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

Lucía Ormaechea[1,2], Pierrette Bouillon[1], Maximin Coavoux[2], Emmanuelle Esperança-Rodier[2], Johanna Gerlach[1], Jerôme Goulian[2], Benjamin Lecouteux[2], Cécile Macaire[2], Jonathan Mutal[1], Magali Norré[1], Adrien Pupier[2], and Didier Schwab[2]

[1] TIM/FTI, University of Geneva, 1205 Geneva – Switzerland
{firstName.lastName}@unige.ch

[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble – France
{firstName.lastName}@univ-grenoble-alpes.fr

Abstract

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.

1 Introduction

Alternative and augmentative communication (AAC) devices have taken an increasingly important role among people with disabilities and their relatives. However, usage of these technologies (*i.e.*, communication boards or electronic media) may be cumbersome (Vaschalde et al., 2018). To surmount this problem, we argue that Speech-to-Pictograph (S2P) translation systems can be helpful for AAC users. In addition, we believe that they can improve the accessibility of health services for patients not speaking the local language. Developing such tools requires in-depth research on several areas of natural language processing (NLP). In this paper, we present a research project aimed at creating systems that automatically translate spoken French into pictographs.

Launched in early 2021, PROPICTO[1] (the acronym stands for *PROjecting spoken language into PICTOgraphs*) is a French-Swiss four-year project, funded by both the French National Research Agency[2] and the Swiss National Science Foundation.[3] It is conducted as a collaboration between the Department of Translation Technology at the University of Geneva and the Study Group for Machine Translation and Automated Processing of Languages and Speech, affiliated to the Grenoble Informatics Laboratory.

For the purpose of this project, we will examine several NLP-related areas; namely, speech recognition, syntactic parsing, sentence simplification as well as pictograph generation. By integrating this series of tasks into different workflows (depending on the target scenario), we propose novel cross-modal machine translation systems that convert spoken language into pictographic units. Using this approach, we intend to tackle societal and communicative needs in the fields of: (1) *disability*, where an individual seeks to communicate with a person having a cognitive disorder, and (2) *medical settings*, where a language barrier exists between patient and practitioner.

2 Architectural Overview

PROPICTO aims at improving the usability of AAC devices by leveraging NLP-based solutions for greater accessibility. We will design new methods and corpora so as to enable spoken utterances to be transcribed directly into sequences of pictographs, either *general-purpose* like ARASAAC,[4] or *specific-purpose* (*i.e.*, SantéBD[5] for health-related concepts). The project will face two major challenges:

- The *scarcity of parallel speech-pictographs corpora*, which constitutes a high hurdle to

[1] https://www.propicto.unige.ch/

[2] https://anr.fr/en

[3] https://www.snf.ch/en

[4] https://arasaac.org/

[5] https://santebd.org/

Nurminen, Brenner, Koponen, Latomaa, Mikhailov, Schierl, Ranasinghe, Vanmassenhove, Vidal, Aranberri, Nunziatini, Escartín, Forcada, Popovic, Scarton, Moniz (eds.)
Proceedings of the 24th Annual Conference of the European Association for Machine Translation, p. 515–516
Tampere, Finland, June 2023.

the implementation of state-of-the-art machine learning (especially end-to-end-based);

- The *need for extensive human and automatic evaluation* to assess the comprehensibility of the output sequences with diverse target groups.

To better address them, we will adopt a cascade approach for our S2P processing workflow, which will be adapted according to the target setting. Thus, a first approach will favor a *concept-based* strategy to address pictographic generation, and will be integrated within a medical-purpose S2P architecture,[6] consisting of an automatic speech recognition (ASR) system and a neural text-to-UMLS[7] module that will define the pictographs to be produced and the syntax. An alternative pictograph generation strategy will rely on a *word-based* approach, and will be preceded by the next stages (as shown in Figure 1): ASR, Dependency parsing (DP) and sentence simplification.

Using a cascade approach is motivated by the expected benefit of one phase over the next. Additionally, it helps to ensure greater model explainability. Our second proposed cross-modal architecture will start from an ASR module, relying on state-of-the-art Wav2Vec2.0 models. The DP task will be addressed with an end-to-end parser whose input is the raw signal for a given utterance. Using the raw signal instead of transcriptions enables us to use prosodic information to better predict syntactic boundaries (Pupier et al., 2022). Extracting a syntactic-based representation can in turn provide key information for a more effective sentence-level simplification. Reducing the linguistic complexity of the input transcript is likely to help the subsequent step, where the translation into pictographs will also be governed by expert grammar rules.[8]

3 Contributions

PROPICTO will make available to the scientific community methods and resources enabling a translation from spoken French into pictographs. The licenses will be as permissive as possible and conform to those of the pictographic sets being used. Furthermore, several prototypes for different target audiences will be put into production at the end of the project: (1) in emergency settings at the Geneva University Hospitals,[9] and (2) in institutions for children and adults with multiple disabilities. These will be tested in real conditions and evaluated using human and automatic methods.

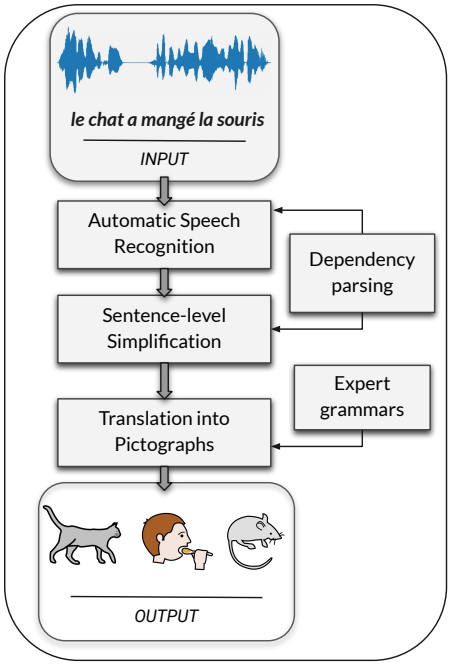

Figure 1: An overview of the Speech-to-Pictograph cascade architecture using a *word-based* approach.

Acknowledgements

This work is part of the PROPICTO project, funded by the Swiss National Science Foundation (N°197864) and the French National Research Agency (ANR-20-CE93-0005).

References

Mutal, Jonathan, Pierrette Bouillon, Magali Norré, Johanna Gerlach, and Lucía Ormaechea Grijalba. 2022. 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*, pages 252–263.

Pupier, Adrien, Maximin Coavoux, Benjamin Lecouteux, and Jerome Goulian. 2022. End-to-End Dependency Parsing of Spoken French. In *Proc. Interspeech 2022*, pages 1816–1820.

Vaschalde, Céline, Pauline Trial, Emmanuelle Esperança-Rodier, Didier Schwab, and Benjamin Lecouteux. 2018. Automatic Pictogram Generation from Speech to Help the Implementation of a Mediated Communication. In *Proc. Swiss Centre for Barrier-Free Communication 2018*, pages 97–101.

[6] For further details on this architecture, see (Mutal et al., 2022).

[7] This acronym refers to Unified Medical Language System (UMLS) concepts.

[8] Like multi-word expressions, verb tenses or proper names.

[9] A demo is available on: `https://propicto.demos.unige.ch/pictoDrClient/translate/`