

Description des mouvements des sourcils pour la génération automatique

Emilie Chételat-Pelé (1, 2), Annelies Braffort (2)

Université de Provence, Aix-en-Provence, France

emilie.chetelat@limsi.fr

LIMSI/CNRS, Orsay, France

annelies.braffort@limsi.fr

Résumé : La présente communication s'inscrit dans le cadre de la génération automatique de Langue des Signes Française (LSF) et plus particulièrement la génération des Gestes Non Manuels (GNMs). La génération nécessite une description précise des GNMs. Actuellement, nous ne disposons que de notations de nature symbolique. Nous proposons une nouvelle méthodologie d'annotation permettant une description fine et précise des positions, tout en tenant compte de la structure temporelle des GNMs. Sur chaque image constituant la vidéo nous positionnons des points sur les éléments à annoter afin d'obtenir leurs coordonnées. Ces coordonnées sont utilisées pour déterminer les positions exactes de chaque élément, dans chaque image. Ces positions servent ensuite à animer un visage synthétique à des fins d'évaluation de l'annotation, à analyser la structure des GNMs, mais également à définir de façon numérique chaque symbole du système d'annotation par flèches que nous avons précédemment établi.

Abstract : Our study tackles Non Manual Gestures (NMGs) annotation within the context of Sign Language (SL) research and more particularly within the context of automatic generation of NMGs. The generation needs a precise description of NMGs. At the present, we have only symbolical notation. Thus, we propose a new annotation methodology, which allows us precise description of movements and which takes into account the temporal structure of NMGs. On the video corpus, we position points on elements to be annotated, to obtain their coordinates. These coordinates are used to obtain precise position of all NMGs frame by frame. These data are used to evaluate the annotation by means of a synthetic face, for numerical analysis (by using curve), and, finally, to obtain numerical definition of each symbol of our previous annotation system based on arrows.

Mots-clés : Langue des Signes, Gestes Non Manuels, formalisation, annotation.

Keywords: Sign Language, Non Manuals Gestures, formalization, annotation

1 Introduction

Dans le contexte de la génération automatique de LSF et de l'animation de signeurs virtuels, nous avons besoin d'accroître les connaissances que l'on a des Gestes Non Manuels (GNMs) aussi bien au niveau de leurs structures internes que de leurs rôles. Acquérir de telles connaissances implique de disposer de descriptions fines des GNMs. Ces descriptions sont obtenues à partir de l'observation et l'annotation de corpus vidéo. En premier lieu, il s'agit donc de mener une réflexion sur la méthodologie à adopter pour l'annotation, selon le degré de précision nécessaire. Nous proposons dans cet article une méthodologie permettant une annotation numérique des GNMs pour une description précise et formelle de leur structure (interne et temporelle).

La première partie de l'article est consacrée à préciser la problématique et plus particulièrement à la présentation des notations et transcriptions actuelles puis à la définition des objectifs fixés. Dans une seconde partie nous proposons une nouvelle méthodologie d'annotation des GNMs permettant l'étude de leur structure interne, spatiale et temporelle en fonction de leur contexte. Nous présentons ici une application de cette méthodologie aux mouvements des sourcils et des paupières.

2 Problématique: une nouvelle méthodologie d'annotation

Cette section présente notre problématique de recherche et les limites des notations habituellement utilisées ainsi que les objectifs de notre étude.

Une observation de l'état des connaissances portant sur les GNMs permet de mettre en évidence, aux travers d'études telles que Jouison (1995), Cuxac (2000) et Vergé (2001), l'importance des GNMs en tant qu'éléments structurels de la LSF, mis en œuvre dans la transmission de contenus linguistiques et indispensables à la compréhension des énoncés (Chételat-Pelé, 2007).

Par exemple les sourcils, parfois associés à d'autres GNMs, permettent entre autres d'exprimer la modalité (interrogation, injonction, conditionnel...) dans les langues signées telles que la langue des signes française (Cuxac, 2000), américaine (Lidell, 1980), suédoise (Bergman, 1984) et néerlandaise (Coerts, 1992).

