Génération de pictogrammes à partir de la parole spontanée pour la mise en place d'une communication médiée

Céline Vaschalde, Benjamin Lecouteux, Didier Schwab

▶ **To cite this version:**

Céline Vaschalde, Benjamin Lecouteux, Didier Schwab. Génération de pictogrammes à partir de la parole spontanée pour la mise en place d'une communication médiée. 50 ans de linguistique sur corpus oraux : Apports à l'étude de la variation, Nov 2018, Orléans, France. hal-01876781

HAL Id: hal-01876781
https://hal.science/hal-01876781

Submitted on 18 Sep 2018

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.

Génération de pictogrammes à partir de la parole spontanée pour la mise en place d'une communication médiée

Céline Vaschalde, Benjamin Lecouteux, Didier Schwab

Équipe GETALP, Laboratoire d'Informatique de Grenoble, Université Grenoble Alpes

Dans le cadre d'un travail de recherche sur la **génération de pictogrammes à partir de la parole spontanée**, nous avons choisi de travailler avec le corpus ESLO pour aborder les problèmes rencontrés dans ce type de conditions de reconnaissance vocale (**conditions acoustiques non optimales**, **superpositions de locuteurs, disfluences**…) [Dufour, 2010], et ainsi **améliorer les performances** de nos modèles actuels.

L'objectif de notre recherche est en effet de **développer un outil basé sur KALDI (système de reconnaissance de la parole libre) [**Povey et al., 2011] **permettant une communication médiée** lorsque l'oralisation et la capacité de communiquer ses besoins, ses idées, vitale chez l'être humain, sont mis en difficulté voire impossibles. Ces troubles langagiers peuvent avoir des **causes multiples** (dysarthrie, dysphasie/aphasie, troubles cognitifs et communicationnels...), et peuvent être **renforcés par des troubles moteurs aggravants** qui empêchent l'accès aux langues signées et au langage non verbal. [Cataix-Nègre, 2017]

Plusieurs méthodes de **Communication Alternative et Améliorée** (CAA) existent et permettent à ces personnes de communiquer, et beaucoup ont recours à un encodage multimodal de l'information avec des pictogrammes. Ces pictogrammes permettent une représentation plus **iconique** de l'information et réduisent donc le coût cognitif d'encodage et de décodage. [Duboisdindien, 2014]. Néanmoins, la manipulation des classeurs et tableaux de pictogrammes utilisés par l'entourage des utilisateurs de CAA pour transmettre un message n'est pas aisée (matériel peu maniable, navigation chronophage…) et entraîne une **perte de spontanéité**. Or l'acquisition et l'utilisation de ces méthodes de communication nécessitent un **bain langagier en contexte pour les apprenants**. La production de messages «sur le vif» représente un réel enjeu pour une compréhension optimale. [Beukelman & Mirenda, 2017]

Un outil de génération de pictogrammes à partir de la parole permettra de faciliter l'encodage **interaction quotidienne, mais également** un gain de **sociabilité,**

d’autonomie ainsi que de confort cognitif et communicationnel pour les utilisateurs de CAA.

La méthodologie de création de cet outil s’axe autour du développement de quatre modules. Le premier est un module neuronal de **reconnaissance vocale** [Elloumi & al, (2018)] qui permet d’obtenir une **transcription orthographique** de l’énoncé produit. Le deuxième est un module de simplification automatique à base de règles qui permet, après un **pré-traitement** de la transcription (tokenisation, étiquetage morpho-syntaxique, lemmatisation), un paramétrage de la complexité du niveau de traduction voulue en fonction des capacités cognitives de l’utilisateur de CAA auquel on s’adresse. Il existe **deux niveaux de traduction**, un premier dans lequel seuls les mots pleins sont traduits, excluant les mots grammaticaux dont la représentation en pictogrammes peut être jugée trop symbolique pour nos utilisateurs, et un deuxième niveau dans lequel tous les mots sont traduits. Le module de simplification va également permettre d’implémenter des règles de **simplification syntaxique** inspirées du Facile À Lire et à Comprendre (FALC), un guide de règles pour rendre accessibles les supports d’information au plus large public possible (transformer les structures passives à la voie active pour éviter les confusions concernant l’agent par exemple). Après cette étape de simplification, un module de **désambiguïsation lexicale** [Vial & al, (2018)] va annoter en sens la transcription pour éviter l’affichage d’un pictogramme non pertinent (exemple : afficher le pictogramme « souris » représentant une « souris d’ordinateur » comme traduction pour l’énoncé « Le chat a mangé la souris »). Enfin, un module d’**affichage** va concaténer tous les pictogrammes de la séquence pour afficher le message complet traduit.

