



HAL
open science

Plateformes pour la création de données en pictogrammes

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

► To cite this version:

Cécile Macaire, Jordan Arrigo, Chloé Dion, Claire Lemaire, Emmanuelle Esperança-Rodier, et al.. Plateformes pour la création de données en pictogrammes. 18e Conférence en Recherche d'Information et Applications, 16e Rencontres Jeunes Chercheurs en RI, 30e Conférence sur le Traitement Automatique des Langues Naturelles, 25e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues, 2023, Paris, France. pp.6-9. hal-04208581

HAL Id: hal-04208581

<https://hal.science/hal-04208581>

Submitted on 15 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Plateformes pour la création de données en pictogrammes

Cécile Macaire¹ Chloé Dion¹ Jordan Arrigo¹ Emmanuelle Esperança-Rodier¹
Benjamin Lecouteux¹ Claire Lemaire^{1,2} Didier Schwab¹

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

(2) LAIRDIL, IUT, Univ. Paul Sabatier, 115 B rte de Narbonne, 31077 Toulouse, France

prénom.nom@univ-grenoble-alpes.fr

RÉSUMÉ

Nous présentons un ensemble de trois outils en ligne pour la création de données en pictogrammes dans le cadre du projet ANR PROPICTO. Chaque outil a été créé pour un objectif précis : (1) annoter des données textuelles en pictogrammes, (2) créer un dictionnaire "bilingue" français-pictogrammes, et (3) post-éditer des phrases annotées en pictogrammes. Ces trois outils nous serviront à développer des outils de traduction automatique vers les unités pictographiques. Il s'agit de créer un nouveau langage élaboré destiné à des personnes en situation de handicap communiquant déjà à partir de pictogrammes. Ce langage sera également proposé à des personnes ne parlant pas la langue du pays dans lequel elles se trouvent et souhaitant véhiculer un message dans un contexte d'urgence médicale.

ABSTRACT

Platforms for creating pictogram data

We present a set of three online tools for the creation of pictogram data in the framework of the ANR PROPICTO project. Each tool has been created for a specific purpose : (1) annotating textual data in pictograms, (2) creating a "bilingual" French-pictogram dictionary, and (3) post-editing sentences annotated in pictograms. These three tools will be used to develop tools for automatic translation into pictogram units. The aim is to create a new elaborate language for people with disabilities who already communicate using pictograms. This language will also be proposed to people who do not speak the language of the country in which they are and who wish to convey a message in a medical emergency context.

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

KEYWORDS: Pictographs, Corpora, Augmentative and Alternative Communication.

1 Introduction

Le projet PROPICTO (PROjection du langage Oral vers des unités PICTOgraphiques)¹ est situé dans le contexte de la Communication Alternative et Augmentée (CAA). La CAA fournit aux personnes ne pouvant communiquer via les canaux traditionnels (langue des signes, parole, gestes) un moyen de s'exprimer de manière efficace et autonome par l'utilisation de tablettes, classeurs de pictogrammes, etc. La mise en place d'outils de CAA répond à de nombreux besoins sociétaux, spécifiquement dans le domaine médical (communiquer avec des personnes ne pouvant s'exprimer dans la langue du pays

1. <https://anr.fr/Projet-ANR-20-CE93-0005>

où elles se trouvent), et du handicap (communiquer avec des personnes en situation de handicap). Dans le cadre de ce projet, il est attendu de proposer des méthodes pour transcrire automatiquement la parole vers des pictogrammes. Le projet comporte également une partie de création de ressources spécifiques à cette tâche, et ensuite distribuées à la communauté scientifique. C'est donc, dans ce but, que ces trois plateformes ont été créées pour (1) produire des données textuelles annotées en pictogrammes, (2) créer un vocabulaire pictographique, et (3) post-éditer des phrases traduites en pictogrammes pour l'évaluation des outils de traduction.

2 Présentation des plateformes

2.1 AnnotPicto : Annotation de phrases en pictogrammes

La première plateforme² permet, à partir d'une phrase renseignée, de rechercher, pour chaque terme, les pictogrammes associés dans deux banques, Arasaac³ et Mulberry⁴. Une version annotée en pictogrammes est alors créée. Une phase de tokenisation et de lemmatisation est appliquée pour récupérer le lemme associé à chaque terme par l'utilisation d'un modèle spacy⁵.

