

Influence de la segmentation temporelle sur la caractérisation de signes

F. Lefebvre-Albaret J. Segouat

WebSourd, 99 route d'Espagne, 31 100 Toulouse

francois.lefebvre-albaret@websourd.org

jeremie.segouat@websourd.org

RÉSUMÉ

La segmentation temporelle des unités signifiantes des Langues des Signes est un problème délicat car il demande de se baser sur de nombreux indices. Pourtant, la délimitation du début et de la fin du signe est souvent une étape nécessaire, préalable à leur caractérisation. Nous explorons dans cet article la façon dont la caractérisation des signes peut être influencée par une variation de la segmentation. En prenant l'exemple de la caractérisation du mouvement, nous montrons comment la définition de critères de segmentation basés sur le mouvement ou les configurations manuelles peut influencer la robustesse des caractéristiques aux variations de frontière temporelle. Nous montrons aussi comment la nature de la mesure effectuée sur le segment (maximum, moyenne, valeurs aux frontières temporelles) influe sur la sensibilité à la segmentation temporelle.

ABSTRACT

Influence of the temporal segmentation on the sign characterization

Temporal segmentation of meaningful units in sign language utterances is a difficult problem because it requires a combination many informations. However, it is often necessary to find the beginning and the end of signs before making their characterization. In this article, we show how the characterization of the signs may be influenced by a variation of their segmentation. Taking the example of the movement characterization, we indicate how the definition of segmentation criteria based on movement or manual configurations can influence the robustness of the characterization to variations of the segment temporal boundaries. We also show how the nature of the measurement made on the segment (maximum, average, values on temporal boundaries) affects the sensitivity to temporal segmentation.

MOTS-CLÉS : Langue des Signes Française, caractérisation, segmentation.

KEYWORDS: French Sign Language, characterization, segmentation.

1 Introduction

Dans le domaine de l'étude des Langues des Signes (LS), la délimitation du début et de la fin des signes est une étape nécessaire, préalable à l'annotation d'unités signifiantes. Cette étape de segmentation est très difficile à expliciter car elle s'appuie sur une prise en compte simultanée de nombreux critères que chaque annotateur peut définir et prioriser différemment.

Dans cette contribution, nous montrons dans un premier temps pourquoi ce problème de segmen-

tation des signes est si délicat. En nous basant sur le corpus proposé pour ce défi gestuel, nous décrivons comment la segmentation peut être utilisée pour caractériser les mouvements utilisés en LS et nous discutons de l'impact d'une variation de segmentation sur cette caractérisation.

2 Quelques critères de segmentation

Lorsqu'on demande à plusieurs annotateurs de délimiter approximativement les signes manuels composant un énoncé en LS, on note généralement une bonne concordance des segmentations à l'exception des signes composés et des unités de grande iconicité. Des études comme (Brentari et Wilbur, 2008) tendent à montrer que la simple régularité phonologique des signes peut mener à une segmentation en signe indépendante de la compréhension des signes. Une expérimentation présentée dans (Lefebvre-Albaret, 2010) montre également que dans de nombreux cas, la seule connaissance du mouvement des mains du signeur permet d'effectuer une segmentation cohérente en signe, sans avoir besoin de comprendre l'énoncé segmenté. Dans les travaux qui précèdent la question est davantage la délimitation approximative d'unités signifiantes que la localisation précise des frontières temporelles du signe. La formulation de critères objectifs permettant de segmenter précisément les signes avec une méthode partagée par plusieurs annotateurs reste encore un problème ouvert.

Ce problème de segmentation pourrait paraître relativement aisé si les productions signées suivaient un modèle phononétique comme celui proposé dans (Liddell et Johnson, 1989) dans lequel un signe est réalisé comme une suite de plusieurs événements :

- **postures** durant lesquelles les paramètres manuels (position, configuration et orientation) sont stabilisés,
- **transitions** qui permettent de passer d'une posture à l'autre
- **tenues** durant lesquelles l'ensemble des valeurs des paramètres est conservé.

