



**HAL**  
open science

## Accessibilité et langue des signes : modélisations, méthodes, applications

Patrice Dalle, Annelies Braffort, Christophe Collet

► **To cite this version:**

Patrice Dalle, Annelies Braffort, Christophe Collet. Accessibilité et langue des signes : modélisations, méthodes, applications. Conférence Internationale sur l'Accessibilité et les Systèmes de Suppléance aux personnes en situations de handicap "Pour une meilleure insertion dans la société" (2007), Jan 2007, Toulouse, France. pp.1–8. hal-01633759

**HAL Id: hal-01633759**

**<https://hal.science/hal-01633759>**

Submitted on 22 Feb 2024

**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.

# Accessibilité et langue des signes : modélisations, méthodes, applications

Patrice Dalle  
IRIT-UPS  
118 route de Narbonne  
31062 Toulouse cedex 9  
(+33)5 61 55 63 09  
dalle@irit.fr

Anneles Braffort  
LIMSI-CNRS  
BP 133  
91403 Orsay  
(+33)1 69 85 81 68  
braffort@limsi.fr

Christophe Colet  
IRIT-UPS  
118 route de Narbonne  
31062 Toulouse cedex 9  
(+33)5 61 55 63 20  
collet@irit.fr

## ABSTRACT

Dans le but de développer des méthodes d'analyse et de synthèse de la langue des signes (LS), nous proposons une modélisation commune de l'espace de signation sur laquelle s'appuie chacun des deux processus. Nous décrivons ensuite une modélisation des gestes du signeur et des méthodes d'analyse des différents composants du signe. Enfin nous présentons des règles d'élaboration d'un énoncé et une représentation du lexique, utilisées pour piloter un signeur virtuel.

## Categories and Subject Descriptors

K42 [Computing Milieux]: Computers and Society –Social Issues– *Assistive technologies for persons with disabilities.*

## General Terms

Algorithms, Human Factors, Languages.

## Keywords

Langue des signes française (LSF), modélisation, analyse de la LSF, génération de la LSF, signeur virtuel, communication en LSF, accessibilité, surdité

## 1. INTRODUCTION

L'accessibilité pour les personnes sourdes passe en grande partie par l'accessibilité aux technologies de l'information et de la communication. Ces technologies doivent donc pouvoir être utilisées en langue des signes (LS), ce qui suppose des interfaces inter-langues. Ces applications seront utilisées si elles sont fiables et respectueuses des spécificités de la LS. En amont des applications de communication, il faut donc construire une modélisation des caractéristiques de la LS aussi bien pour analyser et comprendre des productions en LS que pour en générer via des signeurs virtuels.

Nos recherches visent, à terme, soit à permettre des échanges entre la LS et la langue française par un processus de

traduction, soit à permettre l'usage d'un dispositif à l'aide de la langue des signes. En analyse, à partir de la production en LS d'une personne sourde, le système analysera le message en vue d'en fournir une représentation permettant de le traduire en français, ou interprétera le message pour reconnaître une requête adressée à un dispositif (interrogation d'un système d'information par exemple). En synthèse, à partir d'une entrée en français ou dans un langage spécialisé, le système analysera le message en vue de générer les commandes pour animer un avatar délivrant le message équivalent en LS (annonce par écran dans les lieux publics par exemple).

Les applications envisageables sont très nombreuses. Dans le domaine de la communication, on peut citer tout ce qui relève de la communication à distance (messagerie avec signeur virtuel, MMS bilingue, écriture en LS), l'accès à des systèmes d'information, l'usage d'internet, soit pour émettre des requêtes en LS, soit pour disposer de services internet incluant la LS. Dans le domaine de la formation, des outils et méthodes d'apprentissage de la LS permettraient d'étendre sa diffusion, dans les familles concernées et dans la société ; des outils pédagogiques basés sur la LS ou des aides à la fabrication de documents bilingues faciliteraient l'enseignement bilingue des jeunes sourds ou la formation d'interprètes.

Ces applications présentent des degrés de complexité très variables suivant leur contexte d'utilisation. La variabilité des entrées à analyser ou le niveau de qualité acceptable de la LS à générer imposent de concevoir des systèmes basés sur une représentation interne de la LS et donc sur une modélisation qui respecte les spécificités de cette langue : l'exploitation de l'espace de signation pour représenter les éléments du discours, l'usage de l'iconicité et la visée illustrative de cette langue comme principe structurant de la langue [10], l'utilisation en parallèle de plusieurs composants corporels (mains, épaules, buste, tête, expressions du visage, regard), et enfin l'imbrication de la forme et du sens à tous les niveaux linguistiques.

Dans cette perspective, le couplage de l'analyse et de la génération de la LS semble très pertinent. Au niveau opérationnel, l'analyse des signes peut intervenir en phase d'apprentissage de la génération (capture de geste, utilisation de l'espace, description des expressions du visage...). Inversement, la génération est utilisée pour valider des modèles de gestes ou une interprétation de message. De nombreuses applications peuvent faire appel aux deux modes. Enfin de nombreux aspects du fonctionnement des LS restent encore à découvrir ou à expliciter. La coopération de l'analyse et de la synthèse permet d'affiner la modélisation de la LS et la



compréhension de son fonctionnement, en collaboration avec la recherche linguistique.

Nous présentons les modèles communs à ces deux modes, les méthodes d'analyse et de génération basées sur ces modèles et les premières applications qu'elles permettent de réaliser pour développer l'usage de la LS, faciliter sa diffusion et favoriser ainsi l'insertion sociale des personnes sourdes.

