



HAL
open science

Modélisation, traitement automatique et outillage logiciel des langues des signes

Michael Filhol

► **To cite this version:**

Michael Filhol. Modélisation, traitement automatique et outillage logiciel des langues des signes. Informatique et langage [cs.CL]. Université Paris-Saclay, 2021. tel-03197108

HAL Id: tel-03197108

<https://hal.science/tel-03197108>

Submitted on 13 Apr 2021

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.

Mémoire

présenté pour obtenir

l'habilitation à diriger des recherches

de l'Université Paris–Saclay

Modélisation, traitement automatique et outillage logiciel des langues des signes

par
Michael Filhol

soutenu le 12/03/2021

Jury :

- Marion Blondel, chargée de recherche CNRS, Univ. Paris 8
- Gilles Falquet, professeur associé, Université de Genève (Suisse)
- Olivier Ferret (prés.), ingénieur-chercheur, CEA-LIST
- Sylvain Kahane (rapport.), professeur des universités, MoDyCo, Univ. Paris ouest
- Laurence Meurant (rapport.), chercheur qualifié, FRS-FNRS, Univ. Namur (Belgique)
- Pascale Sébillot (rapport.), professeure des universités, IRISA, INSA de Rennes

Mémoire

Table des matières

I	Introduction	3
II	Activité de recherche	5
1	Modélisation linguistique	6
1.1	Lexique et <i>Zebedee</i>	7
1.1.1	État de l’art précédent : la description « paramétrique »	7
1.1.2	Thèse de doctorat : remise en cause de l’approche	8
1.1.3	Naissance de <i>Zebedee</i>	9
1.2	Discours et <i>Azalee</i>	12
1.2.1	État de l’art précédent : rigidité des niveaux linguistiques et des frontières lexicales	12
1.2.2	Post-doc : ébauche d’une nouvelle approche et premières explorations	15
1.2.3	Extension <i>Azalee</i>	17
1.3	Approche « non nivelée » et formalisme <i>AZee</i>	18
1.3.1	« Dénivellation » par fusion des formalismes	18
1.3.2	Formalisation de la méthodologie d’extraction de règles	19
1.3.3	Des règles isolées au système récursif	21
1.3.4	Spécification du langage <i>AZee</i> et compilateur	23
1.4	Montée en puissance du modèle	24
1.4.1	La séquence comme forme	24
1.4.2	Expressivité en sens et précision des formes	26
1.4.3	Productivité et couverture	27
1.4.4	Impact et reconnaissance internationale	28
2	Synthèse par signeur virtuel	30
2.1	État de l’art précédent : la première génération	30
2.2	Apport d’ <i>AZee</i> pour une deuxième génération	31
2.3	Nouvel état de l’art : animation à granularité décroissante	32
2.4	Collaboration vertueuse TAL–synthèse–linguistique	33

3 Traduction	34
3.1 Corpus parallèle français–LSF	34
3.2 Traduction automatique	36
3.3 Traduction assistée par ordinateur (TAO)	38
4 Schématisation et lien avec représentation formelle	41
4.1 Déverbalisation, schématisation et régularités spontanées	41
4.2 Constitution d’un corpus de schémas	43
4.3 Représentation éditable et écritures	44
4.4 AZee comme piste pour une forme graphique éditable	46
III Projet	47
5 Grammaire formelle	47
6 Synthèse d’AZee	49
7 Traduction assistée	50
8 Schémas et forme graphique	51
Références	52

Première partie

Introduction

Langues à part entière, les langues des signes (LS) sont des langues *orales*, et non écrites, de *modalité gestuelle*, et non phonatoire. Leur histoire est liée à celle des sourds qui comme tout groupe d'humains ont naturellement mis à profit leur faculté de langage à leurs fins de communication. L'émergence des langues des signes apparaît ainsi comme un produit nécessaire et naturel de la condition humaine et non comme une compensation construite pour une communauté handicapée qui serait autrement privée de langage.

Aussi, comme pour nombre de minorités sociales, l'histoire des sourds – et donc celle de la langue des signes – est entachée de tumultes mêlant oppression, isolement, médicalisation, ainsi que questionnements et revendications identitaires, combats pour une reconnaissance, une inclusion, des droits, etc. Un siècle d'interdiction frappe par exemple la LS dans l'éducation en France ainsi qu'ailleurs en Europe suite au congrès de Milan (1880). Après des années de prise de conscience, de structuration et de lutte de la communauté signante, la langue des signes française (LSF) est reconnue légalement comme langue de France, et ayant notamment sa place pour l'accès des sourds à l'information et à l'éducation, depuis 2005 seulement.

Les LS ne se développent donc que depuis tout récemment comme objets d'étude scientifique, et on explique aisément qu'elles soient généralement *peu dotées*, c'est-à-dire que l'on ne dispose que de peu ou pas de ressources ou connaissances linguistiques telles que des corpus, des grammaires, des dictionnaires, etc. Ceci pose un premier problème pour ce qui est de leur étude puisque les données restant rares, elles sont souvent à construire en amont du travail de recherche. La recherche académique sur les LS et la consolidation d'une première communauté scientifique associée date seulement des années 1960 aux États-Unis. En France, ce n'est que peu avant les années 2000 qu'un réseau scientifique se stabilise.

Par conséquent, le domaine d'étude est jeune et différents modèles contradictoires co-existent encore sur plusieurs questions de description linguistique, pourtant élémentaires si rapportées à la linguistique générale. De nombreux consensus restent encore à trouver, et par ailleurs de nombreuses inconnues à identifier. Souvent, les oppositions se font entre deux tendances, l'une assimilatrice et l'autre plus différenciante entre les LS et les langues vocales/écrites, plus étudiées. Chacune de ces approches a engendré des pans entiers de recherche, lancé et nourri des discussions encore en cours pour beaucoup, et contribué à la création et à l'exploration de corpus au fil des années.

Les approches assimilatrices visent à identifier dans les LS les objets, concepts et opérations établis comme universels en langue par la linguistique traditionnelle. En effet, certains principes ont été éprouvés sur de nombreuses langues, se sont avérés robustes aux nouvelles considérées par la suite, et se sont donc imposés comme universels, voire sont devenus des éléments de définition pour caractériser les langues naturelles. On peut par exemple citer la notion de *signe linguistique* associant forme (signifiant) et sens (signifié) et dont le lien est réputé arbitraire (F. de Saussure), la *double articulation du langage* qui distingue le niveau des monèmes ayant un sens de celui des phonèmes [39], l'existence d'un ensemble d'unités dénombrables formant un *lexique*, ainsi que de règles de *syntaxe* pour en contraindre les arrangements en phrases ou syntagmes, etc. En supposant les théories justes vu leur caractère prédictif confirmé, accepter qu'une LS est une langue conduit à lui attribuer les propriétés universelles attendues. Selon ces termes, les contester conduirait en effet à remettre en question le statut de langue de la LS considérée. Ainsi, on a compilé des inventaires de signes qualifiés de « lexicaux », proposé des modèles « phonologiques », mis en lumière une double articulation

similaire à celle identifiée pour les langues vocales, dessiné des arbres syntaxiques pour expliquer des séquences lexicales...

Au contraire, les approches différenciantes remettent en cause l'application de principes établis uniquement sur la base de langues vocales et, faut-il l'admettre, souvent même par le prisme de leur modalité écrite ou d'une transcription. Les LS pourraient en effet être différentes par nature, et leur appliquer les modèles établis sans regard sur elles biaise a priori l'observation de ce qui constitue un objet d'étude potentiellement nouveau. Par exemple, l'expression de relations spatiales entre deux entités du discours se fait typiquement par une projection directe dans l'espace de signation, utilisé de façon continue, chaque main représentant par exemple une entité en jeu et leur position relative respectant la topologie de la relation exprimée. Ce type de construction ne fait pas l'objet de lexiques car leur variabilité est infinie, alors qu'elles s'insèrent pourtant aussi bien dans le flot du discours que les unités répertoriées comme lexicales, parfois même en proportion plus forte. La nature, lexicale ou non, de certaines unités pose même problème. L'existence d'une variabilité continue remet en question la possibilité de décrire convenablement la langue avec une liste de phonèmes discrets quels qu'ils soient. Par ailleurs, l'iconicité omniprésente en LS, c'est-à-dire la propriété pour un signifiant de ressembler au signifié, amène à questionner l'arbitraire du signe. Ainsi, en opposition aux thèses assimilatrices, on a aussi placé l'iconicité au cœur du système linguistique, proposé que la totalité des observables aurait une racine iconique [3, 4], ouvert l'inventaire des composants des syntagmes à des unités dites « non lexicales » ou « partiellement lexicales » [32, 33]...

Avec une vision assimilatrice, une phonologie est construite et un lien arbitraire peut être posé entre un signe et son sens, qu'il s'agisse d'un objet concret du quotidien ou d'un concept élevé en abstraction comme « logique » ou « liberté ». On montre donc ainsi que ce n'est pas du mime, collant au réel et sans capacité d'abstraction comme le proféraient les détracteurs des LS par le passé. On observe que les unités de base que sont les signes inventoriés s'alignent d'une certaine manière comme on forme des phrases avec les mots du dictionnaire, et on peut expliquer que des éléments sont reliés dans l'espace comme on accorde des mots dans une phrase. En montrant des phénomènes en langues des signes similaires à ceux des langues vocales, on justifie dans le fond de leur décerner le même statut. Ces arguments ont sans nul doute contribué à l'acceptation de la langue des signes comme langue et, donc, à son entrée dans le champ de la linguistique. Peut-être une étape nécessaire.

Une fois ce statut accepté, on peut observer plus ouvertement que des unités iconiques sont bien créées au fil du discours de manière extrêmement productive sans être jamais inventoriées a priori. Elles sont parfois même proches du mime mais restent linguistiquement régies et ne peuvent être entièrement improvisées. Elles sont efficaces pour les relations spatiales et la description de formes notamment, mais l'utilisation symbolique de l'espace est tout aussi récurrente pour les concepts plus abstraits. L'iconicité fait donc véritablement partie intégrante de la langue et devait être prise en compte. Cette position différenciante permet de prouver que l'iconicité est structurante et puissante, pas un phénomène dégradant ou insignifiant. Non seulement la LS est une langue, mais elle a une valeur propre, à ajouter au paysage linguistique connu, intimement liée à la pensée visuelle, donc à l'identité sourde et il convient d'en revendiquer les particularités plutôt que de regretter les écarts qu'elle présenterait avec les langues étudiées. Indéniablement, le « réveil sourd » des années 1970 a à la fois participé à forger et bénéficié de ce discours scientifique.

Ainsi chaque approche a-t-elle contribué au positionnement socio-politique de la LS et de la communauté sourde, et tant mieux. Mais si on accepte partout en science que des modèles non-unifiés co-existent lorsqu'ils présentent des domaines d'application disjoints ou prédisent les mêmes effets là où ils se recoupent, on ne peut logiquement considérer que des fondements opposés puissent être valides en même temps dans le même

champ d'application. Et poussées à l'extrême, les deux positions présentées sont irréconciliables. Vu l'histoire tumultueuse mentionnée plus haut et le déni historique du statut même de langue naturelle des LS, il me semble qu'on peut comprendre (et accepter) que la science ait avancé des arguments pour rationaliser des positions à caractère, au fond, politique. Mais aujourd'hui, bénéficiant d'au moins deux décennies de présence conjointe de ces approches, je pense qu'un certain recul est devenu possible pour décrire les LS formellement dans un contexte moins sociologiquement tendu.

C'est ce que j'ai entrepris en début de carrière, et dont une grande partie sera présentée dans ce document. Il m'a paru essentiel de proposer un modèle formel adapté à l'utilisation en informatique de ces langues naturelles, qui prenne en compte leurs dimensions orale – et même exclusivement orale donc fondamentalement non linéaire – et iconique dans l'espace gestuel. Sans me ranger a priori derrière une approche théorique, je considère, pour commencer, les *deux* propositions suivantes vraisemblables à la fois :

- les langues humaines ont des caractéristiques fondamentalement communes qui font précisément qu'on les regroupe sous un nom commun (dans leur genèse, leur fonction, leur utilisation, leur évolution...), donc les LS ont par essence des points communs avec les autres langues ;
- le canal gestuel/visuel dans lequel les LS évoluent – c'est-à-dire le médium supportant toute les réalisations dans ces langues – est *physiquement* différent des autres, donc les LS ont par nature des différences avec les autres langues.

Dans une première partie consacrée à mon activité de recherche, je présente mes réflexions et contributions scientifiques, en respectant en général la chronologie, celle-ci rendant bien compte de mon parcours et du recul progressif pris sur mon domaine. Celles-ci seront traitées en quatre sections, représentant le panel des principaux thèmes de mon activité telle que je la considère aujourd'hui. D'abord, la modélisation linguistique de la LSF et la représentation formelle d'énoncés (§1), thème avec lequel j'ai commencé en thèse et resté le fil rouge de tout mon travail. Ensuite, la synthèse par avatar 3D des représentations proposées, thème toujours envisagé de pair avec le précédent, et nous verrons comment les deux ont profité l'un de l'autre. Puis, l'outillage de la traduction, aussi bien automatique qu'assistée, plus récemment investie mais d'intérêt majeur pour l'outillage des langues, et presque fondateur du traitement automatique des langues au milieu du siècle dernier. Enfin, la représentation graphique pour l'édition logicielle de contenus en LS, un thème menant à l'exploration des écritures du monde et à proposer une piste pour celle de la LS. Une deuxième partie du mémoire présente ensuite quelques projets pour la suite de mes recherches.

Note subsidiaire Ce document fait état de mon parcours et de mes réalisations scientifiques en matière de recherche. Ceci étant, j'entretiens également une participation à l'enseignement et à la vulgarisation par le biais de conférences et rencontres, auprès d'étudiants et du public. En particulier, j'ai récemment coordonné l'écriture et monté une vidéo¹ pour la fête de la science 2020, ayant eu lieu exclusivement en ligne cette année, expliquant la langue des signes et les activités de l'équipe M&TALS² dont j'ai pris la responsabilité en 2019. L'objet de cette note est de la proposer au lecteur novice en langue des signes, celui-ci pouvant y trouver une utilité directe avant d'entamer le contenu qui suit.

1. <https://le-village-des-sciences-paris-saclay.fr/les-langues-des-signes-et-si-les-ordinateurs-savaient-les-comprendre/>
2. Modélisation et Traitement Automatique des Langues des Signes

Deuxième partie

Activité de recherche

Mon domaine de recherche est le traitement automatique de la langue (TAL) appliqué aux langues des signes. Il se situe donc en informatique tout en présentant une intersection importante avec la linguistique (sciences du langage). Cette dernière est d'autant plus importante que nous avons vu, dans le cas des LS, que les descriptions et formalisations linguistiques sont rares, non consensuelles, et incomplètes. Les efforts sérieux en TAL sur les LS deviennent donc rapidement impossibles sans un recul et des questionnements, et donc une contribution, linguistiques. Nous verrons comment, au fil de mes travaux, j'ai constamment tenu non seulement à proposer des méthodes ou modèles compatibles avec leur utilisation dans un programme, mais aussi avec une lecture linguistique des objets proposés, en analogie ou en contraste avec les définitions connues sur ce terrain.

Aussi, cette symbiose linguistique–TAL s'est déjà illustrée dans les modalités écrite puis vocale de la langue. L'informatique pousse à la formalisation non ambiguë des phénomènes, ce qui comme toute mathématisation en science, permet une utilisation prédictive des modèles, et en principe de progresser en les invalidant ou en découvrant de nouvelles explications. Le travail interdisciplinaire présenté dans ce document participe, je l'espère, à un bénéfice réciproque comparable pour la modalité signée.

Comme le TAL porté sur les modalités écrite ou vocale, le TAL en LS recouvre plusieurs domaines liant la LS aux modèles et applications informatiques : synthèse (génération automatique par un signeur virtuel), reconnaissance, traduction... Mes recherches ont principalement porté sur deux domaines en particulier : la modélisation des LS pour la synthèse d'une part, et la traduction de texte à LS d'autre part. Plus récemment, j'ai orienté une part de mon travail vers la question des représentations graphiques éditables, que je présente également dans cette section.

1 Modélisation linguistique

La communauté s'est d'abord efforcée de caractériser et décrire un niveau lexical d'analyse des LS. Du côté linguistique, ont été compilés des inventaires de signes pour former des « dictionnaires » à partir de vignettes (cf. fig. 1), aujourd'hui souvent remplacées par des vidéos et présentées en ligne. Par analogie assimilatrice aux séquences de mots des phrases écrites, on décompose et transcrit le discours en LS comme une suite de signes (unités lexicales) plus ou moins figés mais dont un inventaire (lexique) est a priori possible. En TAL, on conçoit des systèmes de représentation et d'animation basés sur de telles séquences que l'on synthétise en concaténant (donc en générant une coarticulation entre) les unités qui les composent. Ainsi, à défaut de meilleurs modèles, la langue des signes a peu été étudiée autrement que comme un alignement de signes qualifiés à un niveau lexical.

Dès lors qu'on se dote d'un inventaire d'unités réputées lexicales dont la vocation est de construire des énoncés séquentiels, deux questions importantes émergent :

- de quelle façon, avec quel modèle décrire une unité lexicale isolée, par exemple pour la synthétiser ?
- comment agencer ces différentes unités en séquence pour obtenir des énoncés complets et bien formés ?

Dans une lecture par niveaux de construction du langage, ces questions visent les niveaux morpho- ou phonologique (la distinction restant à spécifier) pour la première, et syntaxique pour la deuxième.

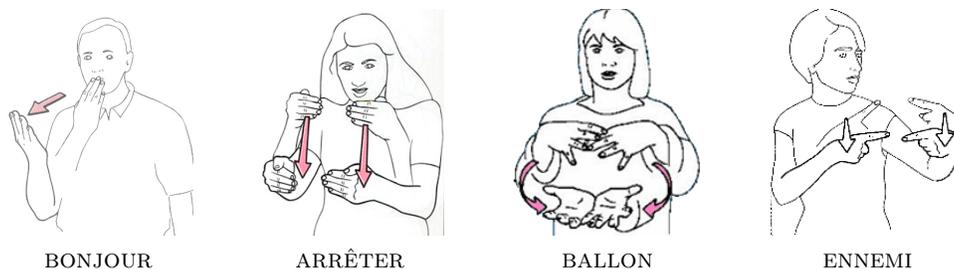


FIGURE 1 – Exemples de vignettes de signes dans un dictionnaire [45]

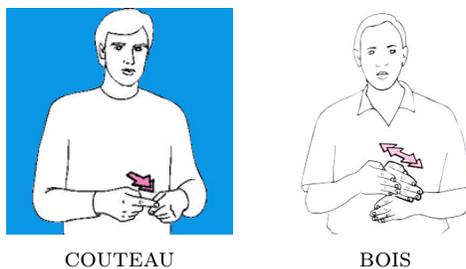


FIGURE 2 – Paire minimale par la « configuration manuelle »

1.1 Lexique et *Zebedee*

Cette section se consacre à la première de ces questions, qui a mobilisé la plupart des recherches de la communauté linguistique à ses débuts sous forme académique (Stokoe dans les années 1960). Elle m’a aussi mobilisé pour mon travail de thèse.

1.1.1 État de l’art précédent : la description « paramétrique »

Les signes sont supposés d’emblée comme les unités de première articulation. Dans la mouvance assimilatrice présentée plus haut, leur description s’est faite en cherchant un équivalent aux unités de seconde articulation, à savoir des unités minimales discrètes se combinant pour former les signes et non interprétables isolément. Ces unités revêtent alors a priori un caractère phonologique, comparable au phonème. Pour en établir la liste, le principe fut d’identifier des *paires minimales* de signes, c’est-à-dire des paires de signes perçus comme distincts et leur différence la plus petite possible. Par exemple, en définissant un paramètre « configuration manuelle », et en considérant « index » et « main plate » comme deux valeurs atomiques distinctes pour ce paramètre, les unités « COUTEAU » et « BOIS » en LSF forment une paire minimale car elles ne varient que par la configuration manuelle (fig. 2).

Ces recherches ont conduit à proposer puis augmenter et stabiliser l’idée d’une représentation du signe en 4 *paramètres* manuels, à savoir la configuration, l’orientation, l’emplacement, le mouvement des mains. Le signe est donc présenté comme un n-uplet de valeurs prises chacune dans un ensemble prédéfini pour le paramètre qu’elle instancie, et leur conjonction spécifiant le signe. Alors, à l’image des phonèmes en langue vocale :

- changer de signe conduit à nécessairement modifier la valeur d’au moins un paramètre ;
- changer un paramètre conduit a priori à changer de signe (que la nouvelle combinaison fasse ou non partie du lexique).



FIGURE 3 – Configurations manuelles « main plate » et « angle droit »

Cette représentation s’est imposée dans toute la littérature, et a toujours principalement cours aujourd’hui. Même l’approche très différenciante de Cuxac sur la question phonologique la suppose en partie [3] : il conteste certes la valeur de phonème non interprétable aux paramètres, et va jusqu’à proposer une hypothèse du tout morphémique [4]. Mais l’inventaire des valeurs morphologiques qu’il dresse part du même découpage a priori : configuration, emplacement, etc.

On note tout de même une différence avec la phonologie classique : ces paramètres sont réalisés de manière simultanée. Les phonèmes dans les langues vocales sont principalement pensés comme des unités mises en séquence, alors que les paramètres de Stokoe supposent une simultanéité par construction. Pour éviter la confusion, Stokoe proposa l’appellation de *chèreme* pour les valeurs paramétriques. Mais à cette simultanéité près, c’est bien un statut phonologique qui est globalement attribué à ces unités.

1.1.2 Thèse de doctorat : remise en cause de l’approche

Mes débuts en thèse ont été motivés par l’observation de nombreux problèmes du modèle à paramètres face à plusieurs situations.

Premièrement, les paramètres sont principalement manuels, les mains étant vues comme l’articulateur privilégié a priori pour porter les unités lexicales. Aucune possibilité n’est offerte, même à la marge, pour décrire les articulations des coudes ou des avant-bras par exemple, alors qu’ils peuvent être les seuls actifs, donc bien porteurs du signe entier. Le modèle est donc d’emblée incomplet.

