

Analyse multimedia vidéo-texte. Contributions à la traduction de la Langue des Signes Française vers le français

Michèle Gouiffès et Annelies Braffort

Durée : 4-6 mois (dates flexibles), possible continuation en doctorat

Niveau : Stage de M2 recherche

Mots clés : Langue des Signes, LSF, Vision par ordinateur, Traitement automatique des Langues.

Gratification : 600-700 € par mois + participation abonnement transport.

Il existe de nombreuses applications autour de l'écrit ou de la parole : des logiciels de traduction (DeepL), des concordanciers (Linguee), des agents conversationnels (ChatGPT) ou tout simplement des moteurs de recherche. À ce jour, très peu d'outils ont été proposés pour les Langues des Signes. Il est nécessaire de concevoir des applications en ligne pour enrichir les ressources en LSF (données, annotations, codes) qui seront utilisées pour développer des applications visant à améliorer l'accessibilité des sourds aux contenus multimedia, l'apprentissage en langue des signes, ou faciliter la communication entre sourds et entendants. C'est dans ce contexte que se situe cette proposition de stage.

1 Contexte de travail

Le stage sera réalisé au laboratoire [LISN](#) (Laboratoire Interdisciplinaire des Sciences du Numérique), au sein du département Sciences et Technologies des Langues dans l'équipe LIPS. Le thème [M&TALS](#) (Modélisation et Traitement Automatique des Langues des Signes) de l'équipe LIPS produit des ressources linguistiques et aborde des problématiques d'analyse, de représentation et de traitement de la Langue des Signes Française (LSF) de manière interdisciplinaire, avec des points de vue de plusieurs domaines de l'informatique (TAL, vision par ordinateur, informatique graphique), ainsi que des sciences du langage, du mouvement et de la perception.

La personne recrutée renforcera l'équipe réunie au LISN autour du projet ANR Gest-ToSay et sera amenée à collaborer avec des chercheurs du laboratoire Gipsa-lab à Grenoble.

2 Les Langues des Signes

Les langues des signes (LS) sont des langues naturelles pratiquées au sein des communautés de personnes sourdes. À la différence des langues vocales qui sont audio-phonatoires, les LS sont visuo-gestuelles. L'information est transmise par différents articulateurs (les mains, les bras, le buste, les épaules, la tête, les éléments du visage, le regard) et leurs mouvements. De plus, le discours se structure dans l'espace, dans le sens où l'espace sert à contextualiser un signe, à placer des objets ou des concepts, à créer des relations visuelles entre ces entités. Ainsi, on ne peut pas réduire un discours en LS à une simple séquence de signes qui auraient un équivalent dans la langue vocale. Notons également que la forme des signes varie en

fonction des transitions avec ce qui précède et ce qui suit (co-articulation) et des contraintes linguistiques (spatialisation) et que des signes très illustratifs peuvent être créés au fil de l'eau.

3 Vers la traduction de la LSF vers le Français

Le projet proposé vise à développer des méthodes associant vision par ordinateur et traitement automatique des langues, allant de l'encodage des vidéos et de la langue vers les modèles de traduction de la langue des signes (vidéo) vers le texte.

Ce stage se concentrera particulièrement sur cette tâche de traduction. Les données multimedia utilisées seront d'une part celles de Mediapi-RGB[3], issues du media sourd [médiapi](#) offrant un bon alignement des deux modalités vidéo et texte, d'autre part celles du corpus Matignon-LSF, constitué de discours en français (données audio et transcriptions écrites) interprétées en LSF, avec un décalage temporel.

Les contributions proposées au sein du laboratoire à base de modèles *Transformers* [3, 2] montrent certaines limites que nous nous attacherons à surmonter, notamment en portant les efforts sur :

- l'amélioration de l'alignement automatique entre vidéo et texte pour faciliter l'entraînement,
- l'amélioration de la prise en compte des spécificités de la LSF (multi-articulation et spatialisation) par l'enrichissement des représentations vidéos
- l'intégration de connaissances issues de modèles linguistiques de la LSF, comme le modèle formel Azee [1].

4 Profil des candidat.e.s

La/le candidat.e idéal.e a une solide formation en **mathématique** et en **informatique** (M2 ou Ingénieur) avec une spécialisation en vision par ordinateur et/ou en apprentissage automatique. La personne recrutée sera amenée à reprendre et à développer du code en Python pour l'analyse de données et l'apprentissage. Elle devra également travailler en équipe. Un bon niveau d'anglais est requis pour la rédaction d'articles scientifiques.

References

- [1] C. Challant and M. Filhol. Extending AZee with Non-manual Gesture Rules for French Sign Language. In *Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 7007–7016, Turin, Italy, May 2024.
- [2] D. Fabre, J. Lascar, J. Halbout, Y. Ouakrim, A. Braffort, T. Hueber, M. Gouiffès, and D. Beutemps. Exploring Sign-level Strategies to Enhance Automatic Translation of French Sign Language. In *IVA 2025 - 25th ACM International Conference on Intelligent Virtual Agents*, Berlin, Germany, Sept. 2025.
- [3] Y. Ouakrim, H. Bull, M. Gouiffès, D. Beutemps, T. Hueber, and A. Braffort. Mediapi-RGB: Enabling Technological Breakthroughs in French Sign Language (LSF) Research through an Extensive Video-Text Corpus. In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: VISAPP*, volume 2, Rome, Italy, Feb. 2024.