## 2. MODÉLISATION DES ESPACES

Dans la grande majorité des recherches menées jusqu'alors sur les langues des signes, l'énoncé est considéré comme une succession de signes isolés, éventuellement coarticulés. Les modèles utilisés sont parfois très éloignés de la réalité du fonctionnement des LS (dactylogologie, d'américain signé...). De plus les éléments non manuels sont encore rarement considérés alors que leur rôle est essentiel. Aucun système ne les prend tous en compte [25], même si quelques projets, essentiellement dans le domaine de la génération, introduisent certains aspects syntaxico-sémantiques spécifiques aux LS (projet européen Visicast [23], [18]).

En France, la plupart des études se réfèrent au modèle linguistique de C. Cuxac, fondé sur l'utilisation pertinente de l'espace et l'iconicité [10],[11] : modélisation de l'espace de signation pour la reconnaissance des verbes directionnels et l'utilisation de proformes [5], pour l'analyse par traitement d'image (TI) [21] ou pour la génération par avatar [8].

Partant de l'hypothèse que les énoncés sont structurés spatialement, à l'aide de l'espace de signation, nous présentons d'abord une modélisation de cet espace, commune à l'analyse et à la synthèse, en précisant ensuite les spécificités de ces deux domaines.

### 2.1 L'espace de signation, fondement de la structuration des énoncés

L'espace de signation est le volume situé devant le signeur. Celui-ci utilise cet espace pour placer les entités du discours et y faire référence. Cette représentation peut posséder des niveaux de granularité plus ou moins fins en fonction des besoins.

#### 2.1.1 Modélisation des entités

Selon la nature des entités et le contexte, il est nécessaire de manipuler des informations plus ou moins complexes sur l'entité. En synthèse, chaque propriété devra avoir une valeur pour pouvoir être réalisée alors qu'en analyse on ne cherchera à extraire que ce qui est utile pour la compréhension de l'énoncé.

##### 2.1.1.1 Types et fonctions

Un premier niveau de représentation de l'espace de signation peut être décrit sous la forme du schéma UML (Figure 1-a) et visualisé sous forme analogique (Figure 1-b). Les différents types d'entités sont définis par les fonctions que peut remplir l'entité : locatif, référence temporelle (relative au temps de l'énoncé ou au temps de l'énonciation) ou référence actancielle. Chaque entité possède un numéro d'ordre de création dans l'énoncé et une liste des fonctions qu'elle a remplies depuis sa création. Le signeur peut ainsi créer des *dates*, des *lieux*, des *objets*, des *animés* (dont nous avons distingué l'entité *personne*) et des *actions* qui sont des relations entre entités. En analyse, le système vérifie, d'après leurs fonctions, la cohérence des créations.



Figure 1 : Modélisation (a) et visualisation (b) de l'espace de signation pour analyser la langue des signes

#### 2.1.1.2 Caractéristiques des entités

Outre son emplacement par rapport au signeur, il peut être nécessaire de définir explicitement l'orientation d'une entité, par exemple pour montrer qu'une personne se dirige vers une voiture, ou s'en éloigne. Ce n'est pas toujours nécessaire et même, pour certaines entités, non pertinent (entités non orientables, comme le ballon, ou de nature « massive » telles que le ciel).

De même, la forme et la taille de l'entité peuvent être précisées. Lorsqu'on a besoin de se référer à une partie de l'entité dans la suite du discours, on utilise une description hiérarchique. La partie la plus générale est décrite en premier. Puis l'attention est portée sur des parties de plus en plus précises, avec autant d'étapes qu'il est nécessaire pour pouvoir communiquer avec précision l'emplacement de la partie dont il est question par rapport à l'entité de base. On peut ainsi décrire successivement une maison, puis, avec la main dominée, reprendre le toit et, avec la main dominante, représenter la cheminée et la placer sur le toit. Ce processus est réitéré si l'on veut ensuite représenter la fumée sortant de la cheminée.

Dans certains cas, il suffira de donner une indication vague de l'étendue de l'entité pour définir une relation spatiale avec une autre entité (livre sur une table).

#### 2.1.1.3 Caractéristiques sémantico-cognitives

Une autre famille de caractéristiques concernent les aspects sémantiques ou cognitifs attachés aux entités tels que la nature « massive », pour lesquelles on n'aura pas besoin d'une description hiérarchisée (la mer, le beurre), le caractère « individualisable » (les objets, les personnes), le type « collectif » (la foule). Ce typage des entités, qui vient s'ajouter aux types « date », « lieu » et « action », permet d'appliquer un filtrage sur le niveau de description des entités, ainsi que sur leur utilisation dans l'énoncé.

Des propriétés de nature topologiques sont utilisées lors de la manipulation des entités dans l'espace de signation à l'aide de proformes, qui vont reprendre un de ses traits saillants choisis en fonction du contexte. Ainsi, on associera à chaque entité un ensemble de valeurs sur ses propriétés topologiques (forme plane, creuse, non délimitée...).

Enfin des propriétés relationnelles, telles que les relations spatiales ou temporelles, sont déduites par le système.

#### 2.1.2 Modélisation de la scène de signation

La scène de signation est constituée d'entités, placées devant le signeur en des emplacements définis, ainsi que des relations

entre ces entités. Ces entités sont repérées par un système de coordonnées centré sur le bassin du signeur. Certaines de ces entités sont prédéfinies et génériques, telles que celles employées pour représenter le passé (demi-droite horizontale dirigée vers l'arrière du signeur), le présent (au niveau du signeur) et le futur (demi-droite horizontale dirigée vers l'avant du signeur). Ces informations vont être indispensables pour structurer des énoncés temporels spatialisés.

#### 2.1.2.1 Relations entre les entités

En LS, les relations sont explicitement montrées dans la scène de signation. A partir des emplacements des entités, on peut déduire des relations spatiales simples, telles que « au dessus », « à droite », ou plus complexes si la configuration de la main intervient (notion d'« intérieur »). Sur l'axe du temps (entités de type date), les relations seront de nature temporelles et suivront les ordres temporels absolu ou relatif au temps de l'énoncé.

