Un corpus multimodal alignant parole, transcription et séquences de pictogrammes dédié à la traduction automatique de la parole vers des pictogrammes

Cécile Macaire, Chloé Dion, Jordan Arrigo, Claire Lemaire, Emmanuelle Esperança-Rodier, Benjamin Lecouteux, Didier Schwab

Un corpus multimodal alignant parole, transcription et séquences de pictogrammes dédié à la traduction automatique de la parole vers des pictogrammes (Accepté à LREC/Coling 2024)

Cécile Macaire[1] Chloé Dion[1] Jordan Arrigo[1] Claire Lemaire[1,2]
Emmanuelle Esperança-Rodier[1] Benjamin Lecouteux[1] Didier Schwab[1]
(1) Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
(2) LAIRDIL, IUT, Univ. Paul Sabatier, 115 B rte de Narbonne, 31077 Toulouse, France
first.last@univ-grenoble-alpes.fr

RÉSUMÉ

La traduction automatique de la parole vers des pictogrammes peut faciliter la communication entre des soignants et des personnes souffrant de troubles du langage. Cependant, il n'existe pas de formalisme de traduction établi, ni d'ensembles de données accessibles au public pour l'entraînement de systèmes de traduction de la parole vers des pictogrammes. Cet article présente le premier ensemble de données alignant de la parole, du texte et des pictogrammes. Ce corpus comprend plus de 230 heures de parole. Nous discutons de nos choix pour créer une grammaire adaptée à des séquences de pictogrammes. Cette dernière s'articule autour de règles et d'un vocabulaire restreint. La grammaire résulte d'une étude linguistique approfondie des ressources extraites du site Web d'ARASAAC. Nous avons ensuite validé ces règles à l'issue de multiples phases de post-édition par des annotateurs experts. Le corpus proposé est ensuite utilisé pour entraîner un système en cascade traduisant la parole vers des pictogrammes. L'ensemble du corpus est disponible gratuitement sur le site web d'Ortolang sous une licence non commerciale. Il s'agit d'un point de départ pour la recherche portant sur la traduction automatique de la parole vers des pictogrammes.

ABSTRACT

A Multimodal Corpus of Aligned Speech, Text, and Pictogram Sequences for Speech-to-Pictogram Machine Translation

The automatic translation of spoken language into pictogram units can facilitate communication involving individuals with language impairments. However, there is no established translation formalism or publicly available datasets for training end-to-end speech translation systems. This paper introduces the first aligned speech, text, and pictogram translation dataset ever created in any language. We provide a French dataset that contains 230 hours of speech resources. We create a rule-based pictogram grammar with a restricted vocabulary and include a discussion of the strategic decisions involved. It takes advantage of an in-depth linguistic study of resources taken from the ARASAAC website. We validate these rules through multiple post-editing phases by expert annotators. The constructed dataset is then used to experiment with a Speech-to-Pictogram cascade model, which employs state-of-the-art Automatic Speech Recognition models. The dataset is freely available under a non-commercial licence. This marks a starting point to conduct research into the automatic translation of speech into pictogram units.

MOTS-CLÉS : Pictogrammes, Reconnaissance Automatique de la Parole, Traduction Automatique.

KEYWORDS: Pictograms, Automatic Speech Recognition, Machine Translation.