



HAL
open science

**Actes de CORIA-TALN 2023. Actes de la 30e
Conférence sur le Traitement Automatique des Langues
Naturelles (TALN)**

Christophe Servan, Anne Vilnat

► **To cite this version:**

Christophe Servan, Anne Vilnat. Actes de CORIA-TALN 2023. Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN): volume 5: démonstrations. CORIA - TALN 2023, 2023. hal-04462998

HAL Id: hal-04462998

<https://hal.science/hal-04462998>

Submitted on 16 Feb 2024

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.



Distributed under a Creative Commons Attribution 4.0 International License



*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
(CORIA-TALN) ¹*

Actes de CORIA-TALN 2023.

Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles
(TALN),
volume 5 : démonstrations

Christophe Servan, Anne Vilnat (Éds.)

Paris, France, 5 au 9 juin 2023

1. <https://coria-taln-2023.sciencesconf.org/>

Avec le soutien de



Préface

Organisée conjointement par les laboratoires franciliens sous l'égide de l'Association francophone de Recherche d'Information et Applications (ARIA) et l'Association pour le Traitement Automatique des Langues (ATALA), la conférence CORIA-TALN-RJCRI-RECITAL 2023 regroupe :

- la 18ème Conférence en Recherche d'Information et Applications (CORIA)
 - la 30ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN) ;
- ainsi que les deux conférences associées, destinées aux jeunes chercheuses et chercheurs :
- Les 16ème Rencontres Jeunes Chercheurs en RI (RJCRI)
 - la 25ème Rencontre des Etudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL)

La conférence TALN (Traitement Automatique des Langues Naturelles) est un rendez-vous annuel qui offre, depuis 1994, le plus important forum d'échange international francophone aux acteurs universitaires et industriels des technologies de la langue. Cet événement, qui accueille habituellement près de 250 participants, couvre toutes les avancées récentes en matière de communication écrite et parlée et de traitement informatique de la langue notamment la recherche et l'extraction d'information, la fouille de textes, le dialogue homme-machine, la fouille d'opinions, la traduction automatique, les systèmes de questions-réponses, le résumé automatique...

Cette année, ont été soumis 51 articles longs et 12 articles courts pour la conférence principale, dont respectivement 29 ont été acceptés pour une présentation orale (dont 2 prises de position) et 9 pour une présentation sous forme de posters. 19 présentations courtes, sous forme de posters, d'articles déjà publiés lors de conférences internationales complètent le programme de la conférence, ainsi que des démonstrations et des présentations de projets en cours. L'alternance de sessions communes entre TALN, CORIA et RJC et de sessions plus spécifiques devraient permettre de susciter des échanges fructueux.

En complément de la conférence principale, se tiennent les ateliers "Défi Fouille de Texte" (DEFT), "Atelier sur l'analyse et la recherche de textes scientifiques" (ARTS), "Humain ou pas humain ? : les nouveaux défis pour les humains" (hOUPSh) et le tutoriel "Apprentissage Profond pour le TAL français pour les débutants" (TutoriAL). Ces ateliers et tutoriel illustrent à la fois des tendances nouvelles présentes dans la communauté et des activités récurrentes.

Un grand merci à toutes celles et tous ceux qui ont soumis leurs travaux, ainsi qu'aux membres du comité de programme et aux relectrices et relecteurs pour le travail qu'ils ont accompli. Ce sont eux qui font vivre la conférence. Merci au comité d'organisation réparti sur la région parisienne, et aux sponsors qui nous ont permis d'organiser cet événement.

Christophe Servan et Anne Vilnat, co-présidents de TALN

Comités

Comité de programme

Présidents

- Christophe SERVAN
- Anne VILNAT

Membres

- Rachel BAWDEN
- Caroline BRUN
- Marie CANDITO
- Rémi CARDON
- Pascal DENIS
- Yannick ESTEVE
- Benoît FAVRE
- Amel FRAISSE
- Thomas GERALD
- Natalia GRABAR
- Lydia-Mai HO-DAC
- José MORENO
- Vassilina NIKOULINA
- Yannick PARMENTIER
- Sylvain POGODALLA
- Solène QUINIOU
- Didier SCHWAB
- Iris TARAVELLA-ESHKOL

Comité d'organisation

- Marie CANDITO
- Thomas GERALD
- José MORENO
- Benjamin PIWOWARSKI
- Christophe SERVAN
- Laure SOULIER
- Anne VILNAT