Un flou est aussi entretenu sur la dualité des mains. On parle « du » paramètre de configuration par exemple alors qu’un signe bimanuel peut en faire intervenir deux différentes. Le cas échéant, on dédouble alors le paramètre mais ceci se fait à la carte sans réflexion sur le statut de cette opération du point de vue formel. Changer une configuration identique sur les deux mains constitue-t-il une paire minimale ou a-t-on changé deux valeurs ?

Par ailleurs, la liste des valeurs inventoriées pour chaque paramètre est réputée exhaustive, alors qu’on trouve toujours des valeurs partielles ou des combinaisons. Par exemple, les configurations « main plate » et « angle droit » de la figure 3 sont tantôt interchangeables (c’est le cas dans « BONJOUR », cf. fig. 1), tantôt discriminantes (angle droit obligatoire sur la main dominée dans « ARRÊTER »). Le cas de « BONJOUR » est un classique : toute configuration est en réalité acceptable pourvu que les doigts (pouce exclus) soient droits (deux articulations distales en extension pleine) et en contact. De ce fait, aucun consensus n’a jamais vu le jour sur le nombre de configurations d’une LS donnée car leur séparation en valeurs distinctes est trop souvent problématique, irrégulière, ou exceptionnelle. La perception catégorielle est même une question ouverte.

Le paramètre « mouvement » souffre également d’une définition approximative : on y range en gros toute la dynamique du signe. Ceci pose plusieurs problèmes. D’abord, si M_1 et M_2 sont deux valeurs possibles, la réalisation « M_1 suivi de M_2 » est nécessairement un mouvement possible aussi, et l’idée d’une exhaustivité des valeurs paramétriques devient donc caduque. Par ailleurs, certains mouvements, dits « secondaires »

sont composables de manière indépendante et simultanée (p. ex. un pianotement des doigts). Ensuite, on y range aussi les « mouvements d'ouverture » ou « de fermeture » de la main, ce qui a pour effet de modifier la configuration. La valeur du paramètre modifié doit-elle alors être la configuration initiale ? finale ? Les déplacements modifient de même l'emplacement, les rotations l'orientation, etc. On voit donc d'une part que la dynamique, intégralement rangée dans un paramètre à effet puits, est assez mal intégrée au modèle, et d'autre part que les valeurs sont fortement interdépendantes voire certaines dérivables des autres.

Ainsi le caractère minimal, couvrant, indépendant, nécessaire et discret de ces valeurs, posés lors de leur définition ou mis en avant pour favoriser l'analogie phonologique, est en réalité très largement discutable. En particulier pour une utilisation en informatique, le modèle global est donc fortement incohérent.

De plus, certaines opérations typiques en LS comme la relocalisation de signe sont rendues très difficiles par le choix, imposé par le modèle, de valeurs discrétisées dans l'espace paramétrique. Par exemple, l'emplacement canonique pour le signe « ENNEMI » (représenté fig. 1) est : chaque main de son côté dans « l'espace neutre » ; l'orientation : index droit tendu vers la gauche, et le gauche vers la droite. Or lorsqu'un des opposants mis en jeu, iconiquement représentés par les mains, est ancré ailleurs dans l'espace, l'emplacement de la main correspondante doit changer. Cet exemple montre une dépendance du signe au contexte de la signation, ici l'ancrage spatial d'une entité du discours, avec lequel il peut en quelque sorte s'accorder. Le problème est que tous les signes ne se modifient pas de la même manière selon le signe, les emplacements représentant des choses variées et plutôt propres à chaque signe.

Ce n'est pas tout, suite à une telle relocalisation, l'orientation de chaque main dans la réalisation « ennemi » doit également changer pour que les doigts restent orientés selon des vecteurs de même direction et de sens opposés. Il faut donc changer trois valeurs paramétriques (sur 8 en comptant deux mains, et sans compter les mouvements qui pourraient être considérés comme affectés également), alors qu'un unique effet – appelons-le grammatical pour l'instant – y est appliqué au départ. En listant à plat les valeurs paramétriques dès la description du signe (hors contexte) au lieu d'en conserver les dépendances internes, l'approche ne peut s'adapter aisément aux variations dans les discours, alors qu'elles seraient invariantes et réactives aux changements contextuels si elles étaient spécifiées en tant que telles, par exemple : « orientation index main droite = vers main gauche ». Le modèle était donc trop rigide pour une utilisation dynamique en synthèse avec une couverture suffisante des phénomènes.

1.1.3 Naissance de *Zebedee*

Ma thèse a donc consisté à proposer une nouvelle approche permettant la synthèse de signes en contexte, sans sur- ni sous-spécification, tout en restant linguistiquement pertinente. J'ai pour ce faire développé plusieurs idées pour la représentation de signes, intégrées ensuite en un modèle formel baptisé *Zebedee*.

D'abord, **tout articulateur peut participer au signe**, qu'il soit manuel ou non, cette distinction n'ayant pas de pertinence a priori. Ainsi le signe « ARBRE » fera participer l'avant-bras pour l'orienter et le coude pour le placer, « ÉCOSSE » le bras entier pour effectuer le mouvement caractéristique du signe, etc. La granularité de description est celle de la réalisation, à savoir que chaque phalange, le dos, les épaules, la tête, etc. peut voir ses degrés de liberté contraints indépendamment ou conjointement à d'autres.

Ensuite, la spécification d'un signe doit comporter tout ce qui ne peut pas être omis pour une réalisation acceptable, mais exclure tout ce qui n'est pas obligatoire à celle-ci. Une description est donc la donnée d'un ensemble de **contraintes nécessaires et suffisantes**, toute contrainte exprimée devant être nécessaire et leur ensemble devant être suffisant. Ainsi, on contraindra les doigts dans « BONJOUR » pour les mettre en

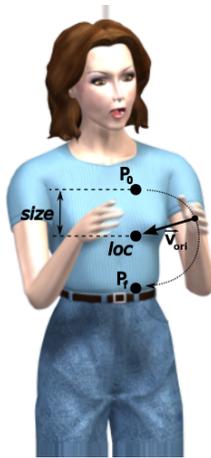


FIGURE 4 – Objets géométriques et dépendances contextuelles pour « BALLON »

contact et dans un même plan (sans quoi la réalisation ne sera pas acceptable), mais pas l'angle formé entre le plan des doigts et celui de la paume, celui-ci variant sans incidence sur son acceptabilité. On contraindra de même d'orienter l'intérieur des doigts vers le signeur, mais pas la direction de pointage des doigts précisément, celle-ci étant finalement libre et, en fait, fonction de facteurs externes comme la hauteur du coude au moment de la production du signe.

Troisièmement, l'espace de signation est à considérer comme un espace géométrique affine, et un signe comme un enchaînement de postures dans celui-ci. Sa description est donc la spécification d'une **figure géométrique dynamique**, dont les contraintes géométriques s'expriment avec les éléments standards de la géométrie euclidienne (points, vecteurs, symétries, rapports de distances, etc.). Par exemple, on peut désigner le milieu de deux points identifiables, comme l'extrémité de chaque index, pour ensuite y diriger le regard. Ainsi, l'espace des descriptions est continu, et les valeurs non discrètes et en nombre infini. Ceci couvre les dépendances internes entre articulateurs du signe comme l'orientation des index dans « ENNEMI », ici celle du regard sur la position des index.

Une idée supplémentaire est de permettre l'expression des objets géométriques non seulement à partir de points du corps (dépendances internes vues ci-dessus), mais sur des **objets contextuels** variables, non connus au moment de la description mais fournis en contexte lorsque le signe est réalisé (dépendances externes). Par exemple, un pointage contraint la direction de l'index vers sa cible, qui est un point du même espace, mais celle-ci n'est connue qu'à la réalisation du geste, non lors de la description générique de l'entrée « pointage ». Ainsi la spécification de « pointage » présentera un *argument* « cible » dont on suppose que la valeur sera fournie une fois le signe en contexte. De même, la réalisation de « BALLON » (fig. 1) peut être modifiée en fonction de sa taille et de sa position (« size » et « loc » sur la figure 4), un immeuble de sa hauteur entre autres, etc. On voit là comme on encapsule, dans la figure géométrique qu'est devenue le signe, les premiers éléments d'iconicité dans les descriptions. Ce faisant, grâce à leurs arguments, les entrées deviennent des *fonctions* du contexte, leur forme s'exprimant en fonction d'éléments contextuels supposés disponibles pour pouvoir être réalisée.

Ces propositions résolvait la plupart des problèmes identifiés plus haut. Mais l'aspect dynamique du signe restait en partie non réglé puisqu'il devenait nécessaire de fournir pour chaque mouvement une fonction du temps décrivant la variation dans son intervalle. Or la liberté totale de description permise par une telle

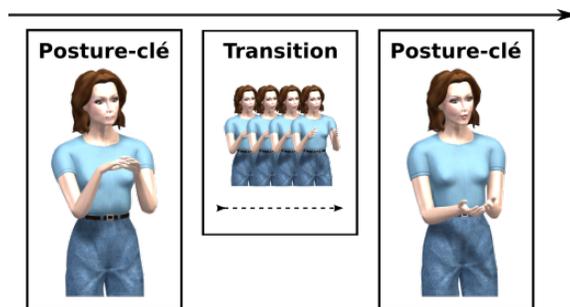


FIGURE 5 – Structure temporelle de « BALLON » en *Zebedee*

approche rendait le système beaucoup trop libre formellement, car autorisait une infinité de mouvements asynchrones physiologiquement impossibles à coordonner. La découverte de la proposition de Liddell et Johnson [38] et la rencontre avec ce dernier au cours de ma thèse puis en première année de post-doc à Gallaudet m’apporta une solution, que l’on verra insuffisante par la suite mais dont je me suis inspiré pour compléter *Zebedee*.

Leur approche découpe le signe en une suite d’unités temporelles, alternant des unités statiques et des unités dynamiques. En résumé, les unités statiques spécifient des postures motivées, et les dynamiques couvrent le temps pendant lequel le corps passe d’une posture à la suivante. Les premières sont les moments où les trois paramètres configuration, orientation et mouvement prennent une valeur définie. Qu’elles soient stables (posture avec durée) ou transitoires, ces moments de rendez-vous synchronisent les articulateurs et rythment les signes d’une manière simple et efficace. En intégrant une structure temporelle à base de « postures clé » et de « transitions » inspirée de cette proposition de Liddell & Johnson, *Zebedee* put finalement adopter un schéma plus contraignant, et limiter la synchronisation « à la main », trop libre car de trop bas niveau de contrôle.

Par exemple, « BALLON » comporte une posture-clé initiale où les mains se trouvent au point nommé P_0 sur la fig. 4, une finale avec les mains en P_f , ainsi qu’une transition entre les deux pendant laquelle un arc est appliqué à la trajectoire de chaque main (le vecteur $\overrightarrow{v_{ori}}$ orienté vers *loc*, emplacement du ballon, à tout instant). Cette structure temporelle à trois unités, représentée figure 5, permet de faire évoluer les articulations dans le temps (axe horizontal) entre postures où ils se synchronisent obligatoirement (axe vertical).

Après ma thèse [13], *Zebedee* a un peu évolué [14], mais seulement à la marge. J’ai également développé un éditeur pour permettre d’outiller les travaux de description, et environ 2000 signes ont par la suite été décrits avec *Zebedee* grâce aux ingénieurs d’études Nadège Devos et Flora Badin sur le projet DictaSign (2009–2012) [8], formées au modèle et encadrées par mes soins. Ces réalisations ont permis d’évaluer quantitativement la couverture du modèle (aucun signe ne résistant à la description) et qualitativement sa prise en main par l’humain. En parallèle, *Zebedee* était également testé du point de vue de l’animation avec la thèse de Maxime Delorme dont j’étais l’encadrant scientifique (cf. §2.2).

Avant d’aller plus loin, soulignons que *Zebedee* n’a pas intégré le paysage des descriptions linguistiques des signes sans contraste. Il entrait même directement en opposition avec certains fondements des modèles précédents. Ceci est vrai à plusieurs titres, mais les deux principaux sont à mon avis les suivants :

- les « paramètres manuels » sont rejetés en tant que notion nécessaire et structurante : on ne définit (et synchronise) que les contraintes articulatoires nécessaires et suffisantes, quelle que soit leur granularité
- les paramètres manuels sont la plupart du temps sous-spécifiés ;

- les unités décrites spécifient chacune leur manière d’être utilisée en contexte (en invoquant des « dépendances externes » de types libres et en nombre a priori illimité), ce qui du point de vue formel autorise chaque signe à se faire fonction d’éléments anticipés venant du contexte.

Ce deuxième point peut poser problème au premier abord puisqu’il engendre un effet de « cas par cas », au sens où nombre de variations ne pourront être prises en charge par des règles génériques qui supposeraient un format unique aux descriptions lexicales. Mais pour l’approche Zebedee, cela est précisément une représentation fidèle de l’objet linguistique décrit. Par exemple, changer la taille d’un ballon, d’une boîte ou d’un bâtiment/immeuble fait l’objet d’autant d’opérations ad hoc, chaque signe définissant ses propres dépendances au contexte. Agrandir un ballon affecte nécessairement et uniquement son rayon, et par conséquent les positions de départ et d’arrivée. Au contraire, agrandir un bâtiment/immeuble peut se faire sur deux dimensions à la fois ou séparément, à savoir la largeur de la base et la hauteur, et changer la hauteur change les positions d’arrivée des mains à la réalisation, mais pas les positions de départ. Écrire des règles génériques à un niveau supérieur du système linguistique pour modifier les réalisations des signes est donc impossible sauf à distinguer les signes nommément. Le cas-par-cas est donc pour Zebedee moins un problème de non-généricité qu’une propriété essentielle des éléments décrits, provenant au fond de l’iconicité de chaque signe considéré individuellement.

Une conséquence remarquable est que les signes deviennent fonction de valeurs contextuelles, provenant donc du niveau sémantique. Ils ne sont plus des unités décrites a priori et sans considération de sens, disponibles pour ne construire le sens/discours qu’ensuite. On voit ici de nouveau s’imposer la remise en question de l’arbitraire du signe.

Cela pose en revanche le problème de savoir comment combiner ces entrées pour construire le discours. C’est la deuxième question posée plus haut (p. 6), à laquelle je me suis attaqué dans la suite de mon travail.

1.2 Discours et *Azalee*

Après Zebedee et des descriptions lexicales capables de varier selon le contexte, la question posée est celle du modèle représentant l’organisation du discours signé, dont la communauté manquait toujours. Même avec Zebedee, on ne décrit qu’un ensemble d’unités ayant vocation à être utilisées/combinées par un modèle de niveau supérieur, supposé les agencer sur un axe temporel pour former un discours dépassant l’unité isolée, et de fait aussi rendu responsable de fournir les valeurs de leurs arguments (dépendances contextuelles). Mais celui-ci restait à construire. C’est à ce problème que j’ai ensuite étendu ma recherche, attaquant ce faisant la deuxième question posée p. 6, identifiée alors comme syntaxique.

1.2.1 État de l’art précédent : rigidité des niveaux linguistiques et des frontières lexicales

L’approche standard alors – et on peut dire la principale aujourd’hui encore – est de supposer que des unités (les « signes ») se trouvent mises en séquence sur l’axe temporel à la manière des mots composant les phrases à l’oral. On suppose l’existence de telles phases, on fait des recherches pour en déterminer les frontières, on donne des catégories grammaticales aux signes (verbe, nom...) et des critères pour les distinguer, on suppose que des règles syntaxiques gouvernent les suites possibles, on discute de « l’ordre canonique de la phrase » (est-il SVO ? SOV ?), etc. On avance aussi des critères sémantico-cognitifs pour ordonner les éléments, par exemple « temps, lieu, action »... Et toutes ces propositions postulent ou supposent un ordre, une suite, une séquence à expliquer. Deux propriétés sont néanmoins relativement souvent reconnues, contrastant avec ce que serait une grammaire parfaitement linéaire.



FIGURE 6 – Description d’une scène : déploiement de pans de murs selon une forme complexe

La première est l’existence d’**unités non inventoriées** a priori, car hautement variables et non figées, au sein de la séquence. Par exemple, le déploiement de la forme des murs entourant une pièce particulière comme dans la figure 6. Les annotations spontanées en linguistique autant que les guides d’annotation publiés font de chacune de ce genre de production un élément de la séquence annotée, leur conférant ainsi une place « rythmiquement », distributionnellement lexicale, au sens où elles s’agencent dans la séquence syntaxique chacune en un bloc au même titre que les unités lexicales figées. Mais l’étiquette attribuée ressemble moins à celle d’un élément dans un ensemble (comme « ARRÊTER » est un élément d’un lexique maîtrisé), qu’à celle d’un ensemble d’unités possibles (comme « transfert situationnel » chez Cuxac ou « partly lexical sign » chez Johnston). Aussi, sa composition interne fait parfois elle-même apparaître une sorte de séquence (par exemple les différents pans et sections de mur fig. 6), et sa traduction prend en général plusieurs phrases lorsque traduite en langue vocale. Autant de raisons qui leur confèrent plutôt une qualité de syntagme composé. Ainsi, malgré l’intuition forte de la communauté quant à l’existence d’une part d’unités annotables en séquence, et d’autre part d’une organisation syntaxique qui en contraindrait la combinaison dans le discours, on ne sait pas dire si ce genre d’unité relève du premier type ou du second. Et ces unités peuvent être nombreuses en proportion dans le discours.

La deuxième propriété observée en langues des signes contrastant avec les approches linéaires traditionnelles est la **simultanéité nécessaire** de certaines articulations ayant une valeur grammaticale forte reconnue, c’est-à-dire non reléguée à un supposé canal coverbal, prosodique ou non linguistique. Une expression du visage comme une levée ou un abaissement des sourcils peut par exemple porter sur des unités en place dans la séquence pour donner une valeur interrogative à un énoncé. Ainsi, une simple séquence d’unités est insuffisante pour entièrement caractériser le discours à signer. Une certaine place *doit* être donnée à la simultanéité.

En général, on considère une activité manuelle à laquelle la plupart des effets lexicaux sont attribués, et une activité dite « non-manuelle » en parallèle. Cette dernière a pour but de compléter le signal manuel/lexical avec les éléments grammaticaux indispensables et non attribuables au niveau lexical. Le système de représentation HamNoSys par exemple [43, 29], permet pour chaque unité mise en séquence de spécifier, en plus de sa section manuelle la plupart du temps remplie, une section non-manuelle pouvant contenir une liste de mouvements choisis parmi une liste de primitives existante (positions des lèvres, des sourcils, gonflement des joues...).

Cependant, cette modification correspond en réalité plus à l’ajout d’un paramètre aux unités lexicales, fut-il non-manuel, qu’à celui d’une piste parallèle pourtant nécessaire. En effet, l’étalement temporel de ce paramètre est alors toujours celui de l’unité lexicale qui le contient. Si le système JASigning [11], reconnu de facto comme logiciel d’animation des entrées HamNoSys, a permis ensuite de décaler les éléments non-manuels de la piste manuelle, la séquence reste celle des unités lexicales sans recouvrement possible entre elles. Pour permettre des articulations non-manuelles sur des séquences plus longues, JASigning maintient entre

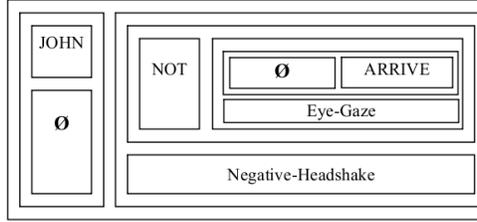


FIGURE 7 – Structure P/C (P dans l’axe vertical, C dans l’axe horizontal)

deux signes consécutifs celles qui sont contenues dans les deux. Mais du point de vue du modèle linguistique sous-jacent, ceci n’est guère satisfaisant puisqu’il spécifie des répétitions là où il faudrait rendre compte d’un geste unique, et il n’est plus possible de distinguer entre une articulation répétée sur chaque signe d’une séquence et une autre étalée sur la séquence entière.

Une formalisation plus générique fut proposée par Huenerfauth : « partition/constitute », ou « P/C » [31]. Dans ce système, une articulation à étaler sur une séquence de signes (un ou plusieurs) peut être mise en parallèle de celle-ci. Le « partitionnement » en deux pistes est opéré par un nœud de type « P », et la « constitution » d’une séquence d’éléments par un nœud « C ». Récursivement, on peut imbriquer parallélisations et sérialisations pour rendre compte d’articulations simultanées autant que séquentielles (voir fig. 7). La simultanéité est donc prise en compte au premier chef. Cela dit, les frontières des unités « constituées » sans « partition » restent les seuls instants repérables pour organiser le discours. Fondamentalement, une séquence lexicale est donc toujours au cœur de l’organisation du discours, et sa granularité temporelle est toujours celle des unités qui la constituent.

Or, en annotant les mouvements des articulateurs séparément, on observe en général que les bornes de nombreuses activités (p. ex. non-manuelles) ne coïncident pas nécessairement avec celles des signes ou des autres partitions. Par exemple, lorsque le regard accompagne le positionnement d’une main dans l’espace, on observe qu’il le précède systématiquement. Lorsqu’une seconde main se déplace ensuite par rapport à la première installée comme repère, on observe que la première s’installe toujours avant que ne commence la seconde, et que si le regard peut décrocher de la cible à tout moment avant la fin du mouvement manuel, les deux mains sont nécessairement relâchées au même moment. Le type de synchronisation du type prévu par P/C est en réalité assez rare, et les écarts entre les bornes ne sont pas constants.

P/C a introduit des « null » nodes dans les séquences constituées pour décaler une borne sur l’axe temporel en insérant des intervalles vides. Mais alors, si les temps d’attente entre les différentes articulations deviennent significatifs, quid des transitions entre les unités « constituées » comme celles de la séquence lexicale ? Ces intervalles sont supposés tous équivalents, conceptuellement identiques à l’espace entre les mots écrits, et la concaténation est supposée être une opération suffisante pour modéliser le discours. Or l’observation montre là aussi des durées variables.

Il m’a ainsi semblé que dès qu’on s’attache à rendre compte de tous les phénomènes articulatoires signifiants en LS, on butte contre les conséquences de ce qui a en réalité été posé a priori sans autre raison qu’un emprunt à des langues d’une autre modalité : la séquence lexicale. J’identifiai donc un besoin de synchronisation plus libre et plus fine, permettant le partitionnement à la P/C mais évitant l’accumulation de null nodes arbitraires. Celle-ci devra coexister avec celle plus contraignante intégrée à Zebedee, identifiée par Liddell et Johnson et reposant sur les « points de rendez-vous » de certaines articulations évoluant en parallèle.



