Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles
Avignon, France, 27 juin au 1er juillet 2022

Une chaîne de traitements pour la simplification automatique de la parole et sa traduction automatique vers des pictogrammes

Cécile Macaire[1,*] Lucía Ormaechea Grijalba[2,*] Adrien Pupier[1,*]
(1) GETALP – Laboratoire d'Informatique de Grenoble – France
(2) Département de Traitement Informatique Multilingue – Université de Genève – Suisse
(1) prenom.nom@univ-grenoble-alpes.fr
(2) Lucia.OrmaecheaGrijalba@unige.ch

RÉSUMÉ

La Communication Alternative et Augmentée (CAA) prend une place importante chez les personnes en situation de handicap ainsi que leurs proches à cause de la difficulté de son utilisation. Pour réduire ce poids, l'utilisation d'outils de traduction de la parole en pictogrammes est pertinente. De plus, ils peuvent être d'une grande aide pour l'accessibilité communicative dans le milieu hospitalier. Dans cet article, nous présentons un projet de recherche visant à développer un système de traduction de la parole vers des pictogrammes. Il met en jeu une chaîne de traitement comportant plusieurs axes relevant du traitement automatique des langues et de la parole, tels que la reconnaissance automatique de la parole, l'analyse syntaxique, la simplification de texte et la traduction automatique vers les pictogrammes. Nous présentons les difficultés liées à chacun de ces axes ainsi que, pour certains, les pistes de résolution.

ABSTRACT

Simplification and automatic translation of speech into pictograms

Alternative and Augmentative Communication (AAC) is becoming an important issue among people with disabilities and their relatives because of the difficulty of its use. To reduce this burden, using speech translation tools in pictograms is relevant. In addition, they can be of great help for communicative accessibility in the hospital environment. Developing such tools requires in-depth research on several axes of automatic language processing. In this article, we present a research project aiming at developing a system for translating speech into pictograms. It involves a processing chain with several axes related to automatic language and speech processing, such as automatic speech recognition, syntactic analysis, sentence simplification, and automatic translation to pictogram units. We present the difficulties related to each of these axes as well as, for some, the avenues of resolution.

MOTS-CLÉS : Communication Alternative et Augmentée, Pictogramme, Analyse syntaxique, Simplification de phrase, Parole, Bout-en-bout.

KEYWORDS: Augmentative and Alternative Communication, Pictogram, Syntactic analysis, sentence simplification, Speech, End-to-end.

∗. Contribution égale.

Volume 2 : 24e Rencontres Etudiants Chercheurs en Informatique pour le TAL (RECITAL), pages 111–123.

1 Introduction

Le projet PROPICTO (*PRojection du langage Oral vers des unités PICTOgraphiques*) vise à développer un axe de recherche autour de la Communication Alternative et Augmentée, en se focalisant sur la transcription automatique de la parole française sous forme pictographique. Il répond ainsi à de nombreux besoins sociétaux dans les domaines du médical (communiquer avec des patients qui n'ont pas la même langue que le praticien) et handicap (communiquer avec des personnes ayant des problèmes cognitifs). Il répond également aux exigences légales adoptées en France relatives au handicap et à l'inclusion (loi du 2 janvier 2002, renforcée par la loi du 11 février 2005). Le projet PROPICTO relève de nombreux défis de recherche autour du TALN. Sa finalité est de proposer des méthodes et des corpus permettant de transcrire directement la parole vers une suite de pictogrammes libres (ARASAAC) ou spécialement créés pour les besoins (médical, familial, etc.). Le projet devra faire face à deux problèmes : la faible quantité de données qui est un frein à la mise en œuvre des techniques état de l'art à base d'apprentissage automatique et la nécessité d'évaluer les méthodes avec des groupes cibles diversifiés. Le projet PROPICTO mettra à disposition de la communauté scientifique l'ensemble des ressources développées : corpus audio associé à sa traduction en pictogrammes, base de données liant les pictogrammes et leur signification sémantique, systèmes de Reconnaissance Automatique de la Parole (RAP), système de simplification pour le FALC, système de Traduction Automatique (TA) parole/pictogramme et métriques d'évaluation humaine ou automatique de la traduction parole/pictogramme. Des prototypes destinés à des publics cibles différents, seront proposés :

— en institution auprès d'enfants et d'adultes polyhandicapés et dans le cadre familial/quotidien auprès de volontaires, par exemple issus, de l'Association Française du Syndrome de Rett (AFSR),
— dans le cadre de la communication médicale, et plus notamment, pour les urgences aux Hôpitaux Universitaires de Genève.

Nous présentons dans cet article les principaux axes actuellement développés autour du projet PROPICTO : la Section 2 présente la problématique de la traduction automatique de la parole vers des pictogrammes. Nous présentons ensuite dans la Section 3 son application au milieu médical, puis nous détaillons un élément essentiel à ce projet dans la Section 4 : l'analyse syntaxique de la parole spontanée. Nous tenons à préciser le caractère préliminaire / prospectif de ces axes de recherche.