Table des matières

Génération automatique de jeux de mots à base de prénoms	1
<i>Mathieu Dehouck, Marine Delaborde</i>	
Éditeur logiciel pour une représentation graphique de la langue des signes française	3
<i>Michael Filhol, Thomas Von Ascheberg</i>	
Plateformes pour la création de données en pictogrammes	6
<i>Cécile Macaire, Jordan Arrigo, Chloé Dion, Claire Lemaire, Emmanuelle Esperança-Rodier, Benjamin Lecouteux, Didier Schwab</i>	
Voice2Picto : un système de traduction automatique de la parole vers des pictogrammes	10
<i>Cécile Macaire, Emmanuelle Esperança-Rodier, Benjamin Lecouteux, Didier Schwab</i>	

Génération automatique de jeux de mots à base de prénoms

Mathieu Dehouck¹ Marine Delaborde^{1,2}

(1) Lattice (UMR 8094), CNRS / ENS / Université Sorbonne Nouvelle

(2) LT2D (EA 7518), CY Cergy Paris Université

mathieu.dehouck@sorbonne-nouvelle.fr, marine.delaborde@cyu.fr

RÉSUMÉ

Nous présentons un système automatique de génération de blagues au format « Monsieur et Madame ». Ces blagues seront ensuite rendues accessibles sur un site internet où les visiteurs seront invités à les noter. Le tout servira ensuite à créer un corpus pour des études ultérieures.

ABSTRACT

Automatic generation of first name based wordplays

We present a system to automatically generate "Mister and Misses" type of jokes. These jokes will then be displayed on a website where visitors will have to opportunity to give feedback in order to generate a corpus.

MOTS-CLÉS : Génération automatique d’humour, jeux de mots, corpus.

KEYWORDS: Computational humor generation, puns, wordplay, corpus.

1 La détection et la génération automatique de traits d’humour

L’humour n’est pas toujours évident à déceler pour un humain et c’est donc aussi le cas pour une machine. Il existe pourtant des systèmes de classification automatique consacrés à la détection de traits d’humour, notamment avec les challenges SemEval 2017 et 2021 (Potash *et al.*, 2017; Meaney *et al.*, 2021) sur des tweets en anglais ou encore les challenges HAHA sur des tweets en espagnol (Castro *et al.*, 2018; Chiruzzo *et al.*, 2019, 2021). La génération automatique de traits d’humour est quant à elle, encore relativement peu étudiée. Amin & Burghardt (2020) ont réalisé un état-de-l’art dénombrant 12 systèmes publiés entre 1994 et 2020.

2 Des jeux de mots avec des noms propres

Nous souhaitons tester la capacité des modèles de langues à générer et/ou résoudre des blagues basées sur la prononciation¹. En effet, les grands modèles de langues pré-entraînés sur de larges corpus sont entraînés à prédire des mots manquants (mots suivants ou mots masqués) mais ne sont à priori pas sensibles à la prononciation, qui est l’un des ressorts principaux de ce type de blagues. De plus, le passage de la prononciation de la séquence Prénom-Nom à celle de l’énoncé attendue présente parfois des changements de sons (voisement, altération des voyelles), rendant la blague d’autant plus difficile

1. Des blagues du type : « La famille *Unetelle* a trois enfants. Comment s’appellent-ils ? »

à résoudre. Pour ce faire, il nous faut un corpus de référence. L'on se propose donc dans un premier temps de générer des séquences Prénom(s)-Nom qui correspondent à des énoncés valides en français.

On utilise le lexique Morphalou (ATILF, 2023) comme source de prononciation (XSAMPA) et d'information morphologique ; les prénoms quant à eux, ont été extraits du Wiktionnaire. On crée ensuite des tries (arbres de préfixes) annotés morphologiquement, associant à leur prononciation les formes orthographiques. Ainsi l'on peut trouver les mots dont la prononciation est compatible avec celle d'un prénom. Enfin, on extrait des séquences de parties-du-discours avec de l'information morphologique de treebanks Universal Dependencies (Zeman *et al.*, 2022) pour servir de colonnes vertébrales syntaxiques aux énoncés.

Les blagues ainsi générées sont rendues accessibles sur un site internet² où les visiteurs sont invités à les noter. Leur retour permettra d'ajouter des traits aux blagues produites en sélectionnant des critères comme « contenu explicite », « incompréhensible » ou un degré d'humour.