De notre point de vue, la réalité d'énoncés en LS est bien différente et met en œuvre des phénomènes de coarticulation (Segouat, 2010) résultant de l'influence mutuelle des signes. Chacun des paramètres du signe peut alors subir une modification en fonction des signes qui sont situés avant ou après lui. En particulier, la configuration d'une fin de signe peut varier en fonction de la configuration du signe suivant. Un autre phénomène qui se produit même dans la production de signes isolés est la stabilisation de la configuration avant le début du mouvement (Koech, 2007) qui avait déjà été observée depuis longtemps pour la LSF par (Jouison, 1990) cité par (Hanke *et al.*, 2011), et qui commence à être prise en compte dans les modèles permettant la synthèse de LS par les signeurs virtuels (Hanke *et al.*, 2011). D'autres facteurs peuvent également compliquer la définition de critères de segmentation. Parmi eux, citons les hésitations, les signes réalisés partiellement dans lesquels la réalisation du signe diffère largement de sa version isolée. Pour les raisons que nous venons d'évoquer, le fait de définir systématiquement le début et la fin d'un signe en contexte comme une stabilisation simultanée des paramètres qui le définissent est vouée à l'échec dans le cas général.

Il peut alors être tentant de définir le début et la fin des signes à partir d'un seul paramètre qui serait le plus saillant (par exemple : les frontières d'un signe sont marquées par un changement de configuration). Ceci n'est pas non plus réaliste car il arrive qu'un paramètre d'un signe soit identique au même paramètre du signe suivant (par exemple, un signe s'achève par la configuration *poing fermé* et le signe suivant commence par cette même configuration). Pour cette raison, il est souvent nécessaire de définir des critères applicables dans la majorité des cas et d'autres critères de substitution qui permettent de couvrir les autres cas.

Une fois que des critères de segmentation ont été définis précisément, il n'est pas non plus aisé de les appliquer systématiquement aux vidéos en LS car l'information nécessaire n'est pas toujours disponible dans les enregistrements. Ainsi, il arrive fréquemment qu'une vue de face du signeur, prise isolément, conduise à sous-estimer les mouvements hors-plan (dans le sens de la profondeur) et à conclure, à tort, à une absence de déplacement de la main. D'autre part, si la résolution temporelle ou spatiale de la vidéo s'avère insuffisante, il sera difficile de localiser d'une manière fiable l'instant où l'un des paramètres du signe reste stable.

3 Segmenter pour caractériser

Il n'est pas inutile de se poser la question de l'utilité, ou de l'utilisation ultérieure d'une segmentation car celle-ci peut varier selon les contextes d'étude et va influencer à la fois les critères utilisés et la précision de la délimitation des frontières des signes. Il est par exemple tout à fait acceptable d'un point de vue méthodologique de travailler avec une segmentation approximative (à $1/10^e$ de seconde près) d'unités si le but est uniquement d'en effectuer une reconnaissance automatique dans la mesure où seule une quantification des insertions, substitutions et délétions sera effectuée in fine.

En revanche, il peut être utile de définir beaucoup plus précisément les critères de segmentation dans les cas où la délimitation des frontières temporelles du signe est utilisée ultérieurement pour effectuer des mesures de caractérisation de l'unité (ouverture de la main, emplacement, vitesse de déplacement, durée etc.). Cela permet à la fois d'utiliser des critères de caractérisation comparables entre les différents signes et de rendre les mesures reproductibles par la communauté scientifique. Dans cette optique, il est tout à fait envisageable d'utiliser des critères de segmentation différents en fonction de l'application visée. En particulier, il peut être judicieux d'utiliser comme critère principal de segmentation, une mesure du paramètre sur lequel va porter la caractérisation (configuration, si on cherche à caractériser des angles entre les phalanges, ou bien mouvement, si on cherche à mesurer des amplitudes de signe par exemple).