Ces études offrent des indications globales sur le rôle linguistique des GNMs. Cependant, l'observation des corpus nous laisse penser que des propriétés plus fines interviennent (amplitude des mouvements, durée...) dont l'étude implique une annotation plus précise que celles actuellement utilisées.

Les notations actuelles des GNMs sont de nature symbolique plus ou moins iconique. Les systèmes d'écriture ou de transcription, comme HamNoSys (Prillwitz et Zienert, 1989), D'Sign (Jouison, 1995) ou SignWriting (Sutton, Gleaves, 1995) décrivent la position d'un GNM de manière plus iconique qu'une simple notation textuelle (voir par exemple la notation de « sourcils levés » dans les différents systèmes, figure 1).

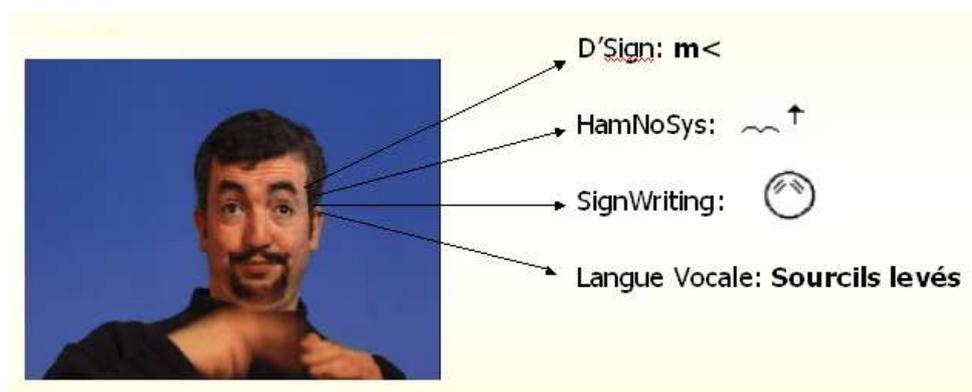


Figure 1: Différentes notations de « sourcils levés ».

La notation d'HamNoSys inclut une flèche qui semble représenter un mouvement (hausse de sourcil). En réalité, cette flèche représente une position figée : sourcil levé.

Ces notations ne sont pas directement exploitables au sein des systèmes de génération automatique car elles ne comportent aucune indication de nature numérique.

De plus, ces notations relèvent un instant figé et ne permettent pas d'observer l'amplitude et la structure temporelle du mouvement. Par exemple, une notation telle que « sourcils levés » n'indique rien sur l'amplitude (quel degré de hausse des sourcils) ou encore la durée de la hausse et du maintien... On ne peut donc étudier l'importance de ces éléments dans la transmission du sens.

Nous proposons dans un premier temps de définir une nouvelle méthodologie d'annotation, permettant d'étudier la structure des mouvements des GNMs afin d'en proposer une description précise et formelle (*quand* interviennent les GNMs : c'est à dire le contexte; *comment* ils interviennent : c'est à dire l'amplitude du mouvement et sa durée selon le contexte). Notre méthodologie permet de mesurer, de noter et de générer l'amplitude et la durée des GNMs.

3 Méthodologie : une annotation directe sur les images

Nous présentons ici l'application de notre méthodologie sur les mouvements des sourcils. L'annotation a été faite sur le corpus LS-Colin (Braffort et al, 2001; Segouat, Braffort, Martin, 2006) dont la qualité de l'image et les différents plans ont été particulièrement précieux pour notre étude. Nous avons utilisé le logiciel Anvil (Kipp, 2004), qui permet d'inclure les symboles que l'on souhaite dans l'annotation. De plus, il offre la possibilité d'ajouter des couleurs dans l'annotation afin de mettre en avant certains phénomènes que l'on souhaite étudier. Enfin, Anvil permet d'annoter directement les images constituant la vidéo à l'aide de points dont on peut ensuite récupérer les coordonnées pour de futures analyses numériques.

La première section (3.1) décrit la façon dont les images de la vidéo ont été annotées, et en particulier le système FACS (Facial Action Coding System) dont certains aspects ont été exploités. Dans une seconde section (3.2) nous détaillons le traitement des données numériques résultantes de l'annotation. Enfin, les trois dernières sections (3.3, 3.4, 3.5) présentent trois utilisations de ces données permettant la validation et l'analyse structurelle et linguistique de l'annotation.

