USA, 2005. ACM Press.
- [Flo02] C. M. Flood. *How do Deaf and Hard of Hearing Students Experience Learning to Write Using Sign Writing, a Way to Read and Write Signs ?* PhD thesis, Long Island University, 2002.
- [FMJ02] A. Fod, M. J. Mataric, and O. C. Jenkins. Automated derivation of primitives for movement classification. *Autonomous Robots*, 12 :39–54, 2002.
- [FvdPT01] P. Faloutsos, M. van de Panne, and D. Terzopoulos. Composable controllers for physics-based character animation. In *SIGGRAPH'01 : Proceedings of the 28th Annual Conference on Computer Graphicsand Interactive Techniques*, pages 251–260, New York, NY, USA, August 2001. ACM Press.
- [Gal92] P.E. Gallaher. Individual differences in nonverbal behavior : Dimensions of style. *Journal of Personality and Social Psychology*, 63 :133–145, 1992.

- [Gle98] M. Gleicher. Retargetting motion to new characters. In *Proceedings of SIGGRAPH'98*, Computer Graphics Proceedings, Annual Conference Series, pages 33–42, July 1998.
- [GLS97] M. Gleicher, J. Lee, and S. Y. Shin. Motion editing with spacetime constraints. In *SI3D '97 : Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 139–ff., New York, NY, USA, 1997. ACM Press.
- [GM94] S. Gibet and P.-F. Marteau. A self-organized model for the control, planning and learning of non-linear multidimensional system using a sensory feedback. *Journal of applied intelligence*, 4 :337–349, 1994.
- [GMHP04] K. Grochow, S. L. Martin, A. Hertzmann, and Z. Popovic;. Style-based inverse kinematics. *ACM Trans. Graph.*, 23(3) :522–531, 2004.
- [Gra98] F. S. Grassia. Practical parameterization of rotations using the exponential map. *Journal of Graphics Tools*, 3(3) :29–48, 1998.
- [Gro79] F. Grosjean. Psycholinguistique et langue des signes. *Langages*, 56 :35–57, 1979.
- [Ham44] W. R. Hamilton. On a new species of imaginary quantities connected with a theory of quaternions. *Proceedings of the Royal Irish Academy*, 2 :424–434, 1844.
- [Hic03] M. HickmannHeloir. Coherence, cohesion and context in narrative development : some comparative perspectives. *Relating events in narrative : Typological and contextual perspectives*, pages 281–306, 2003.
- [HK07] A. Heloir and L. Kervajan. Hiérarchisation des variations pour la génération de gestes expressifs de communication. In *Proc. of Traitement Automatique des Langues des Signes (Atelier TALS 2007)*, 2007.
- [HMP06] B. Hartmann, M. Mancini, and C. Pelachaud. Implementing expressive gesture synthesis for embodied conversational agents. In *gesture in human-Computer Interaction and Simulation*, 2006.
- [HP97] J. K. Hodgins and N. S. Pollard. Adapting simulated behaviors for new characters. In *Proceedings of ACM SIGGRAPH'97*, Computer Graphics Proceedings, Annual Conference Series, pages 153–162, August 1997.
- [HPP05] E. Hsu, K. Pulli, and F. Popović. Style translation for human motion. *ACM Trans. Graph.*, 24(3) :1082–1089, 2005.

- [Hue06] M. Huenerfauth. *Generating American Sign Language Classifier Predicates For English-to-ASL Machine Translation*. PhD thesis, University of Pennsylvania, 2006.
- [HWBO95] J. K. Hodgins, W. L. Wooten, D. C. Brogan, and J. F. O'Brien. Animating human athletics. *Computer Graphics*, 29 :71–78, 1995.
- [Jak73a] R. Jakobson. *Essais de linguistique générale - 1 - Les fondations du langage*. Editions de Minuit, 1973.
- [Jak73b] R. Jakobson. *Essais de linguistique générale - 2 - rapports internes et externes du langage*. Editions de Minuit, 1973.
- [Jou95] P. Jouison. *Écrits sur la langue des signes française*. 1995.
- [KB79] E. S. Klima and U. Bellugi. *The signs of language*. Harvard University Press, 1979.
- [Ken93] A. Kendon. Human gesture. In T. Ingold and Gibson K., editors, *Tools, Language and Intelligence*, 1993.
- [Ken01] Richard Kennaway. Synthetic animation of deaf signing gestures. In *Gesture Workshop*, pages 146–157, 2001.
- [KG03] L. Kovar and M. Gleicher. Flexible automatic motion blending with registration curves. In *ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, pages 214–224, August 2003.
- [KGP02] L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pages 473–482. ACM Press, 2002.
- [Kip04] M. Kipp. *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*. Dissertation.com, 2004.
- [Kit90] S. Kita. The temporal relationship between gesture and speech : A study of japanese-english bilinguals. Master's thesis, University of Chicago, 1990.
- [KM05] R. Kulpa and F. Multon. Fast inverse kinematics and kinetics solver for human-like figures. In *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*, pages 38–43, Dec. 5, 2005.
- [Kor85] J. U. Korein. *A geometric investigation of reach*. MIT Press, Cambridge, MA, USA, 1985.
- [KSK00] T. Komura, Y. Shinagawa, and T. L. Kunii. Creating and retargeting motion by the musculoskeletal human body model. *The Visual Computer*, 16(5) :254–270, 2000.
- [KvGvdH98] S. Kita, I. van Gijn, and H. van der Hulst. Movement phase in signs and co-speech gestures, and their transcriptions by human

