Langue des Signes Française : Etat des lieux des ressources linguistiques et des traitements automatiques

Annelies Braffort

Langue des Signes Française : État des lieux des ressources linguistiques et des traitements automatiques

Annelies Braffort
Université Paris-Saclay, CNRS, LISN, Campus bât 507 - Rue du Belvédère, 91405 Orsay, France
annelies.braffort@lisn.upsaclay.fr

RÉSUMÉ
Cet article présente un état des lieux sur les ressources disponibles pour la recherche sur la Langue des Signes Française (LSF), ainsi que sur son traitement automatique. Après une mise en contexte sur les langues des signes et plus particulièrement la LSF, l'article recense les ressources disponibles pour la recherche sur la LSF en général, puis plus précisément sur leur utilisation pour les recherches en traitement automatique des langues des signes, en s'appuyant sur des exemples de projets représentatifs récents ou en cours.

ABSTRACT
French Sign Language : Overview of language resources and automatic processing

This article presents an overview of the resources available for research on French Sign Language (LSF), as well as on its automatic processing. After a background on sign languages and more specifically on LSF, The article lists the resources available for research on LSF in general, and then more specifically on their use for research in automatic sign language processing, based on examples of recent or ongoing representative projects.

MOTS-CLÉS : Langue des Signes Française, LSF, Corpus, Traitement Automatique.

KEYWORDS: French Sign Language, LSF, Corpus, Automatic Processing.

1 Les langues des signes : langues naturelles visuo-gestuelles

Les langues des signes (LS) sont des langues naturelles pratiquées au sein des communautés de Sourds [1]. Les LS sont des langues sans écriture, relevant, à ce titre, de la modalité orale, du face-à-face, même transmis en différé. Elles sont visuo-gestuelles, produites par de nombreux articulateurs corporels, les mains et les bras bien sûr, mais aussi le visage, les épaules, la tête, le regard et plusieurs composantes faciales, qui peuvent être activés plus ou moins *simultanément* et perçues par les yeux. Cette modalité de perception est bien adaptée à l'interprétation de mouvements organisés dans l'espace. De fait, cet espace joue un rôle fondamental dans la structuration du discours et est nommé *espace de signation*. Les personnes nées sourdes, qui ont une expérience du monde basée sur la perception visuelle et non sonore, expriment avoir un mode de pensée visuel qui leur fait privilégier des constructions linguistiques qui permettent de dire mais aussi de *montrer*.

Ainsi, les LS comportent des constructions linguistiques de nature diverse. Certaines d'entre elles sont considérées comme des unités spéciales appelées *signes lexicaux* ou simplement *signes*, qu'on peut

1. "Sourd" avec un "S" majuscule désigne une identité culturelle, historique et linguistique.

lister dans un dictionnaire. D'autres types de constructions, très illustratives, exploitent les spécificités mentionnées ci-dessus et peuvent être présentes à hauteur de 20 à 80% selon le type de discours (Sallandre *et al.*, 2019).

Les LS se sont développées au fil du temps et sont très liées à la culture. Tout comme les langues parlées, elles possèdent de multiples formes, des dialectes et des variations locales. L'édition courante de l'Ethnologue [2] répertorie 157 LS, mais il en existe cerftainement d'autres qui n'ont pas encore été documentées ni identifiées. Les différences se situent surtout au niveau du lexique, mais il existe des similitudes sur le plan grammatical car toutes les LS exploitent les capacités de multilinéarité, d'utilisation de l'espace et d'iconicité.

La langue des signes française (LSF) est pratiquée en France et dans la partie francophone de la Suisse. Le nombre de locuteurs de LSF, qui peuvent être des personnes sourdes ou entendantes (les membres de la famille et les proches de personnes Sourdes) n'est pas connu avec précision. On cite souvent un nombre variant de 100 000 à 300 000 locuteurs. Il existe également des langues des signes tactiles utilisées par les personnes avec surdicécité. Elles diffèrent significativement des LS visuelles en ce que des éléments comme l'expression faciale sont remplacés par des informations tactiles.

La présence de la LSF dans les médias a augmenté ces dernières années, notamment depuis l'adoption de la loi de 2005 qui la reconnaît comme langue à part entière [3]. Malgré cette loi qui impose aux établissements qui reçoivent du public de rendre accessible les informations quel que soit le type de handicap, encore très peu d'informations sont disponibles en LSF. Cela pose des problèmes d'accessibilité aux informations pour les personnes Sourdes qui peuvent avoir une maîtrise limitée du français, même écrit, puisque c'est souvent pour eux une langue seconde. Des outils tels que la traduction automatique ou du moins l'aide à la traduction, du texte vers la LSF pour améliorer l'accessibilité, mais aussi dans l'autre sens, pour le sous-titrage de vidéos de LSF, ainsi que l'outillage des vidéos de LS en général, seraient d'une grande utilité pour ces langues et la communauté concernée.