2 Traduction automatique de la parole vers des pictogrammes

L'objectif principal du projet dans lequel s'inscrit cet axe de recherche est de permettre aux personnes en situation de handicap de communiquer avec leur environnement via une méthode de Communication Alternative et Augmentée (CAA), ici les pictogrammes. La CAA est une approche utilisant des signes, tableaux de communication avec des symboles et des dispositifs informatiques pour permettre à une personne de transcrire de façon précise son message (Romski & Sevcik, 2005). Il a été notamment démontré que l'utilisation de pictogrammes comme outil d'aide à la communication permet de visualiser la syntaxe, de manipuler des mots et ainsi faciliter l'accès à la langue (Cataix-Nègre, 2017). Mais la prise en main d'un outil de CAA est longue et difficile, et l'utiliser efficacement dans une conversation peut être fastidieux (temps d'adaptation, apprentissage de son fonctionnement) (Cataix-Nègre, 2017).

Dans ce premier axe de recherche, la Traduction Automatique (TA) de la parole vers des pictogrammes, l'objectif est de permettre aux aidants de transformer un message audio en une séquence de pictogrammes. Les aidants englobent le personnel médical (se référer à la Section 3), les familles, et toute personne échangeant avec des individus afin de faire comprendre un message clair et précis aux utilisateurs de CAA. Un premier outil de génération de pictogrammes à partir de la parole spontanée a été proposé dans (Vaschalde *et al.*, 2018). Cette première étude est le point de départ de ce travail de recherche. Nous pouvons découper cet axe en trois parties, formant un système général présenté dans la Figure 1 ci-dessous.

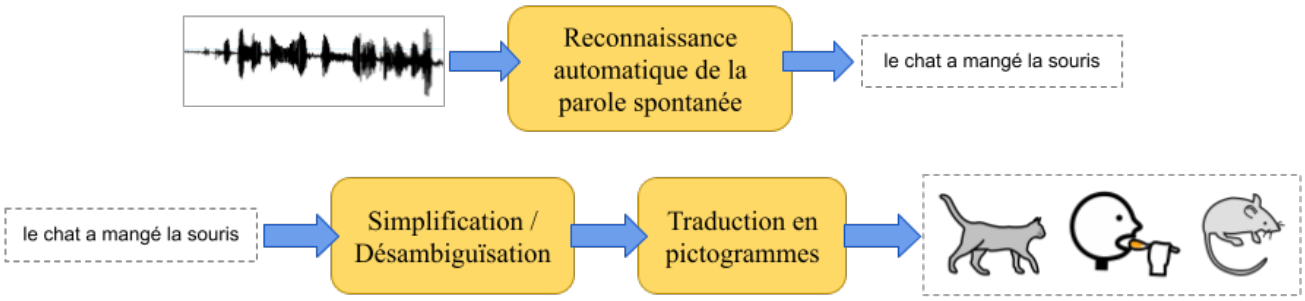

FIGURE 1 – Schéma général du système proposé pour la traduction automatique de la parole vers des pictogrammes.

2.1 La reconnaissance automatique de la parole spontanée

Le premier module (cf. Figure 1) consiste en un système de reconnaissance de la parole spontanée. À partir d'un segment audio enregistré par un aidant, le module génère la transcription associée. La parole spontanée représente plusieurs défis : une qualité sonore peut être faible en entrée (dépendante des outils utilisés pour capturer la voix), différents locuteurs peuvent être entendus dans le même segment audio (situation de chevauchement), des silences et autres hésitations sont généralement observables, mais ne sont pas porteurs de sens. Ce module utilise des modèles de reconnaissance de la parole auto-supervisés, tel Wav2Vec2.0 (Baevski *et al.*, 2020). L'apprentissage auto-supervisé apprend des représentations générales à partir d'une quantité importante de données non étiquetées, cette phase étant appelée le pré-entraînement (*en : pre-training*). Ces représentations sont ensuite utilisées pour répondre à un problème précis, ici la reconnaissance de la parole spontanée, par l'utilisation de données étiquetées via une phase de réglage fin (*en : fine-tuning*). Ces données étiquetées (paires <audio,transcription>) sont récupérées auprès de corpus oraux accessibles gratuitement (ORFEO (Benzitoun *et al.*, 2016) [1]) et en situation écologique, c'est-à-dire auprès des aidants et institutions médicales impliqués dans le projet.

2.2 La simplification et la désambiguïsation lexicale

Le deuxième module du système met en jeu deux traitements, comme illustré dans la Figure 1. Le premier consiste en la simplification du texte donné en entrée en `Facile A Lire et à`

1. https://repository.ortolang.fr/api/content/cefc-orfeo/11/documentation/site-orfeo/index.html

Comprendre (FALC). La simplification en FALC a pour but de transcrire un texte d'un langage classique vers un langage compréhensible par tous, afin de rendre le message plus accessible sans pour autant en perdre le sens original. La simplification consiste à réduire :

(1) la complexité du vocabulaire utilisé (niveau lexique), c'est-à-dire utiliser des mots d'usage courant,

(2) la structure de la phrase sans en altérer le sens original (niveau syntaxe), à savoir transformer la voix passive en voix active, découper le texte initial en plusieurs phrases courtes.

