



**HAL**  
open science

## Modélisation linguistique

Michael Filhol, Annelies Braffort

► **To cite this version:**

Michael Filhol, Annelies Braffort. Modélisation linguistique. La langue des signes française (LSF) : modélisations, ressources et applications, ISTE Editions, 2016, Collection Sciences Cognitives, 9781784050504. hal-01849174

**HAL Id: hal-01849174**

**<https://hal.science/hal-01849174>**

Submitted on 2 Oct 2023

**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.

# Chapitre 3

## Modélisation Linguistique

Michael Filhol et Annelies Braffort

### 3.1 Introduction

La recherche en langue des signes en est encore à ses balbutiements (voir chapitre 1), et d'autant plus dans le domaine de l'informatique, dans lequel le *traitement automatique des langues des signes (TALS)* n'est apparu que très récemment. À l'heure actuelle, trois axes principaux sont étudiés : la reconnaissance, la production, et la traduction automatiques.

Le premier axe, la *reconnaissance de LS*, a émergé il y a une vingtaine d'années. Il implique l'utilisation de dispositifs de capture tels que les caméras ou les systèmes de capture de mouvement, ainsi que des méthodes de traitement du signal pour extraire les données pertinentes. En France, la première étude a débuté en 1993 (Braffort, 1996a).

La *génération de LS*, parfois appelée *synthèse de LS*, a débuté quelques années plus tard, suite à l'augmentation des capacités de calcul des ordinateurs. Elle implique l'utilisation d'un modèle de personnage virtuel destiné à être animé, nommé *signeur virtuel* (voir section 3.3.1), et des techniques de programmation en infographie graphique pour produire une vidéo contenant l'animation (voir section 3.3.2). L'axe de recherche la plus récente est la *traduction LS-Texte ou texte-LS*, pour lequel les entrées et/ou sorties des systèmes de reconnaissance ou de génération peuvent être utilisés. Une problématique encore rarement explorée est celle des *applications pédagogiques*, un type d'applications qui mériterait pourtant d'être développé (voir chapitre 4).

Toutes ces études doivent intégrer des connaissances linguistiques, à mettre en œuvre dans des modèles informatiques nommés ici *modèles linguistiques*, complétés par d'autres liés aux tâches de traitement d'images et d'animation. En raison de l'absence de corpus au début des recherches en linguistique des LS (voir chapitre 2), les modèles linguistiques ont souvent été construits à partir des connaissances intuitives des chercheurs complétées par des entretiens avec des

locuteurs natifs de LS. Avec le développement des corpus de LS, les méthodes ont évolué vers des études fondées sur des corpus et permettant aux chercheurs de créer des modèles représentant des connaissances de nature plus statistiques.

Dans ce chapitre, nous donnons un aperçu de la modélisation linguistique des LS. La section suivante 3.2 décrit le type de données à la disposition des chercheurs de TALS et comment les plus récentes peuvent être utilisés pour des études en linguistique ainsi qu'en informatique. Nous introduisons ensuite les signeurs virtuels et leur animation (section 3.3). La section 3.4 présente les grandes tendances et les approches nouvelles explorées pour la modélisation linguistique des LS. La section 3.5 présente une étude de cas sur l'axe de recherche le plus récent : la traduction automatique. La dernière section (section 3.6) conclut sur les limites actuelles et de leurs raisons possibles, et propose quelques considérations pour l'avenir.

## 3.2 Les données en Langue des Signes

Le but de cette section est de présenter les types de données disponibles pour les chercheurs en TALS et des exemples type d'études qui les utilisent. Nous les présentons selon trois degrés d'abstraction : les données brutes (section 3.2.1), un degré d'abstraction intermédiaire (section 3.2.2) et les représentations abstraites (section 3.2.3).

### 3.2.1 Données brutes

L'expression *données brutes* se réfère aux données qui n'ont pas été soumises à un traitement ou autre manipulation. Ces données, collectées à partir de dispositifs d'entrée variés (caméras par exemple), peuvent être plus ou moins proches des mouvements d'origine. Elles peuvent être utilisées aussi bien pour les études en linguistique qu'en informatique.

#### 3.2.1.1 La vidéo

Le type le plus courant de données brutes est la vidéo. Les plus grands corpus construits pour la recherche linguistique ou l'archivage patrimonial utilisent cette technologie, qui est facile à mettre en place.

Les caméras encodent des données visuelles dans des matrices de pixels de couleur. Un traitement est nécessaire pour distinguer, par exemple, les pixels qui représentent la silhouette humaine de ceux qui composent l'arrière-plan non pertinent d'un point de vue linguistique. En outre, l'ensemble de pixels pertinents ne couvre pas toute l'information véhiculée par le mouvement du corps, parce que la profondeur n'est généralement pas capturée. Le cas échéant, un traitement doit être effectué pour reconstituer la troisième dimension.

Dans le domaine du TALS, les vidéos sont utilisées pour entraîner les systèmes de reconnaissance ou de traduction LS-texte (Cooper *et al.*, 2011).

La vidéo est également utilisée dans des outils pédagogiques, ou pour construire des ressources telles que les lexiques de LS (voir chapitre 4). La figure 3.1 montre

une capture d'écran du portail de lexique multilingue intégré dans le projet européen DictaSign. Il contient une liste de plus de mille entrées exprimées en langues des signes britannique, allemande, française et grecque<sup>1</sup>.

Comme décrit au chapitre 2, la plupart de ces données se présentent sous la forme de vidéo numérique, mais les systèmes de capture récents, tels que la capture de mouvement ou la Kinect®<sup>1</sup>, sont de plus en plus utilisés.

### 3.2.1.2 La capture de mouvement

Les systèmes de *capture de mouvement*, ou *mocap*, sont utilisés pour collecter des données de LS à la fois pour l'analyse linguistique (chapitre 2 section 2.2.3) et pour le TALS. Cette technologie est plus complexe à mettre en place que la vidéo, plus contraignante, généralement intrusive et reste très coûteuse. À l'heure actuelle, peu de corpus sont créés en utilisant la mocap, même si cette technique possède l'avantage de fournir des données 3D avec une meilleure résolution et moins de traitement que ne le nécessite la technologie vidéo.

Il existe différentes technologies de mocap. L'une d'elle consiste à équiper le signeur de marqueurs réfléchissants positionnés près des articulations, et un ensemble de caméras infra-rouge capture leurs positions et suit leurs mouvements au fil du temps. Les caméras sont placées autour du signeur et fournissent les coordonnées 2D de chaque marqueur. Les logiciels qui gèrent ces systèmes permettent de calculer leurs coordonnées 3D respectives. Un post-traitement est nécessaire pour nettoyer et compléter les données (certains marqueurs peuvent être occultés à certains moments de la captation). On peut ajouter à ces systèmes des dispositifs fixes ou embarqués pour capturer la direction du regard.

Dans les premières études (Braffort, 1996a; Sagawa & Takeuchi, 1999; Tyrone, 2002), les données mocap de LS ont été principalement utilisées comme entrée pour les systèmes de reconnaissance, ainsi que de rares études sur la phonologie des LS (la plupart du temps en utilisant des gants numériques, un dispositif d'entrée qui capture la flexion des doigts). Son utilisation pour l'animation des LS est plus récente et comprend la capture des mouvements de l'ensemble du corps. La capture des mouvements des mains et des éléments du visage reste plus complexe et nécessite l'utilisation d'un ensemble de caméras haute résolution très coûteuses rarement disponibles dans les laboratoires de recherche.

Bien que ces données brutes jouent un rôle important dans les applications où des productions réalistes sont nécessaires (voir la section 3.3.2), des représentations plus abstraites peuvent être plus adaptées pour la synthèse en temps réel.

## 3.2.2 Degré intermédiaire d'abstraction

Dans cette section, nous considérons des représentations qui décrivent les données d'une manière plus abstraite, en réduisant les détails pour ne garder

1. <http://www.sign-lang.uni-hamburg.de/dicta-sign/portal/>

[Back to overview](#)
[Zurück zur Übersicht](#)
[Retour au sommaire](#)
[Πίσω στην επίσημη](#)



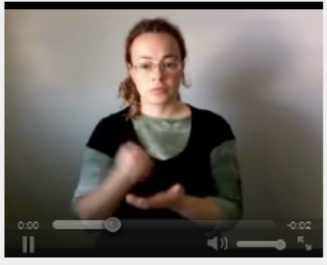

<p>information <span>BSL</span></p>  <p>information(n)#2</p> <p>ḡ ɔ̄ [r 0 ɹ ɔ] [2 0 ɹ 2] [X ɔ] (ɔ̄ ɔ̄ ɔ̄ ɔ̄)</p>	<p>Information <span>DGS</span></p>  <p>INFORMATION2B</p> <p>ḡ ɔ̄ [r 0 ɹ ɔ] [X ɔ] (ɔ̄ ɔ̄ ɔ̄ ɔ̄)</p>
<p>information <span>LSF</span></p>  <p>INFORMATION</p> <p>[ɔ̄ ɔ̄] [ɔ̄ ɔ̄] [ɔ̄ ɔ̄] [ɔ̄ ɔ̄] (ɔ̄ ɔ̄) [X ɔ] [ɔ̄ ɔ̄]</p>	<p>ΠΛΗΡΟΦΟΡΩ <span>GSL</span></p>  <p>ΠΛΗΡΟΦΟΡΩ</p> <p>ɔ̄ ɔ̄ ɔ̄ ɔ̄ ɔ̄ ɔ̄ ɔ̄ ɔ̄ (X ɔ̄) [ɔ̄ ɔ̄ ɔ̄ ɔ̄]</p>

FIGURE 3.1 – Le lexique DictaSign comportant plus de 1000 entrées, exemple : le concept INFORMATION.

que les caractéristiques pertinentes en fonction de l'objectif visé, à savoir les images et les animations « rotoscopée ».

### 3.2.2.1 Les images

La représentation la plus ancienne est l'*image*. Cette représentation a été utilisée très tôt pour les entrées de dictionnaire (Bébian, 1827), à savoir des signes en forme de citation<sup>2</sup> (voir le chapitre 1, section 1.4.2.1). Ce type de représentation est composé au mieux de la posture initiale de l'ensemble du corps, de la posture finale des mains et d'une indication du mouvement entre les deux, généralement sous forme de flèches de différentes formes ajoutées sur le dessin. Ces compositions peuvent être créées semi-automatiquement avec des logiciels récents tel que PhotoSigne<sup>3</sup> (voir chapitre 4, section 4.4.1.4). Ce logiciel peut aider à décrire la trajectoire des mains entre leur postures initiale et finale (voir figure 3.2). Ce type de représentation peut être utile pour l'ingénierie des LS, lorsque une représentation statique est suffisante. Cette représentation est utilisée en France dans des applications pédagogiques<sup>4</sup>.

### 3.2.2.2 La rotoscopie

Le procédé appelé *rotoscopie* est à l'origine une technique cinématographique qui consistait à projeter les images d'un film sur un verre dépoli et qui étaient ensuite redessinées par un animateur, en utilisant un dispositif appelé rotoscope (technique utilisée par exemple par les studios Disney® pour le film *Blanche Neige* en 1937).