Cet article présente un état des lieux sur les ressources disponibles pour la recherche sur la LSF en général, puis plus précisément sur leur utilisation pour les recherches en traitement automatique des langues des signes.

2 Les corpus de LSF : peu nombreux et de petite taille

Les corpus de LSF répertoriés sur les plateformes d'archivage et de diffusion de corpus tels que Cocoon [4] ou Ortolang [5] sont peu nombreux. On dénombre à l'heure actuelle 22 entrées pour la LSF (il existe aussi des dépôts relatifs à d'autres LS). La plupart des dépôts sur Cocoon sont des numérisations de conférences ou séminaires datant des années 1990 et donc à visée patrimoniale. On y trouve aussi LS-COLIN, le premier corpus de LSF enregistré en studio spécifiquement pour la recherche en 2002. Les dépôts sur Ortolang ont tous été créés dans le cadre de projets de recherche. Presque tous sont des corpus de laboratoire, c'est-à-dire enregistrés en studio avec un objectif de recherche.

Le tableau 1 recense uniquement les corpus de LSF conçus pour la recherche et disponibles actuelle-

2. https://www.ethnologue.com/subgroups/sign-language
3. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000809647/
4. https://cocoon.huma-num.fr/
5. https://www.ortolang.fr/

ment sur les plateformes Ortolang et Cocoon. Ils sont de nature très variée selon le type de contenu (production très contrôlée, monologue, dialogue, ou encore repas en famille), les lieux d'enregistrement (en studio, chez des particuliers, dans la rue) ou les conditions techniques (de 1 à 5 caméras synchronisées, système de capture de mouvement). Certains corpus ne comportent que des données primaires et d'autres sont traduits ou sous-titrés en français et annotés. Là encore la nature et le format des annotations est très variée, depuis de simples gloses pour identifier les signes lexicaux jusqu'à des descriptions fines des constructions linguistiques ou du mouvement des articulateurs. La plupart des corpus sont de très petite taille (moins de 10h), à l'exception de CREAGEST et MEDIAPI-SKEL.

CREAGEST [6] est un corpus de laboratoire dont la partie déposée sur Ortolang comporte 156h de LSF produites par plus de 82 signeurs. Il a été créé avec pour objectif la constitution et la documentation de vidéos de productions gestuelles incluant des productions en LSF d'enfants et d'adultes sourds et des productions de gestualité naturelle pour des études en linguistique et en acquisition (Garcia *et al.*, 2013). Il est a l'heure actuelle le plus gros corpus de LSF, mais à ce jour seule une partie des données primaires a été déposée sur Ortolang.

MEDIAPI-SKEL [7] a été élaboré à partir de contenus fournis par Média'Pi !, un média en ligne bilingue en LSF et en français. De nombreux thèmes y sont abordés, sous forme d'actualités présentées par des journalistes ou présentateurs sourds, d'interviews ou reportages avec plusieurs intervenants, ou d'autres formes plus atypiques (photos-reportages, BDs). La langue première de ces contenus est la LSF. Dans un second temps, un sous-titrage en français est produit. La version actuelle déposée sur Ortolang contient 27 heures de données préparées pour des études en TALS (Bull *et al.*, 2020) : des sous-titres alignés avec des vidéos comportant une représentation simplifiée des signeurs sous forme de squelettes 2D avec des points clés sur le visage, les mains et le corps (figure 1). Les vidéos originales sont accessibles par le biais d'un abonnement auprès de Média'Pi.

FIGURE 1 – Extrait du corpus MEDIAPI-SKEL

Ces deux exemples illustrent bien la grande diversité de ces corpus, élaborés dans le cadre de projets avec un objectif précis. Si certains pourraient être utilisés dans d'autres contextes que ceux prévus initialement, dans la pratique cela semble peu courant, sauf de rares cas où le corpus a été prévu dès le départ pour permettre des études pluridisciplinaires, tel que le corpus MOCAP1 qui a été utilisé pour des études en sciences du mouvement, en linguistique et en informatique (Benchiheub *et al.*, 2016; Collomb *et al.*, 2018; Bigand, 2021).