3.1 Annotation des images de la vidéos

Pour l'annotation du mouvement des sourcils, nous nous sommes basés sur un système conçu au départ pour décrire l'expression des émotions du visage, FACS (Ekman et Friesen, 1978). FACS est un modèle formel de codage de l'expression faciale qui prend en compte l'action des muscles faciaux. En effet, Ekman et Friesen définissent tous les mouvements possibles du visage à partir d'un ensemble de muscles. Ils décomposent ces mouvements possibles en 46 actions unitaires (AUs) produites par les muscles.

Pour les sourcils, Ekman et Friesen considèrent quatre muscles, permettant trois actions : hausse de l'intérieur¹ du sourcil (AU 1), hausse de l'extérieur² du sourcil (AU 2) et rapprochement des sourcils (AU 4). Le muscle frontal (figure 2) permet la hausse de l'intérieur et de l'extérieur des sourcils.

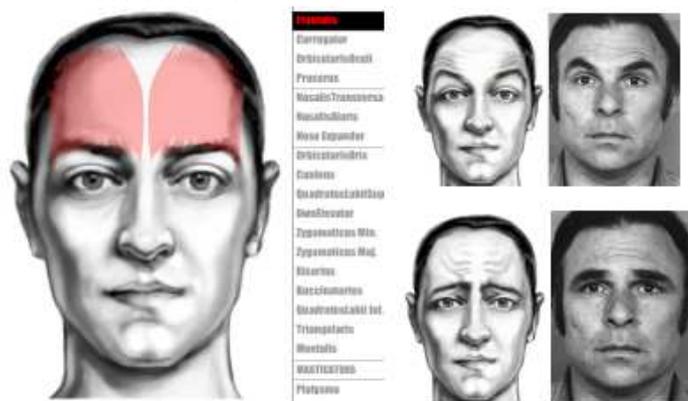


Figure 2 : Muscle frontal et ses AUs associées (hausse de l'extérieur (AU 2) et de l'intérieur des sourcils (AU 1)). Images extraites du site Arnatomy³ (Contreras Flores, 2005) et du manuel de FACS⁴.

Les muscles sourcilier et orbiculaire des paupières, ainsi que le muscle pyramidal du nez permettent un mouvement latéral des sourcils qui se caractérise par une variation de la distance entre les sourcils par exemple lorsqu'ils sont froncés (Figure 3).

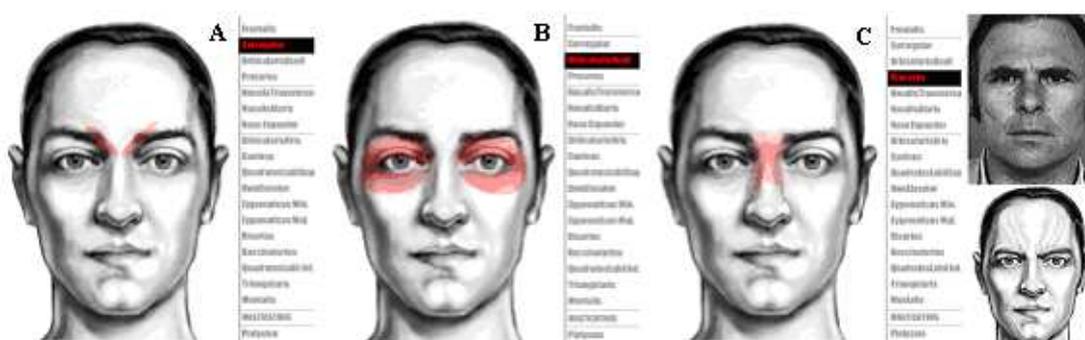


Figure 3 : Muscles sourcilier (image A), orbiculaire des paupières (image B) et pyramidale (image C) et leurs AUs associées : rapprochement des sourcils (AU 4). Images extraites du site Arnatomy et du manuel de FACS.

¹ Par intérieur du sourcil nous entendons le point du sourcil le plus proche de l'arête du nez.

² Par extérieur du sourcil nous entendons le point du sourcil le plus proche de l'oreille.