Maintenant, les rotoscopes sont remplacés par des logiciels spécialisés et le processus peut servir à créer directement des animations à partir de vidéos de LS. Il consiste à sélectionner des postures "clés" à partir d'une source vidéo d'un véritable signeur, par copie de la posture du signeur virtuel sur celle du signeur réel, de manière à ce que la posture résultante virtuelle soit aussi proche que possible de celle de la vidéo de référence. En général, l'infographiste sélectionne un certain nombre de postures à recopier, le logiciel reconstruisant les mouvements complets par interpolation entre ces postures. Du fait de cette sélection des postures pertinentes, la représentation de l'animation résultante peut être considérée comme ayant un certain degré d'abstraction par rapport à de la capture de mouvement. Afin d'obtenir une sortie représentative des contenus linguistiques de la vidéo, ce qui est loin d'être une simple duplication image par image, l'infographiste doit non seulement maîtriser les techniques d'animation mais aussi la LS.

Le principal avantage d'une vidéo rotoscopée est sa haute qualité, c'est-à-dire l'impression d'une expression naturelle, mais aussi la production d'un message linguistiquement correct. L'inconvénient est qu'il s'agit d'un processus coûteux

---

2. Cela signifie qu'une seule instance du signe est décrite, non pas toutes les variations possibles.

3. <http://www.irit.fr/presto/resultats.html>

4. <http://www.irit.fr/GlossaireDD-LSF/>

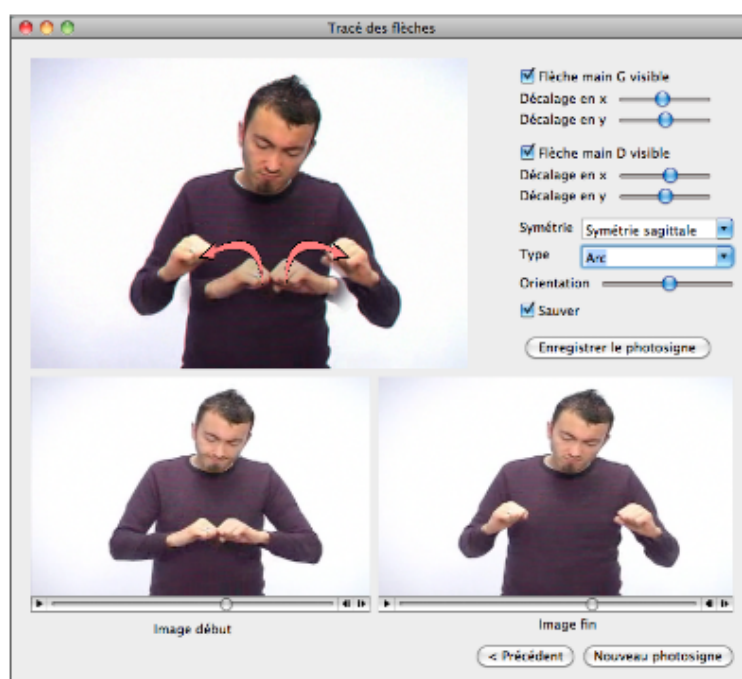


FIGURE 3.2 – Exemple d’image produite par le logiciel PhotoSigne.

en temps et qu'il nécessite une expertise peu courante. Les experts de Web-sourd ont comparé les deux procédés, mocap et rotoscopie, pour des animations « courtes » (< 1 min) et « longues » (> 3 min). Ils ont constaté que la rotoscopie est plus rapide que la mocap pour les clips courts, tandis que de longues vidéos ont tendance à être créées plus rapidement avec la mocap. Les animations 3D peuvent également être créées directement, sans l'aide d'une vidéo comme modèle (Lombardo *et al.*, 2010; Wolfe *et al.*, 2011). Ce processus peut être moins coûteux en temps que la rotoscopie, mais la qualité de l'animation est plus difficile à garantir. La rotoscopie est actuellement utilisée en France dans des applications à destination du grand public (voir chapitre 5, section 5.3.3).

Les données brutes et intermédiaires peuvent être directement utilisées par les systèmes de reconnaissance, les dictionnaires en ligne ou les animations basées sur les données par exemple. D'autres, plus abstraites, ne peuvent pas être utilisées seules et sont associées à des outils de traitement appropriés. Ils sont décrits dans la section suivante.

### 3.2.3 Représentations abstraites

Nous considérons ici des formalismes basés sur des représentations graphiques et des formalismes purement abstraits.

#### 3.2.3.1 Les représentations graphiques

Deux représentations proposent une description abstraite des mouvements des articulateurs : SignWriting et HamNoSys (voir chapitre 1, section 1.4.1.3). Ils sont tous deux basés sur le modèle paramétrique classique, où un signe est décrit avec un ensemble plus ou moins limité de valeurs pour chaque paramètre.

- *SignWriting* : Il a été conçu dans le but d'offrir un moyen d'écrire en LS. Il est composé d'un ensemble de symboles iconiques, combinés au sein d'une représentation en 2D. Dans ce système, la position des éléments du corps est déduite de la position relative des symboles associés à ces éléments dans la représentation. Ceci est la représentation la plus iconique qui existe à l'heure actuelle.
- *HamNoSys* : Il a été conçu comme un système de représentation compositionnelle pour le lexique des LS, permettant aux lexicographistes de créer des dictionnaires et aux linguistes d'annoter les corpus de LS. Il est composé d'un ensemble de symboles combinés de manière linéaire. La description est plus linéaire, donc moins iconique, mais plus facilement utilisable dans les logiciels tels que les traitements de texte, en tant que police de caractère Unicode.

En plus de leur objectif initial (respectivement l'écriture et la lexicographie), ces deux représentations ont un équivalent informatique, respectivement SWML et SigML, et sont utilisés dans divers systèmes de TALS, par exemple des logiciels d'édition (figure 3.3) (Bianchini *et al.*, 2012), de sortie d'un système de reconnaissance, ou d'entrée pour un système de synthèse (Elliott *et al.*, 2004,



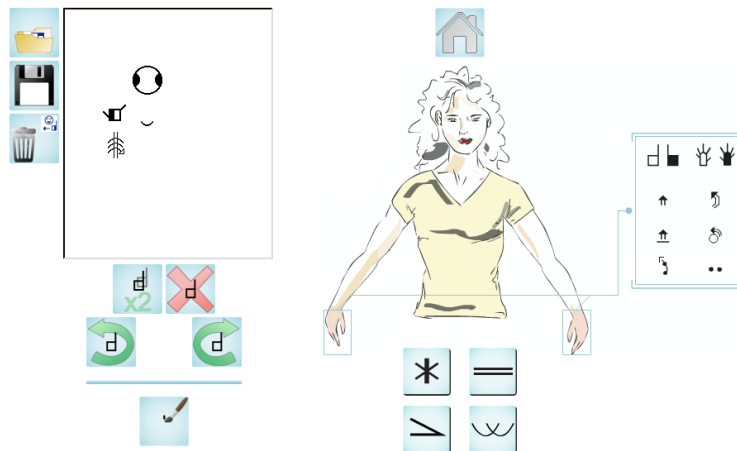


FIGURE 3.3 – Le logiciel SWift pour le système SignWriting.

2011; Bouzid *et al.*, 2012). Par ailleurs, une étude a été menée sur les différents méthodes pour encoder SignWriting en Unicode (Aznar, 2008).

L'utilisation des représentations abstraites a des avantages. Il est possible de reconfigurer ou éditer l'entrée de l'animation pour recalculer la sortie et de construire de nouveaux énoncés. De plus, cela permet de produire des énoncés dans un style assez neutre, en d'autres termes des énoncés anonymes, empêchant l'identification du locuteur, ce qui est généralement impossible avec les approches basées sur les données, même après un processus de rotoscopie.

Toutefois, des inconvénients existent aussi. À ce jour, ces systèmes sont limités pour ce qui est des éléments non manuels, en particulier le visage. Cela rend le message difficilement, voire pas du tout, compréhensibles, car tous les éléments linguistiques ne sont pas présents. En outre, les mouvements générés par ces systèmes sont généralement simplifiés ou purement géométriques et n'incluent pas les contraintes biomécaniques, donc produisent des mouvements irréalistes (Delorme, 2011; Delorme *et al.*, 2012). Enfin, aucune étude approfondie ou de modèle complet n'existe à ce jour sur la coarticulation en LS (Segouat & Braffort, 2010; Segouat, 2010).

Nous notons que ces modèles ont été utilisés, et même conçus en premier lieu, pour décrire les paramètres manuels des unités lexicales des LS. Ces signes sont ceux établis dans le cadre du lexique standard non-illustratif (voir chapitre 1, section 1.4.2) et servant d'entrées de dictionnaire de base, généralement glosé avec les mots d'une langue différente dont la signification est supposée équivalente. Ils sont basés sur les paramètres traditionnels, décrivant ainsi une réalisation donnée d'un signe (la forme de citation ou l'une des variantes possibles) (voir le chapitre 1, section 1.4.2.1).

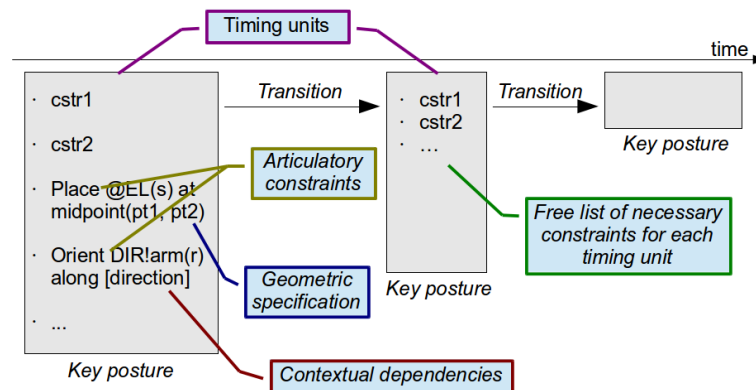


FIGURE 3.4 – Zebedescription

### 3.2.3.2 Zebedee

Il y a quelques années, un modèle alternatif appelé *Zebedee* a été proposé (Filhol, 2008, 2009). Il décrit les signes comme une séquence d’unités temporelles, dont chacune précise un ensemble de contraintes nécessaires et suffisantes à appliquer au squelette virtuel. L’espace de signation est considéré comme un espace géométrique euclidien où des objets géométriques auxiliaires peuvent être construits. Les dépendances entre les éléments des descriptions, voire aux éléments contextuels, ne sont pas seulement possibles, mais également rendus pertinentes. Par conséquent, les descriptions rendent mieux compte de la nature iconique des signes. Les « zebedescriptions » sont basées sur des aspects articulatoires, sémantiques et cognitifs, pas seulement sur l’observation d’un seul énoncé hors contexte, qui rend les modèles difficilement réutilisables ou généralisables. Ainsi, une zebedescription permet au chercheur de décrire un signe et toutes ses variantes possibles avec une unique description. Cette représentation a été initialement conçu pour la synthèse de LS (Delorme, 2010), mais est désormais également utilisé pour faciliter le traitement d’image (Gonzalez *et al.*, 2012).

Ainsi, toutes les représentations abstraites peuvent être utilisées pour générer des animations. Nous explicitons ce processus dans la section suivante (3.3).

## 3.3 Signeur virtuel et animation

Une des utilisations en TALS des données en LS est l’animation de signeurs virtuels. Nous présentons ici les grands principes sur lesquels ceux-ci sont élaborés (section 3.3.1) et animés (section 3.3.2).

