Actes de CORIA-TALN 2023
Paris, France, 5 au 9 juin 2023

Traduction à base d'exemples du texte vers une représentation hiérarchique de la langue des signes

Élise Bertin-Lemée[1] Annelies Braffort[2] Camille Challant[2]
Claire Danet[2] Michael Filhol[2]
(1) SYSTRAN, 5 rue Feydeau, Paris, France
(2) LISN, Univ. Paris-Saclay, bât. 507, rue du Belvédère, 91405 Orsay, France
elise.bertinlemee@systrangroup.com, {annelies.braffort, camille.challant, claire.danet, michael.filhol}@lisn.upsaclay.fr

RÉSUMÉ

Cet article présente une expérimentation de traduction automatique de texte vers la langue des signes (LS). Comme nous ne disposons pas de corpus aligné de grande taille, nous avons exploré une approche à base d'exemples, utilisant AZee, une représentation intermédiaire du discours en LS sous la forme d'expressions hiérarchisées.

ABSTRACT

Example-Based Machine Translation from Text to a Hierarchical Representation of Sign Language

This paper presents an experiment in automatic translation from text to sign language (SL). As we do not have a large aligned corpus, we have explored an example-based approach, using AZee, an intermediate representation of the discourse in SL in the form of hierarchical expressions.

MOTS-CLÉS : Langue des signes, traduction à base d'exemples, représentation intermédiaire.

KEYWORDS: Sign Language, example-based translation, intermediate representation.

1 Introduction

Le travail présenté ici a été réalisé dans le cadre d'un projet qui visait à étudier des solutions d'accessibilité pour les contenus audiovisuels pour les personnes sourdes à l'aide de sous-titrage et de traduction en langue des signes (LS). Pour ce dernier objectif, les trois principales contributions ont été la constitution d'un corpus aligné de texte et de LS (Bertin-Lemée *et al.*, 2022), un système de traduction automatique (TA) du texte en une représentation formelle de la LS (Bertin-Lemée *et al.*, 2023), et un système permettant de générer des animations d'avatars signants à partir de cette représentation (Dauriac *et al.*, 2022). Après un aperçu des enjeux et des travaux récents dans le domaine, nous expliquons la méthode et les choix de conception et décrivons l'implémentation du système de traduction. Enfin, nous donnons des résultats préliminaires et discutons des questions soulevées pour l'évaluation.

Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN), volume 4 : articles déjà soumis ou acceptés en conférence internationale, pages 28–34.

2 Traduction du texte vers la langue des signes

Les LS sont des langues naturelles pratiquées au sein des communautés de Sourds et la Langue des Signes Française (LSF) est celle utilisée en France. Dans la suite, nous parlerons des LS en général quand les aspects abordés les concernent toutes et de la LSF lorsque cela concerne uniquement cette langue. Ce sont des langues visuo-gestuelles : une personne s'exprime en LS en utilisant de nombreuses composantes corporelles (mains et bras, mais aussi torse, épaules, tête, regard et expressions faciales) et son interlocuteur perçoit le message par le canal visuel. Le système linguistique des LS exploite ces canaux spécifiques : de nombreuses informations sont exprimées simultanément et s'organisent dans l'espace, et l'iconicité joue un rôle central et structurant à tous les niveaux (du sub-lexical au discours). À ce jour, les LS n'ont pas de système d'écriture et les rares systèmes graphiques pour la transcription, tels que HamNoSys (Hanke, 2004) ou SignWriting (Bianchini, 2014), ne décrivent que l'aspect lexical. Ceux-ci ne peuvent représenter pleinement le niveau discursif, la multilinéarité, l'utilisation de l'espace et les structures illustratives. Tous ces aspects contribuent aux défis de la TA depuis ou vers les LS.

La TA d'un texte vers une LS est un sujet de recherche assez récent et encore très peu exploré. L'approche dominante en TA est l'approche neuronale, qui s'appuie sur la disponibilité de grands volumes de données alignées (de l'ordre de plusieurs millions de phrases), non disponibles dans cette quantité pour les LS. Si des tentatives existent avec ces méthodes, elles ne donnent pas encore de résultats satisfaisants, comme l'ont montré les évaluations humaines lors du premier défi portant sur la traduction automatique de LS suisse-allemande (DSGS) vers l'allemand (Müller *et al.*, 2022).