Références

- AMIN M. & BURGHARDT M. (2020). A survey on approaches to computational humor generation. In *Proceedings of the The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, p. 29–41, Online : International Committee on Computational Linguistics.
- ATILF (2023). Morphalou. ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.
- CASTRO S., CHIRUZZO L. & ROSÁ A. (2018). Overview of the haha task : Humor analysis based on human annotation at ibereval 2018. In *IberEval@ SEPLN*, p. 187–194.
- CHIRUZZO L., CASTRO S., ETCHEVERRY M., GARAT D., PRADA J. J. & ROSÁ A. (2019). Overview of haha at iberlef 2019 : Humor analysis based on human annotation. In *IberLEF@ SEPLN*, p. 132–144.
- CHIRUZZO L., CASTRO S., GÓNGORA S., ROSÁ A., MEANEY J. & MIHALCEA R. (2021). Overview of haha at iberlef 2021 : Detecting, rating and analyzing humor in spanish. *Procesamiento del Lenguaje Natural*, **67**, 257–268.
- MEANEY J. A., WILSON S., CHIRUZZO L., LOPEZ A. & MAGDY W. (2021). SemEval 2021 task 7 : HaHackathon, detecting and rating humor and offense. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, p. 105–119, Online : Association for Computational Linguistics. DOI : [10.18653/v1/2021.semeval-1.9](https://doi.org/10.18653/v1/2021.semeval-1.9).
- POTASH P., ROMANOV A. & RUMSHISKY A. (2017). SemEval-2017 task 6 : #HashtagWars : Learning a sense of humor. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, p. 49–57, Vancouver, Canada : Association for Computational Linguistics. DOI : [10.18653/v1/S17-2004](https://doi.org/10.18653/v1/S17-2004).
- ZEMAN D., NIVRE J. & AL. (2022). Universal dependencies 2.11. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

2. <https://apps.lattice.cnrs.fr/aligator/>

Éditeur logiciel pour une représentation graphique de la langue des signes française

Michael Filhol^{1,2} Thomas von Ascheberg¹

(1) LISN, CNRS

(2) Université Paris-Saclay

michael.filhol@cnrs.fr, ascheberg@lisn.fr

RÉSUMÉ

Démonstration d'un logiciel d'édition d'AZVD, un formalisme graphique pour la langue des signes française. Basé sur des productions spontanées de locuteurs, AZVD est conçu pour maximiser son adoptabilité par la communauté signante.

ABSTRACT

Software editor for a graphical representation of French Sign Language

Demonstration of a software editor for AZVD, a graphical formal system for French Sign Language. Based on real spontaneous productions by signers, AZVD is designed to maximise its potential for adoption by the signing community.

MOTS-CLÉS : Langue des signes, représentation graphique, écriture, éditeur.

KEYWORDS: Sign Language, graphical representation, writing system, editor.

Les langues dotées d'une forme écrite sont en général équipées de logiciels d'édition, et plus aisément de traitements automatiques en conséquence. Ce n'est en revanche pas le cas des langues des signes (LS) qui en sont dépourvues. Des systèmes de représentation ont bien été conçus pour les LS, avec des objectifs variés (fig. 1), mais on note qu'aucun n'a été adopté par les communautés de signeurs au sens large.

Le besoin d'une représentation semble pourtant exister. Pour prendre des notes, préparer des supports de discours ou travailler des traductions de manière détachée d'un texte source, les sourds et traducteurs ont régulièrement recours à des "schémas de verbalisation" (fig. 2a), ou VD (pour *verbalising diagram*). Ceux-ci mettent en jeu icônes, dessins, lignes et flèches pour les relier... mais ils sont spontanés, à savoir librement composés par leur auteur plutôt que conformes à des règles apprises ou posées a priori. Il n'est donc pas possible de requêter ou interpréter ces dessins pour, par exemple, en synthétiser automatiquement un rendu en LS par un avatar signant.



FIGURE 1 – Exemples de représentations graphiques conçues pour les LS.

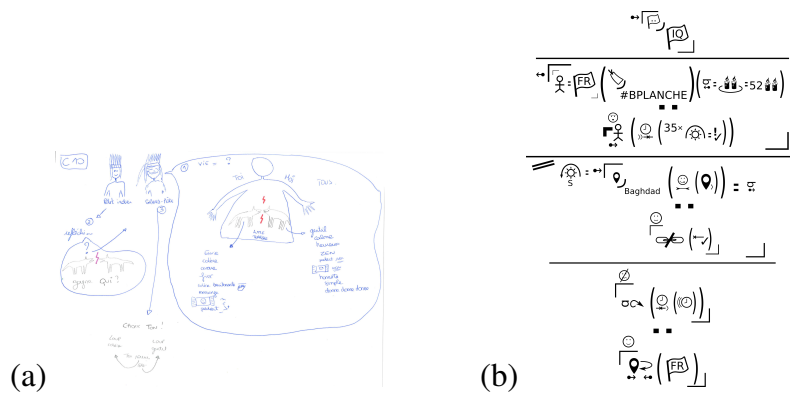


FIGURE 2 – (a) Schématisation VD (spontanée) d’une histoire courte dont la production signée dure environ 1 min – (b) Représentation AZVD (formelle) d’une brève journalistique de 19 s.