Dans tous les cas, la nature de la relation entre les entités mais aussi son orientation interviennent pour déterminer l'ordre dans lequel sont réalisés les signes. En effet, on signe généralement en premier l'entité la plus massive, la plus stable, ou la plus ancienne, puis l'entité la plus mobile, ou la plus récente [10].

#### 2.1.2.2 Pérennité des relations

Au cours du temps, des entités peuvent disparaître, et leurs emplacements peuvent alors être réutilisés par de nouvelles entités. Mais une entité peut aussi être réactivée dans le discours. Un mécanisme doit donc être prévu pour intégrer cette fonction d'oubli ou de mémoire et pour gérer la disponibilité des emplacements.

#### 2.1.3 Représentations informatiques : graphe de scène

La représentation de l'ensemble des entités et de leurs relations spatiales est un graphe comportant autant de noeuds qu'il y a d'entités dans le discours. Ces noeuds contiennent des informations sur les caractéristiques des entités. Certaines valeurs standard peuvent être fournies par défaut (orientation horizontale du plateau d'une table). Les arcs du graphe sont étiquetés et orientés ; ils contiennent les informations sur les relations (spatiales, temporelles). Quand une forme est complexe, elle peut être décomposée en plusieurs noeuds reliés par des arcs (relation partie-tout), de manière à obtenir une description hiérarchique de l'entité. Ce graphe sera instancié de manière différente en analyse et en synthèse. Actuellement, notre représentation est de complexité moyenne. Elle permet de générer des énoncés isolés. Elle devra être enrichie pour permettre de générer des discours plus complexes ou un dialogue

### 2.2 L'espace neutre

Les entités ne sont pas toujours spatialisées. Elles sont parfois simplement réalisées dans l'espace dit « neutre ». Ou encore, il arrive souvent qu'une entité soit tout d'abord signée dans l'espace neutre puis spatialisée à l'aide d'un proforme, du regard, d'un mouvement d'épaule ou de la tête. Dans ce cas, l'entité va d'abord « transiter » par l'espace neutre avant d'être spatialisée dans la scène de signation.

L'espace neutre est représenté par une pile afin de permettre aisément le dépilement d'une entité en vue de son intégration dans le graphe. Cette pile est composée d'éléments comportant les mêmes informations que les noeuds du graphe, sauf les

informations sur l'emplacement et l'orientation qui sont absentes ici.

### 3. MODÈLES ET MÉTHODES D'ANALYSE

Cette section présente les objectifs et les problématiques spécifiques à l'analyse des LS réalisée à l'IRIT (§ 3.1.), ainsi que la description d'une implémentation du modèle de l'espace de signation (§ 3.2.) dans le contexte de l'analyse automatique par traitement d'images (§ 3.3.).

#### 3.1 Objectifs et contraintes de l'analyse de la LS

Dans notre cadre, l'analyse par ordinateur de vidéos en langue des signes a pour but de comprendre le sens du message, soit pour étudier et expliquer le fonctionnement grammatical de la LS, soit pour implémenter une application particulière ayant une entrée en LS. L'analyse des gestes et des expressions (phase d'extraction d'indices visuels et de description) est complétée par une identification des entités et de leurs relations, par la construction du graphe de scène (§2.1.3) et par la combinaison de ces éléments suivant différents niveaux linguistiques, jusqu'au niveau sémantique. Le système doit intégrer une modélisation de la LS et, dans le cas d'une application, la sémantique du domaine. Pour viser des applications réelles, on doit pouvoir traiter des données généralement acquises par une seule caméra et dans des environnements si possible peu contraints.

D'une façon générale, le processus d'analyse et d'interprétation d'image comporte une phase d'apprentissage du modèle de la scène à analyser (ici un personnage et ses gestes) et une phase d'interprétation. La phase d'interprétation consiste à créer une représentation informatique (un graphe par exemple) associant les éléments du modèle à leur réalisation dans l'image. Pour cela, l'image doit avoir au préalable été analysée, en vue de la structurer en composants qui pourront être caractérisés et reconnus. On a donc un double mécanisme de prédiction de ce qu'on s'attend à observer dans l'image (top-down) et de description (bottom-up) de l'image, suivi d'un processus d'appariement entre les éléments prédits et décrits. Cette mise en correspondance peut s'effectuer à des niveaux différents (des pixels à une représentation 3D élaborée) qui varient au cours de l'interprétation.

Dans le cas de vidéos en LS, l'analyse est particulièrement complexe car on étudie un objet composé, articulé et déformable (le corps humain), qui évolue de manière importante et rapide dans le temps, dans un environnement parfois complexe et variable [16, 13, 12, 22]. De plus, dans le cas d'une seule vue, il présente de fréquentes occultations pouvant être importantes (mains-mains, mains-visage...).

Afin de réduire la combinatoire, il est indispensable de pouvoir prédire au maximum ce qu'il faut observer afin de guider l'analyse. Une première solution consiste à délimiter le champ sémantique de l'application, et donc ici le champ lexical. Un deuxième axe concerne la modélisation d'un personnage et de son comportement : une modélisation articulaire du corps humain permet d'inférer la posture des bras par cinématique inverse à partir de la détection des points d'articulation (épaule, coude, poignet) dans l'image ou de décrire les expressions du visage par analyse des déformations. Enfin un modèle de langue permet d'exploiter la logique et les règles de la grammaire pour interpréter les gestes et expressions du signeur au plus haut niveau possible, en

s'affranchissant des fortes variations présentes dans les niveaux proches des pixels et en dépassant les problèmes d'occlusions locales. De plus ces modélisations doivent utiliser des formalismes compatibles afin de pouvoir propager les informations créées par le processus d'interprétation.