4 Exemple de la caractérisation du mouvement

Nous illustrons l'importance du choix de critère de segmentation, pour la caractérisation du mouvement des signes utilisés dans le corpus DEGELS (référéncé sous l'identifiant `oai:crdo.fr:crdo000767` au SLDR d'Aix-En-Provence) capturé à 25 images par seconde.

Nous utilisons comme critère de segmentation du signe, la concordance de tous les paramètres manuels du signe avec sa définition (emplacement des mains par rapport au corps, emplacement relatif des mains, stabilisation de la configuration, stabilisation de l'orientation, pause du mouvement etc.). Ces critères peuvent être interprétés de manière subjective, d'autant plus qu'il

n'existe pas actuellement de définition unique de la notion de signe, problématique que nous laissons volontiers à la communauté des chercheurs en linguistique. Lorsque la configuration du signe est déjà stabilisée, mais qu'on ne se trouve pas entre les instants postures de début et de fin de signe marquées par un mouvement minimal, nous notons également les phases de préparation et de tenue du signe en nous inspirant de la démarche décrite dans (Kita *et al.*, 1998).

Nous visualisons directement les configurations, les orientations et les emplacements grâce aux deux vues dont nous disposons. Un indicateur objectif du mouvement dans une vidéo est la différence entre deux images successives d'une vidéo. Nous avons donc calculé cette distance inter-image en utilisant uniquement la partie droite de l'image correspondant au signeur dont nous segmentons le discours et en pondérant les différences inter-images relatives aux parties supérieures et inférieures de la vidéo (vues de face et de profil). Cette mesure est potentiellement critiquable en ce sens qu'elle prend aussi bien en compte le déplacement du buste que celui des mains, qu'elle surestime les mouvements selon l'axe vertical (pas de vue de dessus) et qu'elle est sensible à l'habillement et à l'éclairage du signeur. Toutefois, cet indicateur s'est révélé suffisamment fiable pour détecter les pauses ou les variations brusques dans les mouvements manuels.

Le processus de segmentation proprement dit a duré 3 heures auxquelles viennent s'ajouter 15 minutes de vérification.

4.1 Schéma d'annotation utilisé

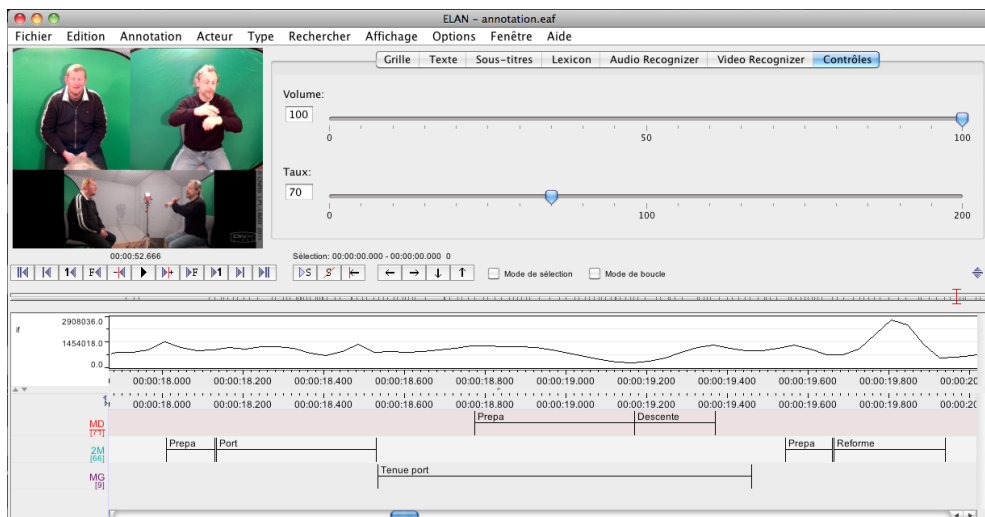


Figure 1: Capture d'écran de la fenêtre ELAN pendant le processus d'annotation.

