



Édition interactive d'énoncés en langue des signes française dédiée aux avatars signeurs

Ludovic Hamon, Sylvie Gibet, Sabah Boustila

► **To cite this version:**

Ludovic Hamon, Sylvie Gibet, Sabah Boustila. Édition interactive d'énoncés en langue des signes française dédiée aux avatars signeurs. TALN - 20ème conférence du Traitement Automatique du Langage Naturel 2013, Jun 2013, Sables d'Olonne, France. 2013. <hal-00911629>

HAL Id: hal-00911629

<https://hal.inria.fr/hal-00911629>

Submitted on 29 Nov 2013

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.

Édition interactive d'énoncés en langue des signes française dédiée aux avatars signeurs

Ludovic Hamon¹ Sylvie Gibet¹ Sabah Boustila²

(1) IRISA, Université Bretagne Sud, Campus de Tohannic, Rue Yves Mainguy, 56000 Vannes

(2) ICUBE, Université de Strasbourg, 300 bd Sébastien Brant, BP 10413, 67412 Illkirch Cedex, France

ludovic.hamon@univ-ubs.fr, Sylvie.Gibet@univ-ubs.fr, boustila@unistra.fr

Cet article a été publié dans les actes de la Conférence TALN-RECITAL 2013

RÉSUMÉ

Les avatars signeurs en Langue des Signes Française (LSF) sont de plus en plus utilisés en tant qu'interface de communication à destination de la communauté sourde. L'un des critères d'acceptation de ces avatars est l'aspect naturel et réaliste des gestes produits. Par conséquent, des méthodes de synthèse de gestes ont été élaborées à l'aide de corpus de mouvements capturés et annotés provenant d'un signeur réel. Néanmoins, l'enrichissement d'un tel corpus, en faisant fi des séances de captures supplémentaires, demeure une problématique certaine. De plus, l'application automatique d'opérations sur ces mouvements (e.g. concaténation, mélange, etc.) ne garantit pas la consistance sémantique du geste résultant. Une alternative est d'insérer l'opérateur humain dans la boucle de construction des énoncés en LSF. Dans cette optique, cet article propose un premier système interactif d'édition de gestes en LSF, basé "données capturées" et dédié aux avatars signeurs.

ABSTRACT

Interactive editing of utterances in French sign language dedicated to signing avatars

Signing avatars dedicated to French Sign Language (LSF) are more and more used as a communication interface for the deaf community. One of the acceptance criteria of these avatars is the natural and realistic aspect of the constructed gestures. Consequently, gestures synthesis methods have been designed thanks to some corpus of captured and annotated motions, performed by a real signer. However, the enlarging of such a corpus, without requiring of some additional capture sessions, is a major issue. Furthermore, the automatic application of motion transformations (e.g. concatenation, blending, etc.) does not guarantee the semantic consistency of the resulting gesture. Another option is to insert the human operator in the utterance building loop. In this context, this paper provides a first interactive editing system of FSL gestures, based on captured motions and dedicated to signing avatars.

MOTS-CLÉS : Langue des Signes Française, édition, geste, base de données sémantiques, signeur virtuel, interaction.

KEYWORDS: French sign language, editing, gesture, semantic data base, virtual signer, interaction.

1 Introduction

La Langue des Signes Française (LSF) est une langue à part entière et constitue l'un des piliers de l'identité et de la culture sourdes. La loi 2005-102 du 11 février 2005 « pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées » a favorisé l'émergence d'applications dédiées à la promotion des moyens de communication et de diffusion de l'information en LSF. Dans le cadre de ces applications, l'utilisation de Signeurs Virtuels (SV) pour produire des messages en LSF semble être une alternative intéressante à la production de vidéos, dans la mesure où elle préserve l'anonymat des personnes sourdes et permet de manipuler, transférer et visualiser de nouveaux énoncés.

Dans ce contexte, les SV dédiés aux langues des signes font l'objet de recherches et d'études avancées, alliant l'analyse linguistique à la synthèse de gestes à partir de langages de construction dédiés. Quelle que soient la précision et l'expressivité du système d'édition, la construction d'un SV soulève des problématiques nombreuses dont l'une est commune à tous les SV : la recherche d'une cohérence linguistique des mouvements produits.