### 3.3.1 Signeur virtuel ou avatar signant

Un *signeur virtuel*, ou *avatar signant*, est un personnage en 3D qui est animé pour produire de la LS. Il est modélisé en trois parties (figure 3.5) :

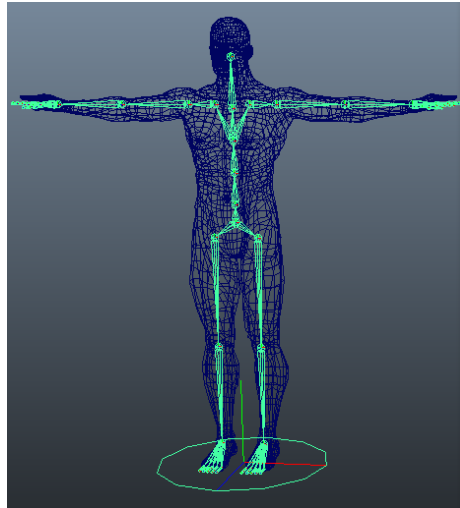


FIGURE 3.5 – Squelette et maillage d’un signeur virtuel (ici, plugin Maya HumanIK)

- la *peau* : un maillage 3D déformable pour représenter la surface du personnage ;
- le *squelette* : un ensemble hiérarchisé de segments interconnectés pour animer la peau attachée à lui.
- la *texture* : les propriétés visuelles du personnage (par exemple la couleur, la luminosité) appliquées sur la peau.

L’animation du corps du personnage est effectuée en spécifiant la posture du squelette (sous la forme d’une liste d’angles de rotation à appliquer aux articulations par exemple) à chaque instant. L’animation des expressions faciales peut être contrôlée soit à l’aide d’un squelette dédié (qui permettra de déformer indirectement la peau), soit avec un ensemble de maillages du visage prédéfinis appelés *morphs* ou *shapes*. En pratique, les approches sont souvent mixtes et utilisent des modèles plus sophistiqués (Noh & Neuman, 1998). L’apparence du signeur virtuel est rendue (calculée) pour chaque image, qui ensemble forment la vidéo complète.

### 3.3.2 Animation de signeur virtuel

#### 3.3.2.1 Les méthodes d’animation

Il existe principalement trois méthodes pour l’animation de signeurs virtuels :

- Le *rejeu de capture de mouvement* (*motion capture replay*, ou *mocap replay*) réutilise l’enregistrement mocap des mouvements, des expressions faciales et du regard d’un signeur pour animer un personnage virtuel (voir la figure 3.6). Les données de mocap peuvent également être utilisées pour constituer des corpus pour les études de linguistique (Malaia



FIGURE 3.6 – Animations en LSF rejouées à partir de capture de mouvement (projet SignCom (Gibet *et al.* , 2011))

*et al.* , 2013), permettant une analyse plus fine des caractéristiques articulatoires dynamiques des productions notamment.

- La *rotoscopie*, présentée précédemment.
- L'*animation à base de règles* consiste à calculer automatiquement les animations à partir des représentations abstraites de l'énoncé à générer (voir la section 3.2.3).

### 3.3.2.2 Avantages et inconvénients

Les méthodes copiant la réalité en utilisant la mocap ou la rotoscopie, aussi appelées *animations basées sur les données* (data-driven animation), permettent de reproduire le message d'origine avec une grande fidélité, à des coûts généralement élevés selon la technologie utilisée et la quantité de soin apportée à la tâche de rotoscopie ou de post-traitement de la mocap (voir chapitre 5).

L'animation à base de règles permet de générer une variété d'énoncés à partir de spécifications génériques à faible coût, mais les mouvements générés sont encore trop robotiques et incomplets.

La méthode à choisir dépend de l'application visée et du budget disponible. Depuis peu (Segouat *et al.* , 2008; Lu & Huenerfauth, 2010; Gibet *et al.* , 2011), des bases de données d'animation de LS sont construites à l'aide de l'approche basée sur les données pour stocker des séquences d'animations qui peuvent être annotées, lues, traitées et combinées dans de nouvelles animations. Dans le même temps, des méthodes sont explorées pour ajouter plus de naturel aux animations générées à partir de règles (Delorme, 2010, 2011; McDonald *et al.* , 2014). Il est important de noter que la fidélité au message d'origine n'est pas toujours une garantie de qualité et que des ajustements artistiques ou linguistiques peuvent rendre un message plus clair. Pour cette raison, il peut être utile de combiner les approches, par exemple le rejeu de séquences de capture de mouvement modifiées à partir de règles (Gibet *et al.* , 2011).

## 3.4 Modèles pour la langue

Cette section présente les différentes approches explorées pour modéliser les LS. Les deux premières sections présentent les points de vue existants dans le monde plus étudié des langues écrites : les grammaires à base de règles (section 3.4.1) et l'apprentissage machine (section 3.4.2). Une troisième partie présente des approches spécifiques aux LS, c'est-à-dire ne partant pas de modèles établis pour l'écrit ou le vocal (section 3.4.3). Dans la dernière section, nous dressons un inventaire des questions ouvertes, tant sur les représentations que sur l'évaluation des modèles (section 3.4.4).

### 3.4.1 À l'origine : les grammaires génératives et catégorielles

Après une décennie de traitement de la langue comme application de la *théorie de l'information*, Bar-Hillel introduit une nouvelle approche des langues écrites en proposant les *grammaires catégorielles* Bar-Hillel (1950). Celle-ci peut être considérée comme le premier pont entre l'informatique et la linguistique, donc comme le début du *traitement automatique de la langue* (TAL) tel que nous l'envisageons aujourd'hui.

#### 3.4.1.1 En TAL

L'objet principal des grammaires catégorielles est la structure de la phrase, analysée en constituants. Ces constituants sont supposés typés et leur type et leur ordre donne un type à la phrase, constituant de plus haut niveau. Les grammaires de constituants, puis la célèbre *grammaire générative* Chomsky (1965), reposent au fond sur deux éléments :

- un ensemble de règles de production récursives, définissant donc un langage infini ;
- un lexique d'éléments terminaux catégorisés (typés).

Les règles contraignent les types des constituants, et permettent ensemble de générer une infinité de séquences d'éléments terminaux (mais pas toutes!). Cela fait émerger la notion de grammaticalité de séquences lexicales, à savoir que l'on peut juger de l'acceptabilité de toute suite de mots donnée.

Par exemple, soit le système simple de règles suivant :

$$\left\{ \begin{array}{l} S \rightarrow NP VP \\ \\ NP \rightarrow PN \\ NP \rightarrow Det N \\ NP \rightarrow Det Adj N \\ \\ VP \rightarrow V \\ VP \rightarrow V Adv \\ VP \rightarrow V NP \\ VP \rightarrow VP Prep NP \end{array} \right.$$

et le vocabulaire d'éléments terminaux :

<i>PN</i> :	Joe, Jupiter, Linux...
<i>N</i> :	enfant, bonhomme, voiture, chocolat, carré...
<i>Det</i> :	le, un, une, ton...
<i>V</i> :	conduisit, vendent, termine, s'approche...
<i>Adj</i> :	ponctuel, content, carré...
<i>Adv</i> :	manuellement, lentement...
<i>Prep</i> :	d', de, avec...

Partant de l'*axiome*  $S$  et en réécrivant successivement les symboles de droite, un algorithme génère aisément les phrases  $S_1$  ou  $S_2$  ci-dessous. Au contraire, aucune suite de dérivation ne peut engendrer  $S_3$ , d'où le fait que l'on juge cette phrase comme étant grammaticalement inacceptable.

- $S_1$  : « un enfant s'approche d'une voiture »
- $S_2$  : « Joe termine ton chocolat »
- \* $S_3$  : « le conduisit Linux content »

En revanche, s'il était basé uniquement sur les catégories des éléments terminaux, cet algorithme générerait  $S_4$ .

- \* $S_4$  : « Joe vendent ton chocolat »

Pour obtenir des grammaires plus rigoureuses, des contraintes d'unification entre constituants peuvent être introduites. Le premier système du genre est LFG (lexical-functional grammar) Kaplan & Bresnan (1982), une autre est HPSG (head-driven phrase structure grammar) Pollard & Sag (1994). Le principe repose sur des traits portés par les éléments lexicaux, pouvant être contraints de s'unifier pour être utilisés conjointement. Par exemple,  $S_4$  pourrait être exclue grâce à un trait binaire *nombre* à valeur singulier ou pluriel sur les nœuds *PN* et *V*, et les valeurs distinctes données aux unités *Joe* et *vendent*.

Tout ce travail suppose la définition d'un vocabulaire d'éléments terminaux, représentant le lexique du langage modélisé, et de l'ensemble de règles de réécriture représentant sa syntaxe. Cette organisation en couches ordonnées d'abstraction de la langue (phonétique, phonologie, lexique, morphologie, syntaxe, sémantique, pragmatique) constitue l'approche traditionnelle de tout le traitement automatique des langues.

### 3.4.1.2 En TALS

Les grammaires génératives couvrent une bonne partie du fonctionnement syntaxique des langues écrites, et quand les premiers projets liés à la synthèse de LS ont démarré, il s'agissait du modèle le plus avancé, et le plus répandu. Aussi, le courant générativiste était largement porté par l'idée que la majeure partie de la *faculté de langage* était innée, et que seuls quelques paramètres spécifiques à la langue devaient être appris pour acquérir une première langue par exemple. Sans qu'il soit certain que les chercheurs s'étant tournés vers les

LS aient vraiment adhéré à cette idée, ces raisons ont sans doute contribué à ce qu'ils se tournent vers la grammaire générative en premier.

Nous l'avons vu, les grammaires en constituants reposent sur un lexique (et en général une certaine compositionnalité de ses éléments) et des règles syntaxiques (et éventuellement des contraintes d'unification). Ainsi, un prérequis pour les utiliser pour la LS est de trouver les équivalents dans la LS observée. Invariablement, cette approche a utilisé un lexique de signes en forme de citation<sup>5</sup>, et les paramètres de Stokoe (chapitre 1, section 1.3.1) pour ce qui est de leur composition interne. Dans une certaine mesure, ces formes de citation permettent les contraintes d'unification. Par exemple, on peut forcer l'unification du paramètre d'emplacement d'un signe dit nominal avec celui d'un autre dit verbal. Alors, de même que l'accord en nombre dans l'exemple plus haut, cette unification des emplacements pourrait jouer le rôle d'accord avec un *verbe directionnel* en LS (voir chapitre 1, section 1.4.3.1).

Cela a d'abord été illustré au sein du projet Visicast (Marshall & Safar, 2004), qui proposait un processus complet de passage de texte à LS. D'autres projets ont emprunté le même chemin, pour la synthèse ou la traduction de textes (San-Segundo *et al.*, 2012), parfois combinés avec les approches basées données Lombardo *et al.* (2010), d'autres encore pour la reconnaissance ou la traduction vers le texte (Wu *et al.*, 2007).

Nous appelons ces modèles *linéaires*, au sens où ils considèrent tous les productions signées comme des séquences d'unités gestuelles, considérées comme lexicales et comparables aux suites de mots composant les phrases. Tous les composants de cette production sont alors repérés à partir des frontières lexicales ainsi posées, qu'ils soient de nature sub- ou supra-lexicale.