Cette tâche prend en entrée la transcription proposée par le système de reconnaissance automatique de la parole spontanée.

Le deuxième traitement porte sur la désambiguïsation lexicale (DL). Dans notre vie quotidienne, nous communiquons par des mots, certains porteurs de plusieurs sens (polysémiques). Cette ambiguïté observée doit être résolue pour véhiculer un message précis et pertinent, et ainsi créer un système fiable de traduction en pictogrammes. La tâche de DL permet d'assigner le sens le plus probable en s'appuyant sur le contexte de la phrase, par l'utilisation d'un inventaire de sens prédéfini (Vial, 2020).

Dans l'exemple de la Figure 1 "le chat a mangé la souris", la souris a deux sens possibles : le premier correspondant à un petit rongeur et le deuxième au dispositif électronique utilisé sur ordinateur. Dans ce contexte précis, le système de DL propose le sens premier du mot.

2.3 De la parole vers les pictogrammes

La sortie simplifiée et désambiguïsée sera le point d'entrée au dernier module de ce système, la génération automatique de pictogrammes. Plusieurs jeux de données en pictogrammes existent, les principaux étant Makaton [2] et Arasaac [3]. Dans un premier temps, nous nous focalisons sur la génération de pictogrammes à partir d'une base de données Arasaac annotée en sens WordNet (Fellbaum, 2010), comme précédemment proposé dans (Schwab *et al.*, 2020).

En reprenant l'exemple précédent :

Référence : le chat a mangé la souris
Sens : chat_1 | manger_1 | souris_2
Pictogrammes : chat manger souris

Ce premier système proposé est une approche modulaire. L'un des défis finaux sera de mettre en place un système bout-en-bout, qui, à partir de la parole, proposera une traduction automatique en pictogrammes.

3 Vers une adaptation de la génération automatique de pictogrammes au domaine médical

Dans les services de santé actuels, et plus particulièrement dans les contextes d'urgence, une communication adéquate et effective entre médecins et patients s'avère cruciale, dans la mesure où elle favorise

2. https://makaton.org/
3. https://arasaac.org/

un diagnostic correct et permet aussi une meilleure adhésion du patient au traitement (Katz *et al.*, 2006). Dans le domaine médical, il s'avère particulièrement important de supprimer les éventuelles barrières à la communication qui peuvent résulter, par exemple, d'un manque de compréhension de la langue parlée par le praticien, d'un faible niveau de littératie en santé (Le Deuff, 2015) ou d'un handicap. Les difficultés qui risquent de découler de ces situations peuvent compromettre la qualité des soins (Eadie *et al.*, 2013) et, par-dessus tout, la santé et la sécurité de la personne traitée (Ku & Flores, 2005). Elles montrent la nécessité de développer des dispositifs de communication appropriés à ces contextes particuliers.

Nous proposons donc d'explorer les solutions issues de la TA qui peuvent être envisagées pour améliorer les échanges communicatifs entre un médecin et un patient lorsqu'une interaction directe n'est pas disponible, en raison de la présence d'une barrière linguistique.

Étant donnée la sensibilité du domaine d'échange médical, où une erreur de traduction est inacceptable, la conception d'un tel système devra être volontairement contraint pour des raisons de sécurité. C'est pourquoi nous proposons l'utilisation d'un système de projection de parole vers des pictogrammes qui repose sur un ensemble de grammaires spécialisées, nous permettant de mieux contrôler les sorties générées. Nous estimons que la conception d'un tel outil pourrait se révéler utile en tant qu'instrument de CAA pour permettre aux personnes allophones d'échanger plus facilement avec le spécialiste traitant lors de l'anamnèse, c'est-à-dire, le recueil d'informations concernant les symptômes et les antécédents médicaux.

3.1 Travaux précédents

Comme évoqué précédemment, le pictogramme est un signe graphique schématique très souvent utilisé dans le domaine de la CAA. L'iconicité fournie par ces symboles, souvent reliée à une plus grande facilité d'interprétation, explique leur présence de plus en plus répandue dans le domaine médical. Partant de l'hypothèse que les signes pictographiques améliorent la compréhension du message véhiculé par le médecin, ils sont souvent utilisés pour rendre les traitements et les instructions médicales plus compréhensibles (Houts *et al.*, 2006; Bandeira *et al.*, 2011; Wołk *et al.*, 2017). Les résultats montrent que l'utilisation de ce type d'outils est avantageuse. Pourtant, la mise en place des systèmes de génération automatique pictographique à partir de la parole est une piste encore très peu explorée pour un cadre d'application tel que le dialogue docteur-patient.

À titre d'exemple, il faut noter le système `Glyph`, développé à l'University of Utah, qui a proposé un outil de conversion automatique du texte en séquences pictographiques (Hill *et al.*, 2016). Visant à améliorer le rappel et la compréhension des consignes de sortie chez le patient (Bui *et al.*, 2012), il s'appuie sur un système basé sur des règles sémantiques, qui rend possible la génération d'une chaîne de pictogrammes suivant le texte donné en entrée.