Cela dit, des régularités ont été observées, aussi bien intra- qu’inter-personnelles (réf. anonymisée), au point où des icônes et dispositions graphiques récurrentes portant un sens identifiable se réalisent parfois de manière identique. On parle de “forme associée”; la lecture devient alors déterminée. Basée sur ces régularités et sur la représentation formelle *AZee* des discours en LS (réf. anonymisée), la piste “AZVD” (fig. 2b) récemment proposée définit un formalisme graphique, éditable et requêtable, visant à maximiser l’adoptabilité par les locuteurs (réf. anonymisée). Contrairement aux VD spontanés, AZVD offre un système graphique régulier qui pour chaque schéma détermine une lecture unique, donc offre la possibilité d’en générer automatiquement un rendu par avatar.

La démonstration proposée est celle d’un éditeur que nous avons développé (fig. 3) pour assister la création et l’édition de schémas AZVD, ainsi que de générer les expressions *AZee* correspondantes, c’est-à-dire produisant la lecture correcte si injectées dans un système d’animation.

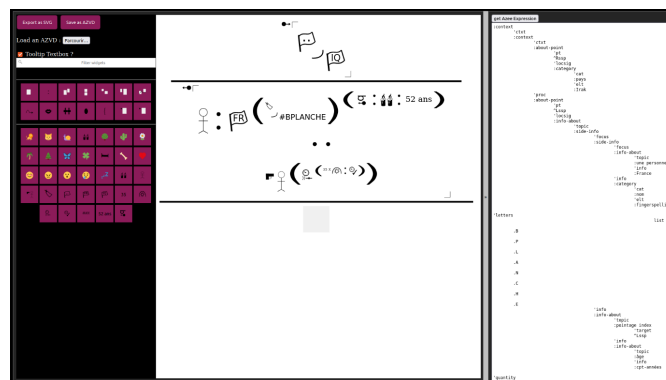


FIGURE 3 – Capture d’écran de l’éditeur AZVD – à gauche : la palette d’icônes combinables dans les schémas – au milieu : le canvas principal d’édition des schémas – à droite : l’expression *AZee* déterminée par le dessin, animable par un avatar.

Nous voyons deux perspectives immédiates de ce travail. D’abord, son utilisation par les experts *AZee* pour l’édition d’expressions en soi, et mesurer le gain de temps apporté à la tâche de création d’expressions. Ensuite, nous aimerions injecter ces expressions générées par l’éditeur dans un système d’animation d’avatar pour produire les énoncés en LS, et à terme mettre le système complet entre les mains de signeurs.

Remerciement Cette recherche a été en partie financée par le projet européen *EASIER (Intelligent Automatic Sign Language Translation)*, accord de financement *Horizon 2020* n° 101016982.

Références

HANKE T. (2004). Hamnosys—representing sign language data in language resources and language processing contexts. In O. S. . C. VETTORI, Éd., *Proceedings of the workshop on the Representation and Processing of Sign Languages*, p. 1–6 : European Language Resources Association (ELRA).

PRILLWITZ S., LEVEN R., ZIENERT H., HANKE T. & HENNING J. (1989). Hamnosys version 2.0, hamburg notation system for sign languages, an introductory guide. *International studies on Sign Language communication of the Deaf*, **5**. Signum press, Hamburg.

STOKOE W. C., CASTERLINE D. C. & CRONEBERG C. G. (1965). *A Dictionary of American Sign Language on Linguistic Principles*. Washington, DC.

SUTTON V. (2014). *Lessons in SignWriting*. The SignWriting Press, 4th édition.

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.