Les pistes d'annotation sont structurées de la manière suivante :

- **IF** : Distance inter-image,
- **MD** : Segmentation des signes effectués par la main droite,
- **2M** : Segmentation des signes effectués avec les deux mains,
- **MG** : Segmentation des signes effectués par la main gauche.

La valeur des annotation est parmi les valeurs suivantes :

- **Glose** correspondant au signe segmenté,
- **Prepa** pour la préparation d'un signe,
- **Tenue** suivi de la glose du signe dont on effectue la tenue.

4.2 Dénombrement des critères utilisés pour la segmentation

Nous avons dénombré *a posteriori* les paramètres qui nous avaient permis d'effectuer la segmentation du début et de la fin des différents signes en notant la fréquence des différents phénomènes utilisés. En d'autres termes, nous observons les changements de paramètre entre l'image située à la frontière du signe et l'image précédente (pour la frontière de début) ou l'image suivante (pour la frontière de fin). Les notations utilisées sont les suivantes :

- **Mvt min** : Mouvement minimal des mains du signeur,
- **Mvt max** : Mouvement maximal des mains du signeur,
- **Emp** : Point de rebroussement ou changement brusque de la trajectoire des mains par rapport au corps,
- **Emp rel** : Point de rebroussement ou changement brusque de la trajectoire d'une main par rapport à l'autre,
- **Ori** : Variation minimale de l'orientation du signe,
- **Conf** : Première ou dernière image où la configuration manuelle est stabilisée.

Nous notons le nombre de fois où chaque critère est utilisé pour effectuer une segmentation de signe. Il est important de noter qu'une frontière de signe peut tout a fait être marqué par plusieurs phénomènes.

	Emp Rel	Conf	Emp	Emp + Ori	Ori	∅
Mvt Min	3	2	2	1	1	31
Mvt Max	0	3	0	0	0	7
∅	0	5	3	0	1	0

Figure 2: Fréquence d'utilisation des critères de segmentation pour détecter le début du signe

	Emp Rel	Emp Rel + Conf	Conf	Conf + Emp	Emp	Ori	\emptyset
Mvt Min	1	1	9	1	1	3	24
Mvt Max	1	0	4	1	0	1	0
\emptyset	3	0	7	0	2	2	0

Figure 3: Fréquence d'utilisation des critères de segmentation pour détecter la fin du signe

Nous voyons d'après les statistiques que les frontières temporelles de début de signe se situent dans plus de 2/3 des cas sur des pauses dans lequel le mouvement est minimal, ou sur des instants où le mouvement mesuré est maximal (qui correspondent en fait à des changements brusques de trajectoire, ou des instant où le mouvement devient plus contrôlé et ralenti). Dans 1/7^e des cas, les changements de configuration sont utilisés surtout comme marqueur de début et de fin dans des unités de très courtes durées. Les emplacements et les orientations sont également utilisés 1 fois sur 7.

Au niveau des marqueurs de fin de signe, les critères de mouvement sont utilisés pour plus de la moitié des signes, les changements de configuration sont utilisés 1 fois sur 4, tandis qu'1/5^e des fins de signes sont détectées grâce à l'emplacement ou à l'orientation des mains. Il n'est pas surprenant que les changements de configuration soient davantage des marqueurs de fin de signe dans la mesure où il arrive fréquemment qu'elles changent avant l'instant de mouvement minimal de la main.

Pour chacun des segments ainsi définis, nous calculons les grandeurs suivantes qui se rapportent au mouvement pour les mains impliquées dans la production des signes :

- La **vitesse maximale** des mains (dans le cadre des signes bi-manuels, il s'agit de la plus grande des vitesses maximales des deux mains),
- La **longueur du déplacement** des mains entre le début et la fin du signe (pour les signes bi-manuels, nous utilisons la somme des longueurs des déplacements des deux mains),
- L'**étalement du signe** calculé à partir de la de la racine carrée de la somme des variances des coordonnées des trajectoires des mains.