Bien que `Glyph` constitue en effet un bon préalable aux objectifs que nous poursuivons, il n'est pas disponible pour la langue que nous traitons et manque en outre d'un module de reconnaissance vocale, permettant une interaction plus rapide. Dans le domaine médical francophone, nous constatons d'ailleurs qu'il n'existe actuellement aucun outil spécialisé de TA parole-pictogrammes qui facilite les dialogues entre les médecins et les patients étrangers. Nous estimons donc que les recherches orientées vers ce but sont particulièrement prometteuses, d'autant plus que ce type de TA améliorerait la maniabilité par rapport aux systèmes pictographiques existants, sujets souvent à un usage chronophage.

3.2 Méthodologie

Pour répondre aux objectifs fixés, tout en tenant compte des particularités d'un domaine spécialisé comme le médical, nous privilégierons tout d'abord une approche suivant une structure en cascade, de manière à mieux contrôler les sorties de chaque module.

3.2.1 La reconnaissance automatique de la parole spécialisée

La première étape vers un tel outil de TA consistera à développer un système robuste de RAP adapté au français médical. Il s'avère en effet d'un défi complexe : son adaptation à une langue de spécialité tel que le discours médical est problématique, surtout dans un contexte aussi sensible que l'anamnèse, où une erreur dans la reconnaissance et, de ce fait, dans la traduction, est susceptible d'entraîner des répercussions indésirables sur la personne traitée (Hacker *et al.*, 2015). Afin de réduire les risques de propagation liés à cette phase, nous proposons d'exploiter les méthodes hybrides développées dans le cadre du projet BabelDr (Spechbach *et al.*, 2019). Sur la base d'une grammaire synchronisée, qui permet l'injection des connaissances linguistiques expertes dans les modèles de langue, il est possible de faire le lien entre les variantes orales possibles et la forme simplifiée traduite. De cette manière, on peut donc générer des corpus parallèles, ensuite utilisés pour entraîner le système de RAP et faire la correspondance entre les variantes orales et la forme canonique traduite et oralisée pour le patient (Mutal *et al.*, 2019).

3.2.2 La génération automatique des pictogrammes en contexte médical

Dans un deuxième temps, nous travaillerons sur la conversion de l'oral vers des suites de pictogrammes issues des bases pictographiques *open source*. Il s'agit d'un défi complexe en termes de compréhensibilité et de disponibilité, et se reflète à deux niveaux :

— En vue d'obtenir une traduction pictographique lisible, il est nécessaire d'avoir recours à une simplification linguistique, comme indiqué dans la Section 2.2. Les traductions trop complexes sur l'axe syntagmatique, calquées sur les structures syntaxiques de la langue source risquent de ne pas contribuer à une meilleure compréhension du message.

— Par ailleurs, une traduction de qualité dépend également de l'adéquation et de la couverture lexico-sémantique des pictogrammes au contexte communicationnel prévu (à savoir, celui du dialogue médecin-patient) ainsi que de la facilité d'interprétation des signes représentés. Certes, les sets de pictogrammes actuellement disponibles abondent. Cependant, les banques destinées au contexte médico-sanitaire sont peu nombreuses et ne font pas toujours l'objet de licences permissives.

3.2.3 L'évaluation des systèmes en termes de compréhensibilité et interprétabilité

En dernier lieu, nous nous attarderons sur la compréhensibilité et l'interprétabilité de notre système de TA parole-pictogrammes dans la situation écologique prévue. Les systèmes de signes iconiques sont souvent destinés à promouvoir une meilleure accessibilité à l'information. À cette fin, une conception particulièrement soigneuse est requise, puisque l'identification d'un pictogramme et l'assimilation de son signifié subjacent n'est pas nécessairement univoque et par conséquent non universelle. A contrario, l'interprétation et les inférences tirées par le lecteur jouent un rôle important

dans le processus d'appréhension, qui découlent principalement de son bagage cognitif et de la vision du monde de la communauté culturelle à laquelle il appartient (Vaillant & Bordon, 2001). Pour ces raisons, nous explorerons les différentes façons d'agencer les traductions pictographiques et d'optimiser l'efficacité communicative de notre système pour les patients allophones. Dans cette perspective, plusieurs visualisations seront proposées et évaluées avec des utilisateurs en conditions réelles d'exploitation ainsi qu'à l'aide des techniques de *crowdsourcing*.

3.3 Résultats prévus

Avec la mise en œuvre de l'ensemble des étapes décrites ci-dessus, nous comptons mettre à disposition au terme de cette recherche un système de TA spécialisé, permettant de projeter la parole vers des unités pictographiques et visant à être utilisé dans des contextes d'échange médical réels. Nous souhaitons pouvoir contribuer, d'un point de vue scientifique, à améliorer et enrichir l'état de l'art qui rattache le domaine du TALN à celui de la CAA.