³ <http://www.artnatomia.net>, outil interactif dédié à l'enseignement des bases anatomiques et biomécaniques des expressions du visage.

⁴ <http://www.face-and-emotion.com/dataface/facs/manual/TitlePage.html>

Description des mouvements des sourcils pour une génération automatique

Ces AUs peuvent se combiner entre elles. Les images ci-dessous (Figure 4) nous montrent les combinaisons des AUs 1 et 4 (hausse de l'intérieur des sourcils et rapprochement des sourcils), AUs 1 et 2 (hausse de l'intérieur et l'extérieur des sourcils) et AUs 1, 2 et 4 (rapprochement des sourcils et hausse de l'intérieur et l'extérieur des sourcils).

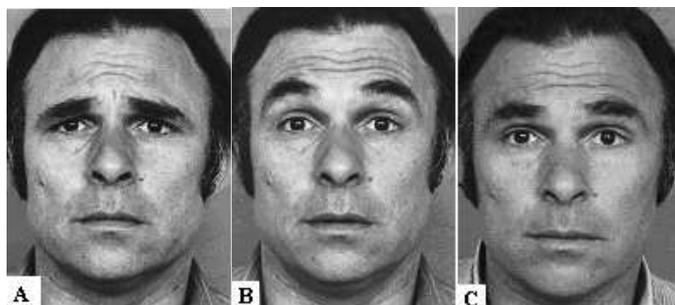


Figure 4 : Combinaison d'AUs : Image A : AU 1 + AU 4 ; Image B : AU 1 + AU 2 ; Image C : AU 1 + AU 2 + AU 4.

Selon les combinaisons, la longueur des sourcils augmente ou diminue. De plus, comme l'ont confirmé les relevés numériques des points, lors d'une hausse du sourcil, l'amplitude de la hausse du milieu du sourcil est plus grande que celle de l'extrémité extérieure du sourcil. Le mouvement est donc perçu d'avantage par la hausse du milieu du sourcil que par la hausse de l'extérieur même si l'AU 2 concerne l'extérieur.

Ce codage formel est efficace pour la description des émotions. Cependant, il ne permet pas de description de la structure temporelle du mouvement sous sa forme actuelle. Nous avons donc utilisé FACS uniquement comme une base initiale à partir de laquelle nous avons élaboré notre propre méthodologie afin de prendre en compte la structure temporelle de la LSF.

Pour les sourcils, FACS considère deux points mobiles (l'intérieur et l'extérieur) pouvant se mouvoir sur un axe horizontal et latéral. Nous avons donc retenu ces deux points pour l'annotation de la vidéo. De plus, comme nous l'avons évoqué, le milieu du sourcil est entraîné par les mouvements des extrémités des sourcils et l'amplitude de son mouvement est parfois plus grande que celle des extrémités. Nous avons donc considéré également le milieu du sourcil et ajouté un point à cet endroit.

Afin de limiter les imprécisions de l'annotation dues à l'épaisseur des sourcils, nous avons doublé les points des extrémités. De même, nous avons triplé le point du milieu du sourcil, qui est plus difficile à positionner avec précision. Enfin, pour déterminer les mouvements des sourcils de manière indépendante des mouvements de la tête nous avons ajouté des points sur un repère fixe : les deux extrémités de chacun des yeux. Notons que les coordonnées x, y de ces points sont relatives à un repère absolu qui est le coin supérieur gauche de l'image. De ce fait, un déplacement de la tête va induire un changement des coordonnées des points même si les sourcils sont immobiles. Il nous faudra donc rendre ces coordonnées indépendantes des mouvements de la tête.

Ainsi, ce sont 18 points que nous positionnons sur chaque image de la vidéo, à raison de 25 images par secondes. La figure 5 récapitule les 18 points positionnés sur chaque image. La figure 6 montre le résultat d'une annotation des 18 points sur une image extraite de la vidéo.