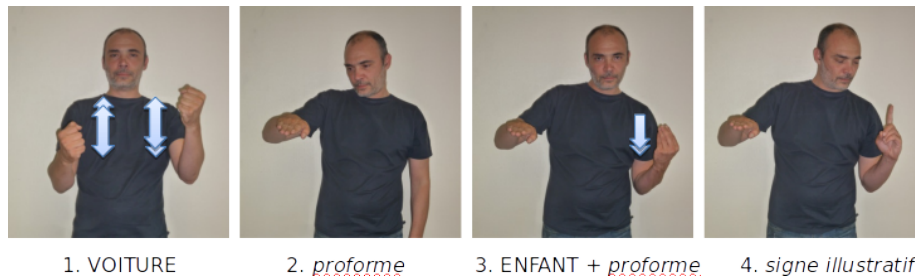
Cela dit, comme expliqué dans le chapitre 1, les LS comportent de nombreuses spécificités où entre en jeu bien plus qu'une simple séquence de signes-mots avec unification de paramètres. Par exemple, la phrase  $S_1$  plus haut ne serait probablement pas traduite en LS par une séquence de signes avec un simple accord paramétrique entre eux. En LSF, nous verrions plutôt quelque chose comme :

— une séquence de quatre unités manuelles comme sur la figure 3.7 :

1. le signe bimanuel VOITURE ;
2. le placement de la main dominée, en forme de main plate signifiant « véhicule » (réalisation nommée *classifieur* ou *proforme* selon les théories, nommée *proforme* dans la suite), dans un emplacement de l'espace de signation (voir chapitre 1, section 1.4.3.2) ;
3. le signe monomanuel ENFANT, réalisé avec la main dominante, durant lequel le placement effectué en (2) est souvent maintenu ;
4. un mouvement de la main dominante en forme d'index tendu (*proforme* « personne debout »), donnant à voir le déplacement de l'enfant vers l'emplacement de la voiture porté par la main dominée, structure appelée *transfert de situation* ou *depicting sign* selon les théories ;

---

5. Signes répertoriés dans les dictionnaires d'images et identifiés par des gloses



1. VOITURE

2. *proforme*3. ENFANT + *proforme*4. *signe illustratif*FIGURE 3.7 – Les 4 unités manuelles composant  $S_1$  traduite en LSF

- un regard décrochant de l’interlocuteur juste avant l’unité (2) et dirigé vers le même emplacement que celui de la main dominée, et un second sur la main dominante pendant son déplacement durant l’unité (4).

Cette production comporte deux placements de proformes (un pour chaque main), ainsi qu’un mouvement illustrant le déplacement du second (la personne) vers le premier (le véhicule). Cette dernière unité est absente des dictionnaires car elle n’est jamais considérée comme pleinement lexicale.

De plus, dans cette séquence, plusieurs éléments sont combinés (les deux mains, le regard). Il n’est donc pas possible de segmenter cette séquence en une suite de signes séparés, ce qui pose problème pour l’analyser avec l’approche discutée ici. De plus pour  $S_5$  ci-dessous, le signeur produira sans doute la même suite, avec une extension de la durée de l’unité (4) en transférant sur lui directement la notion portée par l’adverbe « lentement », c’est-à-dire sans ajouter de signe. Cela crée d’autant plus de simultanéité, qui ne peut être modélisée avec un modèle purement linéaire.

$S_5$  : « un enfant s’approche lentement d’une voiture »

Nous décrivons dans la section qui suit les approches par apprentissage.

### 3.4.2 Approches par apprentissage

Jusque dans les années 1980, la plupart des systèmes de TAL reposaient sur des ensembles complexes de règles écrites à la main. L’arrivée de l’apprentissage machine dans le domaine de la langue doit à la fois à l’augmentation de la puissance de calcul disponible et à la baisse en puissance du courant générativiste pur. De plus en plus, le TAL s’est donc tourné vers ces nouvelles techniques statistiques, généralement plus robustes lorsque confrontées à une entrée langagière imprévue ou erronée comme il en existe dans les données grandeur nature. Mais la quantité et la qualité des données nécessaires pour mettre en œuvre ces techniques peuvent être rédibitoires.

#### 3.4.2.1 En TAL

La plupart des méthodes par apprentissage machine pour le TAL sont basés sur les modèles de Markov cachés (HMM) ou les n-grammes. Un HMM est un



modèle statistique utilisé pour reconnaître des suites de motifs dans les données. À partir de modèles statistiques et probabilistes, celui-ci apprend la variabilité des données au cours du temps, les variabilités intra-signeur (entre différentes productions d'un même signeur) et inter-signeur (entre mêmes productions de signeurs différents). Un n-gramme est une suite contiguë de n éléments dans une séquence donnée. En TAL, ces éléments peuvent être des phonèmes, des syllabes, des lettres, des mots, etc.

Voici quelques exemples d'applications de l'approche statistique en TAL :

- modèles de langue pour la reconnaissance automatique de la parole : trouver la suite de mots la plus probable parmi les suites candidates compatibles avec le signal acoustique ;
- étiquetage morpho-syntaxique : trouver la suite d'étiquettes la plus probable pour une séquence de mots donnée ;
- analyse syntaxique à partir de grammaires probabilistes pour décider de l'arbre syntaxique le plus probable ;
- traduction automatique : trouver la phrase la plus probable pour être une traduction cible de la phrase source.

#### 3.4.2.2 En TALS

Pour les LS, les premières utilisations de ces approches avaient toutes pour but la reconnaissance automatique de données de LS. Les toutes premières utilisaient comme dispositif d'entrée les gants numériques, plus tard remplacés par les caméras vidéo. La reconnaissance automatique comprend deux étapes :

1. une analyse des caractéristiques pertinentes du flux vidéo : cette étape prend en compte les indices de bas niveau, sans considération linguistique donc hors du champ de ce chapitre (pour un état des lieux détaillé, voir Cooper (2011)) ;
2. segmentation et classification : généralement précédée d'une phase d'apprentissage, cette étape effectue la reconnaissance à proprement parler en déterminant la séquence la plus probable d'unités.

Avec l'apparition des méthodes statistiques, les systèmes de reconnaissance de la parole ont significativement évolué, aidés par la constitution de très larges corpus. À cause du manque de corpus de LS, il semble très difficile d'emprunter le même chemin (Cooper & Bowden, 2009) et espérer les mêmes résultats. Malgré cela, beaucoup de travaux se basent sur des HMM ou des variantes ayant pour but de tenir compte de la simultanéité (Vogler & Metaxas, 2001) ou la faible quantité de données.

De plus, ces approches font face aux problèmes suivants, liés à la nature des LS :

1. les signes doivent être découpés en unités de description plus petites, à valeurs phonétiques, dont la définition, la pertinence et la détection demeurent problématiques (Theodorakis *et al.* , 2010), même dans la réflexion linguistique (voir chapitre 1, section 1.4.1) ;

2. les signes lexicaux peuvent présenter de fortes modifications de ces composants internes selon le contexte (iconique notamment) dans lequel ils sont utilisés (voir chapitre 1, section 1.4.2.1), et il n'est pas possible de représenter toutes les variations possibles sans trop généraliser et perdre les catégories pertinentes ;
3. les signes fortement iconiques sont des unités gestuelles construites souvent à la volée, et ne sont listés ni listables dans aucun dictionnaire ce qui rend très difficile de les catégoriser pour un apprentissage classique.

D'autres phénomènes, spécifiques aux LS, doivent être considérés. Une production signée n'implique pas uniquement les mains, mais également d'autres articulateurs du corps, plus ou moins synchronisés sur des axes spatiaux et temporels différents. De plus, le signeur ancre des éléments de son discours dans l'espace de signation et le structure grâce à lui. Les signes, qu'ils soient ou non fortement lexicalisés, sont arrangés dans cet espace topologique, donnant lieu à de nombreux pointages et autres opérations de référencement spatial. Ces phénomènes rendent souvent les systèmes de reconnaissance de la parole impuissants car conçus pour les langues vocalisées (voire transcrites) donc avec une hypothèse de plus forte linéarité (Dalle, 2006). Pour ces raisons, les systèmes de reconnaissance automatique en LS n'ont jamais pu approcher les conditions réelles, en termes de taille de vocabulaire et de couverture de phénomènes de langue.

Des systèmes plus accessibles pourraient voir le jour pour aider par exemple à l'annotation ou au repérage de signes lexicaux dans leur forme de citation. Par exemple, le prototype de wiki en LS du projet DictaSign (figure 3.8) offre une interface de recherche de signe à partir d'une entrée capturée par une Kinect® doublée d'un système de reconnaissance.

Ainsi, par construction, les approches utilisant l'apprentissage sont linéaires, et supposent que les productions signées sont descriptibles en une séquence de signes ou de gloses. Elles souffrent donc des mêmes limites que celles évoquées dans la section précédente. Cette observation rappelle aussi le problème de la transcription (chap. 2), impossible avec le seul emploi de gloses.

Des modèles multi-linéaires, construits pour les LS, ont été élaborés pour dépasser ces limites. Nous les présentons dans la section suivante.

### 3.4.3 Approches dédiées

Certains chercheurs défendent le besoin de développer de nouvelles approches, plus spécifiques aux LS. Les principales critiques sont les suivantes :

- les modèles linéaires, quelle que soit leur nature, ne représentent pas la simultanéité et la synchronisation complexe qui existe entre les multiples articulateurs (manuels et non-manuels) des productions signées ;
- l'apprentissage machine a besoin de quantités de données qui ne sont pas accessibles en LS.

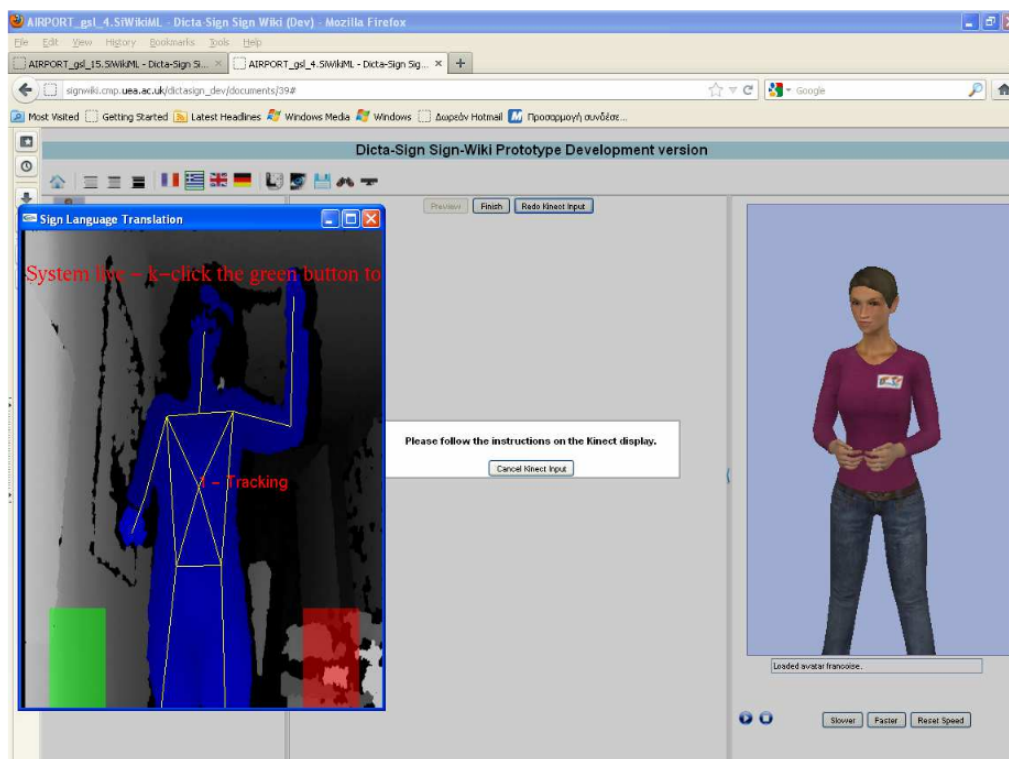
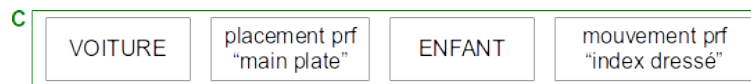
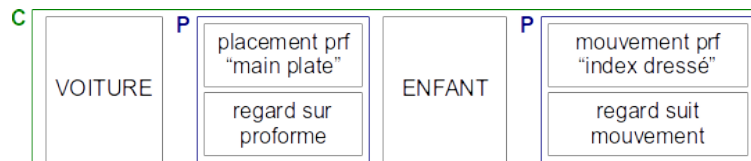


FIGURE 3.8 – Le prototype SignWiki du projet Dicta-Sign utilisant un système de reconnaissance de signes lexicaux.

