

PROPICTO : Développer des systèmes de traduction de la parole vers des séquences de pictogrammes pour améliorer l'accessibilité de la communication

Lucía Ormaechea^{1,2}, Pierrette Bouillon², Maximin Coavoux¹, Emmanuelle Esperança-Rodier¹, Johanna Gerlach², Jérôme Goulian¹, Benjamin Lecouteux¹, Cécile Macaire¹, Jonathan Mutal², Magali Norré^{2,3}, Adrien Pupier¹, Didier Schwab¹ et Hervé Spechbach⁴

(1) Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP*, LIG, 38000 Grenoble, France

(2) FTI, Université de Geneve, Geneve, Suisse

(3) CENTAL, ILC, Université Catholique de Louvain, Louvain-la-Neuve, Belgique

(4) Hôpitaux Universitaires de Genève, Genève, Suisse

RÉSUMÉ

PROPICTO est un projet financé par l'Agence nationale de la recherche française et le Fonds national suisse de la recherche scientifique, qui vise à créer des systèmes de traduction de la parole vers des pictogrammes avec le français comme langue d'entrée. En développant de telles technologies, nous avons l'intention d'améliorer l'accès à la communication pour les patients non francophones et les personnes souffrant de troubles cognitifs.

ABSTRACT

PROPICTO : Developing Speech-to-Pictograph Translation Systems to Enhance Communication Accessibility

PROPICTO is a project funded by the French National Research Agency and the Swiss National Science Foundation, that aims at creating Speech-to-Pictograph translation systems, with a special focus on French as an input language. By developing such technologies, we intend to enhance communication access for non-French speaking patients and people with cognitive impairments.

MOTS-CLÉS : Pictogrammes, Corpus, Communication Alternative et Augmentée.

KEYWORDS: Pictographs, Corpora, Augmentative and Alternative Communication.

1 Introduction

Les dispositifs de communication alternative et augmentée (CAA) jouent un rôle de plus en plus important auprès des personnes en situation de handicap et de leurs proches. Cependant, l'utilisation de ces technologies (*i.e.*, tableaux de communication ou médias électroniques) peut être fastidieuse (Vaschalde *et al.*, 2018). Pour surmonter ce problème, nous partons de l'hypothèse que les systèmes de traduction de la parole vers des pictogrammes (S2P) peuvent être utiles aux utilisateurs de CAA. En outre, nous pensons qu'ils peuvent améliorer l'accessibilité des services de santé pour les patients qui ne parlent pas la langue locale. Le développement de tels outils nécessite des recherches approfondies

dans plusieurs domaines du traitement de la langue et de la parole (TALP). Nous présentons ici un projet de recherche visant à créer des systèmes qui traduisent automatiquement le français parlé en pictogrammes.

Lancé début 2021, PROPICTO¹ (l'acronyme signifie *PRO*jection de la parole vers des *PICTO*grammes) est un projet franco-suisse de quatre ans, financé à la fois par l'Agence nationale de la recherche française² et par le Fonds national suisse de la recherche scientifique.³ Il est mené en collaboration entre le Département de technologie de la traduction de l'Université de Genève et le Groupe d'étude pour la traduction automatique et le traitement automatisé des langues et de la parole, une équipe du Laboratoire d'informatique de Grenoble.

Dans le cadre de ce projet, nous examinerons plusieurs domaines liés au TALP, à savoir la reconnaissance vocale, l'analyse syntaxique, la simplification des phrases et la génération de pictogrammes. En intégrant cette série de tâches dans différents flux de travail (en fonction du scénario cible), nous proposons de nouveaux systèmes de traduction automatique multimodaux qui convertissent le langage parlé en unités pictographiques. En utilisant cette approche, nous avons l'intention de répondre aux besoins sociétaux et de communication dans les milieux suivants : (1) *handicap*, où un individu cherche à communiquer avec une personne souffrant d'un trouble cognitif, et (2) *médicaux*, où une barrière linguistique existe entre le patient et le praticien.

2 Vue d'ensemble

PROPICTO vise à améliorer la facilité d'utilisation des dispositifs de CAA en tirant parti de solutions basées sur le TALP pour une plus grande accessibilité. Nous concevons de nouvelles méthodes et de nouveaux corpus afin de permettre la transcription directe d'énoncés parlés en séquences de pictogrammes, avec un *objectif général* comme ARASAAC,⁴ et un *objectif spécifique* (i.e., SantéBD <https://santebd.org/> pour les concepts liés à la santé). Le projet devra relever deux défis majeurs :

- La *carence de corpus parallèles de pictogrammes vocaux*, qui constitue un obstacle important à la mise en œuvre de l'apprentissage automatique de pointe (en particulier de bout en bout) ;
- La *nécessité d'une évaluation humaine et automatique approfondie* afin d'évaluer la compréhensibilité des séquences de sortie auprès de divers groupes cibles.