La traduction automatique basée sur des exemples (EBMT [1]) est une autre approche fondée sur l'analogie qui utilise un corpus bilingue contenant des textes et leurs traductions (Nagao, 1984). Étant donné un texte à traduire, on sélectionne dans ce corpus des segments contenant des éléments similaires. Ces éléments sont ensuite utilisés pour traduire les éléments du texte original dans la langue cible, et ces phrases sont recombinées pour former une traduction complète. L'approche EBMT peut être mise en œuvre sur des corpus plus petits et donc envisagée dans notre cas. Les capacités de traduction restent liées à la taille du corpus, mais dans un domaine ciblé, on peut espérer obtenir des résultats de meilleure qualité que ceux obtenus actuellement par les approches neuronales.

Comme les LS n'ont pas de forme écrite, une approche courante est de procéder en deux étapes : une première étape consiste à traduire le texte en une représentation intermédiaire, et une seconde étape utilise cette représentation comme entrée d'un système de synthèse pour contrôler l'animation d'un personnage virtuel afin d'afficher le contenu en LS sous forme de vidéo.

Après une première génération d'études basées principalement sur des approches à base de règles (Veale *et al.*, 1998; Zhao *et al.*, 2000; Marshall & Safar, 2004), d'autres ont exploré des approches à base d'exemples (Morrissey & Way, 2005). Elles ont parfois été combinées avec des approches statistiques, comme par exemple De Martino *et al.* (2017) qui a développé un système qui traduit automatiquement quelques textes en portugais brésilien vers la LS brésilienne (LIBRAS), en fonction du contexte et de la fréquence d'apparition dans les traductions précédentes. À notre connaissance, ces projets n'ont donné lieu à aucune suite.

La grande majorité des projets utilisant une représentation intermédiaire de la LS, y compris les plus récents (Egea Gómez *et al.*, 2021), utilisent des séquences de gloses, chaque glose [2] représentant

1. En anglais : *example-based machine translation*.
2. Une glose est une étiquette textuelle, généralement un seul mot, qui reflète la signification du signe qu'elle représente.

une unité lexicale. Les systèmes de traduction traitent alors une séquence de tokens. Cependant, avec ce type de représentation, il est très difficile, voire impossible, de traiter les phénomènes courants de la LS, tels que l'activité non manuelle, les relations spatiales, les structures iconiques ou le rythme des mouvements. Il en résulte des animations de très mauvaise qualité, incomplètes, voire incompréhensibles. Pour cette raison, il parait indispensable d'envisager une représentation intermédiaire plus riche que de simples concaténations de gloses.

Dans certaines approches récentes de bout-en-bout (Stoll *et al.*, 2020), l'utilisation d'une représentation intermédiaire n'est pas présente. Cette approche neuronale, encore très expérimentale, génère directement des vidéos de contenus signés photoréalistes à partir d'entrées textuelles. En plus de nécessiter des corpus de très grande taille, elle n'offre pas les mêmes avantages que les avatars (anonymat dans le rendu, apparence modifiable) que nous avons choisi de privilégier.

3 Approche à base d'exemples et représentation hiérarchique

Comme nous ne disposons pas de corpus aligné de grande taille, nous avons choisi d'explorer l'approche à base d'exemples. Par ailleurs, nous avons retenu AZee comme représentation intermédiaire, une approche formelle de représentation du discours en LS (Filhol *et al.*, 2014). Celle-ci permet de définir des *règles de production*, qui associent des formes à articuler (par exemple, hausser les sourcils) à un sens identifié (par exemple, l'expression d'un doute). En les combinant, on peut construire des *expressions discursives* hiérarchiquement structurées représentant des énoncés complets, déterminant les formes à produire de manière suffisament détaillée pour permettre ensuite de contrôler l'animation d'un avatar (Challant & Filhol, 2022).

Par exemple, considérons les six règles de production suivantes identifiées pour la LSF : les trois sans argument que sont `ministre`, `environnement` et `parler`, ainsi que les trois ci-dessous comportant des arguments (en italique) :

- `info-about`(*topic*, *info*) : *info*, qui est ciblée, est donnée sur un *topic* ;
- `side-info`(*focus*, *info*) : *focus* avec une information supplémentaire (non focalisée) *info* à son sujet ;
- `nerveusement`(*sig*) : *sig* d'une manière nerveuse.

Ces règles peuvent être combinées hiérarchiquement dans l'expression suivante pour former la structure d'un énoncé signifiant "*le ministre de l'écologie parle nerveusement*" et respectant la grammaire de la LSF :

```
:info-about
  'topic
  :side-info          (*)
    'focus
    :ministre
    'info
    :environnement
  'info
  :nerveusement
    'sig
    :parler
```

Afin d'explorer l'utilisation d'une approche à base d'exemples, nous avons utilisé une banque [3] de près de 2000 alignements entre des segments de texte en français et des expressions de ce type, créée à partir du corpus parallèle français-LSF du projet.