Nous avons donc instancié le modèle d'espace de signation en attachant aux entités des types fonctionnels. Ceci permet de modéliser une partie de la grammaire sous la forme d'un modèle de comportement, faisant le lien entre les entités situées dans la scène de signation et les gestes à réaliser pour les créer. Nous utilisons alors cette représentation pour piloter l'analyse de la vidéo.

### 3.2 Visualisation de l'espace de signation

Une première application de cette modélisation a été réalisée sous la forme d'un outil interactif (Figure 2) comportant un lecteur vidéo pilotable, une fenêtre de visualisation de l'espace de signation où chaque entité est affichée avec un widget correspondant à son type et une fenêtre de définition des séquences de création d'une entité. Ces trois fenêtres sont synchronisées, en création comme en lecture. Le logiciel peut traiter des situations de dialogue, ce qui permet d'étudier l'espace partagé par les deux interlocuteurs et les références communes.

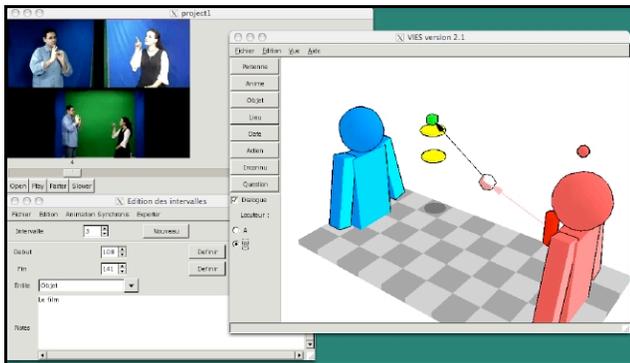


Figure 2 : Visualisation de l'espace de signation

Ce logiciel est actuellement utilisé pour analyser des corpus de LS (il sera intégré dans un outil d'annotation en partition), pour assister un processus de traduction ou pour expliquer comment la LS utilise l'espace, dans des formations à la LS.

### 3.3 Analyse des gestes du signeur

#### 3.3.1 Modèle de comportement

Pour pouvoir construire automatiquement l'espace de signation par analyse d'image, nous avons besoin d'un second modèle, qui représente les règles grammaticales de création des entités. La représentation illustrée figure 3 est une première implémentation de ce modèle.

Il prend en compte quatre composants corporels, les mains (dominante et dominée), le buste et le regard. Chaque action conduisant à la création d'une entité (ACTS) est décrite par un emplacement, un intervalle de temps et une séquence de gestes, chaque type de geste pouvant avoir différentes réalisations (signes localisés ou non, ...). Enfin les gestes sont décrits par les composants qu'ils impliquent et les contraintes sur les valeurs associées à ces composants. Des règles de propagation des informations assurent la cohérence de la création. Ce système est implémenté à l'aide du formalisme des logiques de description [2].

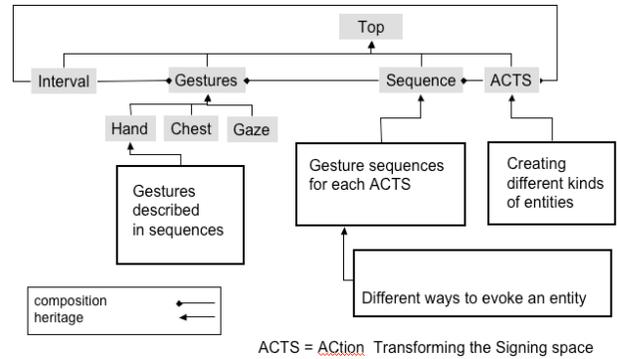


Figure 3 : Modèle de comportement pour créer des entités.

#### 3.3.2 Analyse d'image

Ce modèle est utilisé pour développer une démarche descendante dans la stratégie d'interprétation, complémentaire d'une analyse ascendante guidée par les données. Nous avons implémenté les algorithmes de détection de composants et de segmentation spatiale et temporelle : détection des pixels de couleur peau, détection du mouvement, détection et suivi des mains et du visage, détection d'évènements (contact, changement de configuration), détection de la pose du bras [20], analyse des déformations du visage [24]. Les algorithmes de suivi ont été particulièrement étudiés pour prendre en compte les fréquents phénomènes d'occlusion et les fortes variations de la dynamique (Figure 4).



Figure 4 : Suivi des déformations du visage. Suivi des mains et du visage

Du fait du bruit, des pertes dues à la projection 3D -> 2D, et des problèmes d'occlusion, au fur et à mesure que l'on combine ces opérateurs de traitement d'image (TI), il devient nécessaire d'ajouter de nombreux traitements de contrôle pour limiter les erreurs, ce qui rend ces opérateurs composés très complexes, lents et peu fiables. En exploitant les connaissances sur la grammaire de la LS et sur la réalisation des signes, on peut prédire les évènements visuels qui doivent être observés pour valider l'hypothèse de réalisation d'un signe créant une entité dans l'espace de signation. Dans ce cas, les indices visuels à détecter sont beaucoup plus simples et leur détection est moins coûteuse et beaucoup plus fiable.

Nous avons implémenté l'ensemble du modèle de comportement sur une application de type « requête à une base de renseignements » et montré l'intérêt de cette approche sur quelques opérateurs de détection. Nos recherches actuelles à l'IRIT ont pour but d'implémenter le module de prédiction des évènements visuels, de généraliser cette approche afin de la valider sur l'ensemble des opérateurs de TI et d'étudier l'articulation entre les approches ascendantes et descendantes.