Pour mieux les aborder, nous adopterons une approche en cascade pour notre flux de travail de traitement S2P, qui sera adapté en fonction de l'objectif visé. Ainsi, une première approche privilégiera une stratégie *basée sur les concepts* pour traiter la génération de pictogrammes et sera intégrée dans une architecture S2P à vocation médicale,⁵ composée d'un système de reconnaissance automatique de la parole (ASR) et d'un module neuronal de conversion texte-UMLS⁶ qui définira les pictogrammes à produire et la syntaxe. Une autre stratégie de génération de pictogrammes reposera sur une approche basée sur les mots et sera précédée des étapes suivantes (comme le montre la figure 1) : ASR, analyse des dépendances et simplification des phrases.

1. <https://www.propicto.unige.ch/>

2. <https://anr.fr/en>

3. <https://www.snf.ch/en>

4. <https://arasaac.org/>

5. Pour plus de détails sur cette architecture, se référer à [Mutal et al. 2022](#)

6. Cet acronyme fait référence aux concepts du système de langage médical unifié (UMLS).

L'utilisation d'une approche en cascade est motivée par l'avantage attendu d'une phase par rapport à la suivante. En outre, elle permet d'assurer une meilleure explicabilité du modèle. Notre deuxième proposition d'architecture multimodale partira d'un module ASR, s'appuyant sur des modèles Wav2Vec2.0 à la pointe de la technologie. La tâche d'analyse en dépendance sera traitée avec un analyseur syntaxique de bout en bout dont l'entrée est le signal brut d'un énoncé donné.

L'utilisation du signal brut au lieu des transcriptions nous permet d'utiliser l'information prosodique pour mieux prédire les frontières syntaxiques (Pupier *et al.*, 2022).

L'extraction d'une représentation syntaxique peut à son tour fournir des informations clés pour une simplification plus efficace au niveau de la phrase. La réduction de la complexité linguistique de la transcription d'entrée devrait faciliter l'étape suivante, au cours de laquelle la traduction en pictogrammes sera également régie par des règles de grammaire expertes⁷.

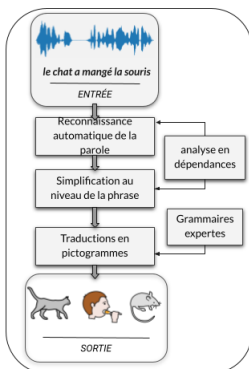


FIGURE 1 – Vue d'ensemble de l'architecture en cascade Parole vers pictogrammes utilisant une approche basée sur *les mots*.

3 Contributions

PROPICTO mettra à la disposition de la communauté scientifique des méthodes et des ressources permettant une traduction du français parlé vers des pictogrammes. Les licences seront aussi permissives que possible et conformes à celles des ensembles pictographiques utilisés. En outre, plusieurs prototypes destinés à différents publics cibles seront mis en production à la fin du projet : (1) dans les situations d'urgence aux Hôpitaux Universitaires de Genève, une démonstration est disponible sur : <https://propicto.demos.unige.ch/pictoDrClient/translate/> et (2) dans des institutions pour enfants et adultes souffrant de handicaps multiples. Ils seront testés en conditions réelles et évalués à l'aide de méthodes humaines et automatiques.

Références

MUTAL J., BOUILLON P., NORRÉ M., GERLACH J. & ORMAECHEA GRIJALBA L. (2022).

7. expressions multi-mot, temps des verbes, noms propres

A Neural Machine Translation Approach to Translate Text to Pictographs in a Medical Speech Translation System – The BabelDr Use Case. In *Proc. Association for Machine Translation in the Americas*, p. 252–263.

PUPIER A., COAVOUX M., LECOUTEUX B. & GOULIAN J. (2022). End-to-End Dependency Parsing of Spoken French. In *Proc. Interspeech 2022*, p. 1816–1820. DOI : [10.21437/Interspeech.2022-381](https://doi.org/10.21437/Interspeech.2022-381).

VASCHALDE C., TRIAL P., ESPERANÇA-RODIER E., SCHWAB D. & LECOUTEUX B. (2018). Automatic Pictogram Generation from Speech to Help the Implementation of a Mediated Communication. In *Proc. Swiss Centre for Barrier-Free Communication 2018*, p. 97–101.