L'approche à base d'exemples s'appuie sur l'analogie avec des exemples existants pour traduire de nouveaux contenus. Cela signifie que nous pouvons tenter de traduire une nouvelle phrase en trouvant des exemples suffisamment proches, et en remplaçant ce qui est différent. Par exemple, pour traduire la phrase "*la présidente parle nerveusement*" qui ne figure pas dans la base d'exemples, on peut partir de l'exemple "*le ministre de l'écologie parle nerveusement*" qui, lui, est présent dans la base, et substituer une traduction de "*la présidente*" à celle de "*le ministre de l'écologie*" dans le segment aligné. Dans cet exemple, "*présidente*" n'a pas de correspondance dans la phrase et est nommé "anti-match", "*ministre de l'écologie*" est nommé sa "correction". L'hypothèse est que si nous trouvons les parties correspondant à chaque anti-match dans la traduction alignée, nous pouvons tenter de les remplacer par les traductions de leurs corrections.

4 Implémentation

Pour une phrase donnée à traduire, on va chercher dans le corpus des alignements dans lesquels le segment de texte est exactement identique et on récupère les expressions alignées correspondantes. En cas d'échec, on considère tous les alignements de texte qui sont "proches" et dont les différences sont les "anti-matchs". La structure globale de la traduction dans la représentation intermédiaire est ainsi conservée, dans laquelle on va pouvoir faire les substitutions.

Dans l'exemple précédent, on peut considérer que le segment "*le ministre de l'écologie parle nerveusement*" est proche de "*la présidente parle nerveusement*". L'anti-match unique $\bar{m}_1$ est "*le ministre de l'écologie*" et sa correction c_1 est "*la présidente*". Notre approche est alors : (1) d'identifier la sous-expression marquée `(*)` ci-dessus comme le nœud correspondant à la traduction de $\bar{m}_1$; et (2) d'y substituer une traduction de c_1. Pour ces deux tâches, on utilise récursivement le même algorithme. Pour (1) on génère les traductions possibles de $\bar{m}_1$ en vue d'y trouver `(*)`, et pour (2) on génère directement celles de c_1, qui pourront être substituées à `(*)`.

Un des problèmes de cette approche est celui de l'échec de la traduction, qui est d'autant plus susceptible de se produire que le corpus d'alignements d'exemples est petit. Dans de tels cas, nous avons recours à une solution de repli où nous décomposons la requête en une partition de plus petits morceaux de texte, que nous traduirons séparément et concaténerons dans le résultat avec pour seule règle le suivi de l'ordre en source. Cette stratégie de repli produit une LS de moins bonne qualité, et équivaut en fait à une traduction littérale (mot à mot) si elle est utilisée systématiquement. Mais elle permet de juxtaposer des morceaux de LS plus importants et donc plus complets et plus fluides sans recourir à la simple concaténation d'unités uniquement lexicales.

L'implémentation pratique de l'algorithme s'appuie sur plusieurs modules de traitement de texte pour trouver les meilleures correspondances dans le corpus existant.
Pour permettre la mise en correspondance, l'anti-match et les partitions sous-phrastiques, la tokenisation au niveau du mot est d'abord effectuée par Open-NMT Tokenizer [4] et une certaine flexibilité est permise lors de la recherche de segments correspondants avec la ponctuation et les articles.

3. L'ensenble du corpus est disponible ici : https://www.ortolang.fr/market/corpora/rosetta-lsf
4. https://github.com/OpenNMT/Tokenizer

Ensuite, le défi principal est de définir quel type de "similarité" dans la langue source peut produire les meilleurs candidats pour la génération de la langue cible. La sémantique et la syntaxe entrent en jeu pour déterminer les éléments similaires à remplacer ou à traduire séparément. En pratique, nous nous appuyons sur deux types d'analyse de texte à différentes étapes de l'algorithme.

Pour trouver les meilleurs anti-matches dans la base de données actuelle et les remplacer par des corrections, nous utilisons l'appariement de chaîne de caractères et considérons comme candidats tous les alignements qui ont des tokens en commun avec le texte soumis. Les meilleures correspondances ont été définies empiriquement comme étant celles qui ont le maximum de tokens en commun, ainsi que la longueur minimale en nombre de tokens ou la meilleure proportion de tokens similaires dans l'ensemble des tokens. Pour l'instant, ce choix reste arbitraire et mériterait une étude comparative.