Nous choisissons volontairement trois grandeurs qui portent respectivement sur une seule valeur située à l'intérieur du segment (des maxima), les bornes du segment (des déplacements) et l'ensemble des valeurs du segment (des variances).

Les différentes grandeurs sont déterminées à partir d'un suivi des positions des mains du signeur, effectué à la main image par image, de manière à éviter les potentiels décrochages dus à un suivi automatique.

5 Influence de la segmentation sur la caractérisation

Nous mesurons ensuite les mêmes grandeurs en décalant les frontières temporelles du signe d'une image vers l'avant et vers l'arrière, et en effectuant successivement cette opération pour le début puis la fin du signe (Nous rappelons que la vidéo est échantillonnée à 25 images par seconde).

Nous notons ensuite la variation des trois grandeurs causée par le changement de segmentation. Nous groupons les différents signes suivant les critères de segmentation utilisés pour délimiter le début ou la fin du signe (étant donné qu'il s'agit des critères les plus utilisés pour segmenter, nous distinguons uniquement les stratégies de segmentation utilisant les mouvements minimaux d'une part, et la stabilisation de la configuration d'autre part).

Pour noter les résultats, nous adoptons les conventions suivantes :

- **début-** : Erreur d'estimation entraînée par un décalage de la position du début du signe d'une image vers l'arrière,
- **début+** : Erreur d'estimation entraînée par un décalage de la position du début du signe d'une image vers l'avant,
- **fin-** : Erreur d'estimation entraînée par un décalage de la position de fin du signe d'une image vers l'arrière,
- **fin+** : Erreur d'estimation entraînée par un décalage de la position de fin du signe d'une image vers l'avant.

La formule utilisée pour déterminer l'erreur d'estimation sur une mesure est la suivante :

$$\frac{|Mesure - Mesure\ avec\ erreur\ d'estimation|}{Mesure}$$

5.1 Erreur dans l'estimation de la vitesse maximale

	Mvt min	Conf
début-	2,2%	9,4%
début+	1,1%	4,2%
fin-	1,5%	6,4%
fin+	0,7%	2,8%

Figure 4: Erreur d'estimation de vitesse maximale en fonction du type de variation des frontières du segment et du critère de segmentation

Dans le cas de l'estimation de la vitesse maximale, les erreurs rapportées sont des erreurs moyennes sur l'ensemble des signes de notre corpus, car l'erreur mesurée est relativement indépendante de la durée des segments.

5.2 Erreur dans l'estimation du déplacement entre le début de la fin du signe

Nous représentons dans le graphe qui suit l'erreur d'estimation du déplacement consécutif à un décalage de la position du début du signe de $-1/25^e$ s, en fonction de la durée du signe en

milliseconde. Comme nous pouvons le voir, il y a une influence significative de la durée (d) sur l'erreur d'estimation (ϵ) que nous avons modélisé par une relation de la forme $\epsilon = ad^{-\beta}$.

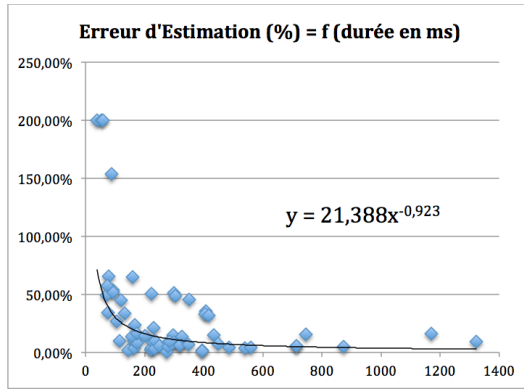


Figure 5: Erreur d'estimation du déplacement causé par une variation de la frontière de début du signe de $-1/25^e$ s

Nous effectuons donc systématiquement des régressions de manière à déterminer, d'après notre modèle, l'erreur d'estimation moyenne pour un signe d'une durée de 250 ms (valeur proche de la durée médiane de nos échantillons). Les résultats obtenus avec cette méthode sont les suivants :