FIGURE 8 – Montage 3 vues du corpus créé dans le projet *DictaSign*

1.2.2 Post-doc : ébauche d’une nouvelle approche et premières explorations

Pour appréhender la LS sans supposer d’organisation en séquence ni définir de catégories ou de contraintes syntaxiques a priori, j’ai voulu revenir à un cadre linguistique adoptant le moins de suppositions possibles, quant à la concaténation d’unités d’une part, mais aussi quant à l’existence-même de niveaux hiérarchisés, de distinction manuel–non-manuel, etc. Ainsi, l’unique principe posé au départ pour structurer toute la suite de mes travaux repose sur le fait qu’une est langue est le système que partagent ses locuteurs, permettant une correspondance entre :

- des *formes articulées*, que l’on peut produire en tant que locuteur, ou observer et annoter – p. ex. en LS la direction du regard, les mouvements des épaules, du buste, la durée d’une pause, les rotations de la tête, la position des mains, la flexion d’un doigt ;
- et des *fonctions sémantiques*, interprétées à la réception du message – p. ex. la datation d’un événement, la séparation de deux parties d’un discours, le concept de « maison » ou de « mathématiques », l’intensification d’un jugement ou d’une quantité, la marque d’objectivité d’une mesure.

L’objectif est alors d’identifier empiriquement, à partir de données attestées de LS (corpus de locuteurs dont la première langue est la LS), les associations systématiques entre les deux (forme et sens), et de les formaliser. Toute fonction interprétée systématiquement pour un critère de forme donné constitue pour nous une *règle d’interprétation*. Inversement, la description des invariants de forme dans les occurrences d’une fonction donnée permet la formalisation d’une *règle de production*. Mon but, du moins à ce stade, étant en particulier la synthèse et non la reconnaissance automatique, ce sont ces dernières qui ont constitué mon principal objectif. Je postulais alors qu’en travaillant à formaliser de telles règles de production (en particulier pour la LSF dans mon cas), on permettrait d’animer des personnages virtuels en synthétisant les formes correspondant aux interprétations associées tout en construisant un système de règles non ambiguës ayant une pertinence linguistique.

Pour s’appuyer sur des données attestées, nous nous sommes grandement nourris au départ du corpus vidéo de LS créé dans DictaSign [40], projet européen terminé en 2012. J’y ai été associé dès son démarrage en février 2009, majoritairement impliqué dans le « work package » dédié aux modèles linguistiques. C’est là que la démarche d’association sens–forme a été initiée, et qu’a été créé le premier corpus comparable multi-langue de cette ampleur (4 langues x 5 heures de LS par 16 locuteurs différents sur 8 tâches variées), cf. fig. 8.

Ce travail a consisté pour moi d’une part à :



FIGURE 9 – Forme repérable : pointage de numéral sur main opposée

- encadrer les tâches d’annotation du corpus (CDD de Nadège Devos et Flora Badin), au niveau lexical manuel contractuellement dû pour la livraison du corpus, et au niveau de formes supplémentaires pour l’aspect suivant ;
- une exploration du corpus en vue de la formalisation de règles linguistiques pertinentes (liens sens–forme) ;
- en parallèle, élaborer le modèle de description formelle pour spécifier (et animer) les règles de production en question, ce que je décris dans la section qui suit.

Le projet ayant posé comme objectif le développement d’un prototype de wiki en LS, et les listes à puces ou numéros étant souvent utilisées dans ce type d’application, les premières explorations du corpus se sont concentrées sur les formes telles que l’utilisation d’une balise numérale sur une main avec pointage de l’index de l’autre main (indiquant a priori une énumération, cf. fig. 9), et les constructions de listes dans le discours en général.

Cette première recherche s’est soldée par un double résultat intéressant :

- les productions interprétées comme listes non-exhaustives d’éléments non mutuellement exclusifs (autrement dit les énumérations « ..., ..., etc. » où les items peuvent être vrais en même temps) faisaient toutes apparaître un mouvement identique sur la fin de chaque élément énuméré, à savoir un mouvement en avant du menton vers la fin de chaque élément de la liste, et un retour en position initiale plus lent après l’élément ;
- les énumérations de choix mutuellement exclusifs (« soit ..., soit ... ») faisaient apparaître un repositionnement du menton dans l’espace sur chacun des items de la liste, avec abaissement et une brève tenue en place sur la fin de chaque item.

Ainsi, pour chacun de ces types d’énumération, si l’on nomme $item_1, item_2, \dots$ les éléments listés, nous pouvons écrire une règle de production pour représenter l’association sens–forme identifiée, et représenter graphiquement la synchronisation des formes réalisées en fonction de ces $item_i$ sur un axe temporel. Par exemple, la figure 10 illustre le premier type de liste ci-dessus (ouverte à éléments non-mutex). Sur celle-ci :

- les rectangles délimitent sur l’axe horizontal les intervalles temporels durant lesquels la forme indiquée est produite (« HOLD » signifie une tenue de la posture du bloc accolé, « hd :p-swp » représente le mouvement d’aller-retour de la tête, et les items en italique représentent les arguments) ;
- l’arrangement des rectangles sur l’axe vertical est arbitraire et n’est pas à interpréter comme les pistes d’une annotation où l’on trouverait un articulatoire annoté par ligne (ex. fig. 11) : deux rectangles partageant un même intervalle sur un diagramme de forme AZee (ex. fig. 10) montrent seulement une simultanéité dans leur réalisation, mais leur ordre vertical sur le schéma est sans objet.

En conduisant ce type de repérage en corpus, nous avons pu ainsi faire émerger de premières observations de ce genre, rapportées dans le livrable 4.2 du projet [12]. J’ai aussi encadré à la suite un stage (M1 de Mathieu Martinet) visant à identifier des règles de production d’expressions temporelles (la période coïncidant avec le

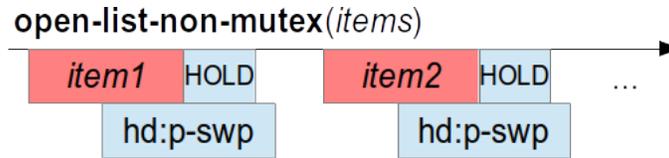


FIGURE 10 – Agencement temporel de la forme associée à « open-list-non-mutex »

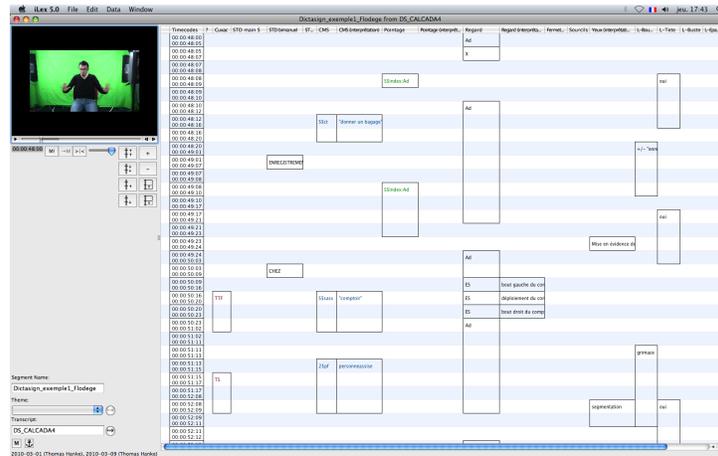


FIGURE 11 – Annotation en pistes (ici avec iLex sur le corpus DictaSign, et les pistes défilant de haut en bas) : un articulateur ou une fonction donnée par piste

début de mes explorations en traduction, ce choix était motivé par les collaborations locales au laboratoire et un corpus en cours de construction, cf. §3).

À ce stade et en vue de l’animation automatique d’un signeur virtuel, il fallait décrire formellement les articulations et synchronisations à générer, les représentations temporelles graphiques ne suffisant pas. C’est l’objet de la section suivante.

1.2.3 Extension Azalee

Les articulations qui nous intéressent dans ces recherches au niveau du discours échappent en général à la seule séquence ou alternance de postures et transitions. On voit qu’il s’agit de positionner des portions de l’activité signée sur un axe temporel de manière libre. Le début du mouvement de tête visible en figure 10 sur chaque item i par exemple n’implique pas de début, fin ou changement de posture dans la réalisation de $item_i$. À part les contraintes datant les début et fin du mouvement de tête (exprimées par rapport à la date de fin de $item_i$), chaque bloc se synchronise indépendamment. Donc même l’approche de Liddell & Johnson ne peut suffire dans ces cas car les bornes d’un intervalle ne matérialisent plus nécessairement des moments où tous les changements de posture s’alignent.

Les synchronisations représentées graphiquement ci-dessus ayant vocation à être animées par un programme, il convient d’en proposer une spécification formelle, ce que j’ai fait avec Azalee dans ma deuxième année de post-doc. Vu la nature des objets à synchroniser, je me suis d’abord tourné vers la logique temporelle de Allen qui formalise spécifiquement l’arrangement topologique d’intervalles à partir de 7 opérateurs de comparaison. J’ai intégré ces opérateurs au premier langage proposé [15], mais ai vite remarqué deux limites :

- seule la précédence entre bornes respectives peut être exprimée alors que la durée de séparation entre deux bornes (quelle qu'elle soit) peut être discriminante – il y a donc sous-spécification possible ;
- chaque opération est en réalité un alias pour un ensemble de contraintes sur des bornes, or on peut vouloir n'exprimer qu'une partie de celles-ci – il y a donc sur-spécification possible.

J'ai donc étendu le langage pour spécifier des contraintes de bornes seules mais le formalisme disparaîtra en réalité rapidement au profit de la généralisation AZee présentée en section suivante. Ce qu'a néanmoins apporté cette étape du travail est le retour de la synchronisation libre finalement délaissée par choix dans Zebedee (pour lequel c'était suffisant) alors qu'envisagée au départ (cf. p. 10). Ainsi nous étions en possession à la fois :

- d'une option de synchronisation contrainte, inspirée par le système de Johnson & Liddell, fort de bons arguments d'utilité dès qu'on s'intéresse aux unités plus figées que sont les signes, d'où mon choix au moment d'élaborer Zebedee ;
- d'une option de synchronisation libre qui permet un contrôle plus souple, et pouvant faire appel à des unités décrites en Zebedee dans son contenu.

1.3 Approche « non nivelée » et formalisme AZee

Azalee ayant été conçu dans la continuité de Zebedee, leur compatibilité a très vite permis une fusion en un seul modèle, généralisant leur cadre formel commun et donnant lieu à une « dénivellation » que nous présentons dans cette section.

1.3.1 « Dénivellation » par fusion des formalismes

L'originalité de mes contributions formelles Zebedee et Azalee était de considérer que :

- la description d'une forme à produire en langue des signes est un problème de synchronisation entre des articulateurs choisis ;
- aucun articulateur n'a de statut privilégié ou d'affectation spécifique a priori, tous pouvant donc indépendamment participer ou non à n'importe quelle opération ou entrée linguistique ;
- les spécifications liées à la position ou l'orientation d'articulateurs dans l'espace de signation sont exprimées à l'aide de contraintes géométriques dans un espace affine ;
- les contraintes géométriques et temporelles sont exprimées à l'aide de contraintes nécessaires et suffisantes ;
- les dépendances entre les éléments des descriptions doivent apparaître dans les descriptions ;
- toute description de forme peut être paramétrée par des arguments pour permettre la généralité des règles et des descriptions.

Lors de leur élaboration, Zebedee et Azalee se situaient à des niveaux différents de construction du langage, à savoir lexical (et infra-) pour le premier et supra-lexical pour le second. Mais on remarque que cette distinction n'est essentielle dans aucun de ces modèles. Ils relèvent en réalité de la même approche par contraintes articulatoires synchronisées dans un espace principalement géométrique, et le même objectif de décrire des règles génériques de production par la spécification d'invariants et la paramétrisation de formes à générer. Qu'elle soit infra- ou supra-lexicale, quoi que cela veuille d'ailleurs dire au fond, une description pourrait se faire avec l'un ou l'autre des formalismes, leur seule différence résidant finalement dans le mode de synchronisation proposé.

Ainsi, la fusion des deux fut aisée et a donné lieu au modèle AZee, permettant de représenter toute règle de production en LS, quel que soit le niveau fonctionnel de la règle intuitivement supposé, pourvu qu'un lien sens-forme soit identifié. Les signes jusqu'alors qualifiés de lexicaux ne sont qu'un sous-ensemble d'association sens-forme dans une vision généralisée de la règle de production. En cela, nous avons « dénivelé » l'approche traditionnelle qui avait toujours supposé une organisation en couches : phonème, morphème, lexème, syntagme, etc. Ceci s'avérera un avantage pour couvrir les productions dont on ne sait dire si elles sont à inventorier lexicalement ou comme des productions langagières complexes.

Restait maintenant à asseoir cette proposition en multipliant les exemples et en sécurisant leur valeur empirique pour démontrer sa capacité générative.

1.3.2 Formalisation de la méthodologie d'extraction de règles

Les règles de productions étaient sur le point de se multiplier. Pour des résultats reproductibles et en standardiser la présentation, il fallait formaliser la méthodologie permettant d'extraire les appariements forme-sens apparaissant dans les corpus.

Pour identifier les liens entre sens et forme et spécifier les règles de production, la méthodologie repose sur un va-et-vient entre les deux, associant un critère de l'un et les occurrences correspondantes de l'autre. C'est-à-dire, dans le sens de l'observation, repérer toutes les occurrences d'un critère fonctionnel et en décrire les formes, et dans le sens de l'interprétation, repérer toutes les occurrences d'un critère formel et en décrire la fonction. Dans les deux cas, il s'agit de grouper des éléments communs dans les occurrences repérées, et pour chaque groupe procéder à un embranchement dans le sens inverse à partir du critère définissant le groupe. Ce va-et-vient entre observation et interprétation permet d'affiner les critères de fonction jusqu'à converger vers un invariant de réalisation.

La procédure est la suivante :

1. Choisir un critère F de départ pour cette itération. F est soit un critère de forme (p. ex. la présence d'un gonflement des joues), soit une fonction sémantique (p. ex. la négation d'une clause), et peut être arbitraire.
2. Lister toutes les occurrences de F dans le corpus. Soit N_{occ} le nombre d'occurrences trouvées.
3. À chaque occurrence ainsi listée, associer des traits :
 - d'interprétation si F est une forme observable ;
 - de la forme visible si F est un critère de fonction.
4. Identifier et numéroter $1..N_{gp}$ les ensembles d'occurrences partageant des traits identiques. Soit N_{out} le nombre d'occurrences ne faisant partie d'aucun groupe.
5. Si $N_{gp} = 1$ et $\frac{N_{out}}{N_{occ}} < \mu$ (avec μ le seuil de rejet de la régularité, par exemple 10 %) :
 - si F est une forme et que l'on cherche des règles d'interprétation (cas de la reconnaissance automatique par exemple) : STOP [A], l'interprétation est assez régulière → on écrit une règle associant à F la sémantique interprétée dans le groupe ;
 - si F est une fonction et que l'on cherche des règles de production (c'est notre cas de figure en synthèse) : STOP [B], la forme est assez régulière → on écrit la règle associant à la fonction sémantique F les traits de forme définissant le groupe, cf. ci-dessous.

Sinon : pour chaque groupe $k \in [1..N_{gp}]$, procéder à une nouvelle itération avec pour critère initial $F.k$, défini par les traits constituant le groupe k de l'itération présente. À ce moment, on crée un embranchement dans la recherche et on inverse le sens de la recherche entre interprétation et observation.



FIGURE 12 – La forme « lip :pout » à synchroniser dans « inter-subj »

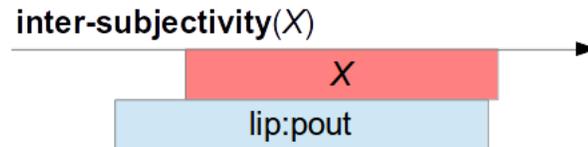


FIGURE 13 – Arrangement temporel des formes pour la règle « inter-subj »

Dans notre cas [B], une fois une forme régulière identifiée, on peut définir une règle de production par la donnée de :

- la **fonction sémantique** F (critère de l'itération ayant atteint [B]), abrégée ensuite pour donner un nom court à la règle ;
- une liste d'**arguments** de la règle, c'est-à-dire les paramètres dont dépendent conjointement la fonction sémantique et la forme (ceux-ci peuvent être optionnels si des valeurs typiques sont spécifiées par défaut) ;
- une description de la **forme** associée, éventuellement paramétrée par les arguments.

Par exemple, on peut définir la règle de production « inter-subj » suivante :

- sens : ajouter un jugement d'inter-subjectivité à un énoncé (jugement partagé : « tout le monde est/serait d'accord que X »), voire d'objectivité ;
- argument unique : X , représentant la chose jugée/objectivée ;
- forme à articuler : geste de la bouche de la figure 12 synchronisé avec X (tel qu'il serait autrement signé seul) de la manière représentée en figure 13.

On remarque que cette méthodologie fonctionne autant pour expliquer les productions traditionnellement qualifiées de lexicales. Par exemple, la règle « BALLON » :

- sens : ballon ;
- arguments (optionnels) : emplacement loc (point de l'espace, devant l'abdomen par défaut), taille r (mesure de distance, demie longueur d'avant-bras par défaut) ;
- forme à articuler : cf. figure 14.

Pour la lisibilité de ces schémas, les boîtes sont colorées en *bleu* lorsqu'elles contiennent uniquement des invariants purs (non influencés par les arguments), *hachurées* lorsqu'elles spécifient des formes qui dépendent d'arguments de la règle (p. ex. dans « BALLON », l'orientation du vecteur normal des deux paumes dépend

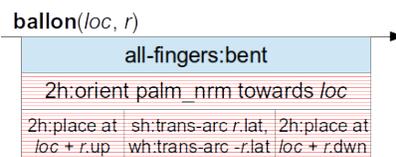


FIGURE 14 – Arrangement temporel des formes pour la règle « ballon »

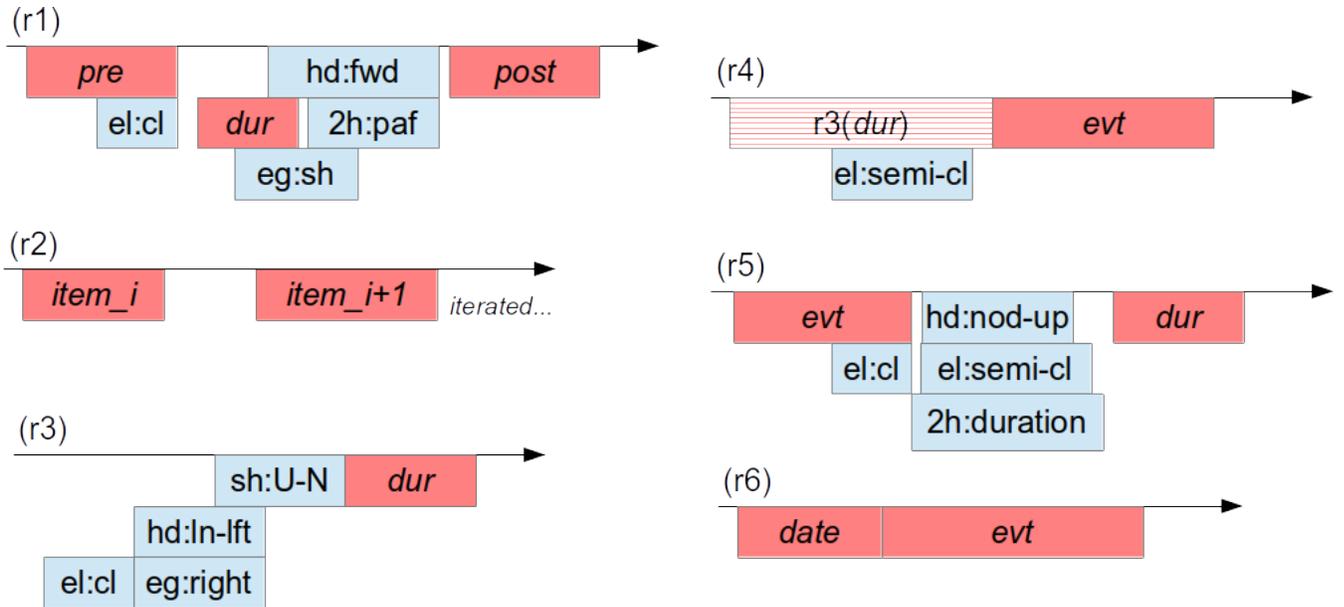


FIGURE 15 – Formes observées pour six règles d’expressions temporelles (r1 à r6)

de *loc*), et en *rouge* uni lorsqu’il s’agit d’intervalles repérés dans le temps mais dont rien n’est connu (reprise complète d’un argument temporel comme le *X* de « inter-subj »).

1.3.3 Des règles isolées au système récursif

Avec l’aide d’un stage (M2 de Mohamed Nassime Hadjadj) qui s’est ensuite poursuivi en thèse, nous avons appliqué cette méthodologie d’abord pour pousser plus loin l’étude des expressions temporelles sur le corpus de 120 brèves en LSF récemment créé (cf. §3.1), particulièrement présentes par construction. Six règles ont émergé, dont les fonctions et arguments sont listés ci-dessous, et dont les formes associées sont schématisées sur la figure 15.

1. Succession de deux événements séparés par une période de moins de 10 jours
3 arguments : la durée *dur* (optionnelle), les événements *pre* et *post*
2. Séquence chronologique de dates, périodes/durées et événements
N arguments : liste des éléments *item_i* dans l’ordre chronologique
3. Période de 10 jours minimum
1 argument : la durée *dur* de la période
4. Événement durant au moins 10 jours
2 arguments : la durée *dur* et l’événement *evt*
5. Événement durant moins de 10 jours
2 arguments : la durée *dur* et l’événement *evt*
6. Événement daté
2 arguments : la *date* et l’événement *evt*

D’autres propositions faisaient également surface dans le champ sémantique des localisations géographiques qui intéressait également l’étudiant. Mais cette liste de règles isolées dans leur construction constitue

en réalité un système de règles pouvant faire appel les unes aux autres pour spécifier leurs arguments. Autrement dit, chaque argument fourni lors de l'application d'une règle peut être lui-même obtenu par l'application d'une règle et produisant un argument du bon type. Par exemple, on trouve dans le corpus utilisé³ la phrase :

Dix ans après l'évacuation musclée de l'église Saint-Bernard, le 23 août 1996 à Paris, les sans papiers et leurs soutiens ne veulent pas être « dans la commémoration » mais dans le « combat »
[...]