Pour faciliter le bain multimodal nécessaire à l’apprentissage, le module de reconnaissance vocale développé doit être particulièrement performant sur la parole spontanée en situation d’interaction quotidienne. De part sa grande taille (800 enregistrements, près de 700h de parole, 758 enregistrements transcrits et validés), la multiplicité des situations d‘interaction explorées (interaction quotidienne comme dans le module « Repas » ou le module « Boulangerie », interaction plus guidée avec les entretiens...) et la diversité des modalités de recueil de données, ESLO est une ressource particulièrement pertinente pour ré-entraîner nos modèles neuronaux. Cela permettra d’améliorer leurs performances de notre outil, de par l’enrichissement considérable de nos corpus d’entraînement via l’ajout de faits de langues propres à l’oral spontané non répertoriés dans les corpus plus normalisés utilisés auparavant (parole médiatique). L’utilisation d’ESLO ajoutera ainsi **de la variabilité dans les modèles acoustiques et les modèles de langues** utilisés jusqu’ici.

De plus, notre étude comportant une **analyse de corpus d’une cinquantaine de textes écrits traduits en pictogrammes, composé** en majorité d’**histoires pour enfants, il nous est apparu particulièrement pertinent de travailler avec le module « Livres pour enfants » de ESLO2** qui regroupe des moments de lecture et d’échange entre des adultes et un ou plusieurs enfants **pour s’appuyer sur des données orales similaires et pour évaluer la performances de notre outil.**

Références bibliographiques

BAUDE, O. & DUGUA, C. (2011). « (Re)faire le corpus d'Orléans quarante ans après: quoi de neuf, linguiste?», *Corpus*, *Varia*, 10 | 2011, 99-118.

BEUKELMAN D. R. & MIRENDA P. (2017). *Communication alternative et améliorée, Aider les enfants et les adultes avec des difficultés de communication*, 1er édition, DE BOECK SUPERIEUR, 400 p, collection Apprendre et Réapprendre

CATAIX-NÈGRE E. (2017). *Communiquer autrement, Accompagner les personnes avec des troubles de la parole ou du langage*, 2ème édition, DE BOECK SUPERIEUR, 336 p, collection Pratiques en rééducation

DUBOISDINDIEN G. (2014). *L'interprétation des pictogrammes. Statut linguistique et limites de l'utilisation des pictogrammes dans la réhabilitation langagière. - Étude de deux groupes d'enfants âgés de 5 à 6 ans – entraînés Versus non entraînés*, mémoire de recherche de Master de Linguistique Générale et Appliquée Spécialité Fonctionnements Linguistiques et Dysfonctionnements langagiers, sous la direction de BOGLIOTTI Caroline, Université Paris-Ouest Nanterre La Défense, 100 p.

DUFOUR R. (2010). *Transcription automatique de la parole spontanée,* Thèse de doctorat en informatique sous la direction de DELÉGLISE P. , Soutenue à l'UFR de sciences exactes et naturelles du Mans

ELLOUMI Z., BESACIER L., GALIBERt O., KAHN J. & LECOUTEUX B. (2018). *ASR Performance Prediction on Unseen Broadcast Programs using Convolutional*

ESHKOL-TARAVELLA I., BAUDE O., MAUREL D., HRIBA L., DUGUA C. & TELLIER I. (2012). *Un grand corpus oral « disponible » : le corpus d'Orléans 1968-2012*., in Ressources linguistiques libres, TAL. Volume 52 – n° 3/2011, 17-46

MONTOYA D. & BODART S. (2009). Le programme Makaton auprès d'un enfant porteur d'autisme : le cas de Julien, *Développements*, 2009/3 (n° 3), p. 15-26. DOI : 10.3917/devel.003.0015. URL : https://www.cairn.info/revue-developpements-2009-3-page-15.htm

PERRIOUX-PERDREAUX M. & SCHMITT A. (2014). *Communication alternative et augmentative auprès d'adultes déficients intellectuels : aménagements de l'environnement et outils de communication adaptés* , mémoire présenté pour l'obtention du Certificat de capacité d'orthophoniste, Université de Lorraine, Faculté de Médecine, Département d'orthophonie

POVEY D., GHOSHAL A., BOULIANNE G., BURGET L., GLEMBEK O., GOEL N., HANNEMANN M., MOTLICEK P., QIAN Y., SCHWARZ P., SILOVSKY J., STEMMER G. & VESELY K., (2011), *The Kaldi speech recognition toolkit,* IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, Hilton Waikoloa Village, Big Island, Hawaii, US

VIAL L., LECOUTEUX B. & SCHWAB D. (2018) Approche supervisée à base de cellules LSTM bidirectionnelles pour la désambiguïsation lexicale. *25e conférence sur le Traitement Automatique des Langues Naturelles*