6. https://www.ortolang.fr/market/corpora/ortolang-000926/
7. https://www.ortolang.fr/market/corpora/mediapi-skel/

Ainsi à l'heure actuelle, pour la LSF, les corpus disponibles pour la recherche sont encore peu nombreux et de taille très limitée. Au niveau international, certaines LS ont pu bénéficier de financements conséquents qui ont permis la création de corpus de grande taille (pour les LS) et bien documentés, comme par exemple le corpus de LS allemande *DGS Corpus* [8] de 560h, qui a été élaboré dans le cadre d'un projet financé sur 15 ans. Deux parties de plus de 50h sont accessibles en ligne : *My DGS*, accessible à tout public et fourni avec des sous-titrages et *My DGS annotated*, à destination des chercheurs. Un tel corpus nécessite le développement d'outils permettant son exploitation et donc des études en traitement automatique des LS (TALS).

La section suivante dresse un état des lieux des études actuelles en TALS sur la LSF, en lien avec les corpus existants.

3 Corpus pour le traitement automatique

La recherche en traitement automatique des LS est beaucoup plus récente que celle dédiée aux langues parlées ou écrites. Elle est particulièrement active dans le domaine de la reconnaissance automatique, mais il existe aussi des projets en génération et depuis peu en traduction automatiques.

Dans le cadre du projet Européen en cours EASIER [9] sur la traduction automatique entre certaines langues écrites et langues des signes d'Europe, un livrable [10] recense les ressources linguistiques qui peuvent être utilisées pour le TALS. Il répertorie en particulier les corpus de LS européennes de grande taille (pour les LS) qui peuvent être utilisés comme données d'entraînement de haute qualité pour la traduction automatique.

On peut distinguer deux types de ressources : les corpus de recherche et les données télédiffusées. Les corpus de recherche, et plus particulièrement ceux créés en vue d'études en linguistique, offrent des données de qualité élevée accompagnés d'une transcription et d'une annotation linguistique, mais ils n'en existent pas de grande taille pour toutes les LS et ils sont malgré tout de taille réduite par rapport aux besoins pour les approches à base d'apprentissage. Les données télédiffusées sont souvent disponibles en quantité relativement importante et comporte généralement des sous-titres synchronisés avec la parole. Cependant il s'agit la plupart du temps de LS produite en direct par un interprète, et soumise aux contraintes temporelles du direct et de la structure du discours oral qui est interprété. De plus, la qualité des alignements entre la LSF et les sous-titres peut être assez mauvaise car les sous-titres ne sont pas alignés avec la LS et il peut y avoir plusieurs secondes de décalage.

Un nouveau type de ressource consiste en des données télédiffusées de LS non interprétées produites par des locuteurs Sourds. Elles sont constituées à partir d'émissions réalisées directement en LS, puis sous-titrées. Ce sont des données d'une très grande qualité à la fois sur la nature de la LS et sur la qualité de l'alignement. Le seul existant à ce jour pour la LSF est le corpus MEDIAPI-SKEL décrit précédemment.

Un autre corpus de ce type a été créé récemment dans le cadre d'une toute première tâche partagée sur la traduction d'une LS vers une langue écrite (LS Suisse Allemande vers l'allemand) dans le cadre

8. https://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/welcome.html
9. https://www.project-easier.eu/
10. https://www.project-easier.eu/wp-content/uploads/sites/67/2021/08/EASIER-D6.1-Overview-of-Datasets-for-the-Sign-Languages-of-Europe.pdf

de la conférence Machine Translation (WMT) [11].

En ce qui concerne le niveau lexical, un des objectifs dans le cadre du projet Easier est constituer une ressource multilingue de type Wordnets pour plusieurs LS. La version actuelle s'est centrée sur les LS grecque et allemande (Bigeard *et al.*, 2022), mais intégrera d'autres LS dont la LSF d'ici la fin du projet.

A l'heure actuelle, une majorité des études, en particulier dans le domaine de la reconnaissance automatique, se concentrent sur les signes lexicaux. Cependant, comme évoqué en section 1, les LS ne sont pas juste une séquence d'unités lexicales discrètes pouvant être répertoriées dans un dictionnaire, mais plutôt des séquences de constructions spatio-temporelles qui combinent des signaux discrets et continus, des composantes manuelles et non manuelles et qui permettent une grande liberté dans la production à la volée d'unités de sens. Les études s'intéressant à ces aspects sont encore très rares.

Pour la LSF, on peut citer les travaux de V. Belissen 2020, dans lesquels l'approche proposée s'est centré sur la détection des unités non lexicales. Pour la partie de son travail portant que la représentation du signeur, au vu du peu de ressources disponibles pour la LSF, il a utilisé un système pré-existantpermettant d'identifier les configurations de mains pour la LS Allemande qu'il a réentrainé pour identifier les configurations de mains sur le corpus DICTA-SIGN (figure 2). Comme son travail n'était pas centré sur le lexique, il est très probablement transférable à d'autres LS.