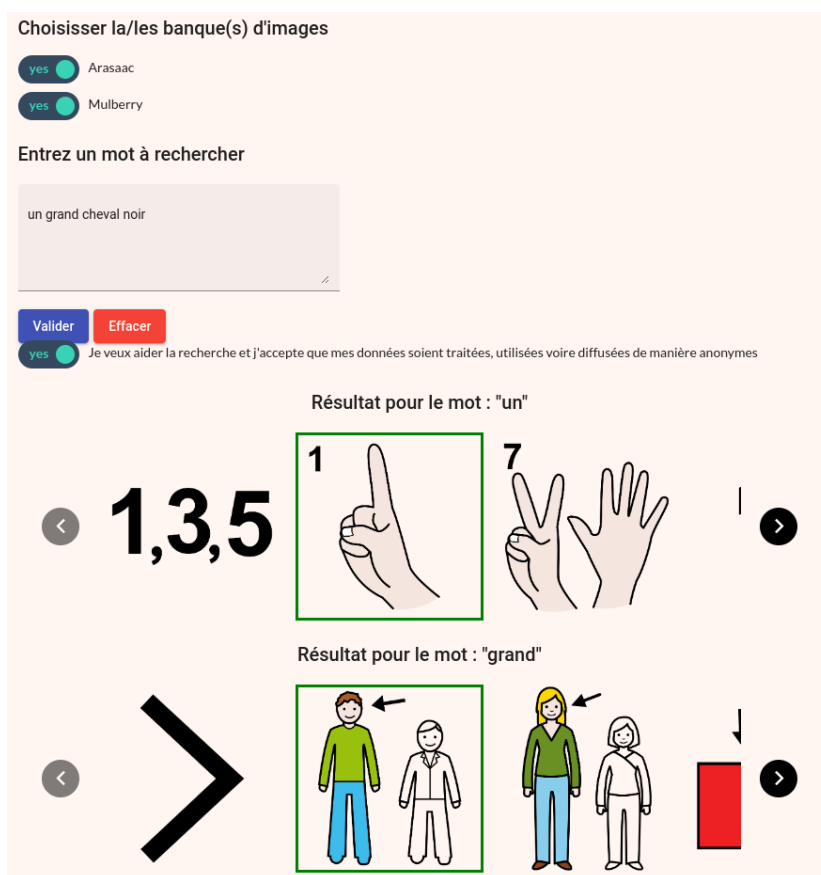


FIGURE 1 – Interface pour l'annotation de phrases en pictogrammes.

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.

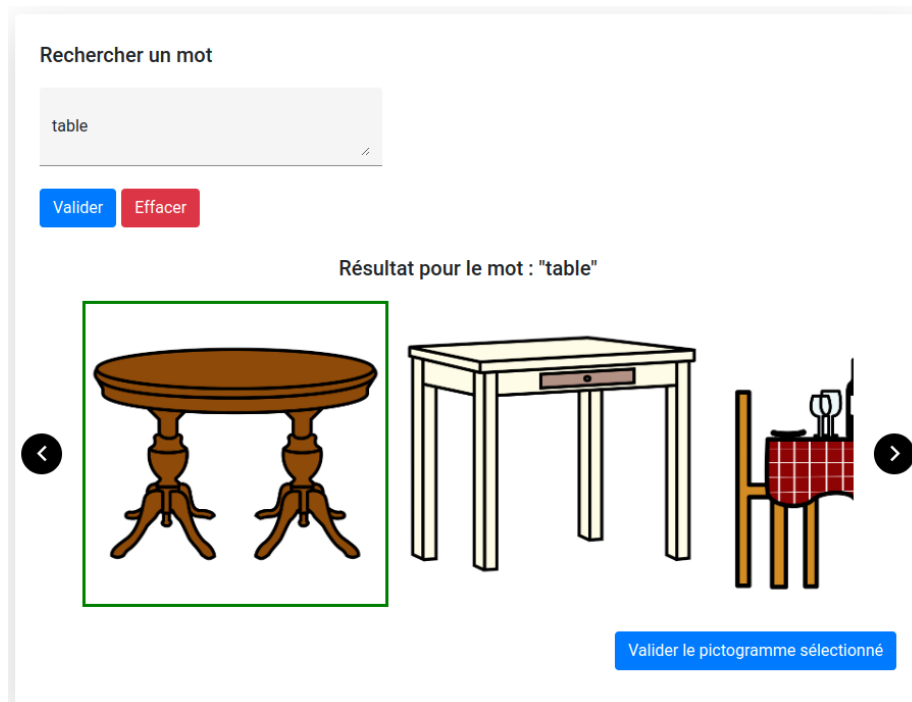


FIGURE 2 – Interface de DicoPicto avec l'exemple du mot *table*.

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).