L'implémentation d'un tel système dans un environnement hospitalier reste une piste très peu explorée malgré son énorme potentiel du point de vue sociétal. Pour un public allophone, nous estimons que l'outil de traduction proposé favoriserait une attention médicale plus efficace et un meilleur accès à l'information lorsqu'aucune langue n'est partagée entre le service médical et la personne soignée, tout en éliminant autant que possible les éventuelles barrières linguistiques dans le cadre d'une anamnèse.

4 Analyse syntaxique du français parlé

Un des axes de recherche important de ce projet est l'analyse syntaxique de l'oral en français. Il existe deux façons classiques pour la réaliser : l'analyse en constituants et l'analyse en dépendances. Dans cette section, nous allons parler plus particulièrement de l'analyse en dépendances. Cette dernière consiste à prédire un arbre pour lequel chaque mot est relié à son gouverneur par une relation typée (sujet, objet, ...).

Ce cadre d'analyse est très étudié sur le texte écrit, notamment grâce à "Universal dependencies" (Nivre *et al.*, 2016) ainsi que la conférence CoNLL et les nombreuses campagnes d'évaluation (*shared task*) qu'elle a proposé par le passé. Cependant, les recherches sur l'analyse syntaxique de la parole sont généralement réalisées sur des transcriptions avec l'ajout de la modalité audio (Pate & Goldwater, 2013). Pour ce projet, nous voulons réaliser cette analyse directement à partir de l'oral sans l'utilisation de transcriptions intermédiaires. Utiliser des transcriptions pose certains problèmes, particulièrement à cause de la propagation d'erreur qu'elle entraîne. Si le module de reconnaissance automatique de la parole fait une erreur, elle va se propager dans le reste du système via les transcriptions. Faire l'analyse syntaxique directement via l'audio permet d'y pallier. De plus, l'accès à la modalité audio permet à l'analyseur syntaxique d'utiliser des informations (telles que la prosodie) qui ne sont pas accessibles via une transcription. Nous faisons l'hypothèse que ces informations sont cruciales pour un analyseur du français parlé.

Le fait de travailler directement avec la modalité audio entraîne des difficultés inédites par rapport à l'écrit. Nous devons effectuer cette analyse de manière jointe avec la RAP. Cela implique donc que nous devons soit utiliser une approche en cascade (*pipeline*) où l'on effectue d'abord la RAP puis l'analyse syntaxique. Ou bien, on peut créer un système de bout-en-bout (*end-to-end*) pour exploiter

les informations présentes dans le signal audio tel que la prosodie ou la présence de pause. Dans ce cadre, l'approche en cascade est notre étalon et l'approche bout-en-bout est celle que nous étudions. Dans la suite de cette section, nous présentons une architecture préliminaire bout-en-bout ainsi que les nouveaux problèmes que cela pose. L'étude étant encore en cours de réalisation, les résultats ne seront pas présentés dans cet article.

4.1 Place de l'analyse syntaxique

Pour obtenir une traduction en pictogramme adapté à la communication alternative augmenté, il est nécessaire d'effectuer une simplification des phrases. En effet, les utilisateurs de CAA mettent en garde contre des traductions trop complexes en pictogrammes issue de la structure syntaxique du langage source. Extraire une représentation syntaxique de l'oral permettra de développer des systèmes de simplification efficace (Brouwers *et al.*, 2014). Ainsi, le système de simplification sera entraîné sur des doublons <phrase, arbre> avec comme objectif de créer un doublon <phrase simplifiée, arbre simplifié>.

De plus, on peut supposer que l'ajout de cette tâche dans un modèle multi-tâche (Caruana, 1997) permettrait d'inciter le modèle à encoder l'information syntaxique de manière plus prononcée dans ses représentations, ce qui pourrait être intéressant pour la traduction vers des pictogrammes.

L'analyse syntaxique occupe donc une place importante dans le cadre de ce projet. Sans cette analyse et au processus de simplification auquel elle contribue, la traduction en pictogrammes risque d'être trop complexe et en conséquence inadaptée à la CAA.

4.2 Les défis du bout-en-bout

Mettre en place un système bout-en-bout pour cette tâche met en lumière certaines différences avec l'écrit. En effet, il est nécessaire d'identifier les parties de l'audio correspondant à un mot spécifique. La segmentation du signal n'est pas explicite contrairement à l'écrit. De plus, utiliser la segmentation de référence du corpus d'entraînement n'est pas souhaitable si on veut avoir un système fonctionnel en situation réelle. Il nous faut donc créer un système capable de trouver une segmentation pertinente du signal pour pouvoir effectuer l'analyse. Une fois ce problème résolu, un second apparaît. Si la segmentation vient du système, elle peut être différente de celle de la supervision. Dans ce cas, il est nécessaire d'adapter la supervision à la segmentation obtenue pour apprendre de manière efficace (Yoshikawa *et al.*, 2016).

Le modèle que nous proposons est basé sur `wav2vec2` (Baevski *et al.*, 2020), plus particulièrement sur la version pré-entrainée dans le projet `LeBenchmark` (Evain *et al.*, 2021). L'architecture est réalisée via la bibliothèque Python `speechbrain` (Ravanelli *et al.*, 2021). Les données que nous utiliserons sont celles de `CEFC-ORFEO` (Benzitoun *et al.*, 2016) annotées en dépendances. L'architecture est décrite dans la figure 2.