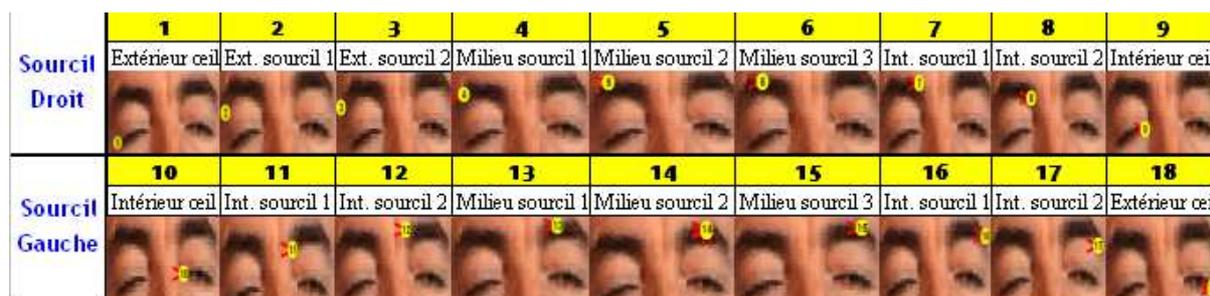


Figure 5 : Emplacement des 18 points

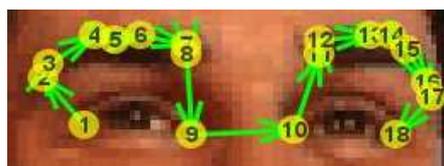


Figure 6: Image extraite de la vidéo avec les 18 points.

Une fois la vidéo entièrement annotée, les coordonnées de chaque point sont exportées dans un environnement de calcul scientifique afin d'établir des données statistiques et numériques.

3.2 Calcul à partir des coordonnées des points

Pour le traitement de nos données nous avons utilisé le logiciel Scilab⁵, logiciel libre de calcul scientifique, qui nous a permis, en créant un script simple, d'automatiser l'ensemble des calculs nécessaires. L'algorithme prend les données brutes en entrée (les coordonnées des 18 points de chaque image) et effectue une série de calculs :

1. Le premier calcul vise à déterminer les moyennes des 2 points des extrémités et des 3 points du milieu des sourcils dans chaque image. Ainsi, on passe de 7 points à 3 points pour chaque sourcil.
2. Les nouvelles coordonnées sont utilisées pour calculer la distance (D) qui sépare chacun des 3 points des sourcils des points extrémités des yeux (par exemple la distance entre le point 1 (x1, y1) et le point moyen 2-3 (x2, y2)) : $D = \text{Racine}((x1-x2)^2 + (y1-y2)^2)$.
3. A partir de la distance (D) on calcule la variation (V) de la position des points en faisant la différence entre la distance à l'image n et la distance à l'image n-1 : $V(n) = D(n) - D(n-1)$. Cette variation peut donc être positive (hausse) ou négative (baisse).
4. Enfin, la variation (V) nous permet de calculer la position (P) de chaque élément (les 3 points de chaque sourcil) rendue indépendante du mouvement de la tête pour chaque image en calculant la somme de la variation de l'image n et de la position de l'image n-1 : $P(n) = V(n) + P(n-1)$.

Les données obtenues sont ensuite utilisées afin d'évaluer l'annotation et de faciliter les analyses.

⁵

<http://www.scilab.org/>

3.3 Evaluation intermédiaire

Les valeurs calculées sont tout d'abord utilisées pour générer automatiquement l'animation des sourcils sur un visage synthétique à partir des données issues de l'annotation. L'animation a été générée à l'aide du logiciel Xface⁶ (Balci, 2004). Xface est une tête parlante en 3D conçue pour la production vocale, et non pour la production de LSF.

La génération automatique nous permet de procéder à une première évaluation qualitative de l'annotation tout en vérifiant si l'ensemble des phénomènes observés se retrouve dans l'annotation. En effet, en jouant simultanément le corpus vidéo et l'animation du visage synthétique, il est possible d'ajuster la méthodologie au fur et à mesure (rajout d'un point par exemple) pour que l'annotation finale soit la plus fidèle possible.

Les figures 8 et 9 illustrent deux configurations des sourcils et des yeux. L'image de gauche est extraite du corpus vidéo, tandis que celle de droite est extraite de Xface.