Lorsque les approches d'appariement et d'anti-match échouent, nous avons recours à des partitions déterminées en naviguant dans l'arbre de dépendance syntaxique obtenu à l'aide de spaCy [5], une bibliothèque avec des modèles prêts à l'emploi et des chaînes de traitement optimisées pour les langues naturelles. L'analyse syntaxique par dépendance n'explore pas toutes les partitions possibles d'une phrase mais limite l'exploration à des morceaux syntaxiquement valides.

5 Test et discussion

Pour tester notre système, nous avons construit un jeu de test en créant des phrases mélangeant des segments de différentes phrases de notre corpus, pour étudier les résultats produits. Notre jeu de test est composé de 15 phrases. Par exemple, la phrase "*Recul de l'âge légal à la retraite : c'est ce que proposent les retraités pour leurs enfants*" a été créée à partir des phrases suivantes du corpus :

— "***Recul de l'âge légal à la retraite :*** *"Il ne faut pas prendre les Français pour des canards sauvages", lance Valérie Pécresse.*"
— "*Des routes nationales bientôt privatisées ?* ***C'est ce que proposent*** *les sociétés d'autoroutes dans une note interne.*"
— "*Solidarité : une ancienne abbaye accueille* ***des retraités***"
— "*Au Japon, des dizaines de pères français se battent désespérément* ***pour*** *voir* ***leurs enfants.***"

Grâce à ce jeu de tests, nous avons pu faire valider, par le biais de focus groups avec des locuteurs de LSF, que notre approche est préférable à une approche basée sur une simple concaténation de signes lexicaux. En effet, les structures spécifiques à la LSF peuvent être trouvées dans les traductions finales, ce qui n'est pas le cas lorsque la langue est réduite à une séquence de gloses.

En outre, l'approche produit des résultats présentant une certaine forme de créativité. En LSF, les paraphrases ou les ajouts sont couramment utilisés, et font d'ailleurs partie de notre corpus tel qu'il a été livré initialement par le traducteur au moment de la création du corpus vidéo. Ces éléments ont été alignés par la suite comme exemples, et apparaissent donc fréquemment dans les traductions générées, même si ce n'est pas toujours strictement nécessaire. Par exemple "*Alsace*" peut être signé par un seul signe, mais il est aussi exprimé dans notre corpus par une expression bien plus complexe impliquant des référencements spatialisés (zone à l'Est de la France placée sur un plan vertical), typique de la LSF quand aucun contexte n'existe encore.

De plus, la sortie de l'algorithme est un ensemble de traductions (construites à partir des différentes

5. https://spacy.io

combinaisons de substitution), et pas nécessairement une expression unique. Cela rend compte, d'une certaine manière, de la réalité de la tâche de traduction. Dans notre jeu de test, le nombre de traductions proposées pour une requête varie de 1 à 12 (moyenne : 4).

6 Conclusion et perspectives

Nous avons présenté une nouvelle idée de système de TA du texte vers la LSF, utilisant une approche à base d'exemples et une représentation hiérarchique de la LS, ainsi qu'un algorithme d'appariement, substitution et concaténation. Le corpus utilisé contient des alignements de textes français et leurs traductions en LSF décrites selon cette représentation. Un prototype a été réalisé et testé sur quelques exemples, fournissant ainsi une preuve de concept. Les capacités de ce système et la taille du corpus doivent encore être étendues avant de pouvoir effectuer de véritables évaluations. Mais nous pouvons d'ores et déjà souligner que l'évaluation d'un tel système ne sera pas facile, puisqu'il propose une traduction d'une langue vers une représentation d'une autre langue, non lisible directement.

Les métriques habituellement utilisées, qu'elles soient quantitatives ou qualitatives, sont adaptées aux cas où les langues source et cible sont textuelles. Dans notre cas, la cible n'est pas directement la LSF, mais une représentation formelle. En outre, cette représentation est utilisée pour générer des animations qui, certes, sont directement "lisibles" par les locuteurs de LSF, mais qui nécessitent de leur côté des phases d'évaluations qui ne sont pas liées aux aspects linguistiques mais plutôt à l'aspect de l'avatar, aux mouvements et à leur degré de bio-réalisme. La mise en place d'un protocole d'évaluation robuste et complet est clairement un sujet d'étude à part entière, qui devra être abordé dans un avenir proche.