En imbriquant récursivement des expressions formées avec les règles, on peut construire l'expression suivante qui non seulement produit les formes observées dans une traduction en LSF, mais compose des opérations sémantiques (chronologie entre événements, datation d'événement, etc.) en parfaite cohérence avec le sens interprété de l'extrait global⁴ :

```
r2(
  item1 = r6(date = "23/08/1996", evt = "évacuation musclée de l'église St-B."),
  item2 = r3(dur = "10 ans"),
  item3 = "sans papiers ..."
)
```

Cette observation est importante car elle fait de l'ensemble de règles un système d'opérations mutuellement récursives, et vérifie pour les expressions une compositionnalité sémantique déterminant directement les formes à produire. L'expression produit en effet une forme (calculable en l'évaluant) dont l'interprétation globale est celle de la combinaison initiale. Autrement dit, le sens interprété de la combinaison des formes (produites par les règles combinées) est la combinaison des sens des règles combinées.

Ainsi, le langage des expressions AZee devient une sorte de représentation sémantique, présentant deux différences essentielles avec les formalismes sémantiques traditionnels. D'une part, les opérations sémantiques élémentaires combinées dans les expressions n'existent que parce qu'une association forme-sens a émergé de l'observation de la langue. Ainsi ces primitives sémantiques ne sont pas réputées universelles ou former un ensemble minimal de construction du sens détaché de toute langue. Au contraire, celles-ci relèvent par essence de la LS, et n'ont pas d'autre visée – en tout cas a priori – que de représenter la structure profonde des discours signés. D'autre part comme nous l'avons vu, toute expression détermine directement les formes à réaliser (dont elle est l'interprétation). C'est donc un langage productif et non seulement une abstraction théorique.

Par ailleurs, la compositionnalité récursive permet de générer un langage infini, propriété partagée avec les grammaires génératives et jouissant d'un quasi-consensus sur le fait qu'elle serait nécessaire à toute langue. Nous insistons néanmoins sur le fait qu'il demeure des différences essentielles avec l'approche générative. D'abord, l'ordre d'apparition des unités atomiques dans l'expression (terminant la récursion) ne représente a priori rien de la séquence finale du discours. Les arguments d'un nœud dans l'expression ne sont même pas ordonnés mais nommés, et ne sont pas nécessairement des formes à signer (ils peuvent être des objets géométriques par exemple). Lorsqu'ils sont ordonnés temporellement à la réalisation, leur position temporelle relative dépend uniquement de ce que dicte la partie forme de la règle qui les contient. Ils pourraient en outre être simultanés. Aussi – et si l'on compare aux structures syntaxiques, surtout – chaque opération dans l'expression est chargée sémantiquement. Il ne s'agit donc pas d'un système qui régulerait la forme finale (de surface) de manière détachée de la sémantique.

3. Il s'agit du corpus parallèle français-LSF des « 40 brèves », créé fin 2012, présenté plus loin en §3.1.

4. Dans l'expression, les noms des règles de production sont soulignés ; ceux de leurs arguments en italique ; entre « guillemets » le sens d'une réalisation dont le détail n'est pas indiqué.

Ces résultats ont été présentés pour la première fois à l’atelier SLTAT de Chicago [21], puis ont été développés et fait l’objet de plusieurs publications jusqu’à aujourd’hui.

1.3.4 Spécification du langage *AZee* et compilateur

Une fois ces fondations théoriques établies et en vue de l’animation automatique d’un signeur virtuel, il fallait définir et implémenter un langage informatique concret, unifiant Azalee et Zebedee et permettant de :

- spécifier formellement les articulations et leur synchronisation en blocs (formes) ;
- définir des règles de production paramétrées (fonctions sémantiques) et les appliquer à des arguments.

Aussi, il fallait en programmer le compilateur/interpréteur capable de calculer (générer) les résultats des structures récursives données en entrée, c’est-à-dire produire les partitions résultantes qu’il faudrait animer. Ce travail a été mené en parallèle et le langage proposé est apparu pour la première fois dans un article à LREC [20], puis développé dans les années qui ont suivi. Il s’agit d’un langage fonctionnel (où tout est valeur, même une fonction), capable de décrire la synchronisation d’articulateurs selon n’importe quel motif.

Dans ce langage, une règle de production se représente par une valeur fonctionnelle ou « AZop » (AZee operator), paramétrée par ses arguments et retournant la forme associée. Celle-ci est souvent de type « SCORE », à savoir une ligne temporelle (cf. partition de musique) où chaque mouvement d’articulateur est synchronisé pour représenter globalement une signation. En tant qu’AZop, une règle peut alors être appliquée à des arguments du type attendu, eux-mêmes des expressions AZee, pour produire une forme construite.

Par exemple, la règle « inter-subj » est spécifiée par l’AZop ci-dessous, dont on reconnaît le paramètre « *x* », et dont la valeur retour est produite par l’opérateur « `sync` ». Celui-ci synchronise les deux blocs « SIG » et « MOUTH » et retourne une « SCORE » arrangeant le contenu de ces blocs sur une ligne temporelle. Cet AZop peut ensuite être appliqué à n’importe quelle valeur de type « SCORE » comme argument « *x* » pour produire la synchronisation représentée en figure 13.

```

azop          % l'opérateur fonctionnel d'AZee (crée une fonction)
'x           % nom d'argument
'nodefault  % pas de valeur par défaut pour "x" donc arg. obligatoire

sync         % l'opérateur de synchronisation (ici deux blocs "SIG" et "MOUTH")
'SIG        % nom de bloc
^x          % contenu du bloc

'MOUTH      % nom de bloc
'start=SIG:start:-.25 % synchro de la borne initiale du bloc
'end=SIG:end:-.1     % synchro de la borne finale du bloc
^lip:pout     % contenu du bloc

```

Je ne développe pas la partie technique et algorithmique du développement de ce compilateur AZee dans ce mémoire. Elle représente un travail conséquent mais n’apporte pas de recul ou de connaissance sur l’objet d’étude ou sur la discipline. Le résultat est en revanche disponible et utilisable sous la forme d’un programme en Python, via un dépôt *git* public⁵.

5. <https://github.com/michael-filhol/AZee-anim>

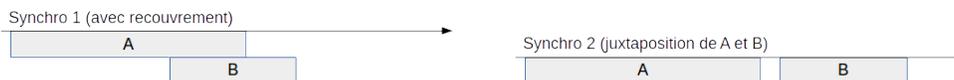


FIGURE 16 – Deux synchronisations possibles en AZee

1.4 Montée en puissance du modèle

En 2015, nous étions forts de résultats prometteurs puisqu’un système productif de règles récursives et à compositionnalité sémantique semblait émerger. Mais une inquiétude demeurait quant au nombre de règles nécessaires à une couverture suffisante de la langue. En effet les fonctions sémantiques des règles identifiées étaient très souvent d’une granularité plutôt fine, qui ne permettait pas de conjecturer un passage à l’échelle sans envisager plusieurs centaines de règles. Par exemple : “deux événements E_1 et E_2 séparés par une période d’une durée d inférieure à $10 \sim 15$ jours”. Or, si une telle proposition peut rester acceptable en vue de l’informatique seule, un système de cette taille paraissait en revanche peu séduisant pour relever d’une modélisation linguistique.

Plus inquiétante encore, la possibilité théorique que nul ensemble de règles ne puisse en réalité jamais générer d’ensemble couvrant de la langue des signes n’était pas écartée. AZee ne légitimerait alors jamais sa place de modèle du côté linguistique en cela qu’il n’expliquerait plus la compositionnalité de la langue dans son ensemble. Il ne serait plus *générateur* de la langue, quand bien même il serait utile à en synthétiser un sous-ensemble.

1.4.1 La séquence comme forme

Depuis le début de mes travaux au CNRS et jusqu’à 2014 inclus, je m’étais essentiellement intéressé à montrer qu’il fallait observer les langues des signes comme des langues non linéaires. En d’autres termes, que le recours systématique à la séquence d’unités devait être abandonné comme premier principe organisateur, et qu’il fallait accepter d’envisager des modèles nouveaux en conséquence. Pour rendre visible et étudier les phénomènes issus de la multilinéarité trop souvent ignorés, j’avais donc surtout investigué les formes simultanées, comme les plissements des yeux en parallèle de mouvements manuels ou du buste, qui m’ont conduit aux résultats résumés plus haut.

Mais ce positionnement scientifique n’implique pas de réfuter l’existence de toute séquentialité en langue des signes. Et quelle que soit la capacité du canal gestuel à paralléliser les informations grâce à l’usage de différents articulateurs, un discours complet a bien entendu recours à une certaine sérialisation. Il était temps de fournir une explication formelle à celle-ci dans le nouveau paradigme proposé.

L’idée originale pour attaquer ce problème de la séquence a été de considérer que la mise en séquence n’est qu’un cas particulier de synchronisation, c’est-à-dire un élément de forme au même titre que la mise en parallèle ou en chevauchement d’intervalles dans une description (cf. fig. 16). En tant que forme, elle devait alors être justifiée de la même manière que tout autre type de synchronisation, à savoir par le jeu des associations forme-sens. Cela contredit l’idée universellement partagée qu’il existe *a priori* une séquence structurant le discours et requérant un système et un niveau de description linguistiques propres (souvent nommé *syntaxe* par analogie celle gouvernant l’ordre des mots en langue écrite).

Nous avons ainsi proposé :

- le *critère de juxtaposition J* suivant, caractérisant les cas de précedence immédiate entre deux éléments identifiables (cf. fig. 16, schéma de droite) :

J : “Deux éléments A et B interprétables séparément dans une production sont *juxtaposés* dans cet ordre si le début de la production B a lieu après la fin de celle de A , et qu’aucun élément interprétable séparément ne se produit entre ces deux bornes.”

- de faire confiance à la méthodologie décrite en §1.3.2, et de l’initialiser avec le critère de forme J , faisant de la première itération une recherche en interprétation.

Dans cette première itération, une fois listées toutes les occurrences (évidemment nombreuses) de juxtaposition de deux éléments interprétables séparément, la tâche est d’associer à chacune l’interprétation si elle existe de cette juxtaposition. Par exemple, la juxtaposition $J(E_1, E_2)$ de l’élément E_1 interprété comme « ville » d’une part, et de E_2 interprété comme « forte puissance touristique » d’autre part, est interprétée en premier lieu comme « ville touristique », et l’opération J a alors comme interprétation de donner la qualité ou l’état E_2 à l’entité E_1 . Autre exemple, $J(E_3, E_4)$ avec E_3 : « trois attentats » et E_4 : « arrestation de 18 personnes suspectes » s’interprète comme deux étapes d’une chronologie où E_3 précède E_4 . Les deux productions répondent au même critère initial de forme J , mais la fonction est différente, E_4 ne s’interprétant pas comme un état de E_3 .

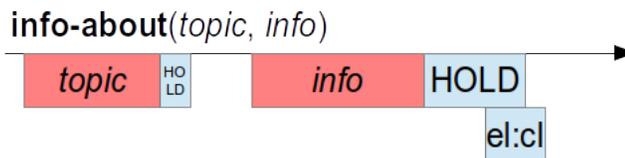
M. N. Hadjadj, dans sa thèse que j’ai encadrée à 90 %, a commencé ce travail sur $N_{occ} = 321$ premières occurrences d’un corpus de brèves journalistiques. La première itération a conduit à des interprétations de J d’une plus grande diversité que les recherches précédentes ($N_{gp} \geq 12$), et on observe au passage une granularité sémantique moins fine (des sens plus génériques) :

1. B donne l’état, le nom ou une qualité/propriété de A ;
2. la lettre B suit la lettre A dans l’épellation d’un nom propre ;
3. A est un contexte (temps, lieu...) pour l’événement B ;
4. A est un repère dans le positionnement de B ;
5. $A, B...$ sont les éléments d’une énumération non-exhaustive, non mutuellement exclusifs ;
6. $A, B...$ forment une liste d’éléments mutuellement exclusifs ;
7. B est un événement qui a suivi A chronologiquement ;
8. A n’est pas vrai, c’est B qui est vrai (« pas A mais B »)...

Pendant un an, nous avons appliqué la méthodologie présentée plus haut au groupe 1 de cette liste, et développé les itérations $J.1$ correspondantes. Les itérations successives ont fini par converger en trois règles stables, toutes à deux arguments de type signation (SCORE), à savoir « info-about », « side-info » et « category », spécifiées ci-dessous.

« info-about »

- Sens : *info*, qui porte sur *topic*, est l’information principale (focus).
- Arguments : deux signations *topic* et *info*.
- Forme : *topic* et *info* juxtaposés dans cet ordre, durée moyenne de ~ 350 ms entre *topic* et *info*, légère tenue de posture en fin de *info*, clignement des yeux (« el :cl » sur la figure ci-dessous) sur la fin de *info*.



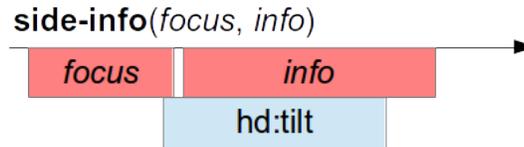
- Exemple : *topic*=PUISSANCE et *info*=TOURISME pour signifier “la/sa puissance est le tourisme”.



FIGURE 17 – Signe « MOUCHOIR », signifiant également “Cholet” (polysémie par métonymie liée à l’histoire de la ville)

« side-info »

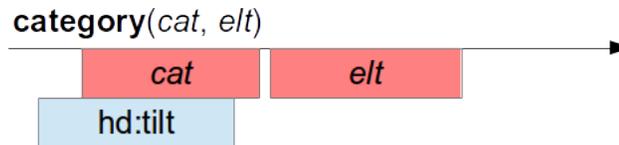
- Sens : *focus* est l’information principale, à laquelle est ajoutée (“entre parenthèses”) l’information *info*.
- Arguments : deux signations *focus* et *info*.
- Forme : *focus* et *info* juxtaposés dans cet ordre, séparation courte (120 ms) entre *focus* et *info*, léger penchement de la tête sur *info*.



- Exemple : *focus*=PERSONNE et *info*=“âge [est] de 23 ans” pour signifier “une personne de 23 ans”.

« category »

- Sens : *elt*, à interpréter comme un élément de la classe *cat*.
- Arguments : deux signations *cat* et *elt*.
- Forme : *cat* et *elt* juxtaposés dans cet ordre, séparation courte (120 ms) entre *cat* et *elt*, léger penchement de la tête sur *cat*.



- Exemple : *cat*=VILLE et *elt*=MOUCHOIR pour signifier “Cholet” – le signe polysémique « MOUCHOIR » peut seul signifier “mouchoir” ou “Cholet” (fig. 17); cette construction interdit l’interprétation “mouchoir”.

Deux observations importantes peuvent être faites de ce résultat : l’expressivité et la précision du modèle d’une part, et la productivité réursive et la couverture d’autre part. Ces points sont développés dans les sections qui suivent.

1.4.2 Expressivité en sens et précision des formes

Avec AZee, on spécifie donc la juxtaposition de la même manière qu’on spécifierait des productions simultanées. Jusqu’ici, les productions en séquence étaient considérées comme des constructions syntaxiques (donc d’un niveau particulier), alors qu’elles apparaissent pour AZee ici au même titre que toute autre règle, et sont identifiées empiriquement en utilisant la même méthodologie. Seul un élément de forme les caractérise, relatif à leur synchronisation interne.

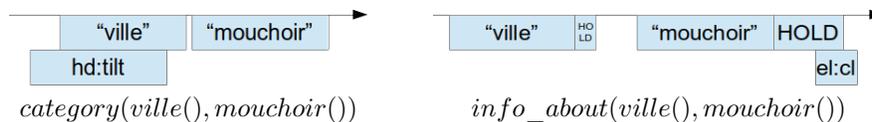


FIGURE 18 – Deux expressions AZee différentes correspondant à la séquence de gloses « VILLE, MOUCHOIR »

De plus, on peut constater que même là où une juxtaposition $J(X, Y)$ est nécessaire à la description de la forme :

- elle est rarement une contrainte suffisante en cela que certains invariants apparaissent aussi dans les durées de transition entre les productions X et Y , par exemple dans « side-info » où la durée séparant les productions juxtaposées doit être aux alentours de 120 ms alors qu’elle est presque triple dans « info-about » ;
- d’autres invariants de forme accompagnent souvent la juxtaposition, comme des penchements de tête ou des clignements des yeux dans les règles présentées.

Un modèle de séquence classique ne peut rendre compte de ces détails, la juxtaposition étant considérée comme opération atomique et souvent suffisante. AZee est plus précis dans l’expression des régularités de la langue, et plus expressif en sens car la raison sémantique d’une juxtaposition vient nécessairement avec son utilisation. Par exemple, l’imbrication fonctionnelle $category(ville(), mouchoir())$, montrée en exemple plus haut, désigne Cholet sans prédicat (en somme un équivalent du groupe nominal “la ville de Cholet” en français) par l’ajout d’une classe (hyperonyme) “ville” à l’unité polysémique de la figure 17 grâce à la règle « category ». Avec un modèle classique à séquence de gloses, la même production est réduite à la suite : VILLE, MOUCHOIR. Mais celle-ci correspondrait tout autant à l’imbrication différente $info_about(ville(), mouchoir())$, qui :

- maintient légèrement le signe VILLE, allonge la durée entre les deux productions mises en séquence, et ajoute un clignement des yeux terminal, donc diffère par la forme (voir comparaison figure 18) ;
- tend à signifier “la/cette ville est Cholet” (ou de manière moins probable : “la/cette ville est un mouchoir”, par polysémie), donc diffère également par le sens interprété.

Les modèles classiques ne considèrent pas les questions de durée entre les unités juxtaposées, ou alors les envisagent comme des éléments de style, prosodiques, hors du niveau de description considéré. Par conséquent, les deux productions ci-dessus seront considérées comme syntaxiquement ambiguës en premier lieu, à résoudre éventuellement sur un autre niveau, par le contexte. Au contraire pour AZee, en considérant toutes les articulations visibles comme pertinentes, ces deux productions sont différentes sur la forme et correspondant à des imbrications sémantiques différentes. Elles sont donc distinctes d’emblée et ne sont à aucun niveau ambiguës, malgré la présence d’un ordre similaire dans la forme finale. En fin de compte, la séquence observée *émerge* du modèle au lieu d’être supposée. Plus qu’un détail cosmétique pour une synthèse plus réussie, une valeur ajoutée existe donc aussi du point de vue de la modélisation linguistique elle-même.

1.4.3 Productivité et couverture

Une fois ces règles trouvées et à des fins d’évaluation de la faculté récursive du modèle à couvrir des discours dans leur ensemble, nous avons écrit les expressions AZee correspondant à 7 premiers mini-discours et 3 min 30 s de vidéo au total, en utilisant les trois règles ci-dessus issues de $J.1$ et en supposant stables les autres groupes $J.k, k \neq 1$. Par exemple, un extrait traduisant l’expression « la ville touristique de Dahab » correspondait en forme et en sens à l’imbrication suivante, les formes étant représentées en partition sur la figure 19 page suivante :



FIGURE 19 – Formes observées pour traduire « la ville touristique de Dahab »

```
category(
  cat = side-info(
    focus = VILLE,
    info = info-about(topic = PUISSANCE, info = TOURISME),
    elt = finger-spell(letters = list(D, A, H, A, B))
  )
)
```

Sur 220 applications de ces fonctions, seules 21 occurrences de *J* n'étaient pas aisément expliquées. Il s'agissait pour la plupart de pauses supérieures à 1.5 s mettant en jeu une rétractation des mains et un repos corporel, difficile à interpréter de manière plus précise qu'une simple continuation de discours. La forme était bien systématique, mais l'interprétation sémantique ou rhétorique était encore incertaine à l'époque. Ceci a fait prendre un pas énorme à notre approche, puisque du petit sous-ensemble de la langue que représentait les expressions temporelles de dates et de durée, nous sommes passés à une couverture dépassant 90%.

À l'instar de *J.1*, on pouvait prévoir que les itérations *J.2*, *J.3*... engendrent une modification des fonctions supposées et probablement de leur nombre, ce qui s'est avéré par la suite. Mais même en supposant un triplement de leur nombre sans aucune fusion au cours des itérations, nous conjecturons un ordre de grandeur d'une vingtaine de règles (plus celles décrivant les « signes » du vocabulaire) au total pour un système couvrant la LSF. Cela a fait basculer nos inquiétudes sur le nombre de règles attendues, et fait espérer un modèle efficace en concision. L'approche devenait attractive car très productive.

Depuis cette expérience, l'ensemble des règles a été complété et de nombreux discours complets ont ainsi pu être transcrits avec des expressions AZee. Tout récemment (2020), un mois de stage (Camille Challant) encadré par mes soins a permis de transcrire l'ensemble des 120 brèves LSF du corpus des « 40 brèves », ce qui constitue une bonne référence d'une heure de discours et confirme l'avancée en couverture. Les 120 descriptions (expressions AZee) seront prochainement ajoutées au corpus déjà en ligne (cf. §3.1).

Par ailleurs, nous avons encore étendu la couverture du modèle pour la LSF à plusieurs constructions très iconiques utilisant l'espace : déplacements de classifieurs, déploiements de formes et surfaces, assez peu employées dans les « 40 brèves ». Des exemples sont référencés plus loin (§2.4).

1.4.4 Impact et reconnaissance internationale

Il existe selon moi un double impact épistémologique à cette approche, en linguistique et en TAL. En effet, AZee donne lieu à un modèle génératif de langue sans supposer l'existence a priori de niveaux de construction du langage. Ceci permet donc de les caractériser *a posteriori* par l'expression de critères formels sur les règles de production elles-mêmes, voire peut-être de clarifier les contours de ce qu'on appelle lexical, syntaxique, etc.