	Mvt min	Conf
début-	9,8%	21,3%
début+	9,0%	24,6%
fin-	10,9%	13,9%
fin+	10,1%	17,5%

Figure 6: Erreur d'estimation du déplacement en fonction du type de variation des frontières du segment et du critère de segmentation

5.3 Erreur dans l'estimation de la dispersion du signe

La relation de dépendance entre l'erreur d'estimation de la dispersion et la durée du signe est de la même forme que celle évoquée pour le déplacement ($\epsilon = ad^{-\beta}$). Les valeurs indiquées dans le tableau qui suit sont donc également des erreurs d'estimation pour des signes de 250 ms.

	Mvt min	Conf
début-	3,7%	7,5%
début+	5,8%	8,1%
fin-	5,4%	6,5%
fin+	3,2%	6,3%

Figure 7: Erreur d'estimation de la dispersion en fonction du type de variation des frontières du segment et du critère de segmentation

5.4 Analyse des résultats

Dans un premier temps, notons que l'erreur d'estimation sur les caractéristiques du mouvement est systématiquement plus importante en utilisant le critère de configuration qu'en utilisant le critère de mouvement minimal pour la segmentation. L'augmentation d'erreur est plus marquée pour l'estimation de la vitesse maximale, ainsi que pour le changement de délimitation du début de signe. Dans ces deux cas, l'erreur d'estimation passe du simple au double en changeant de critère de segmentation.

Les erreurs d'estimations sont fortement liées aux grandeurs caractérisées. Nous pouvons classer les mesures selon leur robustesse à la variation de segmentation :

- **Vitesse maximale** : très robuste au changement de segmentation
- **Dispersion** : moyennement robuste au changement de segmentation
- **Déplacement** : très peu robuste au changement de segmentation

Nous nous gardons de quantifier cette robustesse en raison du peu de données dont nous disposons dans cette expérimentation.

Il apparaît clairement que le sens de l'erreur de segmentation commise (segmentation trop large ou trop courte) peut influencer différemment sur la précision de l'estimation. Cependant, les tendances sont dépendantes de la grandeur calculée, si bien qu'il ne nous est pas possible d'en tirer des conclusions générales pour la segmentation.

6 Utilisation des résultats pour la définition de critères de segmentation

A partir de l'analyse des résultats qui précèdent, nous pouvons mettre en lumière plusieurs éléments qu'il faut garder à l'esprit lors de la définition de critères de segmentation, si celle-ci est utilisée ultérieurement à des fins de caractérisation.

Pour caractériser le plus précisément possible un **segment** défini par deux **frontières** de début et de fin à l'aide de la combinaison de **mesures** prises à chaque instant dans le segment (par

exemple position des mains, rotations des articulateurs), il est possible de jouer sur plusieurs facteurs :

1. La **précision de la mesure** utilisée,
2. Le **critère de segmentation**. La caractérisation sera plus stable si la variation de la mesure aux frontières du segment est minimale,
3. La **résolution temporelle** des mesures (ou fréquence d'échantillonnage) qui permettra d'augmenter la précision de la segmentation,
4. La **construction de la mesure** de caractérisation qui devra donner un poids plus important aux mesures les plus éloignées des frontières du segment (et dont la prise en compte sera la moins affectée par une variation des frontières temporelles du segment),
5. La **taille des segments**. Plus le segment dure longtemps, moins la caractérisation sera affectée par une variation des frontières du signe.

Une des contributions de notre étude est d'avoir permis de quantifier les impacts des facteurs (2), (4) et (5).

7 Pour aller plus loin dans l'analyse de la relation entre segmentation et caractérisation

Nous avons introduit une méthode d'analyse de la sensibilité de la caractérisation à la segmentation appliquée aux mouvements manuels qui a conduit à une estimation de la précision de la caractérisation. Il serait nécessaire de poursuivre cette analyse avec des données de capture de mouvement pour s'affranchir des erreurs d'estimation causées par un suivi approximatif effectué à la main, ainsi qu'un corpus plus important.