- coders. In *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*, pages 23–35, London, UK, 1998. Springer-Verlag.
- [Lab88] R. Laban. *The Mastery of Movement*. Northcote House, 1988.
- [Las87] J. Lasseter. Principles of traditional animation applied to 3d computer animation. In *Proc. of SIGGRAPH*, 1987.
- [LBAM04] C. Lisetti, C. L. Brown, K. Alvaez, and A. Marpaung. A social informatics approach to human-robot interaction with an office service robot. *IEEE Transactions on Systems, man and Cybernetics*, 34 :195–209, 2004.
- [LCP<sup>+</sup>02] J. Lee, J. Chai, P.Reitsma, J. K. Hodgins, and N. Pollard. Interactive control of avatars animated with human motion data. In *Proceedings of SIGGRAPH 2002*, 2002.
- [LD06] B. Lenseigne and P. Dalle. Using signing space as a representation for sign language processing. *Gesture in Human-Computer Interaction and Simulation*, 6 :25–36, 2006.
- [LG99] T. Lebourque and S. Gibet. High level specification and control of communicative gestures : the gessyca system. In *Computer Animation, 1999. Proceedings*, pages 24–35, 26–29 May 1999.
- [Lid89] S. Liddell. American sign language : The phonological base. *Sign language studies*, 64 :195–278, 1989.
- [Lid90] S. Liddell. Structures for representing handshape and local movement at the phonetic level. In University of chicagon press, editor, *Theoretical issues in sign language research*, 1990.
- [Lid93] S. Liddell. Holds and positions : Comparing two models of segmentation in asl. *Phonetics and Phonology*, 3 :189–211, 1993.
- [Lid03] S. Liddell. *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge University Press, 2003.
- [LL04] J. Lee and K.H. Lee. Precomputing human behavior from human motion data. In *Proc. ACM/EUROGRAPHICS Symposium on Computer Animation 2004*, 2004.
- [Los00] O. Losson. *Modélisation du geste communicatif et réalisation d'un signeur virtuel de phrases en langue des signes française*. PhD thesis, Université de Lille1, 2000.
- [Mar83] M. Marey. De la mesure des forces dans les différents actes de la locomotion. *Compte rendu des Séances de l'Académie des Sciences*, 46, 1883.