Par exemple, on peut définir la propriété pour une règle d'être « lexicale » si et seulement si elle génère une forme de type « SCORE » (signation qui prend du temps) sans argument obligatoire. Ce critère permet d'intégrer à l'inventaire « lexical » toute règle définie capable de livrer une production hors contexte (sans

valeur donnée en argument), ce qui est le propre des entrées des dictionnaires. Ainsi, « BONJOUR » et « BAL-LON » se retrouvent qualifiées de lexicales, mais l’inventaire exclut toutes les unités plaçant/déplaçant des proformes/classifieurs⁶, ainsi que les différents types de pointages définis sans cible par défaut.

On peut qualifier de « syntaxique » toute règle de production requérant au moins deux arguments de type « SCORE », et ainsi représenter l’idée que la syntaxe organise les blocs de signation entre eux dans le discours. Sans davantage de contrainte dans le critère, une telle proposition intègre les règles qui parallélisent ou font se chevaucher leurs arguments, c’est-à-dire que l’ensemble « syntaxique » ne serait plus lié à la notion de séquence de ses constituants mais permettrait la simultanéité. Aussi, on note qu’un tel critère exclut également les règles de placement/déplacement de proforme dont les arguments sont géométriques (positions, chemins) et posturaux (proformes), ce qui ferait de ces dernières, si l’on accepte les deux critères ci-dessus, des productions ni lexicales, ni syntaxiques.

On voit que cet exercice permet de revisiter les contours entre les niveaux traditionnels d’analyse du langage, et nourrir la réflexion sur la nature parfois intermédiaire de certaines unités et les doter d’arguments formels. Ces réflexions ont été publiés au workshop international dédié à la représentation des langues des signes [18].

Par ailleurs, la productivité à partir d’un nombre finalement réduit d’opérations laisse entrevoir de structurer des grammaires d’un nouveau genre. La fréquence de certains nœuds, imbrications et motifs peut être mesurée dans les corpus d’expressions, et on peut observer des exclusions entre règles, ou catégoriser des constituants en fonction de leur distribution. De telles grammaires supplanteraient la syntaxe au sens classique d’un ordre de constituants (mots et syntagmes), catégorisés a priori avec des étiquettes supposées (« nom », « verbe »...). Nous l’avons vu aussi, on abandonne nécessairement l’idée d’une syntaxe détachée de la sémantique puisqu’il n’existe *aucune* règle de production ne portant pas du sens, par construction. Une grammaire contraignant les expressions AZee contraindrait donc aussi des opérations sémantiques et non seulement des catégories arbitraires.

Du côté informatique, l’avantage opérationnel d’être directement animable tout en offrant une couverture et une pertinence linguistiques ont, après 2015, séduit les équipes faisant l’état de l’art du domaine de l’animation. En particulier, deux équipes spécialistes en animation de langue des signes par avatar se sont intéressés à l’utiliser en remplacement du modèle lexical classique qu’ils implémentaient. L’une composée de Fabrizio Nunnari et Alexis Héloir et basée au DFKI à Sarrebruck, m’a invité à collaborer pour créer un système d’animation avec Blender (logiciel libre de modélisation 3D) à partir des résultats de mon compilateur.

L’autre, basée à l’université DePaul à Chicago, est composée de John McDonald et Rosalee Wolfe, auteurs de l’avatar « Paula » dont l’objectif est la synthèse de mouvements naturels pour les langues des signes. Pour le faire évoluer et passer d’une suite de signes aux structures AZee en entrée, J. McDonald a d’abord séjourné au laboratoire deux fois l’été entre 2016 et 2018, et nous avons depuis entretenu et poussé la collaboration. Je développe ce travail dans la section suivante sur l’animation (§2).

En outre j’ai accueilli Abdelaziz Lakhfif, professeur de l’Université de Setif (Algérie) et ayant travaillé à un système de traduction de l’arabe vers la langue des signes algérienne. S’intéressant à AZee et bénéficiant d’un financement de son université, celui-ci a séjourné une semaine au laboratoire pour échanger de manière présentielle. Thomas Hanke, qui mène en grande partie l’activité lexicographique à l’Université de Hamburg et père du système HamNoSys principalement utilisé jusque là par les systèmes d’animation, envisage aussi

6. « Proforme » et « classifieur » (ou « classificateur ») sont des termes qui se recouvrent, aux contours et éventuelles différences dépendant des auteurs, désignant une forme prise par une ou deux mains, placée(s) dans l’espace pour représenter un objet dans la scène de signation. En réalité, de même que la configuration, AZee ne se limite pas à considérer des mains : on peut faire usage du bras (un arbre, un poteau...) ou des doigts uniquement (p. ex. personne debout).

un convertisseur vers AZee. Ainsi, plusieurs équipes à l'international aux visions parfois différentes ont le souhait de tirer profit d'AZee, ce qui confirme l'attractivité du modèle.

2 Synthèse par signeur virtuel

La synthèse de langue des signes commence juste avant les années 2000, et a été mon application phare tout au long de mon travail, dès le début de ma thèse en 2005. Toute formalisation devait permettre de générer des commandes déterministes pour animer un avatar.

2.1 État de l'art précédent : la première génération

La tâche de synthèse dépend nécessairement du système formel utilisé pour spécifier l'entrée des énoncés à générer. À ses débuts, elle s'inspire des seuls modèles existant alors, focalisés sur la description lexicale et transcrivant les énoncés comme une suite d'unités ainsi décrites.

Le premier résultat concret, publié et rendu fonctionnel est issu du projet ViSiCAST [1, 28], dont le système d'animation produit est JASigning [11]. Une suite lui est donnée avec eSign puis même DictaSign [9] pendant lesquels des améliorations plus ou moins significatives ont été apportées mais sans changement notable d'approche. Le système suppose toujours une séquence lexicale et manuelle, avec une piste séparée pour l'ajout d'unités non-manuelles mais synchronisées sur les unités de l'autre, éventuellement décalées mais à la main [27], sans abstraction linguistique pour en contrôler l'agencement.

Plusieurs équipes ont utilisé le système ensuite, par exemple pour des systèmes générateurs d'annonces automatiques en gare, une application assez emblématique de la période [7]. Les ressources disparaissent aujourd'hui sur ce projet, et il devient difficile d'en récupérer des informations, a fortiori les livrables et les exécutables produits. Mais celui-ci vit encore sous la forme d'un service en ligne générant des animations⁷. Étant donné une spécification SigML [10] ou HamNoSys [29] – le premier étant une représentation XML du second – il produit et affiche l'animation résultante, le programme s'exécutant côté serveur. Certaines équipes l'utilisent encore comme principale application cible pour leurs démonstrateurs, comme le FTI à Genève [5].

Ce système d'animation marque pour moi un jalon dans l'histoire de ma communauté scientifique. Il est remarquable qu'un tel résultat fut atteint si rapidement après avoir attaqué le problème, et perduré si longtemps en tant que seule solution utilisable pour la communauté en entier. Celle-ci venait presque de découvrir alors les problématiques liées à la synthèse de LS, et les défis conséquents et nombreux posés en termes de technologie pour le rendu graphique et l'animation de personnages virtuels : cinématique inverse, lisibilité des doigts, aspects robotiques du mouvement, appréciation des synchronisations... le tout avec une connaissance linguistique encore très évolutive.

Dans la communauté ensuite, des efforts ont été consacrés à améliorer la génération de postures et de mouvements, par exemple avec le moteur EMBR [30] permettant de résoudre la cinématique inverse avec des méthodes numériques plus souples que celle, analytique à deux segments, utilisée dans JASigning. Mais les fondamentaux en entrée des systèmes restaient les mêmes, et on n'assiste pour ainsi dire à aucune révolution ensuite quant à la synthèse de discours en LS.

7. <http://vhg.cmp.uea.ac.uk/tech/jas/vhg2020/CWASA-plus-gui-panel.html>

2.2 Apport d’AZee pour une deuxième génération

Dès les temps de Zebedee (avant 2010), j’ai moi-même co-encadré (60 %) une thèse [6] pour confronter mes modèles à la synthèse. La problématique majeure au départ était la résolution de cinématique inverse à partir de spécifications Zebedee, donc de postures sous-contraintes sur des chaînes cinématiques à plus de deux segments. Ceci se rapprochait de la problématique étudiée en parallèle avec EMBR, à laquelle l’étudiant avait choisi d’ajouter à la résolution le choix de la posture la plus confortable du point de vue biomécanique pour instancier les degrés de libertés non contraints.

D’avoir remplacé la donnée systématique des paramètres manuels (orientation, etc.) par des contraintes nécessaires et suffisantes en Zebedee complexifie le problème. Nous avons rapidement constaté que les considérations scientifiques pour une animation de qualité relevaient non seulement de la linguistique mais également de la programmation graphique, de l’art de la synthèse 3D, de modèles biomécaniques et des sciences du mouvement. Ces thématiques sont en limite de mon expertise mais grâce à ces réflexions et les publications au cours de ces efforts, nous avons pu rendre visible l’approche dans la communauté internationale, et montré en partie la souplesse et la précision des spécifications données en entrée des systèmes d’animation.

Un an après la publication des extensions et de la généralisation AZee, deux équipes ont souhaité collaborer pour implémenter un système concret d’animation sur AZee, ou adapter le leur aux entrées AZee. Ces nouvelles collaborations furent particulièrement bienvenues comme leur apport (synthèse de mouvement et programmation et rendu graphiques) était tout à fait complémentaire à celui d’AZee qui fournissait l’abstraction linguistique dont ils manquaient. De plus, nous allons voir qu’elles étaient complémentaires entre elles.

D’abord, F. Nunnari et A. Héloir (DFKI) voyaient en AZee de quoi piloter leur projet de système de synthèse pure, c’est-à-dire construisant toute l’animation à partir des spécifications atomiques minimales. D’une part les contraintes articulaires de plus bas niveau sur le squelette en AZee sont les mêmes que celles attendues par les systèmes génériques de cinématique inverse, à savoir des orientations de segments et des placements d’effecteurs. D’autre part, le compilateur AZee génère une frise temporelle situant les instants « clé » où les différentes postures (partielles) sont spécifiées. Ceux-ci correspondent aux « keyframes » propres aux animations synthétiques, entre lesquels des interpolations automatiques construisent les mouvements. La spécification de l’interface entre les deux parties d’un tel système nous semblait donc à portée. Cette approche suppose de tout reconstruire à partir de morceaux minimaux et en relativement petit nombre. Le résultat est un mouvement plutôt robotique mais la garantie est de pouvoir tout synthétiser dès qu’on sait traiter l’ensemble des contraintes atomiques. Cette collaboration a mené à une publication [42], puis la fin du financement dont ils bénéficiaient a mis la collaboration en suspens.

Dans le même temps, R. Wolfe et J. McDonald (Université DePaul à Chicago) souhaitaient également intégrer AZee à leur plateforme d’animation maison « Paula » (fig. 20). Leur approche est celle d’un rendu fin à partir d’animations d’artistes, donc non purement synthétique. À l’inverse de la synthèse à la DFKI, Paula produirait cette fois des mouvements les plus naturels possibles, donc non robotiques, mais serait limité par l’ensemble des briques disponibles (les unités gestuelles fixes ou processus de synthèse implémentés). Pour conduire ce travail, j’ai accueilli J. McDonald deux fois en séjour de courte durée au laboratoire pour débiter les discussions de manière présentielle et nous avons réalisé que la structure interne multi-pistes de Paula était en grande partie compatible avec les synchronisations spécifiées par AZee. J’ai aussi obtenu un financement pour un séjour invité au laboratoire dont J. McDonald a profité pendant un trimestre fin 2017. S’en est suivi une collaboration soutenue, ayant cours encore aujourd’hui, et une adaptation continue de nos systèmes afin



FIGURE 20 – La signeuse virtuelle Paula (voir aussi : <http://asl.cs.depaul.edu>)

de bénéficier chacun des avantages de l'autre : l'expressivité linguistique d'AZee et la qualité de l'animation de Paula (apparence naturelle, humaine des mouvements).

On note une symétrie entre les deux approches, dont nous pouvons mettre la complémentarité à profit. En effet, un rendu naturel est évidemment préférable, mais garantir la possibilité de tout animer quitte à retomber sur une animation plus synthétique a des avantages évidents.

2.3 Nouvel état de l'art : animation à granularité décroissante

Nous partions de deux principes :

1. le principe que plus les morceaux d'animation mis ensemble sont gros, plus l'animation aura l'air naturelle – ceci s'oppose aux systèmes de synthèse spécifiant tout à partir de primitives de mouvement de bas niveau ;
2. l'avantage en AZee que toute expression détermine *in fine* les formes à produire, mais fournit au départ une imbrication sémantique récursive dans les expressions d'entrée, ce qui permet de les considérer à différents niveaux de granularité.

Ainsi nous avons proposé une architecture où chaque expression est considérée d'abord en entier pour être animée en bloc si une animation pré-enregistrée ou un processus de génération est disponible pour la générer, avant d'être découpée en sous-expressions pour être animée par composition de ses parties dans le cas contraire. De cette manière, on ne développe une spécification des formes à produire que pour ce qu'on ne sait (ou peut) pas animer en bloc, et on construit les animations au niveau le plus haut possible de la composition.

La première expression complexe générée avec cet algorithme, en 2017, fut la suivante, signifiant « [il est généralement admis que] le cinéma, le restaurant [et autres activités de ce type] sont bien [divertissantes] », en 2017 :

```

info-about(
  topic = open-list-non-mutex(items = {cinéma(), restaurant()}),
  info = inter-subj(judged = bien())
)

```

Un rendu de l'animation obtenue est disponible en ligne⁸. On observera la différence, notamment en termes de mouvements simultanés (p. ex. manuels/non-manuels) et leur synchronisation, avec les systèmes standard de concaténation régulière de signes lexicaux.

8. <http://asl.cs.depaul.edu/Paula-Azee/Paula-Azee-Example.mp4>

Dans cette animation, un bloc pré-enregistré est utilisé pour le morceau « CINÉMA » par exemple, ce qui a terminé le développement récursif de l’expression. Si toutefois celle-ci était absente de la base de données, les formes générées par l’expression `AZee cinéma()` auraient été développées encore jusqu’à atteindre les contraintes minimales, à savoir ici quatre fois (4 postures clés dans la forme de « CINÉMA ») des contraintes sur la forme et l’emplacement des mains, etc. Ainsi tout est animable en principe, mais on combine des mouvements globaux quand c’est possible, évitant de construire à partir de contraintes articulatoires minimales.

Cette idée et la collaboration entretenue avec DePaul pour l’explorer a grandement fait progressé l’état de l’art et a donné lieu à des résultats bien accueillis chaque année depuis, exposés dans la section qui suit.

2.4 Collaboration vertueuse TAL–synthèse–linguistique

Depuis le début, la collaboration avec DePaul est restée soutenue et nous avons toujours plus mis à l’épreuve le modèle AZee, consolidé son lien avec Paula et augmenté la couverture de la langue des signes. Nos outils théoriques, formels et logiciels apportant des résultats satisfaisants, nous nous sommes attaqués, chaque nouvelle année, à de nouveaux phénomènes linguistiques en terrains de plus en plus propres aux langues des signes et intouchés des autres approches en animation.

En 2017, nous publions notre premier article conjoint sur l’architecture, présentée ci-dessus, du système joignant AZee et Paula [25], et produisons des animations composant des signes canoniques (cf. §2.3). En 2018, nous traitons la question du placement et du déplacement de proformes (classifieurs) dans l’espace de signation [23], une première conquête de notre système sur le terrain très peu foulé des constructions iconiques propres à la LS. En 2019, nous construisons une scène entière [41] en incluant des dynamiques particulières dans le placement d’ensembles d’objets (p. ex. groupe de 4 assiettes localisées) – une revue (numéro spécial de *Machine Translation*) vient par ailleurs d’accepter un article détaillé de ce travail. En 2020, nous avons enfin étendu le champ des constructions fortement iconiques en traitant les déploiements de formes (surfaces, lignes, etc.) [24].

Chacune de ces étapes a donné lieu à une publication dans un événement phare pour la communauté (ateliers SLTAT ou LREC, en alternance chaque année). Dans chacune, nous produisons des animations concrètes permettant à celle-ci d’apprécier non seulement un effort théorique ou algorithmique, mais un potentiel applicatif puisque de réelles animations étaient fournies en exemple à chaque fois. Les voici résumées :

- 2017 : architecture du système et séquence contrôlée de signes stables ;
<http://asl.cs.depaul.edu/Paula-Azee/Paula-Azee-Example.mp4>
- 2018 : placement et déplacements de proformes dans l’espace (unités géométriquement productives, sans formes canoniques) ;
<http://asl.cs.depaul.edu/proforms/proformPlacementAndMovement.mp4>
- 2019 : discours long, décrivant une scène complète dans l’espace avec positionnement d’objets et d’ensembles d’objets (comparer la version filmée originale avec la synthèse AZee+Paula) ;
<http://perso.limsi.fr/filhol/research/files/MTH-table-scene-2019/>
- 2020 : déploiements de formes/surfaces dans l’espace et mains simultanées (observer les synthèses E2, E3 et E4 – films originaux fournis).
<https://zenodo.org/record/3904430>

Après toutes ces années il nous apparaît clairement une collaboration vertueuse, grâce à laquelle non seulement un système d’animation a bénéficié d’un apport linguistique formel, mais un retour de balancier s’est plusieurs fois produit, ayant largement profité à AZee. En effet, c’est en étudiant les indices de formes dans les corpus,

en tâtonnant ensemble sur l’avatar pour reproduire fidèlement des exemples vidéos que nous étudions, en testant les animations au gré de leur synthèse, que nous avons pu conforter ou au contraire infirmer certaines règles, affiner la description de leurs formes et consolider le système dans sa globalité.

On constate aujourd’hui que nous arrivons à produire tous les exemples de toutes les publications de cette liste avec un système d’une douzaine de règles de production seulement. Et cela inclut la reproduction de discours réels dépassant la trentaine de secondes et remplis de productions iconiques (p. ex. 2019), impossibles à synthétiser jusqu’alors. Règles lexicales et celles décrivant les proformes mises à part, cet ensemble est le suivant (on ne donne ici que leur sens) :

- $\text{info-about}(\text{topic}, \text{info})$: info (ayant le focus), à propos de topic ;
- $\text{side-info}(\text{focus}, \text{info})$: focus , à propos duquel est donnée l’information info ;
- $\text{category}(\text{cat}, \text{elt})$: elt à considérer comme une instance de cat ;
- $\text{context}(\text{ctxt}, \text{proc})$: proc dans le contexte ctxt ;
- $\text{inter-subj}(x)$: il est généralement admis que x ;
- $\text{simultaneous}(\text{sig1}, \text{sig2})$: sig1 et sig2 ont lieu simultanément ;
- $\text{each-of}(\text{items})$: chaque élément de la collection items , sans saillance particulière à aucun d’entre eux ;
- $\text{all-of}(\text{items})$: la collection items en tant qu’ensemble, sans focus sur les éléments qui le composent ;
- $\text{place-proform}(\text{prf}, \text{loc})$: placement dans l’espace de signation de prf à l’emplacement loc ;
- $\text{move-proform}(\text{prf}, \text{path})$: déplacement de prf selon path (décrit le mouvement d’une entité) ;
- $\text{deploy-shape}(\text{prf}, \text{path})$: déploiement de prf selon path (décrit une ligne/courbe ou une surface) ;
- $\text{landmark-in-place}(\text{lm}, \text{sig})$: sig autour du repère spatial lm .

Cette économie dans la représentation, la productivité du système de génération et la qualité du rendu possible constituent un point fort de l’approche qui s’intensifiera encore dans le futur proche. Ainsi, après environ une décennie d’existence et grâce à sa collaboration avec Paula, AZee est passé du statut de proposition théorique au mieux séduisante à un système opérationnel qui produit des résultats importants. Notamment, on couvre de nombreuses structures iconiques si particulières aux langues des signes, et qui constituaient le verrou scientifique des systèmes précédents. Associé au système Paula aujourd’hui, AZee fait l’état de l’art du domaine de la synthèse des langues des signes.

3 Traduction

En parallèle des efforts de modélisation de la langue en elle-même, j’ai étudié la question de la traduction de la langue écrite à la langue des signes. En effet les lois sur l’accessibilité et la forte demande sociétale associée motive les projets en sens. J’ai donc voulu explorer les possibilités qu’offrait mon modèle pour outiller la traduction, en commençant par la traduction automatique, dès 2011.

3.1 Corpus parallèle français–LSF

À ce moment, la recherche de règles de production était sur le point de s’intensifier (voir §1.2.2 et suivantes), et je me rapprochais de collègues au laboratoire travaillant en TAL sur le texte afin de démarrer des collaborations sachant que tout effort en traduction impliquerait un travail sur le texte, a priori en tant que langue source. Mais ces deux axes allaient se heurter au même écueil : la faible quantité de corpus exploitable et notamment parallèle, c’est-à-dire bilingue avec alignements de français écrit et LSF, pour les études en traduction.



FIGURE 21 – Montage 2 vues du corpus 40 brèves

La seule ressource parallèle français–LSF de taille suffisante disponible était une archive de 3000 brèves journalistiques de WebSourd⁹ que j’avais récupérée auprès d’eux l’année précédente, alors qu’ils m’avaient accueilli dans leurs locaux pour que j’observe leur travail. Mais les vidéos au format Flash – la vocation était la publication en ligne et non l’observation scientifique – n’offraient pas la qualité suffisante pour des résultats surs : résolutions temporelle et géométrique faibles, vue unique empêchant l’observation sur l’axe de la profondeur. J’ai donc formulé le besoin de constituer un nouveau corpus parallèle au contenu contrôlé. X. Tannier et V. Moriceau, deux chercheurs de mon groupe travaillant sur l’extraction et l’analyse des phénomènes temporels dans les textes, s’étaient montrés intéressés par la traduction en LS, et avaient passé les années précédentes à travailler sur du corpus journalistique (AFP). Cette conjoncture fut idéale pour composer un corpus de genre journalistique avec l’expertise de tous, et d’orienter l’élicitation vers les expressions temporelles.