Parmi les techniques de synthèse de gestes en LSF, celles basées sur des mouvements capturés par un signeur réel, offrent l'avantage, d'un point de vue subjectif, d'obtenir des mouvements plus naturels et acceptables pour la communauté sourde (Gibet *et al.*, 2011), (Héloir, 2008), (Parisot *et al.*, 2010). Néanmoins, ce type de synthèse nécessite la définition d'un corpus préalable, obtenu après plusieurs séances de captures de mouvements coûteuses et fastidieuses. Par conséquent, le corpus initial est généralement réduit. Son enrichissement conduit à des solutions discutables telles que des séances de captures supplémentaires ou l'application de transformations complexes sur des mouvements pré-enregistrés, entraînant des modifications sémantiques incertaines.

Cet article présente une méthode alternative sous la forme d'un premier "framework" d'édition de la LSF à partir de corpus de gestes capturés. À l'aide de ce système, l'utilisateur peut créer de nouveaux énoncés en LSF et identifier/évaluer la sémantique résultant des gestes créés durant la simulation virtuelle. La deuxième partie de ce document présente un état de l'art relatif aux langages de description et de spécification des gestes dans le contexte des SV. La troisième section présente une analyse sur les principaux défis et contraintes d'un système d'édition de mouvements de la LSF. Le système d'édition est décrit dans la quatrième section avec des exemples illustrant ses possibilités et limites. Enfin, des perspectives sur les travaux futurs concluent ce document.

2 État de l'art

Cette section présente, de manière non exhaustive, les langages de description et de spécification des gestes en LSF relatifs aux SV, ainsi que les systèmes d'animation à partir de données capturées, du point de vue de l'édition de gestes en LSF.

Les systèmes de description ou de notation des gestes en LSF ont généralement pour but de décrire des mouvements plus ou moins structurés et codifiés. Des éléments constituant des signes sont généralement identifiés ainsi que, si possible, des règles syntaxiques et sémantiques régissant leur agencement. On distingue ainsi les unités atomiques appelées gestèmes

(Gibet *et al.*, 2001; Vogler, 2003) (phonèmes dans les langues parlées), les morphèmes (plus petites unités porteuses de sens, encore appelées unités phonologiques), les signes ou gloses (combinaisons de signes), les phrases (séquence de signes), les discours. Plusieurs systèmes de notation existent, l'un des plus utilisés aujourd'hui étant *HamNoSys* (Prillwitz *et al.*, 1989). D'autres systèmes de description suivent différentes approches telles que la description paramétrique (Stokoe, 2005), l'approche structurelle phonétique (Liddell, 1989), la visée iconique (Cuxac, 2000) ou la représentation géométrique (Filhol, 2008).

Les langages de spécification permettent de décrire le comportement d'un SV à l'aide d'un formalisme de description de commandes gestuelles. Les données (*e.g.* signes, phrases, suites de symboles, fonctions, etc.) issues de ces formalismes sont généralement interprétées en une séquence de paramètres de bas niveau, qui sont directement utilisés pour produire l'animation du SV. Un langage de spécification peut être vu comme une première Interface Homme-Machine (IHM) où l'utilisateur peut éditer, avant le lancement de l'animation, les mouvements générés par le SV. Les formalismes de spécification des gestes peuvent aller du script basé sur la structure des éléments phonétiques ou morphémiques (Gibet *et al.*, 2001; Elliott *et al.*, 2008) jusqu'à des langages prenant en compte d'autres aspects linguistiques des langues des signes, tels que les classifieurs traduisant l'iconicité des signes (Huenerfauth, 2006), la description de l'espace de signation (Lenseigne et Dalle, 2006) ou la construction syntaxique de phrases en LSF (Losson, 2000; Kervajan, 2011) par exemple.

Les contributions apportées par ces langages sont multiples et vont de la description non ambiguë des signes en termes de structures séquentielles, uni-modales et évoluant dans le temps, jusqu'à la prise en compte de leurs aspects modulateurs. Cependant, les langues des signes transmettent un message exprimé simultanément *via* différents canaux *i.e.* plusieurs parties du corps (*e.g.* mains, expression faciale, regard, etc.) (Huenerfauth, 2006; Vogler, 2003). Cette organisation parallélisée des gestes associée à la signification linguistique de leurs composants est rarement prise en compte par ces langages. De plus, l'édition des phrases en LSF reste complexe et fastidieuse, peu intuitive, elle nécessite de connaître le langage informatique et son paramétrage ainsi que de maîtriser la structure phonétique et morphologique des signes. Il en résulte que les langages de spécification ne permettent pas, dans la majorité des cas, de définir en un temps de spécification satisfaisant le comportement du SV, ni de générer des animations réalistes.