- [Mar94] M. Marey. *Mécanique animale*, volume 119. New-York : Appleton & Co., 1894.
- [Mar05] E. Martin. Étude en linguistique et en traitement d'image, application à un copus de lsf. Master's thesis, Université de Paris VIII, 2005.
- [Mar06] O. Martineau. *Vers une approche géométrique du schéma analyse et synthèse pour la locomotion humaine*. PhD thesis, INSA de Rennes, 2006.
- [MBBT00] J-S. Monzani, P. Baerlocher, R. Boulic, and D. Thalmann. Using an intermediate skeleton and inverse kinematics for motion re-targetting. *Computer Graphics Forum*, 19(3), August 2000. ISSN 1067-7055.
- [MBP05] M. Mancini, R. Bressin, and C. Pelachaud. From acoustic cues to an expressive agent. In *Proc. of Int. Gesture Workshop*, 2005.
- [MHO<sup>+</sup>03] C. Martel, P. Howard, C. Osborn, L. Britt, and K. Myers. Form2 kinematic gesture corpus. In *Video recording and annotation*, 2003.
- [Mil98] A. Millet. Localisation spatiale et paramètres énonciatifs en lsf. In *Actes du colloque international des linguistes*, 1998.
- [ML86] M. Marcotty and H. Ledgard. *The World of Programming Languages*. Springer-Verlag, 1986.
- [Moo98] Bill Moody. *La langue des signes*. International Visual Theatre, 1998.
- [Mor70] Mashariro Mori. Bukimi no tani, the uncanny valley. *Energy*, 7 :33–35, 1970.
- [Mot76] B. Mottez. *À propos d'une langue stigmatisée : la langue des signes*. École des hautes études en sciences sociales., 1976.
- [MR81] CS Myers and LR Rabiner. A comparative study of several dynamic time warping algorithms for connected word recognition. *The Bell System Labs technical journal*, 60 :1389–1409, 1981.
- [MZF06] A. Majkowska, V. B. Zordan, , and P. Faloutsos. Automatic splicing for hand and body animations. In *ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, pages pp. 309–316, 2006.
- [Mé03] S. Ménardais. *Fusion et adaptation temps réel de mouvements acquis pour l'animation d'humanoïdes synthétiques*. PhD thesis, Université de Rennes 1, January 2003.
- [Nei92a] D. Mc Neill. *Hand and Mind. What Gesture Reveals About Thoughts*. University of Chicago Press, 1992.



- [Nei92b] D. Mc Neill. *Hand and Mind. What Gesture Reveals About Thoughts*. University of Chicago Press, 1992.
- [New87] Newkirk. *SignFont Handbook*. Emerson and Associates, 1987.
- [NF05] M. Neff and E. Fiumes. Methods for exploring expressive stance. *Graphical Models*, \_\_ :\_\_, 2005.
- [NMB04] G. Nicolas, F. Multon, and G. Berillon. Inverse kinematics for the calculation of plausible bipedal locomotion using anthropological knowledge. In *Proceedings of CASA*, pages 103–110, Geneva, Switzerland, July 2004.
- [NST94] C.I. Nass, J. Steuer, and E. Tauber. "computers are social actors". In *CHI'94*, 1994.
- [PLZ+89] S. L. Prillwitz, R. Leven, H. Zienert, R. Zienert, T.Hanke, and J. Henning. *HamNoSys. Version 2.0*. International Studies on Sign Language and Communication of the Deaf, 1989.
- [PML04] C. Pelachaud, V. Maya, and M. Lamolle. Representation of expressivity for embodied conversational agents. In *Workshop Balanced Perception and Action, Third International Joint Conference on Autonomous Agents & Multi-Agent Systems*, 2004.
- [PPR02] F. Preteux, M. Preda, and N. Rougon. D4-4. advanced mpeg4 animation system. Technical report, VISICAST - INT, 2002.
- [Rab93] L. Rabiner. *Fundamentals of speech recognition*. Prentice Hall PTR, 1993.
- [RCB98] C. Rose, M. Cohen, and B. Bodenheimer. Verbs and adverbs : Multidimensional motion interpolation. *IEEE Computer Graphics and Applications*, 18 :pp. 32–40, 1998.
- [RDF04] A. Da Rocha, G. Dimuro, and J. De Freitas. A sign matching technique to support searches in sign language texts. In *Proc. of RPSL Workshop, LREC*, 2004.
- [RdLGH04] C. Retoré, O. de Langhe, P. Guitteny, and H. PortineRuttkay. A propos de l'ordre osv en lsf. *SILEXICALES*, 4, 2004.
- [RN01] Z. Ruttkay and H. Noot. Facial expression sculpting with interval constraint. In *Proc. of AA' 01 Ws on Multimodal Communication and Context in Embodied Agents*, 2001.
- [RN05] Z. Ruttkay and H. Noot. Variations in gesturing and speech by gestyle. *International Journal of Human-Computer Studies*, 62 :211–229, 2005. Special Issue on 'Subtle Expressivity for Characters and Robots'.