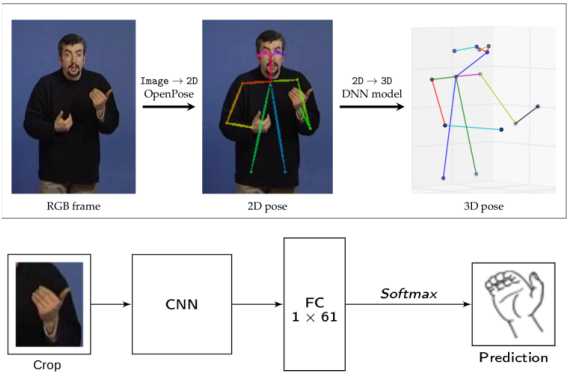

FIGURE 2 – Représentation du signeur en LSF dans la thèse de V. Belissen

En parallèle aux études qui nécessitent des techniques par apprentissage, d'autres approches sont exporées. Par exemple le projet Rosetta [12], terminé récemment, visait à étudier des solutions d'accessibilité pour les contenus audiovisuels. L'un des objectifs consistait à concevoir un système de traduction automatique du texte vers la LSF, affichée via l'animation d'un signeur virtuel. Les trois principales contributions ont été la constitution de ROSETTA-LSF, un corpus aligné de texte et de LS enregistré à l'aide d'un système de capture de mouvement (Bertin-Lemée *et al.*, 2022a), un système de traduction du texte vers une représentation intermédiaire avec une approche à base d'exemples (Bertin-Lemée *et al.*, 2022b), et un système permettant de générer des animations d'avatar à partir de cette représentation et de blocs d'animations préenregistrées (Dauriac *et al.*, 2022). Ce travail a abouti à une preuve de concept fonctionnelle (figure 3) mais limitée par la taille du corpus. Les perspectives

11. https://www.wmt-slt.com/
12. https://rosettaccess.fr/index.php/

de ce travail sont donc liées à la possibilité de développer un corpus aligné de très grande taille ainsi que des outils permettant en particulier de faciliter l'alignement entre la LSF et le français.

FIGURE 3 – Prototype de traduction français vers LSF du projet Rosetta

Un autre aspect important est la possibilité de combiner plusieurs corpus au sein d'un même projet. Par exemple, le projet en cours Serveur Gestuel vise à créer l'équivalent d'un serveur vocal en LSF, intégrant des technologies de reconnaissance, de génération et de dialogue. Pour la partie reconnaissance automatique, les études des laboratoires partenaires du projet sont basées actuellement sur les corpus DICTA-SIGN et MEDIAPI-SKEL. Pour la partie modélisation linguistique qui permet de piloter l'animation de l'avatar, les corpus 40 BREVES et MOCAP1 sont utilisés pour des analyses linguistiques (Martinod *et al.*, 2022).

Ainsi, la possibilité d'utiliser ou produire des systèmes transférables à plusieurs LS et des corpus de différentes LS ou de nature variée permet de dépasser un peu les limites dûes au manque de données.

4 Conclusion

Ces dernières années ont été marquées par des évolutions sur la nature des corpus, la manière de les exploiter ou de les combiner et sur le type de traitement automatique qu'il est possible de mettre en œuvre grâce à eux. En ce qui concerne la LSF, les ressources restent encore très limitées et nécessitent d'être développées et outillée.

Les outils développés en TALS restent à l'heure actuelle des prototypes de recherche et leurs capacités doivent encore être étendues avant de pouvoir procéder à de véritables évaluations. Mais nous pouvons d'ores et déjà souligner que l'évaluation doit elle-aussi être adaptée aux spécificités des LS (multilinéarité, utilisation de l'espace et iconicité), et est un axe de recherche en soi qu'il est nécessaire de développer.