Voice2Picto : un système de traduction automatique de la parole vers des pictogrammes

Cécile Macaire¹ Emmanuelle Esperança-Rodier¹ Didier Schwab¹
Benjamin Lecouteux¹

(1) Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP*, LIG, 38000 Grenoble, France
prénom.nom@univ-grenoble-alpes.fr

RÉSUMÉ

Nous présentons Voice2Picto, un système de traduction permettant, à partir de l'oral, de proposer une séquence de pictogrammes correspondants. S'appuyant sur des technologies du traitement automatique du langage naturel, l'outil a deux objectifs : améliorer l'accès à la communication pour (1) les personnes allophones dans un contexte d'urgence médicale, et (2) pour les personnes avec des difficultés de parole. Il permettra aux personnes des services hospitaliers, et aux familles de véhiculer un message en pictogrammes facilement compréhensible auprès de personnes ne pouvant communiquer via les canaux traditionnels de communication (parole, gestes, langue des signes). Dans cet article, nous décrivons l'architecture du système de Voice2Picto et les pistes futures. L'application est en *open-source* via un dépôt Git <https://github.com/macairececile/Voice2Picto>.

ABSTRACT

Voice2Picto : an automatic speech translation system into pictograms

We present Voice2Picto, a translation system that uses spoken language to propose a sequence of corresponding pictograms. Based on natural language processing technologies, the tool has two objectives : to improve access to communication (1) for people not speaking the local language in medical settings, and (2) for people having a cognitive disorder. It will allow practitioners and families to convey a message in pictograms that can be easily understood by people who cannot communicate via traditional communication channels (speech, gestures, sign language). In this article, we describe the architecture of the Voice2Picto system and the future directions. The application is open-source via a Git repository <https://github.com/macairececile/Voice2Picto>.

MOTS-CLÉS : Traduction Automatique, Reconnaissance de la Parole, Pictogrammes, Communication Alternative et Augmentée.

KEYWORDS: Automatic Translation, Speech Recognition, Pictographs, Alternative and Augmentative Communication.

1 Introduction

Lorsqu'une personne ne peut utiliser les canaux traditionnels de communication (parole, gestes, langue des signes) pour exprimer un message, une communication alternative et augmentée (CAA) peut être mise en place. On retrouve dans la CAA, l'utilisation de pictogrammes, image représentant un concept plus ou moins concret (un mot unique, une entité nommée, une expression polylexicale

par exemple¹). Actuellement, la prise en main d’outils de CAA (par exemple, des classeurs de communication sous forme de pictogrammes) est longue et difficile (Cataix-Nègre, 2017). En effet, un temps d’adaptation et d’apprentissage du fonctionnement de chaque outil est nécessaire. De plus, un manque de soutien institutionnel est présent, entre des équipes médicales non formées, et des lenteurs administratives pour leur mise en place (on parle ici des instituts médicalisés). Pourtant, l’utilisation de la CAA a un impact social positif pour les personnes en situation de handicap. La Croix-Rouge a identifié une réduction du stress, une amélioration de l’autonomie et de l’état de santé, ainsi qu’une meilleure sérénité et plaisir des personnes dans leur vie quotidienne (Croix-Rouge, 2021). L’accès à la communication par tous et pour tous reste donc un défi majeur. Proposer une séquence de pictogrammes à partir de la voix permettra aux personnes des services hospitaliers, et aux familles sans connaissances préalables de véhiculer un message en pictogrammes facilement compréhensible par les utilisateurs de CAA.

2 Présentation du système

Voice2Picto est un système cascade de traduction automatique de la parole vers des pictogrammes combinant différents domaines du traitement automatique du langage naturel. L’utilisation d’une telle architecture permet d’améliorer l’explicabilité du modèle, car nous pouvons évaluer les capacités et donc

l’impact positif ou négatif d’une phase sur la suivante. Son architecture globale est présentée Figure 1 et s’articule entre quatre modules.

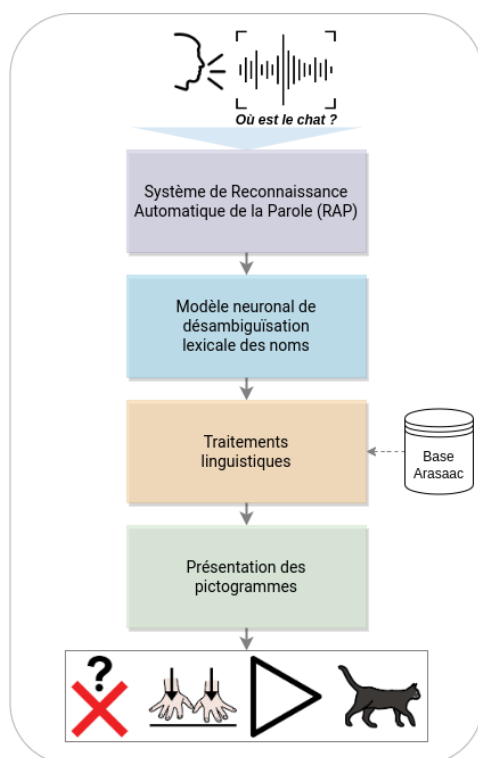


FIGURE 1 – Schéma de Voice2Picto.