- [RPPN04] Z. Ruttkay, C. Pelachaud, I. Poggi, and H. Noot. Exercises of style for virtual humans. *Advances in Consciousness Research Series*, to appear :to appear, 2004.
- [Rut94] Z. Ruttkay. Fuzzy constraint satisfaction. *Proc. of the Third IEEE International Conference on Fuzzy Systems*, pages 1263–1268, 1994.
- [Sal03] M.A. Sallandre. *Tentative de catégorisation dans le cadre d'une grammaire de l'iconicité*. PhD thesis, Université de Paris VIII, 2003.
- [SDO<sup>+</sup>04] M. Stoneand, D. DeCarlo, I. Oh, C. Rodriguez, A. Stere, A. Less, and C. Bregler. Speaking with hands : Creating animated conversational characters from recordings of human performance. In *Proc. ACM/EUROGRAPHICS Symposium on Computer Animation 2004*, 2004.
- [SF06] A. Shapiro and P. Faloutsos. Style components. In *Proc. of Graphics interface*, 2006.
- [snc06] sncf. Montparnasse, gare laboratoire de l'accessibilité. Dossier de presse, mai 2006.
- [snc07] sncf. Gare de l'est, accessibilité pour les personnes sourdes. web-sourd, 2007.
- [SPC04] S. Salvador and Sam C. S. P. Chan. Tastedtw : toward accurate dynamic time warping in linear time and space. In *KDD Workshop on Mining Teporal and Sequential Data*, 2004.
- [Sto60] W. Stokoe. Sign language structure : An outline of the visual communication systems of the american deaf. In *Studies in Linguistic, Occasional Papers*, volume 8, 1960.
- [Sut98] V. Sutton. The signwriting literacy project. impact of deafness on cognition. In *In Proc. of AERA Conference*, 1998.
- [TJ95] F. Thomas and O. Johnson. Illusion of life : Disney animation. In *Proc of SIGGRAPH*, 1995.
- [Tro02] N. Troye. Decomposing biological motion : A framework for analysis and synthesis of human gait patterns. *J. Vis.*, 2(5) :371–387, 9 2002.
- [UAT95] M. Unuma, K. Anjyo, and R. Takeuchi. Fourier principles for emotion-based human figure animation. In *SIGGRAPH 95 : Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 91–96, New York, NY, USA, 1995. ACM Press.
- [UGB<sup>+</sup>04] R. Urtasun, P. Glardon, R. Boulic, D. Thalmann, and P. Fua1. Style-based motion synthesis. *Computer Graphics Forum*, 24 :799–812, 2004.

- [VG04] C. Vogler and S Goldenstein. Adaptive deformable models. *in proceedings of Siggraph 2004*, 2004.
- [VM04] Christian Vogler and Dimitri MetaxasJouison. Handshapes and movements : Miltiple-channel asl recognition. In *Lecture Notes in Artificial Intelligence*, pages 247–258, Genova, Italy, 2004. Springer Verlag.
- [Wal98] HG. Wallbot. Bodily expression of emotion. *European journal of social psychology*, 28 :879–796, 1998.
- [WFB87] A. Witkin, K. Fleischer, and A. Barr. Energy constraints on parametrized models. In *Proceedings of Siggraph'87*, pages 225–232, 1987.
- [WSS94] J. Walker, L. Sprooull, and R. Subramani. Using a human face in a interface. In *Proc. of CHI'94*, 1994.
- [XG00] L. Xu and W. Gao. Study on translating chinese into chinese sign language. *Journal of computer science and technology*, 15 :485–490, 2000.
- [ZH99] V. B. Zordan and J. K. Hodgins. Tracking and modifying upper-body human motion data with dynamic simulation. In *Computer Animation and Simulation '99*, September 1999.
- [ZH02] V. B. Zordan and J. K. Hodgins. Motion capture-driven simulations that hit and react. In *SCA'02 : Proceedings of the ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 89–96, New York, NY, USA, July 2002. ACM Press.
- [Zha01] L. Zhao. *Synthesis and Acquisition of Laban Movement Analysis Qualitative Movement Parameters for Communicative Gestures*. PhD thesis, University of Pennsylvania, 2001.
- [ZMCF05] V. B. Zordan, A. Majkowska, B. Chiu, and M. Fast. Dynamic response for motion capture animation. *ACM Transactions on Graphics : Proceedings of Siggraph'05*, 24(3) :697–701, 2005.
- [ZVRvdS04] I. Zwiterslood, M. Verlinden, J. Ros, and S. van der Schoot. Synthetic signing for the deaf : Esign. In *Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment*, Granada, Spain, July 2004.