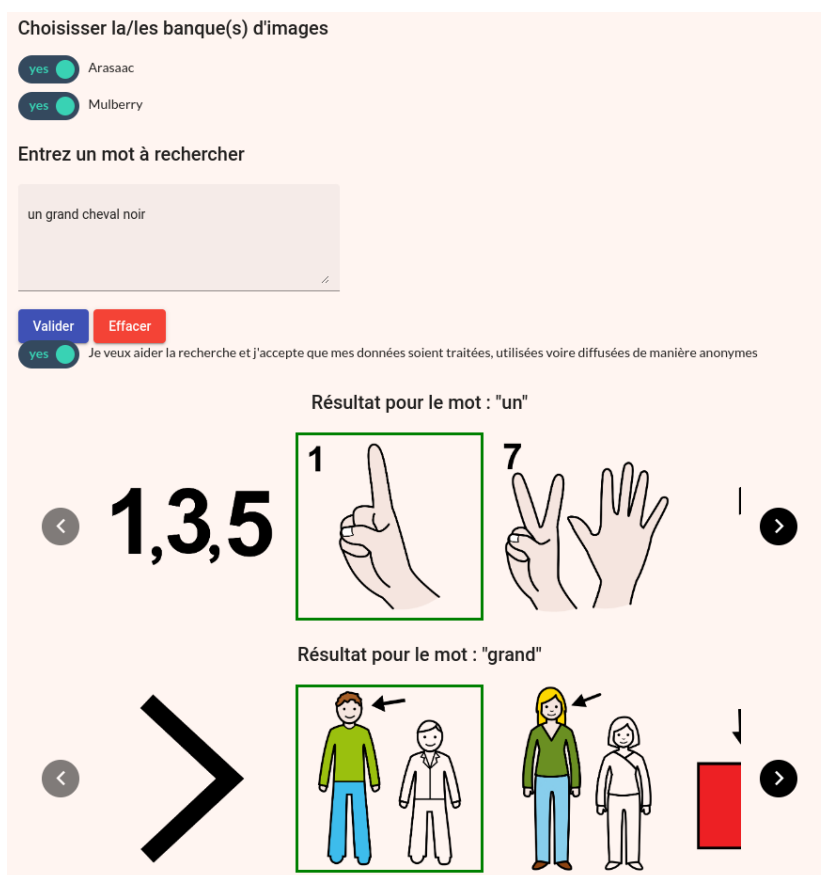


FIGURE 1 – Interface pour l'annotation de phrases en pictogrammes.

2. <http://interaactionpicto.net>
3. <https://arasaac.org/>
4. <https://mulberrysymbols.org/>
5. <https://spacy.io/models/fr>

La personne peut ensuite sélectionner un pictogramme parmi ceux proposés. Il se peut qu'aucune proposition ne soit présentée, car le terme n'existe pas sous forme pictographique dans les banques. La phrase annotée peut ensuite être sauvegardée dans un fichier PDF, avec plusieurs options d'affichage (taille de la police, ajout d'une bordure, et nombre de pictogrammes par lignes). Par cet outil, présenté Figure 1, la création de données textuelles est simplifiée et rapide.

2.2 DicoPicto : création d'un vocabulaire en pictogrammes

La deuxième plateforme est un outil de création d'un vocabulaire en pictogrammes Arasaac. Pour obtenir une traduction précise et fiable de la parole en pictogrammes, il est nécessaire d'avoir un vocabulaire cohérent. Actuellement, Arasaac propose plusieurs pictogrammes pour un même mot (comme le mot *chat* avec quatre pictogrammes⁶ différents), tandis que pour d'autres, aucun n'a été défini (comme pour le mot *écœurer*). Cette variabilité de possibilités et l'absence de certains termes peuvent affecter la qualité de la traduction. De plus, certains pictogrammes peuvent être inappropriés compte tenu du contexte. DicoPicto⁷ (cf. Figure 2) propose donc de rechercher un mot dans Arasaac, et de sélectionner le pictogramme le plus adapté. Son identifiant est ainsi récupéré et associé au terme.