Le premier module est un système de reconnaissance automatique de la parole (RAP), basé sur une boîte à outils de reconnaissance vocale Vosk². L’API permet une réponse presque instantanée grâce à l’utilisation de modèles optimisés, améliorant ainsi les performances de l’application. Le modèle appelé a été entraîné avec la boîte à outils Kaldi (Povey *et al.*, 2011) et obtient 23.95% de taux d’erreurs au niveau des mots sur CommonVoice (Ardila *et al.*, 2020).

À partir de la transcription définie par le module de RAP, un modèle neuronal de désambiguïsation lexicale sur les noms est appliqué, proposé dans Vial *et al.* (2019). Récupérer le sens le plus probable a son importance. Certains termes n’ont pas la même signification compte tenu du contexte (par exemple, *café* peut désigner la graine, la boisson ou le lieu de consommation). Le système neuronal comporte, en entrée, des couches d’encodeur de type *transformers* et utilisent les vecteurs de mots contextuels pré-entraînés de type BERT (Devlin *et al.*, 2019). L’architecture se compose d’une couche de décodage correspondant à un classificateur softmax capable de classer un mot dans l’ensemble des synsets possibles de WordNet (Miller, 1995). Le sens WordNet le plus probable

1. Les pictogrammes sont disponibles sur Arasaac : <https://arasaac.org/pictograms/search>

2. <https://github.com/alphacep/vosk-api>

est alors prédit. Dans Voice2Picto, le modèle utilisé a été entraîné sur deux corpus SemCor et Wngt adaptés au français (Le *et al.*, 2020). Le score F1 sur le corpus d'évaluation SemEval 2013 est de 51.28%.

Un module de traitements linguistiques vient compléter ces deux modules. L'objectif ici est de "nettoyer" le texte, pour ensuite identifier les pictogrammes associés à chaque terme. Ainsi, une traduction pertinente sera proposée. Une phase de tokenisation, de lemmatisation, et d'étiquetage morphosyntaxique est appliquée en premier (à l'aide de la librairie spacy³). Puis, les expressions multi-mots et les entités nommées sont recherchées. Une table d'alignement entre lemmes, sens WordNet et identifiants ARASAAC permet de récupérer l'identifiant associé à chaque lemme ou au sens WordNet dans le cas d'un nom. Enfin, chaque image est récupérée dans la banque d'Arasaac grâce à l'identifiant, puis affichée.

L'interface de l'application est présentée Figure 2. Elle se veut simple d'utilisation, rapide et ergonomique. Les utilisations possibles sont nombreuses, voire infinies puisqu'elles peuvent rentrer dans de nombreuses situations de la vie quotidienne, telles que communiquer un besoin, une envie, une émotion, raconter une histoire ou poser une question à une personne.