4.3 Obtenir la segmentation

Pour extraire la segmentation de l'audio, nous profitons de la connaissance du module de reconnaissance automatique de la parole. En effet, ce module utilise l'algorithme CTC (Graves *et al.*, 2006)

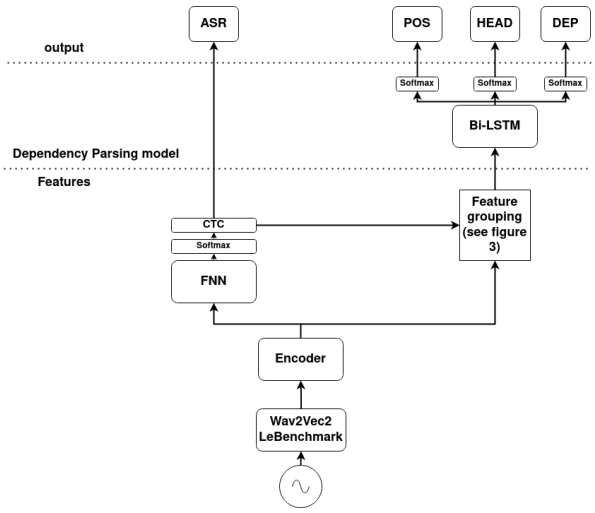

FIGURE 2 – Architecture bout en bout pour l'analyse syntaxique automatique en dépendances à partir du signal.

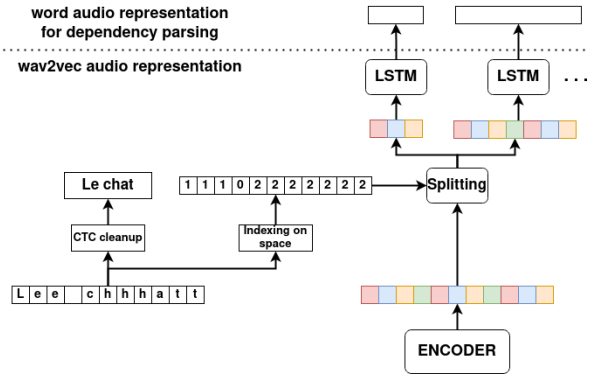

FIGURE 3 – Création des plongements des mots à partir des représentations apprises du modèle et de la segmentation de l'algorithme CTC.

pour créer une sortie caractère par caractère. Pour chaque représentation de l'audio, un caractère est prédit. Ainsi, nous utilisons cette propriété pour segmenter sur les espaces et nous partons du postulat que toutes les représentations audio entre deux espaces (non-vide) appartiennent au même mot. Ces représentations sont ensuite combinées via un LSTM et nous obtenons un plongement pour chaque mot reconnu par le module de RAP. La figure 3 décrit ce processus.

4.4 Adapter la supervision : Oracle

Dans le cas où la segmentation prédite différerait de la vraie segmentation, il y a deux possibilités. Soit des mots sont manquants, soit des mots ont été ajoutés. Dans les deux cas, pour pouvoir créer le meilleur arbre à partir de la segmentation courante, il faut savoir quel mot sont présents dans la phrase reconnue. Pour cela, nous utilisons l'outil "Sclite" de la boite à outil NIST SCTK [4]. Grâce à cette information et à l'arbre de référence, nous calculons le meilleur arbre possible en termes de score d'attachement étiqueté *Labelled Attachment Score, LAS* pour conserver une supervision de bonne qualité.

4.5 Perspectives de recherche future

L'objectif final de ce travail de recherche est de créer un système d'analyse syntaxique pour la parole. La mise au point de ce système soulève de nombreuses questions que nous explorerons dans le futur. Notamment quant à l'architecture du modèle bout-en-bout et les modalités que nous pouvons exploiter.

L'architecture présentée dans cette section dispose d'un décodeur rudimentaire. Le passage à une architecture séquence vers séquence avec un décodeur plus complexe pourrait avoir un impact positif sur l'analyse syntaxique. Une question pertinente est l'ajout de la modalité écrite. En effet, bien

4. http ://www.nist.gov/speech/tools/index.htm

que nous nous concentrons sur l'audio pour le moment, il n'est pas impensable de combiner les plongements de mots audio avec des plongements de mots plus classique venant de FlauBert (Le *et al.*, 2019) ou camemBert (Martin *et al.*, 2019) par exemple. Cela résulterait en un modèle usant à la fois de l'information audio (prosodie, pause, hésitation…) et des plongements de mots entrainés sur une grande quantité de données.

Un autre point intéressant est d'étudier la présence de l'information syntaxique dans les plongements audio via des sondes (Veldhoen *et al.*, 2016; Tenney *et al.*, 2019). Une comparaison avec les plongements de l'écrit serait pertinent dans ce cadre.