FIGURE 3.9 – Enoncé  $S_1$  modélisé avec des nœuds PFIGURE 3.10 –  $S_1$  modelisé avec des noeuds P/C

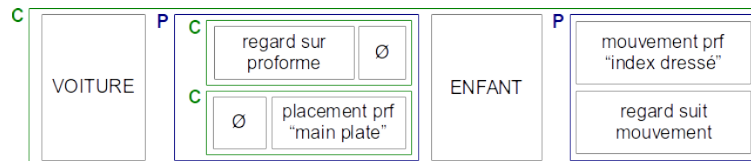
### 3.4.3.1 Multilinéarité

Les modèles multilinéaires prennent en compte la simultanéité des gestes portés par des articulateurs différents sur des empan temporels différents. Un premier travail de description formelle parallélisée a été proposé avec le modèle « P/C » (Huenerfauth, 2006), utilisé ensuite par Lopez-Colino dans sa thèse (López-Colino & Colás, 2012). Il définit deux manières de combiner des parties d’une activité signée : la *partition* et la *constituance* (*partitioning* et *constituting*), respectivement formalisés en P-node et C-node, donnant son nom au système.

Les C-nodes (nœuds *constitués*) créent des séquences de deux ou plusieurs nœuds fils. Par exemple, si tout tient en une séquence d’unités lexicales, celles-ci seront *constituées* sous un C-node parent, qui peut s’illustrer simplement par une ligne temporelle sur laquelle se succèdent les nœuds constitués (figure 3.9). Ceci correspond à ce qui se passe dans une règle de grammaires à constituants, où un nœud parent (non-terminal, équivalent du C-node ici) se réécrit en une séquence de nœuds fils (ici les unités lexicales).

La nouveauté dans le modèle P/C est la possibilité de *partitionner* des nœuds pour créer des axes parallèles (et non séquentiels) partageant le même empan temporel (les mêmes bornes). Par exemple, la phrase  $S_1$  page 89 se traduit en LSF comme le montre la figure 3.7 page 91, avec notamment un regard dirigé vers l’espace de signation sur la deuxième et la dernière unités. Avec le modèle P/C, ceci peut être représenté à l’aide d’un P-node qui parallélise chaque unité avec le regard, tel que l’illustre la figure 3.10.

Cette structure en partition permet la simultanéité de différentes parties de l’activité signée, sans nécessairement placer l’activité non-manuelle à un niveau sous-lexical. On peut dès lors décrire des articulations simples comme un gonflement des joues par-dessus plusieurs unités lexicales à la suite, tout en incluant les phases de transition. Ainsi, ce gonflement pourrait être utilisé sans être appliqué à un signe lexical, et les nœuds terminaux de P/C ne sont plus nécessairement un ensemble de signes au sens d’un lexique. Le lexique (formel) de la grammaire est donc distinct de celui du dictionnaire de la langue. Il est

FIGURE 3.11 –  $S_1$  modélisé avec des noeuds P/C et null

à noter cependant que P/C n'a jamais été proposé comme grammaire pour les LS.

Cela dit pour certains chercheurs dont Huenerfauth lui-même, l'annotation de vidéo de productions signées et l'étude de quelques articulateurs comme les sourcils, le regard et les mains (et il y en a de nombreux autres), révèlent que peu d'entre eux s'alignent exactement avec les bornes des nœuds constitués sur une partition. Au contraire, ils se synchronisent d'une manière complexe et suivant des règles de précedence temporelle systématiques, alors que les bornes des P-nodes servent de bornes pour toutes les partitions qu'ils mettent en jeu.

Au niveau phonétique, à savoir celui qui décrit les articulations d'une production donnée, P/C résout le problème des différences d'ordre ou de durée des nœuds partitionnés en introduisant les « null-nodes » ( $\emptyset$ ), pouvant être constitués comme les autres pour remplir un morceau de partition avec un vide et ainsi en raccourcir la durée relative. La figure 3.11 illustre comment le regard peut être décrit comme précédent le signe glosé ARRIVER.

Si cette approche dépasse les limites des modèles linéaires, elle ne permet pas de généraliser au plus haut niveau grammatical. Les corpus révèlent par exemple qu'un mouvement de la tête en guise de négation dure plus longtemps qu'un signe lexical qu'il pourrait accompagner. Dans le cas d'une suite plus longue, ce mouvement aura tendance à commencer vers la fin du signe, alors que pour un unique signe court, il pourra commencer avant. Selon la forme de la chose à nier, l'ordre d'activation des différents articulateurs en jeu – donc la structure de la description – varie jusqu'à inverser les imbrications des P-nodes, C-nodes et null-nodes. La généralisation en une règle grammaticale pour la négation par le mouvement de la tête s'avère donc difficile à décrire avec ce système.

Il serait finalement plus simple de pouvoir librement positionner les intervalles de description les uns par rapport aux autres, quel que soit l'articulateur mis en jeu. La section suivante présente une approche de ce type.

### 3.4.3.2 Décloisonnement des niveaux

Un problème plusieurs fois formulé est que la plupart des approches présentées ci-dessus font l'hypothèse d'une distinction naturelle entre un niveau lexical et un niveau syntaxique, chacun ayant du coup son modèle de description propre. En d'autres termes, elles sont conformes au modèle à niveaux linguistiques, l'acceptant ainsi tel quel pour les langues des signes. Dans le chapitre 1,

nous avons discuté cette séparation et indiqué qu'elle posait un certain nombre de problèmes (cf. section 1.5.1).

Pour nous, le manque de théorie formelle sur un type de langue peu décrit est une chance pour le TALS de ne pas se fermer aux spécificités inconnues, et pour au contraire formaliser les LS sans le biais de la structure écrite, nécessairement présent en cas de transfert de modèles établis. Pour ce faire, une possibilité est de s'en remettre à une hypothèse de travail moins forte et à une méthodologie dont le supposé de départ n'est ni plus ni moins qu'il existe des liens (linguistiques) entre des *formes* et des *fonctions*, c'est-à-dire entre :

- formes : les productions (états et/ou mouvements) visibles des articulateurs de la langue, par exemple « paupières fermées » ;
- fonctions : leur interprétation (sens, intention), qu'elle soit rhétorique, sémantique, lexicale ou même non identifiée en ces termes, par exemple « changement de sujet » ou « jugement péjoratif sur une personne, un objet ».

Les liens s'élaborent par une annotation et une analyse de corpus de LS. Partant soit d'une forme, soit d'une fonction, un processus de recherche est alternativement lancé, jusqu'à ce que :

- des invariants de forme peuvent être formalisés pour toutes les occurrences d'une fonction identifiée, ce qui donne lieu à une *règle de production* pouvant être utilisée par un système de génération pour animation d'un signeur virtuel ;
- une fonction peut être interprétée pour chaque occurrence d'un critère de forme donné, ce qui génère une *règle d'interprétation* utilisable par des systèmes de reconnaissance automatique de LS.

Par exemple en LSF, il a été montré que le placement d'objets dans l'espace produisait toujours le même patron de forme : l'objet est nommé d'abord, puis un classifieur porté par un bras, une main ou un doigt est ancré à l'emplacement choisi de l'espace de signation avec un léger mouvement descendant, immédiatement précédé d'un décrochement du regard vers le même emplacement (Braffort & Dalle, 2008). Une illustration de ce patron que nous nommerons **R1** est donnée figure 3.12(a).

Dans la figure :

- comme dans une partition d'annotation, les rectangles repèrent les intervalles temporels sur l'axe horizontal, durant lesquels l'articulation qu'ils contiennent doit être produite ;
- cependant, l'arrangement vertical des rectangles est arbitraire – les diagrammes ne sont pas à lire tels que des annotations où chaque piste serait dédiée à un et un seul articulateur ;
- les termes en italique identifient les éléments variables, à considérer comme les arguments de la règle – en bref, les boîtes « à remplir ».

Dans son travail de construction d'un système de description grammaticale applicable aux LS dans un contexte de synthèse de la LSF, Filhol fait trois hypothèses.

La **première** est que toute règle ainsi écrite peut relever de n'importe quel niveau d'abstraction du langage, comme :

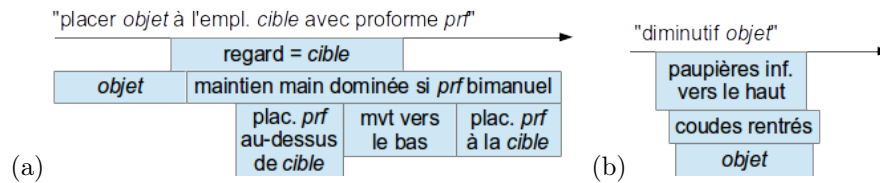


FIGURE 3.12 – Rule time line illustrations (parameter arguments are in italics)

- le niveau lexical, si elle spécifie une unité dont la correspondance entre sa forme articulée et le concept tel que donné par un dictionnaire est stabilisée ;
- la syntaxe, si elle contraint l'ordre de ses sous-éléments ;
- la grammaire au sens plus large si la séquence pure est insuffisante et que des règles plus fines de synchronisation et de recouvrement des parties sont nécessaires ;
- le niveau discursif, si elle décrit comment passer d'un sujet à l'autre, la durée des pauses entre les différents sujets, etc.

La règle *R1* ci-dessus (placement d'entité dans l'espace) serait par exemple située quelque part entre la syntaxe et le discours, sans pouvoir être parfaitement positionnée en ces termes.

La **deuxième** hypothèse est que les règles sont paramétrables, au sens où la forme décrite pour une fonction donnée peut :

- être agencée autour d'un emplacement au contenu nommé mais encore inconnu et devant être précisé au moment de son utilisation (par exemple le proforme à utiliser) ;
- dépendre d'une valeur issue du contexte au moment de la signation, comme le point de l'espace à regarder.

La **troisième** hypothèse est qu'un ensemble suffisant de règles récursivement imbriquables peut être élaboré pour constituer une grammaire complète des productions signées, car plusieurs de ces règles peuvent être récursivement composées, un paramètre d'une règle pouvant par exemple être spécifié par l'application d'une autre. Cette grammaire générerait tout énoncé signé, autrement dit tout énoncé pourrait être dérivé de cette grammaire. Au stade actuel, cette hypothèse assez forte a encore besoin d'être confortée. Un premier ensemble de règles suffisamment grand doit encore être identifié.

Par exemple, à partir de la règle *R1* illustrée en 3.12a et une autre règle ***R2*** dont la fonction serait « qualificatif petite taille » illustrée en 3.12b, la boîte *object* de la première pourrait être remplie avec un appel à la deuxième pour représenter (et produire) le placement dans l'espace d'un petit objet. À son tour, le paramètre de 3.12b peut faire appel à une autre règle, typiquement ici décrivant une entrée de dictionnaire, elle-même également paramétrable. Cela illustre le potentiel récursif du modèle proposé.