## 4. SYNTHÈSE

Pour pouvoir générer automatiquement un énoncé en LS, il faut construire l'énoncé de manière à ce que son organisation respecte le fonctionnement linguistique de la LS, puis synthétiser les gestes qui vont constituer le lexique, et enfin afficher l'énoncé obtenu sous forme d'une animation graphique compréhensible en utilisant un signeur virtuel.

Nous décrivons dans cette section les travaux de recherche menés actuellement au LIMSI. Ils font suite aux premiers travaux menés sur la reconnaissance automatique [3, 5] et reprennent certains des modèles conçus dans ce cadre tout en les appliquant à la génération.

La section 4.1 présente notre approche pour mettre en œuvre les modèles décrits en section 2. La section 4.2 aborde nos pistes de recherche pour la représentation du lexique. La section 4.3 décrit enfin notre démarche pour la constitution d'une plate-forme logicielle d'animation de signeur virtuel.

### 4.1 Règles d'élaboration des énoncés

Après avoir implémenté la représentation de l'espace de signation, nous lui avons associé un ensemble de méthodes permettant de manipuler les entités et leurs relations afin de mettre à jour la scène de signation. Elles sont utilisées au sein d'un ensemble de règles permettant d'élaborer les énoncés.

Ces règles ne permettent pas encore de générer tout type d'énoncé. Un ensemble restreint de règles a été étudié. Il concerne l'utilisation des proformes pour exprimer des relations spatiales entre entités.

Les structures spatio-temporelles proposées partent de l'hypothèse que l'ordre naturel de réalisation est Localisant-Localisé [10].

Ces structures sont exprimées à l'aide d'un langage formel qui est en cours de conception. Il s'agit d'un langage permettant d'exprimer des propriétés temporelles qualitatives accompagnées de contraintes plus ou moins fortes sur les aspects spatiaux. Il est basé sur la logique d'intervalle de Allen [1] et enrichi de types de données et d'opérateurs permettant de manipuler les composantes des signes et les données spatiales.

Nous décrivons ici à titre d'exemple la structure d'élaboration d'une « structure proforme » puis celle d'une relation spatiale.

#### 4.1.1 Génération d'une « structure proforme ».

On appelle « structure proforme » la structure d'unité gestuelles permettant de spatialiser une entité dans l'espace de signation (Figure 5).

Ce type d'énoncé est composé de **signes standards** désignant des entités, notés  $[SIGNE]_{LSF}$ , de **proformes** permettant de spatialiser les entités en un lieu donné, notés  $PF(\text{signe}, \text{lieu})$ , du **regard**, noté  $REG(\text{lieu})$ , qui est utilisé pour « instancier » ou « réactiver » un emplacement dans l'espace de signation, en particulier juste avant d'y placer un proforme [10].

L'entité est tout d'abord réalisée, puis un bref regard est dirigé vers le futur emplacement du proforme, enfin le proforme en lui-même est réalisé à cet emplacement.

Si l'on représente la succession d'évènement correspondant, on observe des phénomènes se produisant en parallèle, que l'on peut représenter dans un système de type partition, où le temps s'écoule de gauche à droite (Figure 5).

- La première piste représente la direction du regard. Le regard est toujours dirigé vers un lieu (partie gris clair), sauf lorsque les yeux sont fermés. La partie gris foncé représente le moment où la direction du regard est contrainte vers un point P de l'espace de signation ( $REG(P)$ ).
- La deuxième piste représente le moment de réalisation du signe standard  $[S]_{LSF}$  représentant l'entité à spatialiser. Il est réalisé au début de la séquence.
- La troisième piste représente le proforme  $PF(S, P)$ , dont la configuration est choisie en fonction de la nature de l'entité S et dont l'emplacement est égal à P.

Les durées des différents événements sont basées sur des valeurs statistiques issues de l'analyse de corpus vidéo [26]. Notons que le regard peut se fixer sur le point P un peu avant la fin de S et se terminer un peu avant la fin du proforme. Il s'agit ici de contraintes souples.

Cette structure peut s'exprimer formellement de la manière suivante :

```
POINT P ;
INTERVAL_TEMP TREG, TSS, TPF ;
(TSS < TREG) v (TSS m TREG) v (TSS o TREG) ;
(TPF e TREG) v (TREG = TPF) v (TREG o TPF) ;
TREG.direction = Vect(yeux, P) ;
TPF.location = P ;
TPF.handshapeList = TSS.proformeList ;
```

Signification des relations de Allen :

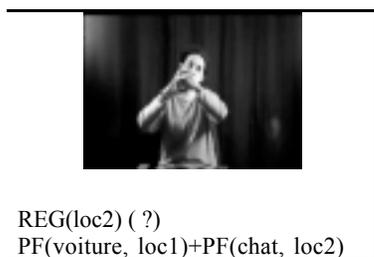
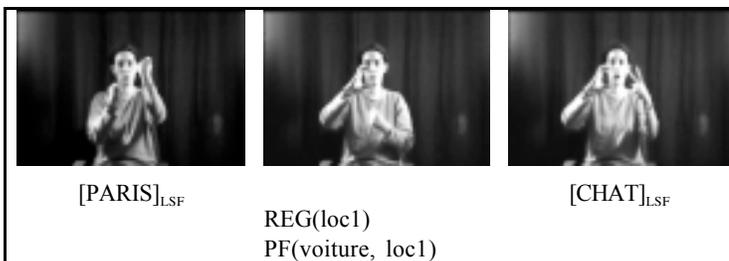
< : précède ; m : précède immédiatement ; o : recouvre partiellement ; e : recouvre totalement à la fin ; v : ou

Les deux premières lignes déclarent un point dans l'espace et trois intervalles temporels. Les deux lignes suivantes expriment les relations temporelles entre les trois intervalles à l'aide des relations de Allen. Les deux lignes suivantes expriment les contraintes spatiales au sein de ces intervalles. La dernière ligne associe au proforme la liste des proformes possibles pour l'entité représentée par S. On ne sélectionne un proforme dans cette liste que lorsque le contexte est suffisamment précisé.