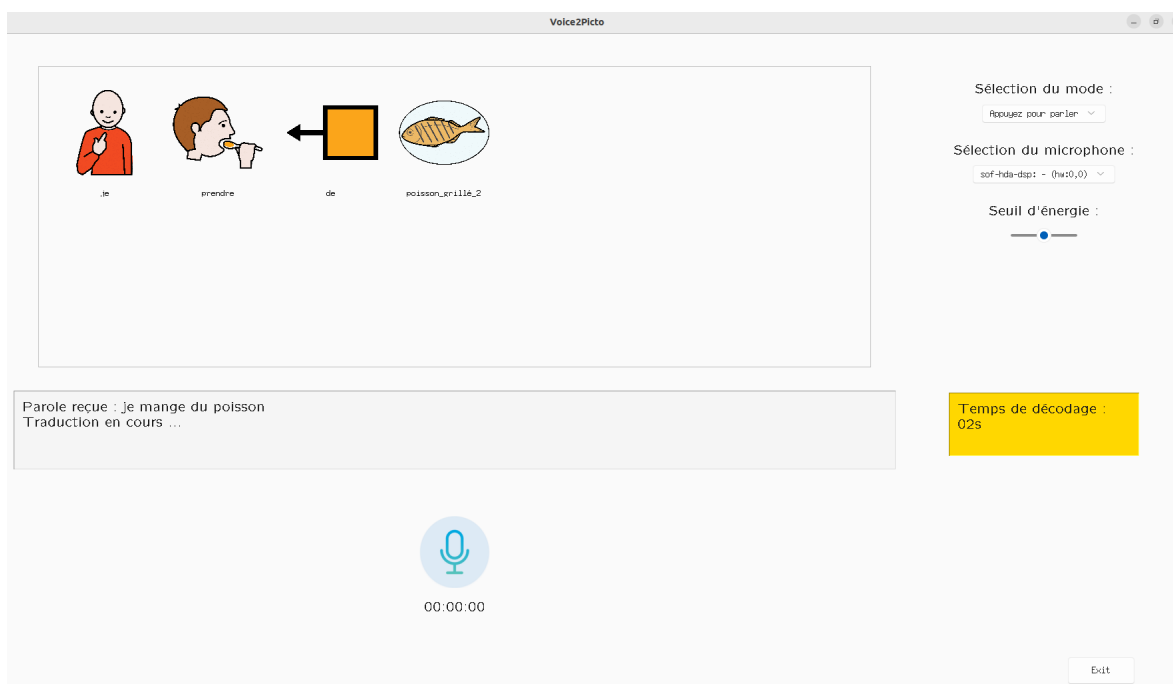


FIGURE 2 – Interface de Voice2Picto.

3 Conclusion & Perspectives

Voice2Picto facilite l'accès à la communication à l'aide de pictogrammes pour un public qui ne connaît pas ou n'utilise pas ce type de support. Les familles et les professionnels de la santé peuvent ainsi disposer d'un outil peu coûteux et facile à mettre en place pour aider les personnes qui utilisent les pictogrammes comme moyen de communication. L'outil étant libre, celui-ci pourra ainsi évoluer grâce

3. <https://spacy.io/models/fr>

aux retours reçus et aux avancées technologiques sur lesquelles il repose. Le code est disponible sur GitHub⁴. Les perspectives futures sont multiples : améliorer la traduction en définissant un vocabulaire précis de pictogrammes à utiliser (actuellement, un terme peut avoir plusieurs pictogrammes associés), gérer les termes non disponibles sous forme pictographique, utiliser une grammaire experte pour orienter les règles de traduction et évaluer la traduction et l’impact de chaque module auprès d’experts.

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). Ce travail a également bénéficié d’un accès aux moyens de calcul de l’IDRIS au travers de l’allocation de ressources 2022-AD011013625 attribuée par GENCI.

Références

- ARDILA R., BRANSON M., DAVIS K., KOHLER M., MEYER J., HENRETTY M., MORAIS R., SAUNDERS L., TYERS F. & WEBER G. (2020). Common voice : A massively-multilingual speech corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, p. 4218–4222, Marseille, France : European Language Resources Association.
- CATAIX-NÈGRE E. (2017). *Communiquer autrement : Accompagner les personnes avec des troubles de la parole ou du langage*. APPRENDRE ET RÉAPPRENDRE. De Boeck Supérieur.
- CROIX-ROUGE (2021). Communiquons autrement - déploiement de la communication alternative améliorée dans les établissements handicap de la croix-rouge française.
- DEVLIN J., CHANG M.-W., LEE K. & TOUTANOVA K. (2019). BERT : Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 1 (Long and Short Papers)*, p. 4171–4186, Minneapolis, Minnesota : Association for Computational Linguistics. DOI : [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423).
- LE H., VIAL L., FREJ J., SEGONNE V., COAVOUX M., LECOUTEUX B., ALLAUZEN A., CRABBÉ B., BESACIER L. & SCHWAB D. (2020). FlauBERT : Unsupervised language model pre-training for French. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, p. 2479–2490, Marseille, France : European Language Resources Association.
- MILLER G. A. (1995). Wordnet : a lexical database for english. *Communications of the ACM*, **38**(11), 39–41.
- POVEY D., GHOSHAL A., BOULIANNE G., BURGET L., GLEMBEK O., GOEL N., HANNEMANN M., MOTLICEK P., QIAN Y., SCHWARZ P. *et al.* (2011). The Kaldi Speech Recognition Toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*, volume CONF : IEEE Signal Processing Society.
- VIAL L., LECOUTEUX B. & SCHWAB D. (2019). Sense Vocabulary Compression through the Semantic Knowledge of WordNet for Neural Word Sense Disambiguation. In *Proceedings of the 10th Global Wordnet Conference*, Wroclaw, Poland.

4. <https://github.com/macairececile/Voice2Picto>