Une telle approche décloisonne ainsi les niveaux linguistiques traditionnels et permet toujours l'identification de niveaux (par exemple lexical, syntaxique...) et

la description systématique de leurs éléments, mais elle ne l'impose pas. Toutes les productions et leurs généralisations, abstractions paramétrées, sont décrites avec les mêmes outils. C'est ce qui rend poreuses les cloisons entre niveaux plutôt que de les séparer a priori.

Dans ce modèle, les productions finales sont le résultat d'imbrications successives de fonctions décrivant des formes associées, jusqu'à produire les articulations non paramétrés (invariants), correspondant à ce que l'on voudrait intuitivement placer au niveau phonétique. Vérifier une telle hypothèse validerait là un type nouveau de grammaire pour les LS, et ce faisant apporterait un éclairage linguistique nouveau à toute la discipline du TAL. En effet, nous remettrions là en question plusieurs propriétés supposées universelles du langage.

Pour décrire une forme (éventuellement paramétrée) à produire, un langage nommé *AZee* est en cours de développement (Filhol, 2012; Filhol *et al.*, 2014). Il est capable de formaliser les aspects suivants :

- synchronisation temporelle : séparation des différents articulateurs en groupes parallèles participant à la forme globale à produire et repérage par contraintes libres sur un axe temporel ;
- description des articulations : postures corporelles (généralement partielles), dont les segments du squelette, les muscles faciaux et la cible du regard, tous spécifiés sans ordre d'importance ;
- géométrie dans l'espace de signation : vus comme dans un espace Euclidien classique, tout emplacement, distance ou orientation s'exprime géométriquement en manipulant points, vecteurs, plans, symétries... certains sites corporels (points) et axes liés au squelette (vecteurs) étant préalablement définis.

Pour ces trois aspects, l'approche consiste à spécifier des contraintes nécessaires et suffisantes (Filhol & Braffort, 2012b). Au lieu de définir une liste a priori d'attributs minimaux (p.ex. les paramètres de Stokoe) pour lesquels il faudrait aussi établir l'inventaire a priori des valeurs possibles, la description inclut tout élément devant être vérifié dans la forme doit être décrit (pas de sous-spécification), et seul ce qui est nécessaire à la production apparaît dans la description de la forme (pas de sur-spécification).

*AZee* est déjà implémenté dans le prototype KAZOO, une application web de génération de langue des signes avec signeur virtuel (fig. 3.13). Le projet est en cours, mais la version courante offre déjà les fonctions de base pour générer quelques énoncés à partir de descriptions *AZee*, et en cela offre un premier élément de validation du modèle (Braffort *et al.*, 2015).

Les approches multilinéaires et décroisées permettent de dépasser les limites des modèles linéaires décrites dans les sections précédentes, en permettant la représentation de n'importe quelle production indépendamment du nombre et du type des articulateurs en jeu. Elles laissent donc présager une couverture plus large des propriétés des langues des signes, mais aucune évaluation n'a encore permis de le mesurer à proprement parler. La section suivante parle de limites supplémentaires et étudie la question difficile de l'évaluation d'un modèle linguistique.



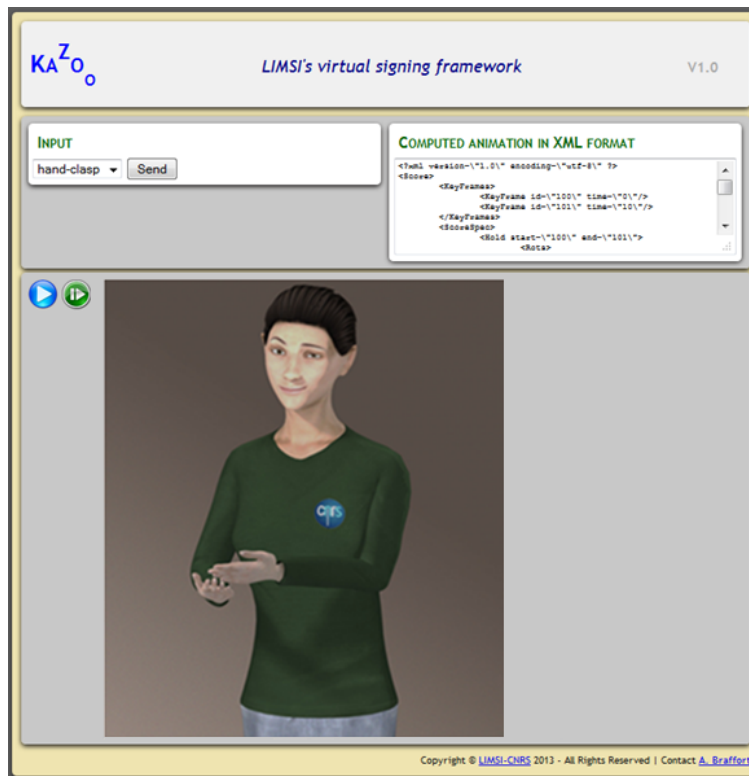


FIGURE 3.13 – Démonstrateur web Kazoo, version 1.0.

### 3.4.4 Limites, évaluation

Un certain nombre de problèmes demeurent, dont la modélisation de l'espace de signation et des structures fortement iconiques. La question de la validation de modèles linguistiques reste également un problème en suspens.

#### 3.4.4.1 Espace de signation

En langue des signes, les entités du discours sont fréquemment associées à des emplacements de l'espace et la plupart des relations entre elles sont alors directement projetées dans le même espace en faisant usage de ces emplacements (voir chap. 1, section 1.4.3.1). Ces relations peuvent être topologiques/spatiales ou temporelles, ou encore marquer une comparaison, une opposition, etc. Les actions sont souvent signées avec l'utilisation de verbes directionnels ou des structures plus complexes très fortement iconiques, toutes s'empilant dans le même espace physique. Étant donnée l'importance de ces utilisations de l'espace, les modèles de représentation se doivent de les intégrer dans leurs outils de description.

Seuls quelques uns de ces modèles intègrent pourtant la structure spatiale des énoncés LS. Soit ils découpent l'espace en un ensemble discret de valeurs d'emplacement, soit ils font le choix d'un espace continu dans lequel tout point de l'espace peut représenter un emplacement linguistique. En reconnaissance, des études se sont concentrées sur les verbes directionnels (Sagawa & Takeuchi, 1999) et les relations spatiales iconiques simples (Braffort, 1997). De même, des travaux en synthèse ont abordé les verbes directionnels (Elliott *et al.*, 2000) et des structures iconiques plus complexes (López-Colino & Colás, 2011). Mais à chaque fois, seul un faible lien est fait entre les outils de ces modèles et les mécanismes linguistiques.

À ce jour, la représentation formelle la plus précise de l'utilisation de l'espace de signation (incluant les entités localisées, les liens entre elles...) est proposée par Lenseigne et Dalle (Lenseigne & Dalle, 2005). Dans ce travail, la représentation symbolique de l'espace de signation consiste en un cube autour du signeur divisé en plus petits cubes égaux matérialisant les différents loci (emplacements linguistiques) possibles. Chaque locus peut être attribué à une entité du discours, et étiqueté sémantiquement avec des rôles tels que *personne*, *date*, etc.

La figure 3.14 donne un exemple d'organisation spatiale possible pour l'énoncé suivant (la traduction est littérale), où les lettres insérées correspondent à un locus : « Dans la ville de Toulouse (A), se trouve un cinéma nommé *Utopia* (B) qui projette un film (C) le jeudi 26 février à 21h30 (D). La personne (E) qui l'a réalisé (F), qui est-elle (G) ? » Les rôles des entités reliées sur le schéma est spécifié par une forme et une couleur, par exemple un cylindre rouge une personne, un disque jaune pour un lieu... La figure 3.15 donne les étapes successives dans l'élaboration de cet énoncé, chacune créant un locus dans l'espace de signation.

Mais même cette représentation ne couvre l'intégralité des relations entre les entités, et davantage d'études sont nécessaires pour construire un modèle complet de l'espace, intégrable à des applications logicielles pour la reconnaissance,

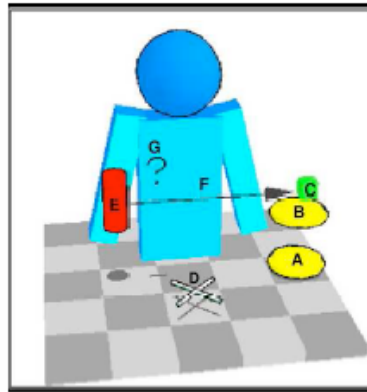


FIGURE 3.14 – Exemple de représentation de l’organisation spatiale d’un énoncé

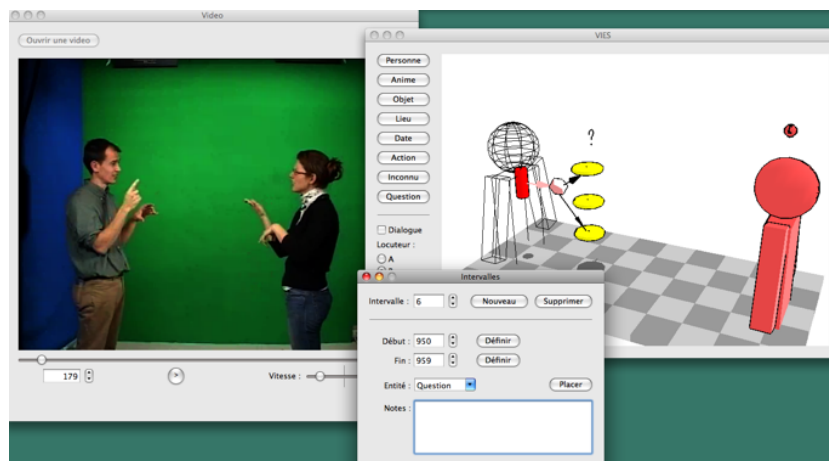


FIGURE 3.15 – Elaboration de l’organisation spatiale d’un énoncé à l’aide du logiciel VIES (chapitre 4, section 4.4.1.3)

la synthèse ou la traduction.

#### 3.4.4.2 Unités fortement iconiques

Comme expliqué section 3.4, les modifications en contexte des signes et les unités iconiques sont rarement intégrées aux modèles, souvent limités à la représentation lexicale en forme de citation. En observant les annotations issues du projet DictaSign dans lequel 5 heures de dialogue en LSF ont été annotées, nous avons dénombré les *fully lexical signs* (FLS, unités lexicalisées) et les *partly lexical signs* (PLS) telles que les définit T. Johnston (Johnston, 2010) et incluant les pointages, des signes iconiques et les ancrages de la main dominée (buoys). Sur l'ensemble de nos 25 000 unités, 10 000 PLS ont été repérées, alors même que les modifications des formes de citation (typiquement l'usage iconique de l'espace dans les verbes directionnels) n'ont pas été annotées, celles-ci comptant donc comme FLS. Cela prouve que l'usage des unités autres que les simples signes lexicalisés dans leur forme de citation sont très largement présentes en LS, et appuie le besoin d'en faire état dans les modèles de représentation.

En reconnaissance, de premières structures iconiques ont été modélisées pour repérer l'utilisation de proformes dans l'espace (Braffort, 1997), ou les relations topologiques entre deux entités (Bossard, 2006). Une étude théorique a aussi produit un inventaire de types sémantico-cognitifs pour la synthèse (Lenseigne & Dalle, 2005), basé sur des schémas permettant plus de complexité iconique, mais l'implémentation reste à faire.

À ce jour, la plus large couverture des structures iconiques pour la synthèse est apportée par López-Colino (2011) dont le système génère ce que l'auteur appelle des *inflective classifier constructions* (modifications d'une ou plusieurs valeurs paramétriques lexicales, p.ex. les verbes directionnels), des *classifier predicates* (relations spatiales entre deux entités) et les *affixial classifiers* qui combinent les deux précédents.