#### 4.1.2 Génération d'une relation spatiale

Les structures spatio-temporelles proposées pour la génération d'une relation spatiale entre deux entités partent de l'hypothèse qu'une relation n'est exprimée que si les entités mises en jeu ont été signées au préalable [10]. L'exemple « Le chat est dans la voiture » est illustré Figure 6.

Nous avons proposé un formalisme basé sur une grammaire sémantico-cognitive afin de représenter les connaissances nécessaires à la génération de ce genre de relations [8]. Cette description formelle permet d'exprimer que l'entité chat est repérée dans un espace formé par l'entité voiture. Notons que, du fait que la voiture est considérée dans ce contexte comme étant un contenant, le proforme utilisé pour la spatialiser est la configuration C, qui exprime bien cette notion.



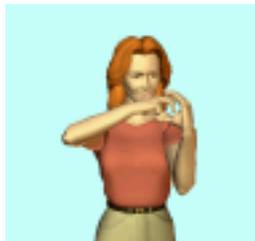
**Figure 6 : Séquence « [VOITURE]<sub>LSF</sub>, là, [CHAT]<sub>LSF</sub>, dedans »**

L'étape suivante consiste à représenter l'énoncé à générer. Il est décomposé en deux parties :

- La séquence « [VOITURE]<sub>LSF</sub>, là » correspond à une structure proforme telle que décrite précédemment (2 premières images de la séquence montrée figure 6).
- La séquence « [CHAT]<sub>LSF</sub>, dedans » est composée aussi d'un triplet (signe standard, regard, proformes), mais la partie « proformes » met en œuvre les deux mains et comporte des contraintes spatiales plus complexes (2 dernières images de la séquences montrée figure 6).

Pour exprimer ces contraintes, on va manipuler des « sous-espaces » et les mettre en relation avec les mêmes relations que pour les relations temporelles, afin de pouvoir exprimer des contraintes permettant de positionner automatiquement les deux proformes représentant la voiture et le chat l'un par rapport à l'autre (voir figure 7).

Ici, pour la relation spatiale DANS, il s'agit de positionner un des proformes « à l'intérieur » d'un autre. On définit deux sous-espaces correspondant aux deux configurations et on indique que l'intérieur de l'un « recouvre » l'extérieur de l'autre totalement.



**Figure 7 : Exemple de génération d'une relation spatiale.**

Cette contrainte s'exprime de la manière suivante :

```
POINT P ; //créé précédemment par le regard
SUB_ESP E1, E2 ;
E1=inside(PF1.handshape) ;
E2=outside(PF2.handshape) ;
E1 d E2 // d : recouvrement total
```

La manière dont s'implémente un recouvrement total d'intervalles spatiaux en animation passe par des procédures d'évitement de collision très complexes. La première version simplifiée réalisée est décrite dans [3]

## 4.2 Représentation du lexique

### 4.2.1 Les limitations des systèmes actuels

Pour pouvoir produire des énoncés de tout type en LS, il est nécessaire de reconsidérer aussi les modèles de représentation du lexique. En effet, la plupart des modèles proposés jusqu'alors décrivent ces lexiques comme des ensembles de signes isolés, chacun spécifiant un ensemble fixe de paramètres réputés suffisant à la définition d'un signe. Le plus souvent, ces paramètres sont la configuration des mains, leur emplacement, l'orientation des paumes, le mouvement et l'expression du visage. Ces unités gestuelles peuvent être transcrites formellement à l'aide de systèmes tels qu'HamNoSys [17] et sa représentation informatisée SigML [19], qui est le plus abouti des systèmes basés sur les paramètres.

Toutefois, ces modèles présentent des limitations ; par exemple les paramètres sont considérés comme étant tous nécessaires et suffisants, ce qui n'est pas toujours le cas. De plus, les dépendances qui existent entre ces paramètres ne sont pas exprimées.

Enfin, la spécification systématique de paramètres pose un problème de taille à l'intégration des lexèmes au niveau syntaxique. En LS, un lexème doit, pour assurer la continuité du discours, pouvoir se déformer selon des règles plus ou moins spécifiques. La juxtaposition de signes standard tels qu'ils peuvent être décrits dans le lexique ne suffit pas à élaborer un énoncé acceptable. Ces adaptations contextuelles sont subtiles et font l'objet de nombreuses études de la part des linguistes des LS [10, 22]. Malgré certaines tendances générales comme la marque de l'interrogation portée sur les sourcils, d'autres influences -qu'elles soient syntaxiques, sémantiques ou même pragmatiques- ne modifient pas tous les lexèmes de la même manière. C'est le cas lorsqu'on ajoute au discours une visée illustrative [11].

Autrement dit, s'il est des règles générales, l'écriture d'une règle universelle pour chaque apport discursif est, dans un système paramétrique, infaisable. Pour cause, les paramètres lexicaux à modifier ou à contraindre ne pourront dans bien des cas être repérés que de manière ad hoc ou par des mécanismes compliqués d'exceptions, en d'autres termes des règles non génériques.

### 4.2.2 Une représentation géométrique et séquentielle

Ainsi l'étude du lexique de la LSF nous a-t-elle conduits à revisiter cette notion trop figée de paramètre, pour permettre un modèle formel de spécification lexicale basé sur une représentation géométrique et séquentielle qui :

- rend compte de l'importance relative des segments corporels dans un signe en décrivant le concours de tous ceux (et uniquement ceux-ci) qui le composent ;
- fait apparaître les interdépendances paramétriques issues de la définition d'un signe ;
- donne aux lexèmes la souplesse nécessaire à leurs déformations en contexte.