La synthèse de phrases en LSF requiert la mise en place de méthodes traduisant un scénario spécifié dans l'un des formalismes décrit précédemment, en une séquence de commandes gestuelles interprétable par le moteur d'animation. Les méthodes basées sur des mouvements enregistrés lors d'une séance de captures avec un acteur réel, permettent généralement de rendre l'avatar virtuel plus "expressif" et "naturel" que les méthodes procédurales et descriptives (Gibet *et al.*, 2011). Néanmoins, dans le cadre d'un signeur virtuel, deux difficultés apparaissent : (i) la combinaison et le parallélisme de plusieurs canaux, *i.e.* la production de gestes exécutés simultanément par plusieurs parties du corps incluant le torse, les épaules, les mains et le visage et (ii), la concaténation d'unités de mouvements qui ne garantit pas la consistance sémantique du mouvement résultant. Dans ce contexte, aucune équipe de recherche ne s'est intéressée à cette double problématique.

3 Besoins et contraintes d'un système d'édition de gestes

L'un des points centraux d'un système d'édition interactive de gestes en LSF est d'utiliser, pour la synthèse des mouvements de l'avatar, des données capturées sur un signeur réel. Nous proposons le schéma conceptuel suivant, centré autour de la gestion de données de la LSF pour construire de nouveaux énoncés et générer en sortie une animation d'un SV (Figure 1). Nous détaillons ci-après les différents modules de cette architecture, en précisant les défis à relever ainsi que les principales contraintes et limites d'un tel système d'édition. Ils comprennent (i) l'annotation sémantique des données, (ii) la base de données hétérogènes associée au moteur de requêtes, (iii) l'IHM permettant la construction d'énoncés en LSF, (iv) le moteur d'animation et de visualisation 3D.

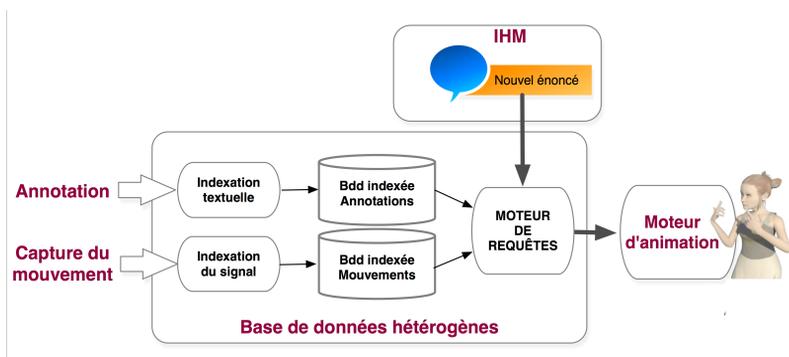


FIGURE 1 – Schéma conceptuel d'un système d'édition de gestes en LSF

Annotation sémantique des données : L'annotation des données capturées à partir des vidéos réelles est une étape située au cœur du processus d'édition. En effet, c'est à ce niveau que s'effectue réellement le lien entre la structuration linguistique des données et la caractérisation fine des éléments servant à contrôler l'animation de l'avatar. Afin de réaliser ce couplage, il est nécessaire de considérer plusieurs pistes d'annotation ou canaux tels que la configuration et l'emplacement des mains, l'expression faciale, la direction du regard, etc. La segmentation temporelle permet de définir les fragments temporels sur chacune des pistes correspondant aux gloses, signes, éléments phonologiques et morphémiques. Chaque fragment comporte une étiquette propre aux valeurs/attributs sémantiques définis suivant un schéma de spécification phonétique/phonologique/syntaxique (Duarte, 2012). Cette annotation, à la fois spatiale et temporelle, réalisée à partir des vidéos enregistrées lors de la séance de captures de mouvements avec le logiciel ELANTM, a notamment été exploitée dans le cadre du projet *SignCom* (Gibet *et al.*, 2011).