Parmi les milliers de brèves disponibles, nous avons opéré un choix équilibrant les valeurs d’une liste de critères syntaxico-sémantiques définie en amont. Ceux-ci font apparaître toutes sortes de combinaison d’événements datés ou non, ayant duré ou non, de relations de précédence entre eux, avec ou sans durée de séparation ("deux jours avant", "peu après", "à la suite de"), de dates absolues et relatives, etc. Le tout était mélangé avec des critères sur les répétitions et la cause, à la fois pour prévoir la matière pour des études futures, et pour diluer la concentration du phénomène afin qu’il ne soit pas identifié par les traducteurs. Le travail d’annotation et sélection a duré les trois mois d’un stage (M1 d’Ornella Wandji). L’acquisition vidéo s’est faite ensuite dans les locaux de WebSourd dans leurs conditions quotidiennes de travail à l’exception d’une caméra supplémentaire pour synchroniser une vue de côté. Le résultat est un ensemble de 120 brèves en LSF (40 traduites chacune par 3 locuteurs), totalisant 1 h de LSF avec deux vues synchronisées et texte parallèle en français (fig. 21).

Ce travail et la méthodologie en amont ont été décrits et publiés [26]. La ressource produite, premier corpus académique parallèle de traductions français–LSF, s’est avérée féconde pendant des années ensuite, que ce soit comme on l’a vu pour la recherche de règles de productions (§1) mais aussi pour appuyer les travaux en traduction. Elle est aujourd’hui disponible sur OrtoLang¹⁰.

9. Fermée depuis 2015, *WebSourd* était une société basée à Toulouse qui offrait des services aux Sourds, dont une page d’information quotidienne en ligne. Employant de nombreux sourds, c’était une structure reconnue du monde signant en France de son époque.

10. <https://www.ortolang.fr/market/corpora/40-brevs>

3.2 Traduction automatique

Comme pour beaucoup de tâches dans l'ère du TAL moderne, deux approches existent principalement en traduction automatique :

- celle à base de règles, modélisant les connaissances linguistiques et les opérations de transferts de langue source à langue cible ;
- celle reposant sur des algorithmes d'apprentissage automatique (méthodes statistiques) entraînés sur des corpus parallèles pré-alignés, c'est-à-dire des corpus bilingues de traductions annotées avec les correspondances entre les deux versions.

Les approches statistiques à la traduction reposent toutes sur :

1. des modèles qui à des séquences de mots associent des probabilités, donc une hypothèse fondamentale de linéarité selon laquelle les mots/morphèmes sont les éléments de base dont la séquence constitue les phrases, dans lesquelles les éléments sont repérés par leur position ;
2. l'accès à des volumes de données suffisamment grands pour permettre l'entraînement des systèmes d'apprentissage.

Or je ne jugeais à cette époque vérifiable aucune de ces conditions. D'une part, nous avons montré que la LS n'était pas efficacement réductible à une suite linéaire de signes et qu'il convenait de l'appréhender de manière multi-linéaire. D'autre part, les ressources sont extrêmement limitées en LS, et quand bien même nous aurions des centaines d'heures de capture vidéo disponibles et leur traduction en parallèle, nous étions encore loin de l'alignement nécessaire. Ainsi me suis-je dirigé vers une approche à base de règles, mettant AZee à profit.

Fin 2013, je proposais à la communauté SLTAT une idée originale d'architecture de traduction à base de règles de texte à signe [21], que j'ai ensuite publiée dans un article début 2015 [22]. Elle est construite une fois de plus sur la dualité forme-sens, et fait l'hypothèse que la description AZee représentant l'énoncé à signer (l'entrée prévue du système de génération) peut servir directement de représentation pivot pour le discours à traduire, puisque les règles contenues et imbriquées dans celle-ci représentent non seulement une structure grammaticale de l'énoncé mais aussi de fait sa structure sémantique en même temps. Avec cette approche, la traduction n'est plus un double problème de dé-verbalisation d'abord (un concept que nous reprenons en section suivante) de la langue source en une représentation tierce faisant pivot avant une re-verbalisation vers la langue cible. En effet, l'expression AZee déterminant **directement les formes à générer** dans la langue cible, aucun effort de verbalisation n'est nécessaire, le résultat à synthétiser étant directement spécifié par l'expression elle-même. Le problème se réduit à celui d'un transfert du texte vers l'imbrication sémantique AZee.

La proposition est, pour chaque règle ou motif d'expression AZee, de **rechercher les formes de la source portant le même sens**. En d'autres termes et avec un point de vue TAL sur le texte, il s'agit pour chaque règle AZee connue en langue des signes de repérer les formes (en français écrit cette fois) pouvant être interprétées de la manière donnée par le sens de la règle. Lorsqu'une forme textuelle est ainsi repérée, la règle peut être "déclenchée" pour une utilisation dans la traduction à proposer en sortie. Le processus est donc guidé principalement par les opérations possibles en langue cible. Par exemple, la règle « open-list-non-mutex », dont la fonction est l'énumération non-exhaustive d'éléments non mutuellement exclusifs, peut être déclenchée dès lors que l'on observe dans le texte une suite d'éléments séparés par des virgules et de même type syntaxique, et introduite par « comme » ou « par exemple » (cette règle est écrite à des fins d'illustration mais des systèmes entraînés par apprentissage de corpus annotés pourraient tout à fait être

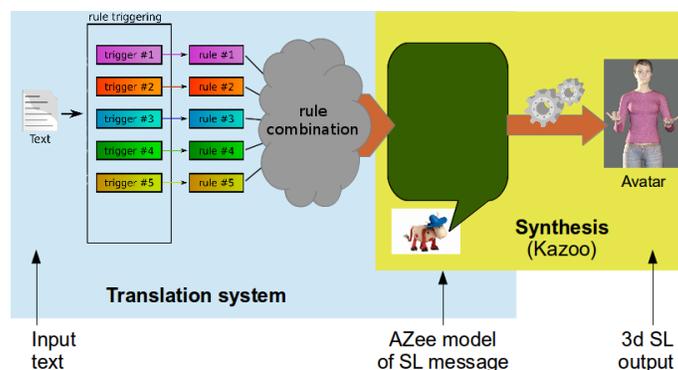


FIGURE 22 – Architecture pour la traduction automatique avec AZee comme pivot

utilisés). Ou encore, ces sous-problèmes peuvent être traités par des déclencheurs entraînés par apprentissage automatique.

Le problème global, illustré sur la figure 22, devient donc la jonction de :

- un ensemble de sous-problèmes TAL de recherche d’information, implémentable par des modules d’analyse textuelle chacun prévu pour déclencher une règle et ses arguments autant que possible (cadre « rule triggering » sur le schéma) et produisant donc chacun des sous-expressions AZee (partielles sans doute) ;
- le problème ensuite de combiner ces sous-expressions (nuage gris) en une sortie AZee (en vert), servant du coup à la fois de sortie du système de traduction (cadre bleu) et d’entrée du système de synthèse (cadre jaune).

En 2014, un petit matériel de départ existait de part et d’autre du nuage du milieu de la figure. D’une part côté droit, un ensemble de plus en plus couvrant de règles (dont celles venant du critère de juxtaposition) se formait, correspondant aux briques de bases pour construire les énoncés à générer. D’autre part côté gauche, de premiers prototypes de déclenchement de règles produisaient en sortie des nœuds ou structures arborescentes partielles correspondant aux règles et arguments déclenchés, peuplant le nuage (stages de L. Patris et B. Testu). L’opération absente et nécessitant une première étude était celle de la recombinaison à l’intérieur du nuage pour fournir les énoncés complets à partir des morceaux déclenchés. À cet effet j’ai encadré un stage de M2 (A. Angerville, 2015) avec l’objectif de proposer une méthode pour définir les principes, indices, priorités ou conditions régissant la combinaison en question.

Le résultat a été limité à des propositions logiques (premier ordre) décrivant des propriétés observées sur un petit corpus de quelques dizaines de combinaisons. Ce travail préliminaire n’a pas été publié car je le juge encore trop spéculatif. Aussi, il nécessitait une identification plus claire des règles AZee utilisables, à l’époque loin de la stabilisation entrevue aujourd’hui. En attendant, je me suis alors tourné vers un autre paradigme lié à la traduction : la traduction assistée par ordinateur (TAO). J’explique cette nouvelle piste en section suivante, mais la traduction automatique à base de règles est revenue sur la table par le biais de deux collaborations.

La première fut avec le Modyco en 2018. C. Bogliotti et D. Battistelli m’ont associé à un projet de psycholinguistique lié à l’apprentissage de la temporalité par les sourds et de son expression en LSF. J’ai co-encadré le stage de S. Bellato, principalement à Nanterre, dont un des objectifs était l’écriture de règles de passage du formalisme de D. Battistelli, spécifiant formellement la sémantique d’expressions temporelles

en français, à AZee. Ce travail assez court a néanmoins pu tester une démarche sur un petit nombre de structures, et l'étudiante a pu publier ses résultats à la conférence RECITAL [2].

Aussi, le projet actuel ROSETTA financé par la BPI implique entre autres le LIMSI et Systran, entreprise française historique du traitement de la langue et de la traduction automatique. Une partie de ce projet prévoit une traduction de français écrit vers langue des signes, avec une traduction à base de règles vers AZee. Nous visons une traduction vers des patrons d'expressions AZee plutôt qu'une détection règle à règle. Ceci est peut-être la clef pour progresser sur ce terrain, mais il est encore un peu tôt pour statuer. Des expériences devraient être menées dans l'année qui vient.

3.3 Traduction assistée par ordinateur (TAO)

Avec cette approche, on n'envisage plus une production automatique quitte à accepter une part d'approximation, mais on impose au contraire une production linguistiquement correcte et équivalente au message source, en conservant un traducteur humain dans le processus. Loin d'être la même approche "en plus soigné" ou "en moins automatique" selon le point de vue, elle correspond en réalité à des applications aujourd'hui encore rigoureusement différentes. Par exemple, l'approche automatique convient à celui qui veut se faire une première idée du contenu d'un site web, et il est impensable d'imaginer un interprète humain près de chaque utilisateur d'internet pour ce cas de figure. Inversement, on ne peut faire l'économie d'un traducteur pour rendre un service ou transmettre des informations fiables à destination d'un public (mairie, état civil...), et l'expertise humaine est donc nécessaire.

En partie suite aux nombreuses lois du monde prévoyant aujourd'hui de plus en plus d'accessibilité aux citoyens, la demande en traduction LS est grandissante mais le monde professionnel concerné n'est pas outillé pour y répondre. De plus, tandis que la traduction de texte à texte est assistée par différents programmes de TAO et traitements logiciels (éditeurs, concordanciers, mémoires de traduction...), les LS ne disposent d'aucun équivalent pouvant assister les traducteurs dans leur tâche. Je me suis donc intéressé à ce qui pourrait être proposé pour réduire ces manques, et ai lancé cette nouvelle thématique de recherche : outiller les métiers de la traduction en langue des signes, en commençant par la constitution d'un réseau d'acteurs impliquant scientifiques et professionnels de terrain.

J'ai d'abord assisté, en mai 2015, au congrès des traducteurs professionnels TAO-CAT à Angers pour faire un état des lieux des pratiques et technologies utilisées dans le domaine de la langue écrite, seule servie en la matière. Aussi, j'entendais approcher les acteurs principaux du marché logiciel pour les sensibiliser à une ouverture à la langue des signes. J'y ai fait la rencontre de la fondation *Olanto* dont le but est de fournir des logiciels libres pour le traitement automatique des langues en général, et de chercheurs de l'*Université de Genève*, tous intéressés par le TAL, la représentation des connaissances et la conception d'interfaces. Pour également associer les praticiens du métier, à savoir les traducteurs et interprètes, j'ai contacté le plus gros service d'interprétation français *Interpretis*, société toulousaine.

J'ai ensuite travaillé à fédérer ce premier groupe en un consortium de recherche Franco-Suisse, notamment en organisant des ateliers et des réunions physiques pour réfléchir ensemble à ce que pourrait devenir le métier de traducteur outillé par des interfaces et systèmes logiciels. J'ai présenté le projet à la Délégation Générale des langues françaises et des langues de France (DGLFLF) qui a octroyé 6000 euros pour financer cette mise en réseau. En avril 2016 nous soumettions collectivement un projet original de 3 ans au guichet de financement conjoint ANR-SNF¹¹ (instrument PRCI). Malheureusement, le projet a essuyé deux refus

11. *Fonds national suisse*, agence nationale de financement de la recherche fondamentale en Suisse.

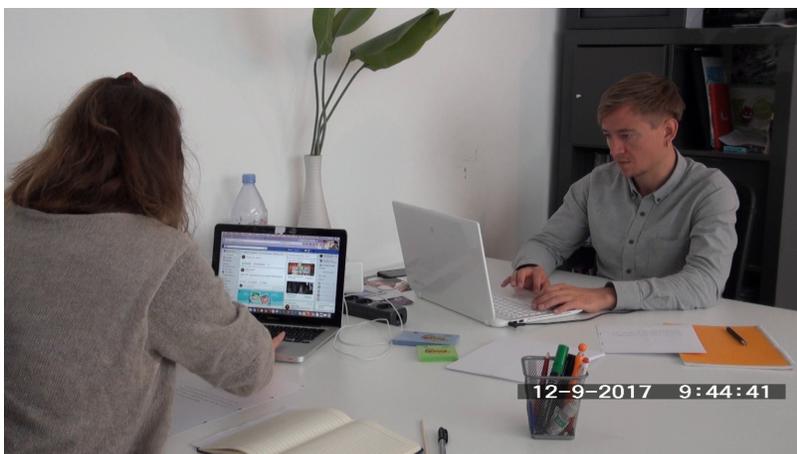


FIGURE 23 – Tournage d'un binôme de traducteurs à la Manufacture Digitale

malgré des relectures et jugement positifs de la SNF, le taux de sélection très bas de l'ANR ayant été avancé comme raison du rejet.

La *Manufacture Digitale*, entreprise adaptée au personnel majoritairement sourd, a ensuite rejoint le rang des collaborateurs professionnels de la traduction en langue des signes. Avec celle-ci, j'ai décroché un projet dans le cadre de l'appel « Langues et Numérique 2018 » de la DGLFLF, qui a renouvelé son soutien cette fois à la hauteur de €40,000 sur 18 mois. Le projet Proto-TAO ainsi financé m'a permis de travailler avec des traducteurs en langue des signes française (LSF) pour étudier leur processus métier, de la phase de préparation face au texte à la livraison du résultat. En effet, un aspect spécifique des traductions LS est qu'en tant que langue visuo-gestuelle, la LS impose le placement d'entités dans l'espace de signation (p. ex. un actant à gauche, un objet à droite...), dont les choix sont souvent naturellement contraints par des réalités extérieures à l'énoncé lui-même. Par exemple, les placements géographiques sont dépendants des cartes géopolitiques ; ceux des personnes sont influencés par leur parenté ou rapports hiérarchiques ; les repères chronologiques fonction des dates, durées et événements en jeu, etc. Le traducteur (p. ex. d'une information journalistique) commence donc typiquement par les recherches encyclopédiques nécessaires et l'élaboration d'un schéma suivant des techniques encore personnelles pour organiser son espace puis son discours.

Pour mieux étudier la pratique du terrain, nous avons filmé des sessions de traduction *in vitro* dans l'entreprise partenaire. Un corpus de textes était préparé, de longueurs et contenus variés selon des critères sélectionnés, et donné à traduire à des binômes de traducteurs de l'entreprise. Cinq sessions d'une demie à une journée chacune ont été filmées dans leurs locaux (figure 23), de la découverte des textes à la livraison d'une traduction face caméra. Ceci nous a fourni des données exploitables pour étudier les problématiques suivantes : que fait un traducteur pour traduire un texte qu'il découvre ? quelles recherches ou questions se pose-t-il systématiquement ? voire, procède-t-il à certaines tâches dans un ordre fixe ? Ces éléments sont utiles à la définition d'une interface de TAO. L'intérêt du travail en binôme est que les traducteurs discutent des problèmes rencontrés, donc la pensée est exprimée et plus observable.

Le projet ROSETTA démarré ensuite au laboratoire a permis de financer une thèse dans ce sillage (Marion Kaczmarek, par ailleurs diplômée interprète), encore en cours et co-encadrée par mes soins (90 %). Pendant son doctorat, l'étudiante a annoté et étudié les vidéos issues de Proto-TAO, et conduit de nouvelles expériences auprès des professionnels (p. ex. entretiens et brainstorming). Trois propriétés centrales dans les interfaces

de TAO texte à texte ont aussi été identifiées comme non transposables aux LS, donc comme nécessitant une solution adaptée aux langue des signes. Ces interfaces reposent toutes en effet sur trois piliers essentiels :

1. un principe de linéarité de la traduction par segments¹² : la traduction de la concaténation des segments d'origine (en langue source) est la concaténation des segments traduits (en langue cible) – autrement dit on peut, à peu près au moins, traduire phrase à phrase en en conservant l'ordre ;
2. une forme écrite pour remplir et éditer les segments cibles ;
3. une mémoire de traduction, automatiquement interrogeable par le système à la sélection d'un segment source et capable d'extraire les segments déjà traduits ou très proches.

Or, en langue des signes, aucun de ces principes ne peut être transposé tel quel :

1. on observe que l'ordre des segments source dans une traduction en LS est rarement conservé, et les traducteurs passent précisément un temps significatif à réorganiser le plan du discours cible (aussi la notion de segment en LSF n'est pas définie) ;
2. la LS n'a pas de forme écrite éditable aujourd'hui ;
3. il n'existe que très peu de ressources alignées français–LSF et les outils d'alignements en TAL prévus pour le texte ne sont pas directement adaptés pour faire intervenir le format vidéo.

Les propositions apportées en réponse sont respectivement :

1. permettre à l'utilisateur de composer un plan librement, sans segments prédécoupés à remplir mais un plan arborescent librement organisable à la souris ;
2. permettre à l'utilisateur d'insérer tout support visuel utile dans les parties définies par son plan (images, zones de texte, mots-clés, schémas...) ;
3. créer une banque d'alignements et un système de requête affichant texte et vidéo en vis-à-vis (requêtable par un concordancier français–LSF, lui-même à adapter).

Un début de prototype d'interface web avait été développé dans un stage (Isabelle Hoxha, été 2018) pour présenter la première proposition ci-dessus à des traducteurs. Un autre stage aujourd'hui en cours (Alix Larroque) lui fait suite, encadré par M. Kaczmarek et moi-même et mieux armé des observations issues de la thèse. Nous espérons faire de premières présentations et expériences auprès des acteurs du métier dans l'année qui vient. La deuxième proposition se limite pour l'instant à des mots (texte), des images ou des vidéos (p. ex. enregistrements webcam), mais pourra inclure à terme des représentations graphiques éditables telles que présentées en section suivante (§4).

La troisième proposition a bien avancé, l'étudiante ayant créé manuellement une première base d'environ 300 alignements dans les 120 vidéos des « 40 brèves », et un concordancier ayant été mis en ligne¹³, ouvert aux tests sur demande (fig. 24). Il a été présenté à une journée d'atelier que nous avons organisée à Toulouse en décembre 2019 auprès de traducteurs et d'étudiants en traduction.

Après deux ans de thèse, six communications avec l'étudiante en premier auteur ont déjà été publiées, dont plusieurs dans des congrès internationaux du domaine : LREC [36], le workshop associé sur le traitement des LS [37], SLTAT [35] ; et une conférence « métier » : *Translating and the Computer* [34]. Une autre est en cours d'évaluation pour le numéro spécial d'une revue.

¹². On appelle « segment » un morceau du document source (à traduire), en général prédécoupé par le logiciel et correspondant à une phrase.

¹³. <https://platform.postlab.fr>

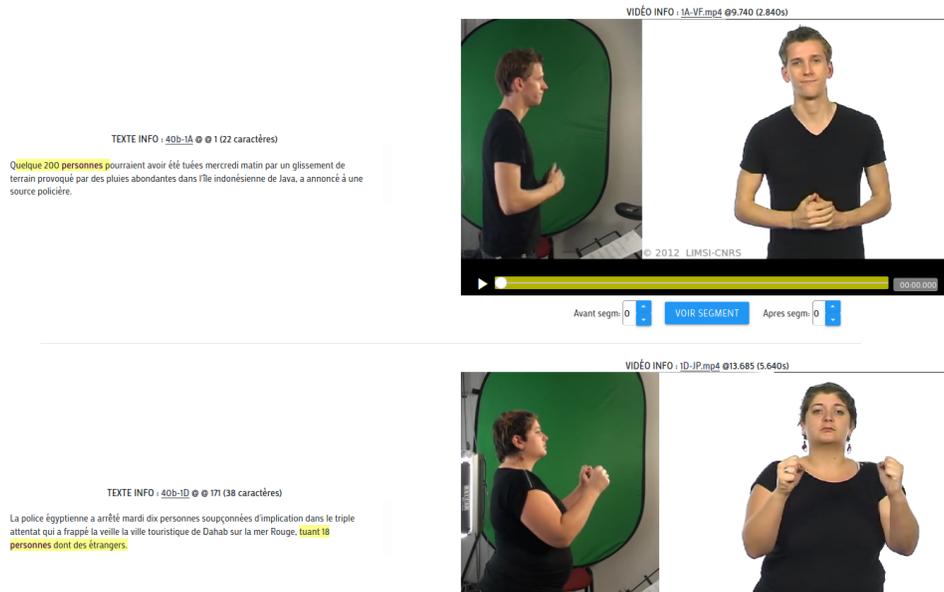


FIGURE 24 – Concordancier français-LSF développé au cours de la thèse de M. Kaczmarek

4 Schématisation et lien avec représentation formelle

Le dernier des thèmes principaux de mes travaux, plus récent, est celui de la standardisation d'une représentation graphique éditable du discours signé. Une raison de considérer cette problématique comme importante est qu'une telle représentation rendrait possible l'édition logicielle, la copie, l'échange et l'économie de stockage de contenus anonymes qui pour autant pourraient rester synthétisables. La « schématisation » étant d'ailleurs une pratique utilisée et enseignée par les professionnels de la traduction notamment, j'ai souhaité explorer ce monde de la représentation graphique, parfois abusivement qualifiée « d'écriture », ce que je présente dans cette section.

4.1 Déverbalisation, schématisation et régularités spontanées

Une tâche essentielle au métier de traducteur en LS est la *déverbalisation*, qui consiste à produire un schéma représentant le sens de l'énoncé source, plus ou moins optimisés directement pour la langue des signes. Le but est de l'utiliser ensuite comme support suffisant (de sens) pour re-verbaliser en LS (forme cible) sans biais induit par le texte (forme source). La place faite à la déverbalisation dans certaines formations d'interprètes et traducteurs est importante, mais les praticiens déplorent qu'aucun standard n'existe pour cette étape qui gagnerait à pouvoir être partagée et son enseignement structuré. Aussi me suis-je intéressé à cette pratique pour entre autres intégrer des modules d'assistance aux schémas dans les environnements de TAO en préparation.