Figure 8 : Position neutre



Figure 9 : Diminution de la distance entre les sourcils et baisse de la paupière supérieure.

Jouer simultanément les deux vidéos nous permet également d'évaluer la qualité du visage synthétique, d'identifier ses défauts, afin de prévoir des conditions minimales de réalisme pour les visages des personnages de synthèse utilisés dans la production automatique de LSF. Par exemple, l'évaluation sur le logiciel Xface a permis d'observer des défauts tels qu'un manque d'amplitude au niveau des sourcils (il est possible de l'augmenter mais le rendu est irréaliste car il n'est pas conçu pour une telle amplitude) ainsi que l'absence de rides. La combinaison de ces deux défauts induit une perception réduite du mouvement des sourcils. D'autres mouvements ne sont pas possibles sur le visage synthétique de Xface (par exemple l'abaissement de la paupière du bas, indispensable pour ouvrir grand les yeux) ce qui, lors de l'évaluation, engendre des problèmes de perception de l'interlocuteur. Il est ainsi possible d'établir une liste des éléments essentiels que doivent posséder les visages synthétiques et les systèmes de génération automatique de LSF afin de prendre en compte les GNMs et augmenter la qualité visuelle et linguistique des messages générés.

Cette première utilisation des données permet une évaluation rapide de la méthodologie. Les données vont alors être utilisées pour l'analyse des GNMs.

⁶

<http://xface.itc.it/>

3.4 Analyse structurelle des GNMs

Les données numériques sont également exploitées pour produire des courbes afin d'analyser la structure des mouvements. Le graphique ci-dessous (Figure 10) nous informe par exemple de l'amplitude du mouvement de l'intérieur des sourcils et permet une classification des hausses.

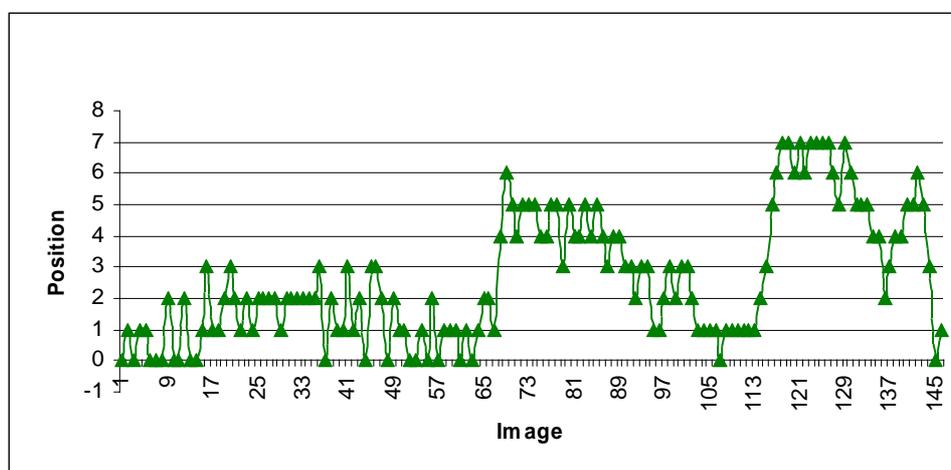


Figure 10 : Positions du point intérieur du sourcil droit.

La courbe nous montre trois amplitudes de hausses pour l'intérieur du sourcil : une petite (de 1 unité chez ce signeur), une moyenne (2 unités) et une forte (3 unités). Ces hausses peuvent être définies de façon précise par rapport à la hausse de niveau 1 (petite hausse) : une hausse de niveau 2 est deux fois plus haute qu'une hausse de niveau 1 tandis qu'une hausse de niveau 3 est trois fois plus haute qu'une hausse de niveau 1. Si la valeur numérique des amplitudes des hausses varie d'un signeur à l'autre, leur nombre et leur proportion les unes par rapport aux autres restent identiques. Une hausse très forte des sourcils (7 unités sur le graphique) va donc se faire par étapes, plusieurs hausses successives de différents degrés avec des tenues d'au moins une image pour chacune d'entre elles. Le constat est le même concernant les baisses. Les courbes peuvent donc nous permettre d'analyser finement la structure du mouvement des GNMs.