Base de données hétérogènes : Afin de générer des animations à partir des données préalablement enregistrées, il est nécessaire de construire une base de données, par nature hétérogènes, constituée de deux parties : (i) une base de données de mouvements, contenant les mouvements bruts, (ii) une base de données sémantiques, permettant d'indexer chaque fragment de mouvement selon une catégorisation linguistique. Le principal défi technique revient ici à définir les meilleures structures d'indexation à la fois pour le texte (données d'annotation) et pour le signal (données issues de la capture), tout en maintenant la cohé-

rence et la consistance entre ces deux niveaux d'information. Il est important à ce niveau de considérer l'efficacité de l'accès aux données et la capacité du langage d'interrogation de la base de données à extraire des informations précises et pertinentes. Un dictionnaire met en correspondance les annotations et les fragments de mouvements représentés par un ensemble de paramètres formels incluant l'identifiant du mouvement, la partie du corps concernée, les postures de début et de fin, etc. Pour extraire/charger un fragment de mouvement, il est possible d'interroger directement la base de données brutes avec les paramètres formels ou d'interroger la base de données sémantiques avec une glose, qui délivrera les paramètres formels, correspondant à un fragment de mouvement de la base de données brutes.

Interface et construction d'énoncés en LSF : La manipulation d'énoncés en LSF nécessite la mise en oeuvre de mécanismes interactifs à la fois intuitifs et efficaces pour caractériser les différents niveaux du langage : choix des signes/gloses, ordonnancement des gloses induisant la syntaxe des énoncés, aspects clausaux (négation, interrogation, etc.) souvent liés aux expressions faciales, informations émotionnelles (liées à la prosodie), etc. Il est nécessaire également de définir la façon de rechercher les mouvements dans le contexte de l'énoncé, ainsi que les paramètres de l'animation caractérisant la fluidité du mouvement. Enfin, on précise à ce niveau les choix d'apparence de l'avatar et les paramètres liés à la visualisation 3D (modélisation 3D, habillage, scène, éclairage, etc.).

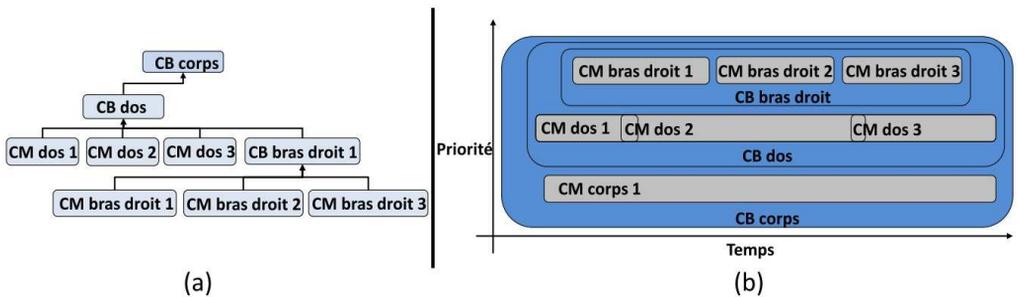


FIGURE 2 – Système de contrôleurs dans le projet *SignCom* (Gibet *et al.*, 2011). (a) arborescence des contrôleurs, (b) structures spatiale et temporelle des contrôleurs. CM : contrôleur de type rejeu *i.e.* "Motion player", CB : contrôleur de type "mélangeur" *i.e.* "Blender".

Animation du signeur virtuel : La création de nouveaux mouvements peut être réalisée par un assemblage de deux types de composants d'animation nommés "contrôleurs" (Figure 2). Le contrôleur de type "rejeu" (CM), associé à une partie du corps, récupère un fragment de mouvement de la base de données. Le contrôleur de type "mélangeur" (CB) mélange les CM temporellement, par interpolation ou concaténation et spatialement, en donnant une priorité dans l'ordre d'exécution du mouvement, selon la partie du corps concernée (Figure 2(b)). Dans le cadre du projet *SignCom*, des contrôleurs ont été définis à l'aide d'une arborescence (Figure 2(a)) spécifiée dans un script textuel simple. Du point de vue de l'édition, chaque CM représentant un geste, un énoncé est ici défini comme une organisation séquentielle et parallèle des gestes en LSF par cette arborescence.