Une étude statistique du lexique de la LSF [5] révèle indiscutablement que les objets géométriques sont récurrents dans la constitution des signes : la plupart des signes bimanuels incorporent une symétrie ou un parallélisme, un grand nombre de signes matérialisent un plan, une droite ou un cercle...

Le modèle considère l'unité gestuelle non plus comme un n-uplet de valeurs discrètes mais comme une réalisation géométrique dynamique dont chaque composant, à la manière d'une figure de géométrie spatiale classique, est construit de manière non ambiguë grâce à un langage formel [14]. Ainsi, pour spécifier un lexème, nous proposons de créer pas à pas les objets à partir desquels est ensuite décrit le comportement des parties du corps impliquées.

### 4.2.3 Exemple

A titre d'exemple, la description complète du signe [IMMEUBLE] (voir figure 8) est donnée ci-dessous. Cette représentation permet en contexte de préciser la taille de l'immeuble (Height), sa largeur (Size) et son emplacement dans l'espace de signation (Loc). Ces trois éléments sont des dépendances externes. La ligne L, le point M et le vecteur V sont des données géométriques à partir desquelles est construit le signe S, ainsi que le vecteur Up qui est défini a priori.

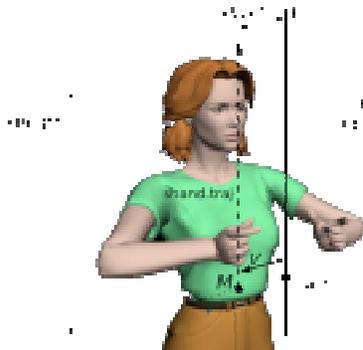


Figure 8: Signe standard [IMMEUBLE]<sub>LSF</sub>.

Les contraintes (de dépendances, géométriques) sont exprimées séquentiellement à l'aide d'opérateurs classiques exprimant le parallélisme ( $//$ ), la perpendicularité ( $\perp$ ), le produit vectoriel ( $\wedge$ )...

Le signe [IMMEUBLE] est décrit à l'aide de la configuration, l'orientation, l'emplacement initial et la trajectoire du mouvement de la main dominante (shand). La main dominée est décrite par symétrie.

Ce ne seront pas toujours les mêmes éléments mis en œuvre dans tous les signes.

```
SIGN S
LINE L
L // Up
L THRU {Loc}
POINT M
VECTOR V
V = Vect({Loc}, M)
V  $\perp$  L
|V| = {Size}
S.shand.config = "BSL C"
```

```
S.shand.ori.palm = -V
S.shand.ori.fingext = Up  $\wedge$  V
S.shand.traj.start = M
S.shand.traj.mvt = {Height}*Up
S.whand SYM S.shand WRT L
REGISTER S "building"
```

Ce modèle reste à valider sur un ensemble conséquent de signes et sera prochainement intégré au système d'animation du LIMSI.

## 4.3 Implémentation et évaluation

Nous développons actuellement une plate-forme logicielle d'animation de signeur virtuel qui va nous permettre d'évaluer les modèles et représentations présentés ici [3]. Nous procédons de manière itérative, afin d'assurer à chaque étape une évaluation sur les différents aspects modélisés.

Ainsi, à court terme, nous voulons évaluer la génération des lexèmes selon le modèle présenté ci-dessus. Nous allons dans un premier temps animer des énoncés simples et isolés de type « phrase à trous », où seuls les lexèmes correspondant aux trous devront être générés. Ces énoncés nécessiteront cependant une spatialisation des lexèmes à générer, afin d'évaluer la pertinence de la représentation pour cet aspect, comme dans « Le train pour *Toulouse* ne s'arrêtera pas à *Massy* », où les différents lieux (qui sont les trous des phrases) sont spatialisés. Les parties fixes sont conçues par un infographiste dont la langue première est la LSF, ce qui permet d'assurer une bonne qualité pour ces parties et de concentrer l'évaluation sur les parties calculées.

La deuxième phase consistera à animer la structure spatio-temporelle présentée au paragraphe 4.1. et à l'animer. Un premier prototype permettant de générer des relations géographiques a été expérimentée précédemment [7], il reste à connecter le moteur de génération à la plate-forme d'animation, pour laquelle il faut encore finaliser le module de gestion des collisions.

Les étapes suivantes permettront d'augmenter progressivement le degré d'automatisation du moteur de génération et les capacités techniques de la plate-forme d'animation

## 5. Conclusion et perspectives

Nous avons présenté une démarche visant à construire les modèles et les formalismes qui permettront de développer des applications en langue des signes. Ces modèles doivent rendre compte de ce qui fait réellement les spécificités de la langue des signes, l'utilisation de l'espace comme support des concepts du discours, l'iconicité comme principe structurant de la langue et l'utilisation simultanée de plusieurs composants corporels, tels que les épaules, le visage ou le regard et non seulement les gestes manuels. Nous avons proposé des modélisations de haut niveau, de l'espace de signation et des gestes qui le construisent, ainsi qu'une représentation du lexique. Ces modèles doivent pouvoir être partagés par les modules d'analyse et de synthèse, dont la synergie nous paraît très fructueuse. Bien que ces travaux soient encore en phase de développement, des premiers outils concrets commencent à être développés, qui apportent une nouvelle façon d'appréhender la langue des signes. Plusieurs projets régionaux (GAM-LS-Ile de France, SESCA-Midi-Pyrénées), nationaux (SignCom RIAM) ou européens (Dicta

Sign), qui débutent ou sont en cours d'expertise, combinent analyse et génération

Actuellement, le modèle d'espace de signation est commun aux deux approches, avec deux implémentations spécifiques. Ce couplage analyse-synthèse suppose encore des développements pour représenter les énoncés. Nous étudions la compatibilité avec les formalismes existants pour les langues orales; d'autre part, le programme de recherche (Ls-Script<sup>1</sup>) a étudié les bases d'un formalisme graphique de la LSF [15] qui pourrait jouer ce rôle.