5 Conclusions

Transcrire et simplifier la parole, que ce soit spontanée ou spécialisée, puis la traduire en unités pictographiques n'est pas une tâche simple. Les modèles pour ce type de tâche sont composés de multiples modules faisant appel à de nombreux axes de recherche. Ainsi, la reconnaissance automatique de la parole rejoint la simplification du texte, elle-même nécessitant l'information syntaxique pour opérer. Ce texte simplifié pourra être ensuite traduit en une séquence cohérente de pictogrammes.

Au-delà de l'aspect purement scientifique de cette tache, les trois axes de recherche présentés ont pour but d'améliorer la vie quotidienne des personnes en leur proposant un outil fiable de communication, que cela soit pour un public en situation de handicap langagier, ou pour un public allophone en milieu hospitalier.

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). Nous souhaitons remercier nos encadrants Pierrette Bouillon, Maximin Coavoux, Emmanuelle Esperança-Rodier, Benjamin Lecouteux, et Didier Schwab.

Références

BAEVSKI A., ZHOU Y., MOHAMED A. & AULI M. (2020). wav2vec 2.0 : A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, **33**, 12449–12460.

BANDEIRA F., FARIA F. & ARAUJO E. (2011). Quality assessment of inhospital patients unable to speak who use alternative and extended communication. *Einstein (São Paulo)*, **9**, 477–482. DOI : 10.1590/S1679-45082011AO2083.

BENAMARA F., HATOUT N., MULLER P. & OZDOWSKA S., Éds. (2007). *Actes de TALN 2007 (Traitement automatique des langues naturelles)*, Toulouse. ATALA, IRIT.

BENZITOUN C., DEBAISIEUX J.-M. & DEULOFEU H.-J. (2016). Le projet orféo : un corpus d'étude pour le français contemporain. *Corpus*, (15).

BROUWERS L., BERNHARD D., LIGOZAT A.-L. & FRANÇOIS T. (2014). Syntactic sentence simplification for french. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)@ EACL 2014*, p. 47–56.

BUI D., NAKAMURA C., BRAY B. & ZENG-TREITLER Q. (2012). Automated illustration of patients instructions. *AMIA*.

CARUANA R. (1997). Multitask learning. *Machine learning*, **28**(1), 41–75.

CATAIX-NÈGRE E. (2017). *Communiquer autrement : Accompagner les personnes avec des troubles de la parole ou du langage*. APPRENDRE ET RÉAPPRENDRE. De Boeck Supérieur.

DIAS G., Éd. (2015). *Actes de TALN 2015 (Traitement automatique des langues naturelles)*, Caen. ATALA, HULTECH.

EADIE K., CARLYON M. J., STEPHENS J. & WILSON M. D. (2013). Communicating in the pre-hospital emergency environment. *Aust Health Rev*, **37**(2), 140–146. DOI : 10.1071/AH12155.

EVAIN S., NGUYEN H., LE H., BOITO M. Z., MDHAFFAR S., ALISAMIR S., TONG Z., TOMASHENKO N., DINARELLI M., PARCOLLET T. *et al.* (2021). Lebenchmark : A reproducible framework for assessing self-supervised representation learning from speech. *arXiv preprint arXiv :2104.11462*.

FELLBAUM C. (2010). Wordnet. In *Theory and applications of ontology : computer applications*, p. 231–243. Springer.

GRAVES A., FERNÁNDEZ S., GOMEZ F. & SCHMIDHUBER J. (2006). Connectionist temporal classification : labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning*, p. 369–376.

HACKER K., ANIES M., FOLB B. L. & ZALLMAN L. (2015). Barriers to health care for undocumented immigrants : a literature review. *Risk Manag Healthc Policy*, **8**, 175–183. DOI : 10.2147/RMHP.S70173.

HILL B., PERRI-MOORE S., KUANG J., BRAY B. E., NGO L., DOIG A. & ZENG-TREITLER Q. (2016). Automated pictographic illustration of discharge instructions with glyph : impact on patient recall and satisfaction. *J Am Med Inform Assoc*, **23**(6), 1136–1142. DOI : 10.1093/jamia/ocw019.

HOUTS P. S., DOAK C. C., DOAK L. G. & LOSCALZO M. J. (2006). The role of pictures in improving health communication : A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling*, **61**(2), 173–190. DOI : 10.1016/j.pec.2005.05.004.

KATZ M. G., KRIPALANI S. & WEISS B. D. (2006). Use of pictorial aids in medication instructions : a review of the literature. *Am J Health Syst Pharm*, **63**(23), 2391–2397. DOI : 10.2146/ajhp060162.

KU L. & FLORES G. (2005). Pay now or pay later : providing interpreter services in health care. *Health Aff (Millwood)*, **24**(2), 435–444. DOI : 10.1377/hlthaff.24.2.435.

LAIGNELET M. & RIOULT F. (2009). Repérer automatiquement les segments obsolescents à l'aide d'indices sémantiques et discursifs. In A. NAZARENKO & T. POIBEAU, Éds., *Actes de TALN 2009 (Traitement automatique des langues naturelles)*, Senlis : ATALA LIPN.