Pour aller plus loin dans l'analyse de la robustesse des grandeurs de caractérisation du signe à la segmentation, il pourrait être intéressant de prendre en compte également d'autres facteurs :

- **La nature du signe caractérisé** : Il faudrait par exemple dissocier les signes impliquant des rotations, des mouvements répétitifs, des mouvements en aller-retour etc. (Lefebvre-Albaret et Dalle, 2008).
- **La nature de la caractéristique mesurée** : Il est probable que les modèles d'estimation d'erreur diffèrent pour des paramètres manuels et non manuel.
- **La prosodie** : Il est également probable que la robustesse de la mesure soit influencée par la manière dont sont exécutés les signes (nerveux, lent, petit espace de signation etc.).
- **Les biais systématiques** : Le changement de critère de segmentation peut entraîner une variation systématique dans l'estimation des paramètres (par exemple, une surestimation systématique de la durée d'un signe ou une sous-estimation de la vitesse moyenne du déplacement des mains).

L'étude de tous ces éléments permettrait de déterminer la meilleure manière de caractériser les différents types de signes, la manière optimale de les segmenter, et la sensibilité de la caractérisation aux segmentations.

L'application concrète de ces résultats fondamentaux permettrait à l'avenir de guider la communauté scientifique dans l'explicitation du protocole de caractérisation des signes afin de mentionner d'emblée les éléments qui pourraient influencer de manière significative sur les résultats obtenus et rendre les caractérisation plus reproductibles. En cas de protocoles divergents pour les caractérisation de signes, la quantification des biais et erreurs permettrait d'évaluer la compatibilité entre des résultats d'analyse produits par plusieurs équipes, voire d'effectuer des méta-analyses en pondérant les résultats compte tenu de leur précision.

References

- BRENTARI, D. et WILBUR, R. (2008). A cross-linguistic study of word segmentation in three sign languages, sign languages: spinning and unraveling the past, present and future. *Quadros (ed.)*. Editora Arara Azul. Petrópolis/RJ. Brazil.
- HANKE, T., MATTHES, S., REGEN, A., STORZ, J., WORSECK, S., ELIOTT, R., GLAUERT, J. et KENNAWAY, R. (2011). Using timing information to improve the performance of avatars. In *Second International Workshop on Sign Language Translation and Avatar Technology*, Dundee.
- JOUISON, P. (1990). Analysis and linear transcription of sign language discourse. In *Current trends in European sign language research. Proceedings of the 3rd European Congress on Sign Language Research*, pages 337–353, Hamburg: Signum-Verlag. Prillwitz, Siegmund / Vollhaber, Tomas (eds.).
- KITA, S., van GIJN, I. et van der HULST, H. (1998). Movement Phases in Signs and Co-speech Gestures, and Their Transcription by Human Coders. *Gesture and Sign Language in Human-Computer Interaction*, pages 23–35.
- KOECH, C. (2007). *A kinematic analysis of sign language*. Thèse de doctorat, New Jersey Institute of Technology's.
- LEFEBVRE-ALBARET, F. (2010). *Traitement automatique de vidéos en LSF, modélisation et exploitation des contraintes phonologiques du mouvement*. Thèse de doctorat, Université de Toulouse.
- LEFEBVRE-ALBARET, F. et DALLE, P. (2008). Une approche de segmentation de la langue des signes française. In *15ème conférence sur le Traitement Automatique des Langues Naturelles*, Avignon.
- LIDDELL, S. et JOHNSON, R. (1989). American sign language : the phonological base. *Sign Language Studies*, 64.
- SEGOUAT, J. (2010). *Modélisation de la coarticulation en Langue des Signes Française pour la diffusion automatique d'informations en gare ferroviaire à l'aide d'un signeur virtuel*. Thèse de doctorat, Université Paris-sud / Orsay.