Malgré les efforts récents pour inclure toujours plus de types d'unités gestuelles dans les modèles, de nombreux problèmes restent en suspens, parmi eux les unités iconiques permettant de décrire les situations ou les taille ou forme des entités, et les prises de rôles, sans doute les structures linguistiques les plus résistantes à la formalisation.

#### 3.4.4.3 Le problème de l'évaluation d'un modèle

Il est toujours possible de lister les limites des modèles proposés, comme nous l'avons fait au long de ce chapitre. Cela dit, ces remarques ne sauraient suffire pour évaluer ces modèles à proprement parler. Globalement d'ailleurs, le domaine manque de techniques et de métriques d'évaluation.

À notre connaissance, l'évaluation des approches et formalismes de description pour les LS est une question qui n'a pas été elle-même approfondie. Ce que l'on trouve dans la littérature comporte :

- des listes d'exemples positifs de représentation de productions signées, appuyant un modèle ;

- des cas d'exemples résistant au modèle proposé, la plupart du temps pour en présenter une extension qui couvre finalement le phénomène ;
- des protocoles et des plateformes pour évaluer la qualité de productions synthétisées, ou des scores de performance de systèmes de reconnaissance.

Selon nous, une distinction doit être clairement faite entre un modèle et les applications l'utilisant. Alors qu'une évaluation positive d'un logiciel de synthèse ou de reconnaissance peut laisser entendre l'aspect prometteur du modèle sous-jacent, il n'en reste pas moins invisible et absent du protocole d'évaluation, et n'est donc pas ce qui est finalement évalué (Morrissey, 2011). Aucun système de synthèse ne génère d'animation parfaite, aucun système de reconnaissance ne traite la langue dans son ensemble, et il est impossible de savoir formellement lequel du modèle ou de l'application en est la cause.

L'évaluation des applications logicielles est discutée plus loin, dans le chapitre 5, section 5.4.4. Nous nous concentrons maintenant sur les propriétés et possibilités des modèles sous-jacents. Cette section traite de la difficulté de cette tâche, et propose quelques pistes de réflexion.

**Couverture et précision** Une première idée pour évaluer un modèle est de se poser la question de sa couverture. Étant donné un élément ou une propriété de la langue modélisée, celui-ci peut-il être représenté ? Le problème avec une telle question est qu'un modèle est toujours associé avec un point de vue théorique sur l'objet à décrire. Cela influe sur son panel d'inventaires, de catégories, etc. voire le façonne complètement. De même, tout jugement quant à la capture d'un phénomène par un modèle donné sera porté avec le point de vue théorique de la personne qui évalue. Ainsi, l'objectivité est impossible puisqu'elle supposerait paradoxalement un modèle de référence validé, donc évalué.

Une façon d'aborder néanmoins le problème de la couverture est de reformuler la question de la façon suivante : « étant donné la propriété  $P$  du modèle  $A$ , peut-on décrire  $P$  dans le modèle  $B$  ? » et de dresser le comparatif pour tous les couples  $(A, B)$  de modèles différents. On notera cependant que cela exclue d'entrée tous les phénomènes connus mais inclus dans aucun formalisme (on pense aux prises de rôle par exemple). Aucun inventaire comparatif ne saurait donc être suffisant.

**Utilisabilité par l'humain** D'autres critères, moins sur le formalisme que sur son utilisation pragmatique, peuvent entrer en jeu lorsqu'il s'agit de choisir une représentation pour créer par exemple des bases de données ou lexiques. La facilité de description, de relecture, de correction et de partage ne sont pas théoriquement essentiels mais créent évidemment des préférences lorsque le temps de travail humain est compté. Certains critères feront qu'un modèle sera choisi par la communauté linguistique parce que plus lisible (même si relativement moins puissant), ou pour en partager les descriptions comme avec le Sign puddle<sup>6</sup> de SignWriting<sup>7</sup>.

6. <http://www.signbank.org/signpuddle/>

7. Base de donnée lexicale ouverte et collaborative

Nous listons ci-dessous quelques critères de cette nature :

- temps moyen nécessaire à la création d'une entrée (un signe, une règle...)
- temps moyen pour le déchiffrement d'une entrée par un humain
- temps moyen d'une formation au modèle
- couche logicielle nécessaire pour manipuler ou créer les entrées (polices de caractères, éditeurs...)
- facilité du copié-collé dans un texte, un courrier électronique, une publication scientifique
- quelles requêtes sont possibles dans une base de données remplie ?

**Utilisabilité par la machine** Il est évident qu'un modèle dont l'implémentation est aisée sera préférée de l'informaticien par rapport à une description ouverte ou non suffisamment formelle pour être mise en œuvre. Cela dit, il reste douteux d'envisager uniquement les sorties possibles d'un traitement automatique pour juger le modèle sous-jacent, et plus l'application logicielle est complexe, moins cela a de sens.

Cela montre à quel point la page est vierge en termes d'évaluation de modèles formels pour les LS, et si l'on peut proposer des critères mesurables de comparaison, il reste épineux d'envisager l'évaluation formelle – positive ou négative – d'un modèle donné.

Après ce passage en revue des modèles de description formels des langues des signes et leurs limites, la section qui suit s'attaque à un domaine plus récemment appliqué aux LS : la traduction automatique.

### 3.5 Une étude de cas : la traduction automatique

Le domaine de la traduction automatique a plus récemment été aussi appliqué aux LS. Le but peut être soit de créer des traductions complètement automatiques, soit d'assister à la tâche de traduction sur laquelle l'humain a la main. Dans les deux cas, les outils de TALS présentés dans les sections précédentes sont utilisés, en entrée ou en sortie selon le sens de la traduction.

En traduction automatique aussi, il existe deux approches, souvent opposées mais parfois hybridées : celles à base de règles et celles statistiques. À présent, ces dernières ont gagné le terrain de la traduction automatique appliquée aux données textuelles et vocales. Pour les LS, on a plus longtemps préféré les premières (3.5.1) malgré le mouvement dominant vers les approches statistiques, mais celles-ci sont finalement apparues également (section 3.5.2). Des travaux plus récents explorent aussi la voie d'approches dédiées (section 3.5.3).

#### 3.5.1 Approches à base de règles

La traduction par règles a émergé dans les années 1970. Plusieurs méthodes sont possibles, dont les niveaux d'abstraction peuvent être illustrés par le triangle de Vauquois (figure 3.16). La première méthode, triviale, est la conversion

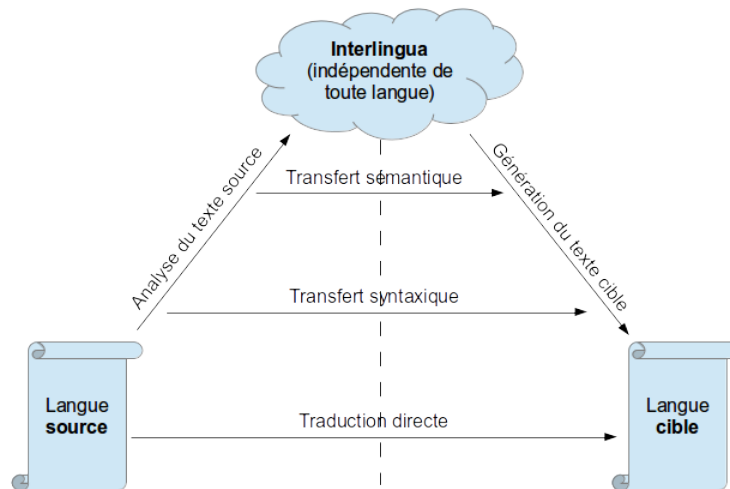


FIGURE 3.16 – Le triangle de Vauquois

lexicale *directe* de chaque mot source à son correspondant en langue cible, suivant la séquence entrée.

Mais les efforts de recherche sont rapidement montés le long des côtés du triangle pour produire des systèmes de *transfert* partant d'une analyse syntaxique de la langue source, et générant les phrases en langue cible à partir d'une représentation syntaxique transférée en cette langue. Avec ces méthodes et au regard de l'illustration, chaque pas vers la pointe du triangle se fait au prix d'un effort grandissant en analyse de la langue source ou en génération vers la langue cible. En revanche il réduit aussi l'écart séparant les deux représentations entre lesquelles se font le transfert. Les approches par transfert ne sont pas parfaitement généralisables à l'ensemble des paires de langues car chaque paire nécessite en théorie son propre module de transfert, ce qui rend la généralisation fastidieuse.

L'idéal serait une analyse complète du sens de la donnée source en une représentation « interlingua », supposée capable de représenter tout sens possible, d'une façon indépendante de la langue source (ou à équidistance entre toutes). L'avantage serait alors immédiat pour généraliser à toutes les paires de langues puisque l'on supprime le besoin de transfert. Cependant, un argument fort contre cet idéal théorique et basé sur l'hypothèse de Sapir-Whorf qu'il n'existe aucune langue universelle possible. Les différences culturelles et dans la manière de voir, d'étiqueter et de penser le monde rendent impossible une telle « langue » intermédiaire. Cela dit, l'on pourrait envisager un système s'élevant suffisamment haut pour approcher une représentation pivot, sorte de description sémantique

enrichie, assez abstraite pour servir d'interlingua. Par extension, ce sont ces approches que nous appellerons par *interlingua*, par contraste avec les approches par *transfert* qui implémentent explicitement l'étape de transfert.

Là encore, la majeure partie de la littérature scientifique traite de traductions de l'écrit ou de l'oral vocalisé transcrit, et rares sont les travaux sur les LS. Morrissey (2008) rapporte huit études sur le sujet, du projet Zardoz (approche par interlingua) (Veale *et al.*, 1998) au plus récent projet espagnol (approche par transfert) (San-Segundo *et al.*, 2007). Six d'entre eux impliquent la langue des signes américaine (ASL); tous ont pour langue source une langue textuelle et la LS pour cible, avec un signeur virtuel pour la production finale.

ViSiCAST (Elliott *et al.*, 2000) est probablement le travail le plus abouti à ce sujet, présentant un processus par étapes de l'anglais à la LS, avec la *discourse representation theory* (DRT) comme représentation sémantique intermédiaire. À notre connaissance, aucun système de traduction n'a été présenté à proprement parler, mais le processus complet a été décrit, basé sur une grammaire à séquence lexicale (principe décrit dans (Marshall & Safar, 2004)). Aucune approche aboutie ne considère la LS comme langue pleinement multi-linéaire, et toutes considèrent les mains comme les principaux, sinon les seuls, articulateurs en jeu, décrits de manière paramétrique au sens de Stokoe.

Le temps de travail humain requis pour élaborer les règles linguistiques utiles à un traitement automatique est prohibitif et disqualifie la plupart du temps les approches à base de règles. Faiblement robustes et trop lentement évolutives, celles-ci ont été progressivement délaissées à mesure que les puissances de calcul ont augmenté et permis d'envisager les approches par apprentissage statistique, présentées ci-après.

### 3.5.2 Approches basées sur les données

Les approches basées sur les données ont commencé à gagner le terrain de la traduction automatique dans les années 1990. On trouve deux courants : le courant statistique et celui basé sur l'exemple.