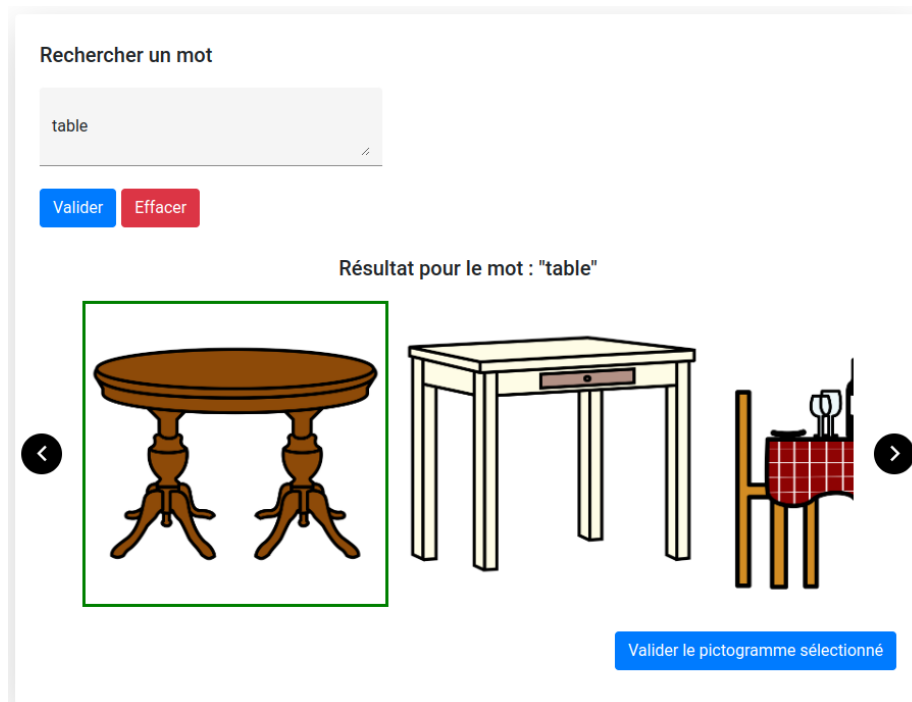


FIGURE 2 – Interface de DicoPicto avec l'exemple du mot *table*.

2.3 PostEditPicto : Post-édition de phrases en pictogrammes

PostEditPicto⁸ est une plateforme de post-édition de phrases traduites en pictogrammes Arasaac. La post-édition a son intérêt puisqu'elle permet de corriger les traductions proposées pour les

6. <https://arasaac.org/pictograms/search/chat>

7. <https://lig-interaactionpicto.imag.fr/#/dicoPicto>

8. <https://lig-interaactionpicto.imag.fr/#/postEditPicto>

rendre intelligibles (Robert, 2010). Ainsi, ce procédé permet d'évaluer rapidement la qualité des traductions, et de donner des pistes d'amélioration concrètes. De plus, les corrections peuvent servir de référence pour des phrases qui n'ont pas de traduction en pictogrammes, et donc créer des ressources supplémentaires. PostEditPicto permet d'ajouter, de déplacer (par une fonction drag and drop) et de supprimer chaque pictogramme présenté. Une fonction recherche permet de récupérer un mot précis dans la banque Arasaac pour l'ajouter au contenu proposé. L'interface est présentée Figure 3.



FIGURE 3 – Plateforme PostEditPicto avec *vous avez beaucoup de chance* comme phrase exemple.

3 Conclusion & Perspectives

Ces trois plateformes présentent un intérêt tout particulier pour la création de ressources pictographiques dans le cadre du projet ANR PROPICTO. La plateforme DicoPicto a notamment permis l'annotation de plus d'une centaine de termes. La plateforme PostEditPicto sera plus précisément utilisée dans le cadre d'une évaluation des traductions proposées auprès d'un public d'aidants (familles et personnels médicaux). La plateforme de création de données textuelles en unités pictographiques, AnnotPicto, est à destination des familles cherchant un outil efficace pour annoter leurs propres ressources. Nous intégrerons d'autres banques de pictogrammes dans les plateformes : Mulberry, Sclera.

Remerciements

Ce travail a bénéficié d'un financement du Fond National Suisse (No. 197864) et de l'Agence Nationale de la Recherche, via le projet PROPICTO (ANR-20-CE93-0005).

Références

ROBERT A.-M. (2010). La post-édition : l'avenir incontournable du traducteur? *Traduire. Revue française de la traduction*, (222), 137–144. DOI : [10.4000/traduire.460](https://doi.org/10.4000/traduire.460).