3.5 Evaluation de la formalisation

Les données numériques permettent enfin de valider et d'instancier, de façon numérique, la description formelle sous forme de flèches que nous avons proposée dans une première approche (Chételat-Pelé, Braffort, Véronis, 2007). Ce système repose sur quatre points :

- Une description du mouvement (et non d'une position), par exemple baisse de la paupière au lieu de paupière basse.
- Une décomposition des mouvements : une diagonale sera décomposée en un mouvement vertical et un mouvement horizontal (et/ou frontal).
- Une décomposition des éléments à annoter : par exemple nous séparons la paupière du bas de celle du haut.
- L'utilisation de symboles au lieu de mots afin d'aider à la perception visuelle des structures récurrentes à l'annotation avant tout traitement informatique (Figure 12). Un symbole peut décrire plusieurs phénomènes (par exemple la couleur varie en fonction de l'intensité, figure 13).

Références

- BALCI K. (2004). MPEG-4 based open source toolkit for 3d facial animation. In *AVI04, working Conference on Advanced Visual Interfaces*. Gallipoli, Italie, 25-28 Mai 2004.
- BERGMAN B. (1984). "Non-manual components in signed language: Some sentence types in Swedish Sign Language". In Loncke F., Boyes-Braem P., Lebrun Y. (eds), p. 49-59, Lisse : Swets & Zeitlinger.
- BRAFFORT A., CHOISIER A., COLLET C., CUXAC C., DALLE P. FUSELLIER I., GHERBI R., JAUSIONS G., JIROU G., LEJEUNE F., LENSEIGNE B., MONTEILLARD N., RISLER A., SALLANDRE M.-A. (2001). Projet LS-COLIN. Quel outil de notation pour quelle analyse de la LS ?. In *Journées Recherches sur la langue des signes*. UTM, Le Mirail, Toulouse.
- CHETELAT-PELE E., BRAFFORT A., VÉRONIS J. (2007). Mise en place d'une méthodologie pour l'annotation des gestes non manuels. In *TALS 2007. Traitement Automatique des Langues des Signes 2007 : atelier de Traitement Automatique des Langues Naturelles 2007*. Toulouse.
- CHETELAT-PELE E., BRAFFORT A., VÉRONIS J. (2008). Signs Language Corpus Annotation: Toward a New Methodology. In *LREC 2008. Conference on Language resources and Evaluation*. Marrakech. (A paraître).
- COERTS J. (1992). *Nonmanual grammatical markers. An analysis of interrogatives, negations, and topicalisations in sign language of the Netherlands*. 234 p. Amsterdam University, Dissertation.
- CUXAC C. (2000). *La Langue des Signes Française: Les voies de l'iconicité*. Faits de Langues. Paris : Ophrys.
- VICTORIA FLORES V. (2005). Artnatomy/Artnatomia. url (www.artnatomia.net). Spain.
- EKMAN P., FRIESEN W. V. (1978). *Facial Action Coding System (FACS)*. Manuel Palo Alto : Consulting Psychologists Press.
- JOUISSON P. (1995). *Ecrits sur la Langue des Signes française*. Garcia, B. (éd). Paris : L'Harmattan, Paris
- KIPP M. (2004). *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*" Boca Raton Florida: Dissertation.com
- LIDDEL S. (1980). *American Sign Language Syntax*. The Hague : Mouton Publishers.
- PRILLWITZ S., ZIENERT H. (1989). *Hamburg Notation System for Sign Language: Development of a sign writing with computer application*. Allemagne : S. Prillwitz & T. Vollhaber (Eds.): Current trends in European Sign Language Research Signum.
- SEGOUAT J., BRAFFORT A., MARTIN E. (2006). Sign language corpus analysis: synchronisation of linguistic annotation and numerical data. LREC 2006. *Fifth International Conference on Language Resources and Evaluation* Genoa Italy : 2006
- SUTTON V., GLEAVES R. (1995). *SignWriter – The world's first sign language processor*. Deaf Action Committee for SignWriting, La Jolla, CA.
- VERGER F. (2001). *Le regard en Langue des Signes Française*. Thèse de doctorat de Sciences du Langage. Université de Toulouse le Mirail.