4 Mise en oeuvre d'un système d'édition de gestes

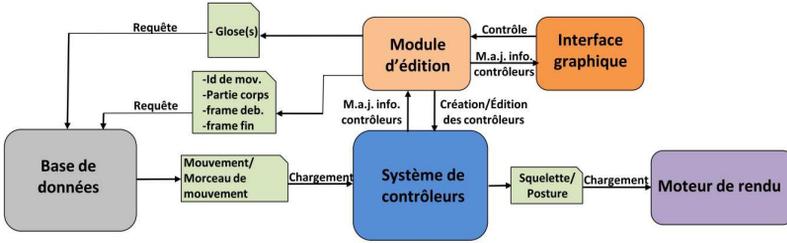


FIGURE 3 – Architecture du système d'édition de gestes

Les capacités du projet SignCom, en termes d'édition de gestes, possèdent certaines limites telles que la non-édition des contrôleurs lors de la simulation, l'utilisation de paramètres formels peu intuitifs et la non-exploitation de la base de données sémantiques (Gibet *et al.*, 2011).

Pour palier ces limites, un nouveau système d'édition a été créé (Figure 3). Ce système repose sur l'évolution de l'architecture du projet SignCom et sur un couple "module d'édition"/"interface graphique". Ce couple offre les moyens de créer et de spécifier les contrôleurs ainsi que d'observer une représentation de leur arborescence (Figure 5). Ce système permet, de plus, d'interroger la base de données avec une glose (*i.e.* interrogation de la base de données sémantiques) ou avec l'ensemble des paramètres formels (*i.e.* interrogation de la base de données brutes) dans le but d'extraire un fragment de mouvement et de le charger dans un CM.

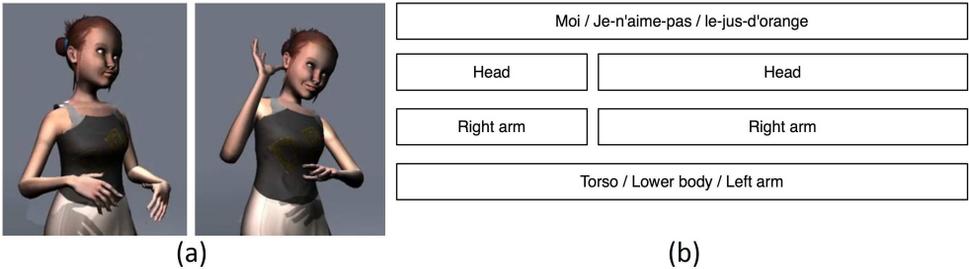


FIGURE 4 – (a) Visualisation de l'avatar signeur (b) Construction d'énoncés à partir de la spécification de la séquence de signes / gloses et des contrôleurs associés sur les différents canaux gestuels

Par conséquent, créer et éditer de nouveaux énoncés revient à spécifier et à éditer, de manière complète, la séquence de signes, à identifier les contrôleurs correspondants sur chaque partie du corps et leur hiérarchie par l'intermédiaire de l'interface graphique ; la qualité de l'animation résultante peut être évaluée en temps réel (Figure 4).

L'interface graphique est illustrée par la partie gauche de la figure 5(a). La partie "Bvh File search" permet de spécifier le type de contrôleurs ainsi que ses paramètres formels pour les CM et son type d'algorithme de mélange pour les CB. La création d'un CM, par une recherche



(a)



(b)

FIGURE 5 – (a) Aperçu de l'interface graphique (b) Extrait d'un scénario construit où l'avatar décrit la préparation d'un cocktail

selon une ou plusieurs glose(s) dans la base de données, peut être effectuée par la section "Search by keywords". La partie "Up to date" est dédiée à l'édition de chaque type de paramètre d'un contrôleur, avec une sous-section propre à chaque type. La hiérarchie des contrôleurs est affichée "textuellement" (Figure 5(a) haut gauche). Enfin, la figure 5(b) montre un extrait d'un scénario construit sur le thème de la description de l'élaboration d'un cocktail.

5 Conclusion et perspectives

Cet article présente un premier système original d'édition interactive de gestes de la Langue des Signes Française (LSF) dédié aux Signeurs Virtuels (SV). Reposant sur une base de données hétérogènes de mouvements capturés et annotés, ce système permet la concaténation et le mélange de gestes temporellement et spatialement à l'aide d'une interface graphique.

L'enrichissement d'un corpus de mouvements capturés de la LSF pour les SV demeure un problème complexe. Les solutions existantes peuvent être coûteuses et laborieuses (e.g. séances de captures de mouvements supplémentaires) ou incertaines (e.g. application de transformations sur des fragments de mouvements modifiant leur sémantique). Là où le mouvement est étroitement lié à une sémantique, les systèmes d'édition de gestes en LSF dédiés aux SV semblent être de plus en plus nécessaires pour construire, éditer et enrichir de tels corpus.