# Table des figures

2.1	Schéma organisationnel de la thèse . . . . .	16
2.2	Passage polymorphémique en langue des signes américaine . .	16
2.3	Contributions et domaines adressés dans ce mémoire. . . . .	19
3.1	Relations entre signe, signifiant et signifié . . . . .	23
3.2	Bifurcation entre la visée illustrative et conceptuelle . . . . .	27
3.3	Modèle iconique, schéma récapitulatif . . . . .	29
3.4	Organisation des différents champs d'études des LS. . . . .	30
3.5	Systèmes de génération de gestes de la langue des signes par ordre chronologique . . . . .	33
3.6	EMOTE modifie un mouvement existant . . . . .	38
3.7	Le style regroupe deux problématiques . . . . .	39
3.8	Shéma illustrant le concept d'uncanny valley, d'après Masahiro Mori et Karl MacDorman. . . . .	41
4.1	Locuteur représentant un boxeur en visée iconisatrice. . . . .	50
4.2	Diagramme de classe représentant l'entité et son référent . . . .	51
4.3	Représentation d'une entité hors visée iconisatrice . . . . .	52
4.4	Spécification d'attributs de taille ou de forme pour les référents	53
4.5	Spécification des relations spatiales en transfert de situation . .	54
4.6	Transfert personnel appliqué à un objet non animé : la théière est "personnifiée" . . . . .	55
4.7	Diagramme de classe représentant les référents impliqués lors d'un transfert personnel et leurs attributs . . . . .	56
4.8	Diagramme de classes reprenant le formalisme de UML et représentant le modèle du discours en LS . . . . .	59
5.1	Relation entre l'entité, son référent et les signes associés . . . . .	64
5.2	Illustration d'une phase de préparation : «Demain, soleil (il y aura)» . . . . .	66
5.3	Illustration d'une phase de rétraction partielle (NUAGES) et d'une phase de rétraction totale en fin de séquence (AU_REVOIR) . .	67

5.4	Organisation de la phrase gestuelle et décomposition en phases	69
5.5	Le signe standard chewing-Gum	71
5.6	Vision générale du modèle prosodique de la langue des signes	71
5.7	Organisation de la branche des paramètres statiques au cours de la production d'un geste	72
5.8	Organisation des paramètres stables pour la configuration manuelle	73
5.9	Quelques configurations décrites par le modèle	75
5.10	Quelques configurations décrites par le modèle	76
5.11	Modélisation de la localisation	77
5.12	Plans de l'espace de signation	77
5.13	Les huit emplacements définis pour la main	78
5.14	L'orientation définie comme une relation entre le vecteur normal à une zone de contact et une localisation ou un plan d'articulation	79
5.15	Hiérarchie des caractéristiques prosodiques du modèle de description	80
5.16	Deux spécifications de mouvement différentes : tracé et pointé.	82
5.17	Représentation des différents canaux intervenants en langue des signes, aspect multi-linéaire	83
5.18	Partition du mouvement décrivant l'exemple tiré de [Cux00] : "Tout en marchant, le boxer regarde le chien loup [...] et bave beaucoup"	84
5.19	Compositionnalité morphémique en visée iconicisatrice : l'exemple du boxer [Cux00]	85
5.20	Le mélange de mouvements ne permet pas de respecter les contraintes nécessaires à la restitution intelligible d'un signe composé. Le signe représenté dans l'image du haut, obtenu par mélange de mouvements ne conserve pas la contrainte de contact entre la partie intérieure des doigts et le menton. Au contraire, le signe représenté dans l'image du bas, obtenu par mélange de contraintes la respecte.	86
6.1	Systèmes d'acquisition des mouvements, répartition des réflecteurs acquis par le système de capture optique	91
6.2	Systèmes d'acquisition des mouvements, ce système fait intervenir un outil de capture optique et des gants de données. Une fois acquises, les trajectoires de mouvements (courbes) sont rééchantillonnées puis fusionnées.	92
6.3	Le signe REGARDER_MOI exécuté par le signeur. La séquence du dessus correspond à une expressivité neutre tandis que la séquence du dessous correspond à une qualité expressive "brusque".	93