Remerciements

Ce travail a été financé par le projet d'investissement Bpifrance "Grands défis du numérique", dans le cadre du projet ROSETTA (RObot for Subtitling and intElligent adapTed TranslAtion). Nous remercions Noémie Churlet, Raphaël Bouton et Media'Pi ! pour leur engagement dans ce projet, qui n'aurait pas eu la même validité et le même impact sans eux.

Références

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

BERTIN-LEMÉE E., BRAFFORT A., CHALLANT C., DANET C. & FILHOL M. (2023). Example-Based Machine Translation from Text to a Hierarchical Representation of Sign Language. In *The 24th Annual Conference of The European Association for Machine Translation*.

BIANCHINI C. S. (2014). *Analyse métalinguistique de l'émergence d'un système d'écriture des langues des signes : SignWriting et son application à la langue des signes italienne (LIS)*. Thèse de doctorat. Thèse de doctorat dirigée en Sciences du langage de l'Université Paris 8.

CHALLANT C. & FILHOL M. (2022). A First Corpus of AZee Discourse Expressions. In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, p. 1560-1565, Marseille, France.

DAURIAC B., BRAFFORT A. & BERTIN-LEMÉE E. (2022). Example-based Multilinear Sign Language Generation from a Hierarchical Representation. In *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology : The Junction of the Visual and the Textual : Challenges and Perspectives*, p. 21–28, Marseille, France : European Language Resources Association.

DE MARTINO J. M., SILVA I. R., BOLOGNINI C. Z., COSTA P. D. P., KUMADA K. M. O., CORADINE L. C., DA SILVA BRITO P. H., DO AMARAL W. M., BENETTI Â. B., POETA E. T. *et al.* (2017). Signing Avatars : Making Education More Inclusive. *Universal access in the information society*, **16**(3), 793–808.

EGEA GÓMEZ S., MCGILL E. & SAGGION H. (2021). Syntax-aware Transformers for Neural Machine Translation : The Case of Text to Sign Gloss Translation. In *Proceedings of the 14th Workshop on Building and Using Comparable Corpora (BUCC 2021)*, p. 18–27 : INCOMA Ltd.

FILHOL M., HADJADJ M. & CHOISIER A. (2014). Non-Manual Features : The Right to Indifference. In *International Conference on Language Resources and Evaluation*, p. 49–54, Reykjavik, Iceland.

HANKE T. (2004). Hamnosys - representing sign language data in language resources and language processing contexts. In O. STREITER & C. VETTORI, Éds., *LREC 2004, Workshop proceedings : Representation and processing of sign languages.* : Paris : ELRA.

MARSHALL I. & SAFAR E. (2004). Sign language generation in an ALE HPSG. In S. MÜLLER, Éd., *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, p. 189–201, Stanford, CA : CSLI Publications. DOI : 10.21248/hpsg.2004.11.

MORRISSEY S. & WAY A. (2005). An Example-Based Approach to Translating Sign Language. In *Workshop on Example-Based Machine Translation, MT SUMMIT*, p. 109–116, Phuket, Thailand : Asia-Pacific Association for Machine Translation, Tokyo.

MÜLLER M., EBLING S., AVRAMIDIS E., BATTISTI A., BERGER M., BOWDEN R., BRAFFORT A., CIHAN CAMGÖZ N., ESPAÑA-BONET C., GRUNDKIEWICZ R., JIANG Z., KOLLER O., MORYOSSEF A., PERROLLAZ R., REINHARD S., RIOS A., SHTERIONOV D., SIDLER-MISEREZ S. & TISSI K. (2022). Findings of the First WMT Shared Task on Sign Language Translation (WMT-SLT22). In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, p. 744–772, Abu Dhabi, United Arab Emirates (Hybrid) : Association for Computational Linguistics.

NAGAO M. (1984). A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. In *Proceedings of the International NATO Symposium on Artificial and Human Intelligence*, p. 173–180. USA : Elsevier North-Holland, Inc.

STOLL S., CAMGOZ N. C., HADFIELD S. & BOWDEN R. (2020). Text2sign : Towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*, **128**(4), 891–908.

VEALE T., CONWAY A. & COLLINS B. (1998). The Challenges of Cross-modal Translation : English-to-Sign-Language Translation in the Zardoz System. *Machine Translation*, **13**(1), 81–106.

ZHAO L., KIPPER K., SCHULER W., VOGLER C., BADLER N. & PALMER M. (2000). A Machine Translation System from English to American Sign Language. In *Conference of the Association for Machine Translation in the Americas*, p. 54–67 : Springer.