Plus largement, les sourds désirant prendre des notes ou préparer des discours en langue des signes sans passer par le français écrit (langue « étrangère ») ont souvent spontanément recours au même type de schémas pour mettre leur langue sur papier. La schématisation apparaît donc comme un outil essentiel pour plusieurs profils de locuteurs, et ma curiosité croissante pour ces productions graphiques m'a mené à en collecter directement auprès des auteurs que je fus amené à rencontrer, jusqu'à en récupérer plusieurs dizaines de pages, parfois avec une vidéo en LSF ou un texte équivalent en français. Un exemple est donné figure 25,

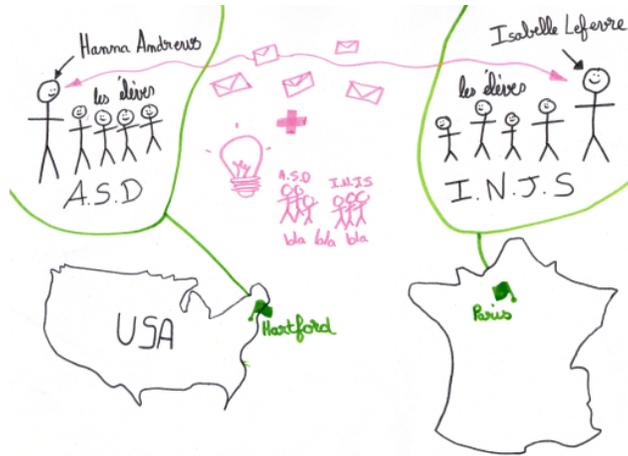


FIGURE 25 – Schéma de verbalisation



FIGURE 26 – Trois captures consécutives correspondant au schéma de la fig. 25

extrait de la production d'un élève de l'Institut National des Jeunes Sourds (INJS) à Paris, schématisant une narration de son professeur.

Mes premières observations ont révélé que ces schémas contenaient de nombreuses régularités spontanées. Par exemple, l'arrangement topologique des entités sur le dessin exactement reproduit dans l'espace de signation (fig. 26), ou l'emploi d'un signe « égal » avec normalisation des tailles de part et d'autre pour signifier que l'un est un état ou une propriété de l'autre (fig. 27). On observe ainsi des constantes entre schémas d'une même personne, mais aussi entre schémas de personnes différentes.

Ceci ouvre la voie à une part de standardisation, déjà souhaitée par certains, tendant vers un format échangeable (non uniquement personnel) et éditable, indexable informatiquement. De plus, les régularités observées correspondaient souvent à des règles de production identifiées en AZee. Un lien pourrait donc également être fait avec ce modèle formel de représentation, ce qui sera exploré plus bas (§4.4).



« le lion est gentil » « Fidel Castro va bien »

FIGURE 27 – Deux exemples d'utilisation du signe « = »

À mesure que je comprenais la valeur de ces productions graphiques, je me suis convaincu qu'un nouveau champ de recherche se révélait, à la fois lié à la représentation profonde (intime) de la langue par ses locuteurs et susceptible d'apporter une solution quant à l'édition de LS dans les interfaces logicielles. Malheureusement, il s'agit en général de productions personnelles et éphémères que les auteurs ne s'intéressent pas à conserver. Ainsi les schémas ont-ils toujours échappé à la recherche scientifique qu'ils mériteraient. Pour démarrer sur ce nouveau terrain de recherche, j'ai décroché un premier financement par le laboratoire (soutien interne dit « action incitative »), puis un projet DIM-RFSI de la région Île-de-France, en 2018.

4.2 Constitution d'un corpus de schémas

Pour une étude plus méthodique et des analyses plus contrôlées, la première étape est de constituer un corpus de référence de tels schémas. Comme tout corpus, ceci requiert en amont la préparation d'une méthodologie et un matériel d'élicitation, afin d'équilibrer les critères pesant sur les schémas. Ensuite seulement, une campagne de collecte à proprement parler peut être conduite et des analyses menées sur des données riches et documentées.

Plusieurs critères ont été utilisés pour calculer l'équilibre des phénomènes attendus dans les schémas. En premier lieu, nous avons identifié trois types de situation pouvant faire usage de schémas, de nouveau inspirés par la dualité forme–sens :

- prise de notes d'un discours signé, où le scripteur est en quelque sorte *sténographe* et ne décide que de la forme sur papier ;
- élaboration d'un discours en vue de le signer, où le scripteur est *auteur* du contenu entier ;
- traduction d'un texte écrit vers la LS, où le scripteur crée une forme résultat dont le contenu sémantique est contraint.

La différence entre ces trois situations sont résumées dans le tableau suivant :

	forme	sens
sténographe	<i>fixée</i>	<i>fixé</i>
auteur	<i>non fixée</i>	<i>non fixé</i>
traducteur	<i>non fixée</i>	<i>fixé</i>

On peut s'attendre à ce qu'un schéma « sténographique » inclue davantage d'éléments de forme puisqu'ils auront été directement observés par le scripteur, tandis que dans le cas « auteur », il peut se limiter à des productions plus sémantiquement motivées et laisser une part d'implicite pour le moment de la réalisation. S'il y a un jour standardisation, c'est sans doute un équilibre de ces tendances qui devra être atteint. Il faut donc les étudier toutes de manière équilibrée.

Ainsi, trois **types d'élicitation** ont été prévus pour recréer les situations ci-dessus, à équilibrer entre les tâches des participants, à savoir respectivement : vidéos en LS, consignes libres, textes en français.

Deuxièmement, un critère de **genre** a été choisi, vu la nature iconique de la langue et l'impact que peut avoir la différence entre les discours neutres et les contes par exemple. Nous avons équilibré les élicitations sur trois genres :

- conte/légende, qui contient nombre d'actants et actions contrastés, avec plein engagement du narrateur ;
- reportage journalistique, qui contient des actions de toutes sortes et des événements bien vécus, mais sans engagement du narrateur ;

- définition, qui impose en général un ton plus détaché, factuel et intemporel, à savoir sans émotion liée à une circonstance.

Ensuite, nous avons varié les **longueurs** (ou durées pour les vidéos) des tâches vu l'impact sur la mémoire au moment de restituer la production en LSF, donc l'impact potentiel sur le niveau de détail des schémas. En effet, l'équivalent d'une phrase de texte peut plus facilement être livré par cœur, au contraire des énoncés longs. Nous avons prévu deux catégories : les courts (3–4 lignes de texte ou moins de 30 s de vidéo), et les longs (1 page ou 1.5 à 2 min). Ces catégories sont issues d'un travail annexe effectué sur l'élicitation de productions longues pour limiter les disfluences dans la captation [19].

Enfin, nous avons ajouté des critères sur le nombre d'**actants**, de **localisations** et de **déplacements**, vu l'impact sur l'utilisation de l'espace en LS. En effet, les actants et localisations sont souvent ancrés dans l'espace de signation, et les déplacements fonction de ancrages précédents, or les schémas en dépendent fortement.

Armés de tous ces critères, nous avons pu établir un tableau de tâches à fournir aux participants selon leur profil (traducteur professionnel, sourd signant, etc.). Une première campagne de collecte a eu lieu, principalement auprès de traducteurs, et plus d'une centaine de schémas ont déjà pu être collectés. La méthode ici présentée et la composition du matériel d'élicitation ont été publiés dans les actes du dernier LREC [16].

4.3 Représentation éditable et écritures

On envisage de standardiser, au moins partiellement, le format de ces schémas pour en permettre une édition logicielle, que ce soit pour de simples échanges entre personnes (traducteurs dans un service ou particuliers par e-mail) ou une animation par un avatar. L'objectif est alors, autrement dit, le codage et le décodage non ambigu de productions signées, dans un format inspiré par les schémas observés.

Ces travaux se rapprochent donc nettement de la question d'une forme écrite pour la LS. Mais je me tiens à l'écart de la volonté de « définir un système d'écriture ». En effet une telle entreprise dépasse selon moi le chercheur seul ou son équipe, engage nombre de situations quotidiennes et d'utilisateurs aux profils différents, et peut impacter la culture et le recul méta-linguistique des locuteurs d'une façon que je ne saurais estimer. J'indique au contraire qu'un code éditable n'a pas nécessairement vocation à servir de script comparable aux écritures humaines, et qu'il n'est a priori pensé que pour les applications envisagées au départ.

Je présume néanmoins que pour peu que la communauté se dote et pratique un jour une écriture effective de sa langue, elle souhaitera naturellement que ce soit celle qui se voit intégrée aux logiciels d'édition, et ne pas avoir à apprendre de code ou langage supplémentaire. À l'inverse, on observe que des scripts proposés pour les LS (p. ex. SignWriting, cf. fig. 28) et dotés de logiciels ont suscité un certain intérêt pratique par endroits, ce qui laisse penser qu'au moins une partie de la communauté puisse être en demande de propositions. Et si celle-ci s'empare d'un outil logiciel disponible et de sa forme graphique sans la rejeter, les réflexions sur la forme écrite idéale seront catalysées et en partie influencées. Ces remarques faites, il est nécessaire d'être prudent car il me semble impossible de postuler une séparation totale avec la question de l'écriture.

Pour un meilleur recul sur cette problématique aussi passionnante que riche et complexe, et un meilleur positionnement du format « schéma » de représentation des signes, j'ai débuté des réflexions il y a plus d'un an sur la nature et le statut de telles productions, et les ai comparées aux différents scripts connus. Le point fort de ce travail a été de les envisager comme productions spontanées (et non des systèmes), et de les comparer

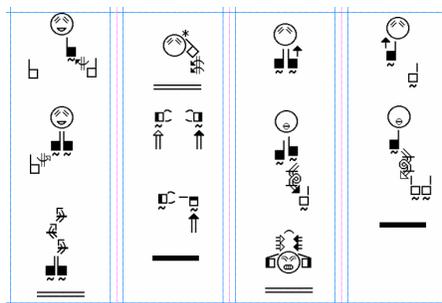


FIGURE 28 – SignWriting [44]

d’une part aux systèmes d’écriture déjà établis pour les langues vocales du monde (mais non signées), et d’autre part aux systèmes conçus (mais non établis) pour les langues des signes.

En résumé, les écritures sont en premier lieu distinguées en deux classes : les scripts logographiques et les scripts phonographiques. Les premiers sont composés de symboles représentant chacun un élément de sens sans indication a priori de sa prononciation, qui ne sera déterminée que par la connaissance de la langue. Les seconds se subdivisent en plusieurs types, mais consistent tous en symboles qui au contraire représentent des sons à articuler, sans indication a priori de leur sens, qui ne sera de même accessible qu’avec la connaissance de la langue. L’exemple le plus clair d’écriture logographique est l’écriture chinoise, tandis que le français écrit se range en système phonographique. Mais aucune écriture n’est en réalité purement d’une classe ou de l’autre. Tous les systèmes mélangent des éléments des deux, en proportions inégales. Par exemple, la différence entre « a » et « à » en français n’a aucune justification phonographique ; elle est la marque d’une différence de sens.

D’autres propriétés peuvent également être attribuées à chaque script, comme le sens de lecture. Ceux-ci sont divers : gauche à droite puis haut en bas (cas du français), droite à gauche d’abord, haut en bas d’abord, voire un choix plus libre... La linéarité en revanche, abstraction faite de son sens de lecture, est un invariant dans les écritures des langues vocales.

En comparaison, j’ai dénombré une douzaine de systèmes conçus pour les LS (exemples fig. 29), et tous sont d’inspiration intégralement phonographique, au sens où tout symbole du système spécifie une articulation à produire, et non directement un concept ou une opération sémantique. Principalement, il s’agit d’articulations manuelles, et c’est la grille d’analyse paramétrique (Stokoe) discutée plus haut que l’on retrouve comme structurant le système (combinaison de symboles, un pour chaque paramètre/main). Aussi, tous supposent la séquence de signes traditionnelle également discutée, chacun faisant l’objet d’une représentation plus ou moins graphique dans une séquence linéaire équivalente. La plupart du temps, celle-ci se lit de gauche à droite et de haut en bas.

Il est frappant de constater l’invariance du choix phonographique dans tous les systèmes proposés, alors que d’autres types de scripts existent et n’ont jamais été considérés. En aucun cas le choix est-il même justifié par leur auteur, sans doute influencé par la culture alphabétique occidentale dont relève tous les auteurs concernés. Par contraste, les schémas spontanés sont principalement logographiques, ce qui explique à mon avis que les sourds ne se soient pas naturellement emparés de tels systèmes. Je postule que si spontanément les locuteurs se tournent vers une stratégie donnée, celle-ci doit inspirer les systèmes proposés pour la standardisation.



FIGURE 29 – Systèmes d’écriture conçus pour les LS

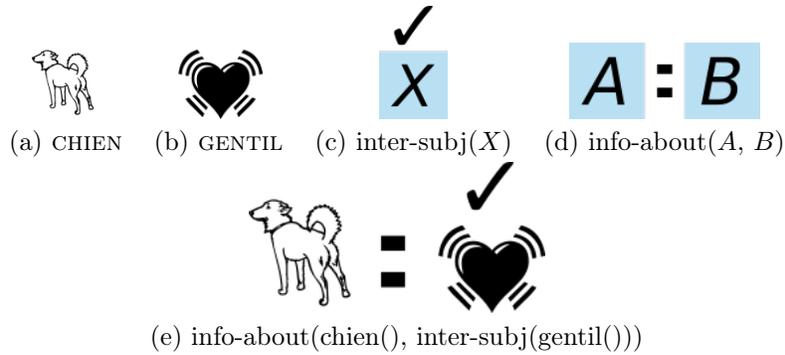


FIGURE 30 – Combinaison de quatre formes graphiques (a–d) pour représenter une expression AZee (e)

4.4 AZee comme piste pour une forme graphique éditable

Comme indiqué plus haut, des régularités ont été observées non seulement entre formes graphiques spontanées, mais aussi entre celles-ci et règles de production AZee ou des expressions simples. Par exemple, le dessin « $A : B$ », dont des exemples sont donnés en figure 27, correspond à l’expression AZee « *info_about(A, B)* ».

Si ces régularités d’associations entre formes graphiques et règles de production AZee se confirment, un pont entre AZee et des schémas standardisés peut être construit en spécifiant des formes ou combinaisons graphiques pour chaque opération AZee connue (règle de production). En les spécifiant de telle sorte que ces formes peuvent récursivement s’imbriquer à l’image des règles qui leur correspondent, on peut construire des schémas complexes équivalents aux expressions AZee, comme illustré en figure 30. De cette manière, toute combinaison graphique peut se résoudre en une expression AZee, et toute expression AZee étant calculable et déterminante de formes à signer, les schémas le sont aussi. En suivant un tel principe, nous définissons donc une représentation graphique des discours en LS directement animable par un système de synthèse.

Pour une visualisation plus large du rendu possible d’une telle proposition, je l’ai matérialisée en proposant une forme graphique à chaque règle de production connue, inspirée des schémas recueillis si une forme régulière y était déjà observée (fig. 31), et les ai appliqué ensuite à une expression AZee complexe représentant un

discours réel. En guise d'exemple complet, le discours *La Bise et le Soleil*, visible sur <https://atlas.limsi.fr/?tab=LNT> en cliquant sur « LSF », est représenté figure 32.

Grâce à une telle visualisation, on peut développer une caractérisation du genre de script potentiel que serait celui proposé ici, en discutant sa place vis-à-vis des propriétés traditionnelles envisagées pour les écritures, et le comparer aux systèmes conçus pour les LS. Par exemple, on note que comme toutes les langues écrites du monde, celui-ci est un mélange déséquilibré d'éléments logographiques (majoritaires ici) et phonographiques, ce qui n'est le cas d'aucun script conçu pour les LS. Inversement, il est largement iconique, au contraire de la moitié des scripts conçus, et de toutes les écritures établies. Sa particularité est d'être récursivement planaire, c'est-à-dire qu'il se développe non linéairement mais en deux dimensions sur la feuille, s'apparentant à aucune écriture connue mais bien à un script existant : la notation arithmétique. Tout ce travail d'étude et de comparaison des scripts existants ainsi que cette proposition de forme graphique a été consacré par une acceptation pour publication dans la revue *Sign Language Studies*, à impact significatif pour ce qui est des études sur les langues signées [17].

À ce stade les hypothèses restent spéculatives et les propositions exploratoires. Mais après avoir défriché ce terrain nouveau, j'ai pu mettre en avant une approche originale pour l'édition logicielle et la définition d'un standard de représentation graphique pour la LS, qui me semble mériter d'être développé. Cela fait partie de mon projet pour les années à venir, exposé dans la suite de ce document.

Troisième partie

Projet

Pour les années à venir, je souhaite poursuivre chacun des quatre axes présentés ci-dessus : la modélisation formelle de la langue et son fonctionnement (un effort continu depuis le début de ma carrière), la synthèse (qui va de pair et permet l'évaluation continue), la TAO et la représentation graphique éditable. Les sections suivantes expliquent tour à tour le projet principal qui portera chacun de ces axes de recherche. Les trois derniers bénéficient déjà d'un soutien conséquent garanti dans le futur proche, à savoir des projets d'envergure, deux BPI et un européen, ce qui me donnera les moyens d'un travail fourni sur chacun, assisté par des recrutements dans tous les cas.

5 Grammaire formelle

Comme nous l'avons montré, le travail d'extraction et de spécification de règles de production en AZee a permis d'atteindre un ensemble relativement stabilisé et couvrant de la LSF. En particulier sur le genre journalistique mais pas uniquement celui-ci, plus d'une heure au total de discours réels ont pu être transcrits en des expressions AZee combinant environ 6500 applications de ces règles.

Par construction, le système formel qu'est AZee permet l'application mutuellement récursive de celles-ci, ce qui permet un nombre infini d'expressions. Du point de vue du langage natif et de son compilateur, la seule condition pour une expression est d'être bien formée, et sans erreur de typage. Ainsi, lorsqu'un argument de type « SCORE » (blocs rouges sur les schémas) est attendu par une règle de production (AZop), n'importe quelle valeur de ce type constitue une expression bien formée. Mais du point de vue de la langue des signes,

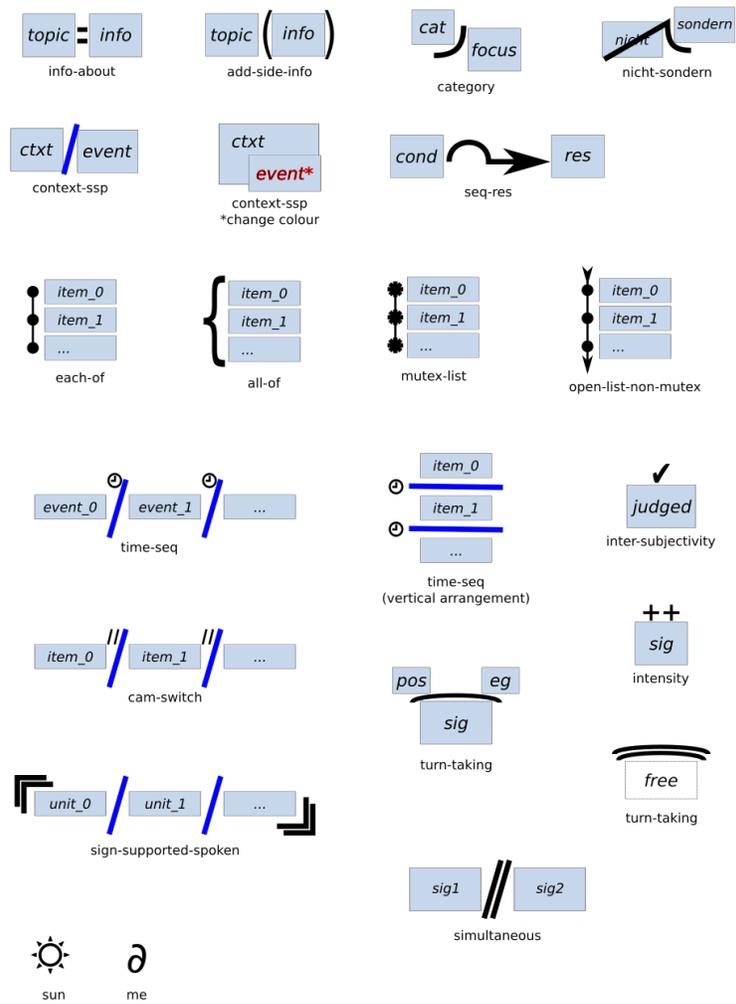


FIGURE 31 – Formes graphiques pour règles (AZops) de production

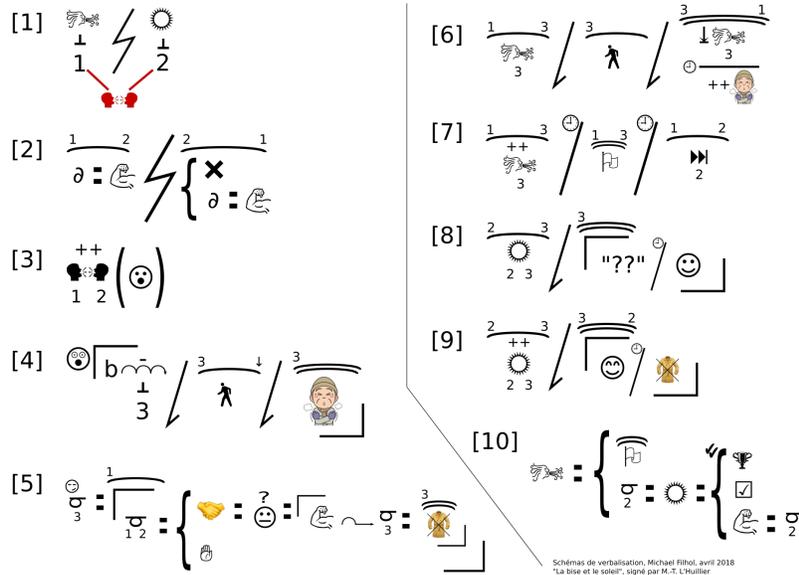


FIGURE 32 – Forme graphique pour *La Bise et le Soleil* en 10 parties

il est raisonnable de conjecturer que toutes les combinaisons ne sont pas équiprobables, que certaines sont peut-être très rarement observées en corpus, voire seraient spontanément rejetées par les locuteurs.