6.4	Proportion de variance explicitée par les composantes des matrices de mouvements projetées sur la base des postures propres du mouvement neutre (mouvement 1) . . . . .	96
6.5	Algorithme de DTW adaptatif . . . . .	99
6.6	A chaque itération, le chemin d'appariement obtenu est projeté sur le niveau supérieur, il fournit ainsi un guide plus contraint lors de l'itération suivante . . . . .	100
6.7	De haut en bas : signaux originaux, alignement par DTW standard, DTW à pente contrainte et DTW adaptative. L'algorithme de DTW adaptative propose un alignement limitant les discontinuités et capable de s'adapter à la fois aux perturbations locales et globales entre séries temporelles . . . . .	101
6.8	Résultats d'alignements temporels illustrés à l'aide de postures. Les deux premières séquences représentent deux réalisations réalisées respectivement de manière neutre et brusque. La troisième séquence illustre l'alignement de la séquence brusque sur la séquence neutre. Les courbes présentées sous les poses représentent l'évolution de l'influence de la première posture principale de la base de représentation du mouvement ( $\Sigma_{ref}^+ U_{ref}^T$ ) . . . . .	102
6.9	Application de l'alignement à la synthèse : une fois alignés, il est aisé de modifier la qualité expressive d'un mouvement de manière continue et fluide par simple interpolation . . . . .	104
6.10	Segmentation automatique réalisée par l'agorithme proposé par [BSP <sup>+</sup> 04]. Une marque de segmentation est posée lorsque la dérivée de l'erreur dépasse trois déviations standard de la moyenne de l'erreur . . . . .	106
6.11	Comparaison entre les résultats de segmentation automatique et manuelle . . . . .	107
6.12	Session d'annotation manuelle à l'aide de l'outil Anvil. Les canaux "lhand" et "rhand" sont extraits de manière automatique et assistent la segmentation . . . . .	109
6.13	Identification des phases de mouvement et relations temporelles observées entre les mouvements du bras et de la main . . . . .	110
6.14	Alignements temporels extraits de gestes simples réalisés selon des qualités expressives différentes . . . . .	112
6.15	Aplication des profils d'alignements inter-styles à la synthèse de gestes expressif . . . . .	112
6.16	Exemple de signes subissant la proximalisation / distalisation . . . . .	113
6.17	Exemples de signe subissant une chute de main dominée . . . . .	114
6.18	Architecture générale du processus de génération . . . . .	116

6.19	Interface interactive pour l'évaluation de la génération sensori-motrice . . . . .	120
6.20	Modèle sensori-moteur . . . . .	121
6.21	Interface interactive pour la compositionnalité multi-linéaire des gestes . . . . .	122
C.1	Représentation hiérarchique du squelette . . . . .	136
C.2	Diagramme de classe représentant notre paramétrisation des rotations . . . . .	138
C.3	Maillage de l'enveloppe externe . . . . .	139
C.4	Représentation graphique de l'influence d'une articulation sur le maillage. L'articulation considérée se nomme 'thorax' . . . . .	139
C.5	La faible densité de points formant le maillage ne permet pas une représentation fine de l'enveloppe de l'agent . . . . .	140
C.6	Le placage de texture renforce le réalisme du rendu des poses de l'agent . . . . .	141
C.7	La prise en compte de cartes de normales raffine les détails locaux de la surface de l'enveloppe . . . . .	141





## Résumé

Les travaux présentés dans cette thèse sont dédiés à la conception et l'animation d'entités virtuelles autonomes et réalistes, capables de générer des gestes de la Langue des Signes. L'accent est mis sur la qualité expressive des gestes.

Nous montrons que les aspects expressifs du geste, en plus de conférer au mouvement un aspect réaliste, participent à la construction du sens dans le cadre de la visée iconisatrice de la langue des signes.

Nous proposons un modèle du discours prenant en compte les visées illustrative et non illustrative de la langue des signes ainsi que ses principales structures de grande iconicité.

Ce modèle appelle une spécification des gestes capable de gérer les différentes modes expressifs.

Une phase d'analyse nous permet ensuite d'identifier certains paramètres d'expressivité. Cette analyse s'applique sur une séquence de gestes en langue des signes française réalisée par un locuteur sourd selon différentes qualités expressives.

Enfin, nous proposons une méthode de génération de mouvement compatible avec le modèle de description des gestes. Cette méthode de génération met en œuvre différentes méthodes de génération de mouvement dont un modèle de contrôle sensorimoteur capable de s'adapter aux variabilités spatiales et temporelles induites par les paramètres d'expressivité.

## Abstract

We present in this Phd report our investigation dedicated toward the design and animation of autonomous and realistic signing agents. The highlight is focused on expressive quality of gestures.

We show how expressive quality of gesture may convey meaningful information in the context of iconicity.

We then propose a discourse model that is able to handle both illustrative and non illustrative intends involved in sign language production.

Thanks to motion analysis performed on a gesture sequence that is captured on a professional deaf sign language instructor, we highlight key expressive parameters.

We then propose motion generation methods that suits our gesture specification model. this motion generation method rely on several gesture generation methods that are able to handle the variability induced by the requested expressive quality of the gesture.