Le développement d'applications utilisant la communication en LS doit pouvoir s'appuyer sur de tels modèles réalistes de la LS pour que ces applications soient réellement acceptées et utilisées par le public visé [9].

Les premières applications émergentes concernent des outils facilitant l'apprentissage de la LS ou son enseignement, l'utilisation en LS de dispositifs d'information (internet, bornes d'information) et la communication par signeur virtuel (annonces SNCF).

## 6. REFERENCES

- [1] Allen JF. Towards a general theory of action and time. In Allen J., Hendler J., Tate A. (eds.) *Readings in planning*, Kaufmann, San Mateo, 1990, 464-479
- [2] Baader F. et al, *The description logic handbook*, Eyrolles, 2003
- [3] Bolot L., Braffort A., Filhol M. Elsi balbutie ! Vers une plateforme d'animation d'avatar signant, WACA 2006, second workshop sur les Agents Conversationnels Animés, Toulouse, 26-27 octobre 2006
- [4] Braffort A. Reconnaissance et Compréhension de gestes, application à la langue des signes. thèse, Université Paris-XI Orsay, 1996
- [5] Braffort A. ARGo: An Architecture for Sign Language Recognition and Interpretation. In: *Proceedings of Gesture Workshop York GB. Mars 1996*, 17-30
- [6] Braffort A. Research on Computer Science and Sign Language: Ethical Aspects. LNCS n° 2298, *Gesture and Sign Language in Human-Computer Interaction: Gesture Workshop, GW 2001*, London, UK, April 18-20, 2001.
- [7] Braffort A., Bossard B., Segouat J., Bolot L., Lejeune F. Modélisation des relations spatiales en langue des signes française, Atelier LS, TALN, Dourdan, 6-10 juin 2005
- [8] Braffort A., Lejeune F. Spatialised semantic relations in French Sign Language: toward a computational modelling. In: *LNAI n°3881 Springer (2006): Gesture Workshop 2005*, Berder, France, May 18-20, 2005, 37-48
- [9] Braffort A, Dalle P., *Sign language applications : preliminary modeling*, UAIS, special issue « Emerging technologies for Deaf Accessibility in the Information Society, Stephanidis C.ed. à paraître 2008
- [10] Cuxac, C. *La langue des signes française – les voies de l'iconicité*, Ophrys. 2000
- [11] Cuxac, C. *Phonétique de la LSF : une formalisation problématique*, colloque "Linguistique de la LSF : recherches actuelles" Lille 23-24 sept. 2004, *Lexicales n° 4*, 2004, 93-114
- [12] Dalle P., Lenseigne B. Vision-based sign language processing using a predictive approach and linguistic knowledge. In: *IAPR conference on Machine Vision Applications – MVA Tsukuba, Japon.*, 2005, 510-513
- [13] Fasel B., Luetttin J. Automatic facial expression analysis : a survey, *Pattern Recognition* 36, 2003, 259-275
- [14] Filhol M., Braffort A. A sequential approach to lexical sign description, *LREC 2006 - Workshop on Sign Languages*, Gênes, Italie. 2006
- [15] Garcia B., Boutet D., Braffort A. Dalle P. Sign language in graphical form : methodology, modellisation and representations for gestural communication, in *Interacting Bodies (ISGS) Lyon, France juin 2005*
- [16] Gavrilin D.M. The visual analysis of human movement : a survey, *Computer Vision and Image Understanding*, vol. 73 no 1, 1999. 82-98
- [17] Hanke T. *HamNoSys - an introductory guide*. Signum Press, Hamburg. 1989
- [18] Huenerfauth, M. Spatial representation of classifier predicates for machine translation into american sign language. In *Workshop on Representation and Processing of Sign Language 2004*, Lisbon Portugal 2004 24-31
- [19] Kennaway R. Experience with and Requirements for a Gesture Description Language for Synthetic Animation. In: *Lecture Notes in Computer Science n° 2915 (2004) GestureWorkshop 2003* Genova, Italy, 2003, 300-311
- [20] Lenseigne B., Gianni F., Dalle P. A new gesture representation for sign language analysis, in *LREC 2004*, Lisbon Portugal, 2004, 85-90
- [21] Lenseigne B., Dalle P. Using signing space as a representation for sign language processing, in *LNAI n° 3881. Springer, Berlin, Gesture Workshop, Berder France*, 2005, May 18-20, 25-36
- [22] Liddell S. *Grammar, gesture and meaning in American Sign Language*. Cambridge University Press, 2003
- [23] Marshall, I. and Safar, E., Sign Language Generation in an ALE HPSG, In: *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar Center for Computational Linguistics*, Edited by Muller, S., Katholieke Universiteit Leuven, 189-201, 2004.
- [24] Mercier H, Peyras J., Dalle P. Toward an efficient and accurate AAM fitting on appearance varying faces, in *7th Int. Conf. on Automatic Face and Gesture Recognition*, Southampton, UK, 2005, 363-368
- [25] Ong S., Ranganath S. Automatic sign language analysis : a survey and the future beyond lexical meaning. *IEEE Trans on PAMI*, vol 2 no 6 2005, 873-891
- [26] Sallandre M.A., *Les unités du discours en Langue des Signes Française. Tentative de catégorisation dans le cadre d'une grammaire de l'iconicité*, thèse, université Paris 8, décembre 2003

<sup>1</sup> <http://lsscript.limsi.fr>