Il serait donc intéressant d'étudier statistiquement les structures et motifs apparaissant dans les expressions AZee qui génèrent ces discours attestés en LSF. En effet, ces expressions sont en quelque sorte l'équivalent AZee des « structures profondes » (quand bien même sémantiques) générant celles « de surface » observées (formes). Avec des arguments distributionnels et statistiques, des contraintes pourraient être formulées sur les motifs d'imbrication qu'elles contiennent, et on pourra voir en celles-ci une valeur grammaticale au sens où elles régissent la structure profonde des énoncés. Ainsi qu'on le fait en syntaxe de l'écrit, des catégories pourront être définies pour regrouper des éléments aux propriétés similaires comme « utilisables comme argument de [...] ». D'autres propriétés formelles peuvent entrer en jeu comme la profondeur de sous-expressions, etc.

Camille Challant, étudiante s'étant intéressée à AZee en juin 2020 et ayant produit la première ressource citée ci-dessus, a souhaité étudier cette problématique pour son stage de M2, que j'encadrerai au laboratoire à partir de février 2021. Elle a le projet de demander une bourse de thèse pour poursuivre ensuite sur cette voie d'une grammaire formelle en termes AZee. Si ce travail trouve des résultats convaincants, par exemple sous la forme de règles prédictives sur la composition des discours en LS, il aura confirmé la capacité d'AZee à outiller non seulement les programmes de synthèse mais également les linguistes avec un langage permettant de formuler des contraintes grammaticales sur la LS étudiée.

6 Synthèse d'AZee

Plus qu'une simple application fil rouge depuis le début de mes travaux en modélisation, la synthèse de LS est parfois motrice de l'effort de modélisation notamment en permettant une évaluation en continu des spécifications de forme. Je souhaite progresser dans cette activité, autant pour ce qu'elle apporte indirectement

à la théorie que pour ce qu'elle permet d'applicatif. En particulier, elle sera soutenue sur les trois prochaines années par deux projets en train de démarrer.

Le premier projet est celui d'un serveur gestuel, objectif comparable à celui d'un serveur vocal par téléphone, porté par l'entreprise grenobloise IVèS (financement BPI). Celui-ci prévoit une thèse en analyse-synthèse de mouvements de la langue des signes, avec un démarrage dans le premier semestre 2021. AZee spécifie les mouvements à produire ainsi que leur synchronisation, mais comme nous l'avons vu, la spécification formelle pure au plus bas niveau de l'articulation, si utilisée telle quelle, pose des problèmes de dynamique du rendu qui tend à paraître plus robotique. La solution de Paula est de combiner des blocs pré-animés, de granularité plus grosse, plutôt que de combiner uniquement des contraintes articulatoires élémentaires. Ces blocs peuvent être de différents types : enregistrés à partir de capture de mouvement, générés par un processus/modèle géométrique dynamique, pré-animés manuellement par un artiste... Pour l'instant, le choix est intuitif ou le résultat d'un tâtonnement, et les pré-animations faites à la main. Le but de la thèse est de déterminer la meilleure façon d'animer chaque élément spécifié par AZee et d'implémenter un système pour évaluation, par exemple avec Blender, plateforme libre d'animation 3D. En amont, des études pourront être menées sur des données réelles de langue des signes fournies, sous forme de capture de mouvement (mocap) ou de vidéo, et d'autres pourront aussi être enregistrées au besoin.

Le deuxième projet est « EASIER », projet européen (ICT57) que j'ai décroché avec un consortium de 14 partenaires et 6 nations représentées, traitant de la traduction automatique et de la post-édition/TAO et démarrant début 2021. L'un des lots est consacré à la synthèse, et prévoit l'extension et l'amélioration du système Paula, prévu pour générer les traductions en sortie. John McDonald de DePaul sera recruté au laboratoire six mois à temps plein au total, en trois fois sur les trois prochaines années pour collaborer de manière rapprochée, ce qui devrait donner un coup d'accélérateur à la jonction Paula-AZee.

7 Traduction assistée

Cette part de mon activité se poursuivra principalement avec l'encadrement de la thèse de M. Kaczmarek (jusqu'à fin 2021) pour le volet recherche. J'espère y donner une réalité logicielle en atteignant un premier prototype d'environnement logiciel permettant des démonstrations auprès des professionnels du métier au plus vite, voire si possible de telle sorte que M. Kaczmarek puisse s'appuyer dessus dans sa dernière année.

Alix Larroque est déjà en stage avec nous depuis septembre 2020 avec l'objectif de fournir une intégration des premières propositions issues de la thèse. Ce prototype est déjà avancé (cf. fig. 33), et un concordancier existe par ailleurs prêt à être intégré. Il est donc envisageable de proposer, par exemple lors d'un nouvel atelier du même genre que nous avons déjà organisé deux fois avec les experts du terrain ces dernières années, des sessions de démonstrations auprès des professionnels.

Dès le début 2021, je vais travailler avec Hannah Bull, doctorante au laboratoire en reconnaissance de LS vidéo dont je participe à l'encadrement (25 %) pour les aspects TAL et linguistique, Michèle Gouiffès sa directrice, et Marion Kaczmarek pour ajouter au concordancier une option de recherche par la vidéo. Nous intégrerons le résultat à la plateforme de TAO en cours de développement pour proposer une recherche bidirectionnelle dans le concordancier, ce qui n'est pas encore le cas vu la différence de modalité. Nous avons obtenu un stage financé par le laboratoire à cet effet. Ce mini-projet est intéressant à la fois pour faire progresser nos outils à l'adresse des traducteurs, et pour valoriser les apports scientifiques provenant des différentes activités de l'équipe et en consolider les échanges et l'interdisciplinarité.

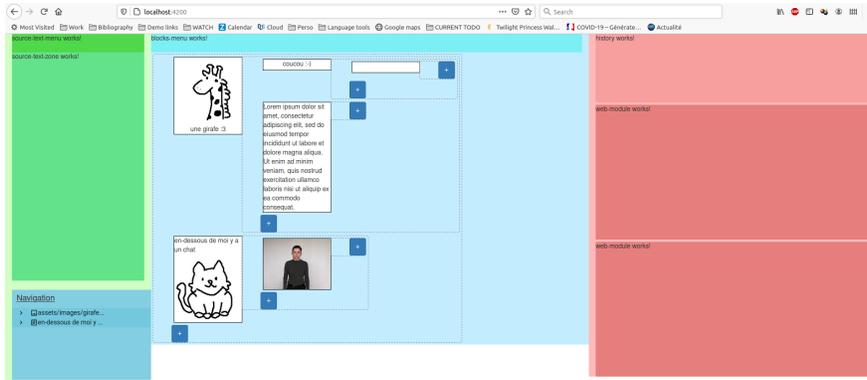


FIGURE 33 – Prototype d’interface TAO en cours de développement

Pour la suite, une question nouvelle que j’aimerais traiter dans les années qui viennent, et dont la solution serait intégrable au prototype, est l’assistance à la livraison du résultat final d’une traduction. En effet dans le cas de traductions longues, une fois la structure (plan) du discours préparée et les choix terminologiques effectués, les traducteurs traduisent souvent par morceaux montés ensuite car il n’est pas possible de livrer la totalité d’un discours par cœur. Le problème vient du fait que les coupes en LS impliquent une co-articulation et une apparence et un fond (visible derrière le locuteur) constant. On ne peut donc procéder comme pour une langue vocale, à savoir compter sur un silence entre les morceaux ou étaler le travail sur plusieurs jours et plusieurs lieux avec un accoutrement et des lieux qui varieraient).

La solution envisagée est comparable à celle utilisée par les présentateurs de télévision en direct qui sont soumis aux mêmes contraintes : par-cœur impossible, regard mobilisé et montage impossible. C’est-à-dire, un « prompteur » qui une fois le plan finalisé et son contenu fourni par le traducteur (par exemple un système graphique associé à des mots-clés), pourrait le faire défiler d’une manière adaptée à la LS et qui permettrait de livrer d’une traite un discours préparé, même en l’absence de forme écrite. C’est une pratique déjà observée par le biais du papier, qui pourrait tout à fait bénéficier d’une assistance logicielle utile. Cela nécessitera en revanche des expériences auprès d’utilisateurs pour ajuster les options et la présentation aux pratiques effectives, et mesurer le gain en confort et en temps de préparation.

8 Schémas et forme graphique

Sur la fin de la période, j’ai exploré un terrain nouveau de la représentation des LS : la schématisation graphique par ses locuteurs et la possibilité d’une standardisation éditable (cf. §4). À partir d’une collecte informelle d’abord puis d’un corpus élicité, j’ai découvert dans ces productions spontanées des régularités permettant déjà certaines conjectures, et d’observer une certaine proximité avec AZee. Ce résultat m’est apparu comme fascinant, aussi je souhaite pousser ce champ de recherche dans les années à venir.

Pour commencer, l’étudiante Chloé Paris a contacté le laboratoire pour effectuer son stage de M2 sur le thème de l’écriture des langues des signes, et a pris connaissance de ces travaux. Elle a souhaité y participer, et sera donc hébergée pour 5 mois à compter de mars 2021. L’objectif sera d’une part d’augmenter le nombre de schémas du corpus en organisant une collecte supplémentaire en suivant la méthodologie présentée plus haut, et d’autre part effectuer une analyse statistique de ce corpus. Chloé Paris souhaite déjà vivement poursuivre en thèse.

Par ailleurs, avec 13 partenaires répartis dans 6 pays de l'UE, j'ai monté le projet européen « EASIER » (EU-ICT57) que nous avons tout juste décroché pour un démarrage en 2021. Celui-ci traite de traduction automatique et d'édition logicielle (et post-édition) par des traducteurs humains, tâche pour laquelle ces recherches sur la schématisation et les formes graphiques trouvent une application directe ainsi par ailleurs que mes travaux en TAO. Un « work package » entier est consacré à la post-édition et à l'édition graphique, et dont je serai responsable scientifique.

Un des objectifs dans cette partie du projet est de construire un éditeur de schémas complexes tels que proposés dans mes travaux récents, et de le relier au compilateur AZee et au système d'animation Paula. Ainsi l'auteur des schémas pourra générer et visualiser directement une animation pour chacune de ses productions graphiques. Un tel éditeur permettra donc à tout utilisateur, sans apprendre le formalisme AZee, de composer son schéma et d'en visionner une lecture en langue des signes. Cette application est comparable à une synthèse vocale de texte, mais opérant à partir d'une représentation graphique, donc humainement accessible, et dont les entités et les liens sont sémantiques !

J'ai l'espoir que ceci constitue à l'avenir un levier pour engager et inclure la communauté signante sur la question de l'outillage logiciel et de la représentation graphique de leur langue, inclusion qui constitue toujours un critère essentiel d'acceptation sociale des recherches en LS. Aussi, il se peut que les propositions faites dans ce travail avant tout d'outillage logiciel soient reprises en partie dans le quotidien de ses utilisateurs. Un tel phénomène marquerait le début d'un échange fécond avec les usages du terrain et la communauté sourde, et pourrait même encourager l'émergence d'une pratique écrite « en langue des signes » dans la communauté signante.

Mes choix d'études et le chemin parcouru depuis le début de ma thèse jusqu'à aujourd'hui a été en grande partie motivé par l'envie de comprendre comment le sens se code en langue humaine – ou plus exactement peut-être comment la langue crée du sens pour l'humain. La découverte de nouvelles langues m'a toujours passionné autant comme fin que comme moyen pour prendre le recul que cela suppose. Celle de la langue des signes m'a révélé une façon élégante de parler du monde, et qui en apprend une se rend vite compte des liens étroits et privilégiés que celle-ci entretient avec le sens de ce qu'elle sert à dire. En outre à mesure que j'en découvrais la puissance à mes débuts, j'en mesurais aussi l'effacement, forcé ou regretté, face à un besoin qu'une communauté entière qualifie pourtant de vital. Ce fut une deuxième raison pour moi de m'attacher à la doter, d'outils de traitement automatique d'une part, pour ses utilisateurs et les professionnels avec les sourds comme premiers intéressés, mais aussi de modèles et de connaissances théoriques d'autre part, pour des découvertes plus fondamentales, avec le débat et la rationalisation scientifiques comme premiers leviers.

Après quinze ans de participation à l'entreprise collective qu'est la science, je souhaite aujourd'hui encore développer les axes de recherche que j'ai présentés plus haut, en y associant et en accompagnant les vocations scientifiques futures. Les différents projets d'envergure en cours ou démarrant cette année sécurisent pleinement l'environnement dont j'aurai la responsabilité pour plusieurs années à venir.

Références

- [1] J.A. Bangham, S.J. Cox, R. Elliott, J. Glauert, I. Marshall, S. Rankov, and M. Wells. Virtual signing : Capture, animation, storage and transmission – an overview of the visicast project. In *IEE Colloquium on Speech and Language Processing for the Disabled and Elderly*, London, United Kingdom, 2000.

- [2] Sandra Bellato. Vers la traduction automatique d’adverbiaux temporels du français en langue des signes française. In *Actes de la conférence RECITAL (TALN-RECITAL@PFIA)*, Toulouse, 2019.
- [3] Christian Cuxac. *Langue des signes française, les voies de l’iconicité*, volume 15–16. Ophrys, 2000.
- [4] Christian Cuxac. Phonétique de la LSF : une formalisation problématique. *La linguistique de la LSF : recherche actuelles, Silexicales*, 4, 2004.
- [5] Bastien David and Pierrette Bouillon. *Traduction automatique de la parole vers la langue des signes de Belgique francophone. Évaluation d’un avatar destiné aux transports en commun par la communauté sourde*, pages 143–148. 10e conférence de l’IFRATH sur les technologies d’assistance. Recherches pluridisciplinaires pour l’autonomie des personnes en situation de handicap. 2018. ID : unige :105673.
- [6] Maxime Delorme. *Modélisation du squelette pour la génération réaliste de postures de la langue des signes française*. phdthesis, Université Paris-sud, Orsay, France, December 2011.
- [7] S. Ebling and J. Glauert. Exploiting the full potential of jasingning to build an avatar signing train announcements. In *Third International Symposium on Sign Language Translation and Avatar Technology*, Chicago, IL, USA, October 2013.
- [8] Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and François Goudenove. Dicta-sign : Sign language recognition, generation and modelling with application in deaf communication. In *4th Workshop on the Representation and Processing of Sign Languages : Corpora and Sign Language Technologies*, Valletta, Malta, 05 2010.
- [9] Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and François Lefebvre-Albaret. The dicta-sign wiki : Enabling web communication for the deaf. In K. et al. Miesenberger, editor, *ICCHP, Part II, LNCS 7383*, pages 205–212, Berlin Heidelberg, 2012. Springer-Verlag.
- [10] Ralph Elliott, John R. W. Glauert, Vincent Jennings, and Richard Kennaway. An overview of the sigml notation and sigml signing software system. In *Proceedings of the 4th Language Resources and Evaluation Conference (LREC)*, pages 98–104, Lisbon, Portugal, 2004.
- [11] Ralph Elliott, John R. W. Glauert, Richard Kennaway, Ian Marshall, and Éva Sáfár. Linguistic modelling and language processing technologies for avatar-based sign language presentation. *Universal access in the information society (UAIS)*, 6(4) :375–391, 2008.
- [12] M. Filhol, A. Braffot, and S. Matthes. Dictasign deliverable d4.2 : Sentence descriptions for BSL, DGS, GSL, LSF, 2010. DictaSign project, EU-FP7.
- [13] Michael Filhol. *Modèle descriptif des signes pour un traitement automatique des langues des signes*. Theses, Université Paris sud, June 2008.
- [14] Michael Filhol. Stabilisation d’un modèle à contraintes pour la représentation lexicale des langues des signes. In *Actes de la conférence TALN, traitement automatique des langues des signes*, 2010.
- [15] Michael Filhol. Combining two synchronisation methods in a linguistic model to describe sign language. *Springer LNCS/LNAI*, 7206, 2012.
- [16] Michael Filhol. Elicitation and corpus of spontaneous sign language discourse representation diagrams. In *Proceedings of the 9th workshop on the representation and processing of sign languages*, 2020.

- [17] Michael Filhol. A human-editable sign language representation inspired by spontaneous productions... and a writing system? *Sign Language Studies*, 21(1), 2020.
- [18] Michael Filhol and Mohamed Nassime Hadjadj. Juxtaposition as a form feature; syntax captured and explained rather than assumed and modelled. In *Proceedings of the workshop on Representation and Processing of Sign Languages*, 2016.
- [19] Michael Filhol and Mohamed Nassime Hadjadj. Elicitation protocol and material for a corpus of long prepared monologues in sign language. In *Proceedings of the workshop on Representation and Processing of Sign Languages*, Miyazaki, Japan, May 2018.
- [20] Michael Filhol, Mohamed Nassime Hadjadj, and Annick Choisier. Non-manual features : the right to indifference. In *Proceedings of the workshop on Representation and Processing of Sign Languages*, Reykjavik, Iceland, 2014.
- [21] Michael Filhol, Mohamed Nassime Hadjadj, and Benoît Testu. A rule triggering system for automatic text-to-sign translation. In *International Workshop on Sign Language Translation and Avatar Technology (SLTAT)*, Hamburg, Germany, September 2013.
- [22] Michael Filhol, Mohamed Nassime Hadjadj, and Benoît Testu. A rule triggering system for automatic text-to-sign translation. *Universal access in the information society (UAIS)*, 15(4), 2015. C. Stephanidis (ed.), special issue on "Recent Advances in Sign Language Translation and Avatar Technology", R. Wolfe, E. Efthimiou, J. Glauert, T. Hanke, J. McDonald, J. Schnepp (guest eds.), Springer.
- [23] Michael Filhol and John McDonald. Extending the azee-paula shortcuts to enable natural proform synthesis. In *Proceedings of the workshop on Representation and Processing of Sign Languages*, pages 45–52, Miyazaki, Japan, May 2018.
- [24] Michael Filhol and John McDonald. The synthesis of complex shape deployments in sign language. In *Proceedings of the workshop on Representation and Processing of Sign Languages*, Marseille, France, May 2020.
- [25] Michael Filhol, John McDonald, and Rosalee Wolfe. Synthesizing sign language by connecting linguistically structured descriptions to a multi-track animation system. In *International Conference on Human-Computer Interaction*, Vancouver, Canada, 2017.
- [26] Michael Filhol and Xavier Tannier. Construction of a french–LSF corpus. In *LREC workshop on Building and Using Comparable Corpora workshop (BUCC)*, 2014.
- [27] John Glauert and Ralph Elliott. Extending the sigml notation. a progress report. In *Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT)*, Dundee, Scotland, 11 2011.
- [28] T. Hanke et al. Visicast deliverable d5-1 : interface definitions, 2002. ViSiCAST project report.
- [29] Thomas Hanke. Hamnosys—representing sign language data in language resources and language processing contexts. In O. Streiter & C. Vettori, editor, *Proceedings of the workshop on the Representation and Processing of Sign Languages*, pages 1–6. European Language Resources Association (ELRA), 2004.
- [30] Alexis Heloir and Michael Kipp. Embr - a realtime animation engine for interactive embodied agents. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents (IVA-09)*. Springer, 2009.
- [31] M. Huenerfauth. *Generating American Sign Language classifier predicates for english-to-asl machine translation*. PhD thesis, University of Pennsylvania, 2006.

- [32] T. Johnston and A. Schembri. On defining lexeme in a signed language. *Sign Language & Linguistics*, 2(2), 1999.
- [33] Trevor Johnston. *The Auslan Archive and Corpus. The Endangered Languages Archive*. University of London, 2008.
- [34] Marion Kaczmarek and Michael Filhol. Assisting Sign Language Translation : what interface given the lack of written form and the spatial grammar? In Tradulex, editor, *Translating and the Computer*, Londres, United Kingdom, November 2019. Tradulex.
- [35] Marion Kaczmarek and Michael Filhol. Computer-assisted Sign Language translation : a study of translators' practice to specify CAT software. In *International Workshop on Sign Language Translation and Avatar Technology*, Hamburg, Germany, September 2019.
- [36] Marion Kaczmarek and Michael Filhol. Alignment Data Base for a Sign Language Concordancer. In *International Conference on Language Resources and Evaluation (LREC)*, Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), Marseille, France, May 2020.
- [37] Marion Kaczmarek and Michael Filhol. Use cases for a Sign Language Concordancer. In *Workshop on the Representation and Processing of Sign Languages*, Marseille, France, May 2020.
- [38] S. K. Liddell and R. E. Johnson. American sign language, the phonological base. *Sign Language studies*, 64, 1989.
- [39] André Martinet. *Éléments de linguistique générale*. Armand Colin, Paris, 1960.
- [40] S. Matthes, T. Hanke, J. Storz, Eleni Efthimiou, Athanasia-Lida Dimou, Panagiotis Karioris, Annelies Braffort, Annick Choisier, Julia Pelhate, and Eva Sáfár. Elicitation tasks and materials designed for dicta-sign's multi-lingual corpus. In *4th Workshop on the Representation and Processing of Sign Languages : Corpora and Sign Language Technologies*, Malta, 2010.
- [41] John Mcdonald and Michael Filhol. Fine tuning dynamics in contextualized proform constructs from linguistic descriptions. In *International Workshop on Sign Language Translation and Avatar Technology (SLTAT)*, Hamburg, Germany, September 2019.
- [42] Fabrizio Nunnari, Michael Filhol, and Alexis Heloir. Animating azee descriptions using off-the-shelf ik solvers. In *Proceedings of the international workshop on the representation and processing of Sign Languages*, Miyazaki, Japan, May 2018.
- [43] S. Prillwitz, R. Leven, H. Zienert, T. Hanke, and J. Henning. Hamnosys version 2.0, hamburg notation system for sign languages, an introductory guide. *International studies on Sign Language communication of the Deaf*, 5, 1989. Signum press, Hamburg.
- [44] Valerie Sutton. *Lessons in SignWriting*. The SignWriting Press, 4th edition, 2014. ISBN 978-0-914336-55-6.
- [45] International Visual Theatre. *La langue des signes, dictionnaire bilingue LSF-français*, volume 2, 3. IVT Paris, 1997.