LANGLAIS P. & PATRY A. (2007). Enrichissement d'un lexique bilingue par analogie. In (Benamara *et al.*, 2007), p. 101–110.

LE H., VIAL L., FREJ J., SEGONNE V., COAVOUX M., LECOUTEUX B., ALLAUZEN A., CRABBÉ B., BESACIER L. & SCHWAB D. (2019). Flaubert : Unsupervised language model pre-training for french. *arXiv preprint arXiv :1912.05372*.

LE DEUFF O. (2015). La littératie digitale de santé : un domaine en émergence. In *Les écosystèmes numériques et la démocratisation informationnelle : Intelligence collective, Développement durable, Interculturalité, Transfert de connaissances*, Schoelcher, France. HAL : hal-01258315.

MARTIN L., MULLER B., SUÁREZ P. J. O., DUPONT Y., ROMARY L., DE LA CLERGERIE É. V., SEDDAH D. & SAGOT B. (2019). Camembert : a tasty french language model. *arXiv preprint arXiv :1911.03894*.

MUTAL J. D., BOUILLON P., GERLACH J., ESTRELLA P. & SPECHBACH H. (2019). Monolingual backtranslation in a medical speech translation system for diagnostic interviews - a NMT approach. *Proceedings of Machine Translation Summit XVII Volume 2 : Translator, Project and User Tracks*, p. 169.

NIVRE J., DE MARNEFFE M.-C., GINTER F., GOLDBERG Y., HAJIC J., MANNING C. D., MCDONALD R., PETROV S., PYYSALO S., SILVEIRA N. *et al.* (2016). Universal dependencies v1 : A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, p. 1659–1666.

PATE J. K. & GOLDWATER S. (2013). Unsupervised dependency parsing with acoustic cues. *Transactions of the Association for Computational Linguistics*, **1**, 63–74.

RAVANELLI M., PARCOLLET T., PLANTINGA P., ROUHE A., CORNELL S., LUGOSCH L., SUBAKAN C., DAWALATABAD N., HEBA A., ZHONG J., CHOU J.-C., YEH S.-L., FU S.-W., LIAO C.-F., RASTORGUEVA E., GRONDIN F., ARIS W., NA H., GAO Y., MORI R. D. & BENGIO Y. (2021). SpeechBrain : A general-purpose speech toolkit. arXiv :2106.04624.

ROMSKI M. & SEVCIK R. A. (2005). Augmentative communication and early intervention : Myths and realities. *Infants & Young Children*, **18**(3), 174–185.

SCHWAB D., TRIAL P., VASCHALDE C., VIAL L., ESPERANÇA-RODIER E. & LECOUTEUX B. (2020). Providing semantic knowledge to a set of pictograms for people with disabilities : a set of links between wordnet and arasaac : Arasaac-wn. In *LREC*.

SERETAN V. & WEHRLI E. (2007). Collocation translation based on sentence alignment and parsing. In (Benamara *et al.*, 2007), p. 401–410.

SPECHBACH H., GERLACH J., MAZOURI KARKER S., TSOURAKIS N., COMBESCURE C. & BOUILLON P. (2019). A speech-enabled fixed-phrase translator for emergency settings : Crossover study. *JMIR Med Inform*, **7**(2), e13167. DOI : 10.2196/13167.

TENNEY I., XIA P., CHEN B., WANG A., POLIAK A., MCCOY R. T., KIM N., VAN DURME B., BOWMAN S. R., DAS D. *et al.* (2019). What do you learn from context ? probing for sentence structure in contextualized word representations. *arXiv preprint arXiv :1905.06316*.

VAILLANT P. & BORDON E. (2001). Le statut du signe iconique entre iconicité et intertextualité. *VISIO, revue de l'Association Internationale de Sémiotique Visuelle*, **6**(4), 57–74.

VASCHALDE C., LECOUTEUX B. & SCHWAB D. (2018). Génération de pictogrammes à partir de la parole spontanée pour la mise en place d'une communication médiée. In *50 ans de linguistique sur corpus oraux : Apports à l'étude de la variation*, Orléans, France. HAL : hal-01876781.

VELDHOEN S., HUPKES D. & ZUIDEMA W. (2016). Diagnostic classifiers : Revealing how neural networks process hierarchical structure. In *Pre-Proceedings of the Workshop on Cognitive Computation : Integrating Neural and Symbolic Approaches (CoCo @ NIPS 2016)*.

VIAL L. (2020). *Modèles neuronaux joints de désambiguïsation lexicale et de traduction automatique*. Theses, Université Grenoble Alpes [2020-....]. HAL : tel-03033118.

WOŁK K., WOŁK A. & GLINKOWSKI W. (2017). A cross-lingual mobile medical communication system prototype for foreigners and subjects with speech, hearing, and mental disabilities based on pictograms. *Comput Math Methods Med*, **2017**, 4306416. DOI : 10.1155/2017/4306416.

YOSHIKAWA M., SHINDO H. & MATSUMOTO Y. (2016). Joint transition-based dependency parsing and disfluency detection for automatic speech recognition texts. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, p. 1036–1041.