Les approches statistiques sont principalement issues de la reconnaissance de la parole. L'idée fondamentale est de maximiser  $P(T|S)$  la probabilité que la phrase  $T$  soit une traduction de  $S$ . Cela est équivalent à maximiser le produit  $P(S|T) \cdot P(T)$ , où  $P(T)$  est la probabilité d'apparition de la phrase  $T$  dans la langue cible (on l'appelle le *modèle de langue*), et  $P(S|T)$  la probabilité que  $S$  soit une traduction de  $T$  (on l'appelle le *modèle de traduction*).

Des corpus bilingues sont utilisés pour estimer ces probabilités. Le modèle de langue ne requiert qu'un corpus monolingue et utilise des modèles n-grammes comme les systèmes de reconnaissance statistiques. Le modèle de traduction est appris à partir de grands ensembles de bitextes de mots et/ou de syntagmes *alignés*, c'est-à-dire des textes parallèles où chaque mot ou syntagme de la langue source est repéré dans l'équivalent dans la langue cible. Cette approche nécessite :

1. de considérer tous les énoncés, source et cible, comme des séquences de mots dans des phrases segmentées en mots ;



2. de disposer de grands volumes de données parallèles pré-alignées pour un entraînement suffisant des systèmes d'apprentissage.

La traduction par l'exemple emploie un principe d'analogie, comparant la séquence d'entrée avec le contenu d'un corpus bilingue de traductions de référence. Elle comporte trois étapes :

- la recherche des extraits les plus semblables à l'entrée dans le corpus traduit et récupération de la portion traduite en langue cible ;
- le repérage des *chunks* (morceaux) alignés ;
- la recombinaison des segments en langue cible pour construire la traduction finale.

Là encore, peu de travaux nourrissent la littérature sur les LS. Les approches statistiques ont d'abord été utilisées pour traiter la traduction de LS à texte (Bauer *et al.* , 1999; Stein *et al.* , 2006; Forster *et al.* , 2013), mais la traduction de texte à LS est aujourd'hui aussi un sujet de recherche. La traduction par l'exemple a été plus récemment proposée par (Morrissey, 2008).

Tous ces travaux utilisent généralement des outils pré-existants et prévus pour des langues textuelles ou vocalisées, par exemple Matrex<sup>8</sup> ou Moses<sup>9</sup>, imposant une vision linéaire des énoncés signés. Dans (Morrissey, 2011), l'auteur compare plusieurs représentations (gloses, HamNoSys, SignWriting) et conclue que « dans le nombre limité de représentations disponibles », les gloses restent les plus pratiques « malgré leur incapacité à porter l'information flexionnelle et linguistique suffisante pour comprendre un énoncé signé ».

Les méthodes basées données ont des avantages sur celles à base de règles (indépendantes de la langue, peu d'intervention humaine), mais leurs performances varient largement selon la paire de langues en jeu. Aussi, contrairement aux systèmes à base de règles qui implémentent une connaissance du système, la grammaticalité des phrases résultant d'un système appris statistiquement variera avec la qualité du modèle appris.

Enfin, le problème de la rareté des données d'apprentissage persiste. Nous avons vu comment la linéarisation du signal linguistique était simplificatrice pour les énoncés signés, et comment le manque de définition claire sur le signe rend problématique la segmentation de de l'énoncé en « suite de signes ». Pour nous, il demeure trop d'obstacles pour satisfaire les conditions (1) et (2) présentées ci-dessus.

### 3.5.3 Approches dédiées

Cette section présente l'approche dédiée aux langues des signes décrite dans la section 3.17 page 110 en vue d'une traduction de texte vers les LS. Comme pour les autres, il n'existe à ce jour aucune implémentation suffisante pour juger proprement de la faisabilité ou de la validité de la technique proposée. Il convient donc là aussi d'y voir plutôt une proposition théorique à explorer.

8. <http://www.openmatrex.org/>

9. <http://www.statmt.org/moses/>

D'après ces travaux, les énoncés en LS mettent en jeu des appels de règles de production générant récursivement les formes à articuler pour la fonction représentée (Filhol, 2011; Filhol *et al.*, 2013). Ces règles de fonction à forme sont établies sans identification obligatoire d'un unique niveau linguistique dont elles relèveraient, et leur ensemble constituent une grammaire de la langue ainsi décrite, où la question des niveaux linguistiques est devenue secondaire.

Travaillant sur des productions en LSF et observant que les fonctions portées par les règles sont lourdement chargées sémantiquement, Filhol suggère d'utiliser ces règles comme pivot d'un système de traduction de français écrit à LSF. Le problème de la traduction se reformule alors en une recherche de formes dans le français source portant les fonctions des règles disponibles constituant le système cible. La production de la langue cible à partir de ces règles est alors triviale puisque c'est exactement ce que contiennent les règles invoquées, et précisément les descriptions de leur contenu qui les ont fait exister.

Ceci divise le problème global en une multitude de sous-problèmes de TAL, à savoir autant que de règles disponibles. (cf. fig. 3.17). Chaque problème devient une tâche d'extraction d'information telle que traditionnellement définie dans le domaine. Pour mettre en œuvre cette architecture, il faut programmer un module d'extraction pour chaque règle de production AZee. Chaque module peut alors indépendamment être :

- plus ou moins complexe, faisant appel aux outils de traitement nécessaires, de l'étiqueteur morpho-syntaxique à l'analyseur morphologique, du plus simple indice lexical voire typographique jusqu'à la plus complexe résolution d'anaphore ;
- plus ou moins automatique (aussi selon l'application visée), étant donné le grand nombre d'éléments relevant du contexte et les ambiguïtés toujours présentes.

Dans une dernière étape, et à l'instar de la traduction par l'exemple où les segments récupérés doivent être combinés pour proposer la traduction finale (cf. la partie grise de la figure 3.17), les règles déclenchées le seront a priori séparément, et nécessiteront donc d'être recombinaisons, les faux positifs écartés, etc.

Contrairement à l'approche « marche avant » des systèmes qui partent de la langue source et effectuent des traitements vers la langue cible en passant des représentations intermédiaires par pas successifs, cette nouvelle architecture propose de chercher dans la langue source des formes correspondant à des fonctions cibles déjà connues et vues comme des candidates potentielles à la traduction. En d'autres termes, le système part plutôt de la langue cible et cherche ce qu'il sait traduire dans la source, ce qui rompt avec la traditionnelle marche avant des approches par transfert. Pour l'auteur, le fait que le pivot ne soit construit qu'à partir de règles identifiées en langue cible et sans regard sur le texte peut s'illustrer par un triangle modifié, dont le sommet du pivot (en haut) se retrouve décalé du côté cible (fig. 3.18). Il n'a pas vocation à être équidistant des deux langues, il est au contraire pleinement porteur de la « pensée signe ».

Si nous ne croyons pas qu'une traduction comparable à celle d'une production humaine soit à la portée d'un algorithme aujourd'hui, nous envisageons toutefois

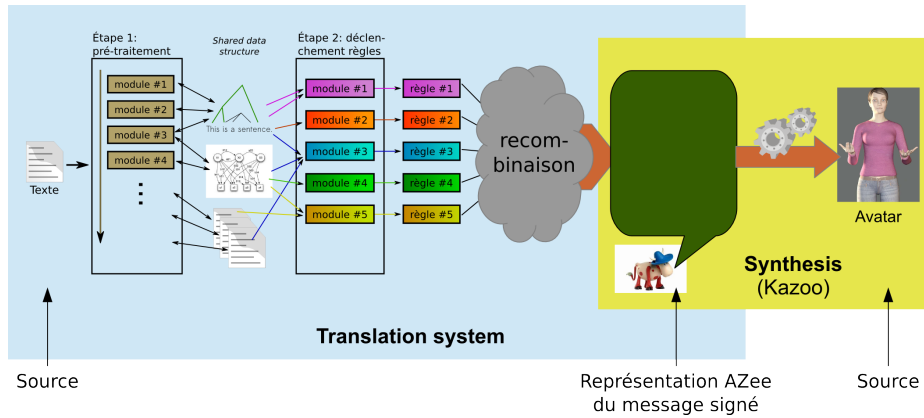


FIGURE 3.17 – Architecture globale de la proposition

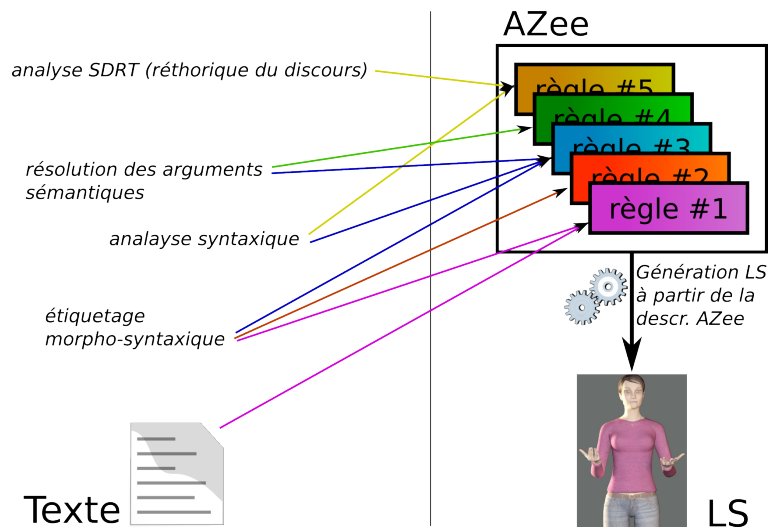


FIGURE 3.18 – Le triangle de Vauquois

des plateformes d'assistance au traducteur. Il serait possible par exemple de proposer des morceaux de traduction (ensembles de règles déclenchées et leurs arguments lorsque cela a été possible), dont la recombinaison serait la tâche du traducteur, comme le remplissage des éléments manquants ou la suppression d'éléments incorrects.

À la manière des systèmes de mémoires de traduction déjà en utilisation dans la profession, nous pourrions mémoriser les imbrications de règles choisies pour la traduction de morceaux de texte sélectionnés et ainsi accélérer la tâche de création de l'animation finale. Une telle application ferait selon nous le premier pas de l'outillage de la traduction numérique pour les langues des signes.

### 3.6 Conclusion et perspectives

Nous avons montré dans ce chapitre qu'il existe trois principales limites aux approches classiques. La première est que les corpus sont encore bien trop rares pour déployer aisément les méthodes basées sur l'apprentissage statistique. La seconde est la linéarisation des productions signées en une séquence de gloses qui représente une trop forte simplification, écartant certaines propriétés essentielles liées à la multi-linéarité. La troisième est la définition problématique de certaines notions comme l'unité lexicale ou la phrase, sur lesquelles reposent essentiellement les approches en question. Si nous pouvons espérer pouvoir collecter de grandes quantités de données dans un futur à déterminer, les autres aspects sont sans doute plus problématiques.

Dans le domaine de la modélisation et du traitement automatique, les LS sont souvent étudiées au travers du prisme des langues vocalisées, souvent même par celui des grammaires indo-européennes. Une raison est sans doute que les résultats dont nous disposons viennent presque tous de projets de recherche à la durée plutôt courte, alors que des travaux à plus long terme permettraient de mieux s'attaquer aux particularités des LS et à leurs propriétés essentielles. Cette situation est aussi vraie en linguistique pour toutes les langues peu décrites (Mettouchi, 2013). Il convient d'être vigilants quant aux méthodologies suivies pour l'analyse et la modélisation de telles langues, qui sont souvent la cause des limites rencontrées.

Les chapitres suivants présentent les logiciels connus à ce jour d'abord dans le domaine de l'enseignement lié aux LS (chap. 4), puis à visée grand public (chap. 5), dans lesquels les défis de modélisation linguistique présentés ici ont toute leur importance.