TABLE 1 – Corpus de LSF sur Ortolang et Cocoon. * : non renseigné

Nom	Description	Taille	Locuteurs	VF	Annot.	Dépôt
LS-COLIN	monologue, narration, récit, explication	1,5h	13	partielle	partielle	Cocoon
CREAGEST	dialogue et acquisition	> 156h *	> 82 *	partielle	partielle	partiel, Ortolang
DEGELS1	dialogue, comparable LSF et gestualité	LSF : 35', gest. : 39'	LSF : 5, gest. : 4	partielle	partielle	Ortolang
40 BREVES	monologue, traduction de brèves journalistiques	1h	3	oui	oui	Ortolang
MOCAP1	monologue, description de photos	2h	8	non	partielle	partiel, Ortolang
DICTA-SIGN-LSF-V2	dialogue, plusieurs tâches avec plusou moins d'élicitation	8h	16	oui	partielle	Ortolang
MEDIAPI-SKEL	très varié, issu d'un média bilingue	27h	>100	oui	non	Ortolang
ROSETTA-LSF	monologue, traduction de phrases de type journalistique	3h	1	oui	oui	Ortolang
CORPUS CATTEAU	poésie en LSF, interprétation en FR et entretiens	*	*	oui	oui	Ortolang
SIGNES EN FAMILLE	Echanges spontanés durant le repas familial	2h à 4h30 par famille	10 familles	non	oui	Ortolang
CLM-MOCAP	monologue et dialogue, capture de mouvement	*	10	non	non	Ortolang
LG-IDF	récit, lexique	1h30	1	non	non	Ortolang

Références

BELISSEN V. (2020). *From Sign Recognition to Automatic Sign Language Understanding : Addressing the Non-Conventionalized Units*. Thèse de doctorat. ED STIC, Université Paris-Saclay 2020.

BENCHIHEUB M.-E.-F., BERRET B. & BRAFFORT A. (2016). Collecting and analysing a motion-capture corpus of French Sign Language. In *Proceedings of the LREC2016 7th Workshop on the Representation and Processing of Sign Languages : Corpus Mining*, p. 7–12, Portorož, Slovenia : ELRA.

BERTIN-LEMÉE E., BRAFFORT A., CHALLANT C., DANET C., DAURIAC B., FILHOL M., MARTINOD E. & SEGOUAT J. (2022a). Rosetta-LSF : an Aligned Corpus of French Sign Language and French for Text-to-Sign Translation. In *Proceedings of the 13th International Conference on Language Resources and Evaluation*, Marseille, France : ELRA.

BERTIN-LEMÉE E., BRAFFORT A., CHALLANT C., DANET C. & FILHOL M. (2022b). Example-Based Machine Translation from Text to a Hierarchical Representation of Sign Language. DOI : 10.48550/ARXIV.2205.03314.

BIGAND F. (2021). *Extracting human characteristics from motion using machine learning : the case of identity in Sign Language*. Thèse de doctorat. ED STIC, Université Paris-Saclay 2021.

BIGEARD S., SCHULDER M., KOPF M., HANKE T., VASILAKI K., VACALOPOULOU A., GOULAS T., DIMOU A.-L., FOTINEA S.-E. & EFTHIMIOU E. (2022). Introducing sign languages to a multilingual wordnet : Bootstrapping corpora and lexical resources of Greek Sign Language and German Sign Language. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages : Multilingual Sign Language Resources*, Marseille, France : ELRA.

BULL H., BRAFFORT A. & GOUIFFÈS M. (2020). MEDIAPI-SKEL - a 2D-skeleton video database of French Sign Language with aligned French subtitles. In *Proceedings of the 12th Language Resources and Evaluation Conference*, Marseille, France : ELRA.

COLLOMB A., BRAFFORT A. & KAHANE S. (2018). L'anatomie du proforme en langue des signes française : Quand il sert à introduire des entités dans le discours. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)*, (34). DOI : 10.4000/tipa.2164.

DAURIAC B., BRAFFORT A. & BERTIN-LEMÉE E. (2022). Example-based Multilinear Sign Language Generation from a Hierarchical Representation. In *Proceedings of the LREC2022 7th International Workshop on Sign Language Translation and Avatar Technology : The Junction of the Visual and the Textual*, Marseille, France : ELRA.

GARCIA B., L'HUILLIER M.-T. & SALLANDRE M.-A. (2013). Creagest : enjeux linguistiques, patrimoniaux et socio-éducatifs d'un grand corpus de langue des signes française. *La nouvelle revue de l'adaptation et de la scolarisation*, (64). DOI : 10.3917/nras.064.0081.

MARTINOD E., DANET C. & FILHOL M. (2022). Two new AZee production rules refining multiplicity in French Sign Language. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages : Multilingual Sign Language Resources*, Marseille, France : ELRA.

SALLANDRE M.-A., BALVET A., BESNARD G. & GARCIA B. (2019). Étude exploratoire de la fréquence des catégories linguistiques dans quatre genres discursifs en LSF. *Revue de linguistique et de didactique des languese (LIDIL)*, (60). DOI : 10.4000/lidil.7136.