Les perspectives de ces travaux reposent sur l'enregistrement, dans la base de données hétérogènes, des nouveaux mouvements créés ainsi que de leurs sémantiques associées de manière optimale. Pour cela, la recherche d'une interface d'édition plus complète et plus intuitive (e.g. édition graphique de la hiérarchie des contrôleurs) sera poursuivie selon deux directions : (i) l'abstraction de tout paramètre numérique pour les non-experts du domaine de l'animation et (ii) l'édition aux niveaux spatial et temporel de la sémantique des mouvements créés à l'image des systèmes d'annotation vidéo actuels (e.g. logiciel ELAN). Des algorithmes d'apprentissage seront, par la suite, étudiés pour construire un système de "suggestions sémantiques" des mouvements créés par l'utilisateur, en fonction des opérations d'édition effectuées.

Références

- CUXAC, C. (2000). La langue des signes française (lsf) : les voies de l'iconocité. In *Faits de langues*, numéro 15–16. Ophrys.
- DUARTE, K. (2012). Motion capture and avatars as portals for analyzing the linguistic structure of signed languages. In *PhD thesis, Université de Bretagne Sud*.
- ELLIOTT, R., GLAUERT, J. R. W., KENNAWAY, J. R., MARSHALL, I. et SAFAR, E. (2008). Linguistic modelling and language-processing technologies for avatar-based sign language presentation. In *Universal Access in the Information Society*, volume 6, pages 375–391.
- FILHOL, M. (2008). Modèle descriptif des signes pour un traitement automatique des langues des signes. In *Thèse de doctorat, Université Paris-Sud*.
- GIBET, S., COURTY, N., DUARTE, K. et NAOUR, T. L. (2011). The signcom system for data-driven animation of interactive virtual signers : Methodology and evaluation. <http://www-irisa.univ-ubs.fr/Valoria/signcom/en/>.
- GIBET, S., LEBOURQUE, T. et MARTEAU, P. (2001). High level specification and animation of communicative gestures. In *Journal of Visual Languages and Computing*, volume 12, pages 657–687.
- HÉLOIR, A. (2008). Système de communication par agent virtuel, aide à la communication des personnes sourdes. In *Thèse de doctorat, Université de Bretagne Sud*, pages 168–171.
- HUENERFAUTH, M. (2006). Generating american sign language classifier predicates for english-to-asl machine translation. In *Thèse de doctorat, University of Pennsylvania*.
- KERVAJAN, L. (2011). Contribution à la traduction automatique français / langue des signes française (lsf) au moyen de personnages virtuels. In *PhD thesis, Université d'Aix/Marseille*.
- LENSEIGNE, B. et DALLE, P. (2006). Using signing space as a representation for sign language processing. In GIBET, S. et al., éditeurs : *Gesture in Human-Computer Interaction and Simulation, GW, Lecture Notes in Computer Science*, volume 3881, pages 256–260. Kluwer Academic.
- LIDDELL, S K et Johnson, R. E. (1989). American sign language : The phonological base. In *Studies in the Linguistic Sciences*, volume 64, pages 197–277.
- LOSSON, O. (2000). Modélisation du geste communicatif et réalisation d'un signeur virtuel de phrases en langue des signes française. In *PhD thesis, Université de Lille*.
- PARISOT, A. M., S, R., S, V. et A, V. (2010). Construire et déconstruire l'espace dans une langue des signes : démonstration d'un protocole d'enregistrement simultané des mouvements des membres supérieurs et du globe oculaire. In *Atelier TALS 2010 dans le cadre du congrès TALN 2010*.
- PRILLWITZ, S., LEVEN, R., ZIENERT, H., HANKE, T. et HENNING, J. (1989). *Hamburg Notation System for Sign Languages - An Introductory Guide*. University of Hamburg Press.
- STOKOE, W. C. (2005). Sign language structure : an outline of the communication systems of the american deaf. In *Linguistics, Occasional Papers 8 (1960), Journal of Deaf Studies and Deaf Education*, volume 10, pages 3–37.
- VOGLER, C. (2003). American sign language recognition : Reducing the complexity of the task with phoneme-based modeling and parallel hidden markov models. In *PhD Thesis, University of Pennsylvania*.