

Synthèse de la parole pour les langues peu dotées

Application au wallon

L'Europe est riche par sa diversité linguistique, même si celle-ci est masquée par le rôle dominant des langues officielles et menacée par l'interruption de la transmission au sein des cellules familiales. Doter ses langues minorisées d'outils de traitement automatique n'est peut-être pas suffisant à leur revitalisation, mais est aujourd'hui indispensable à leur revalorisation. Parmi ces langues, le wallon est reconnu officiellement comme langue endogène de Belgique depuis 1990 et fait partie des « langues de France » inventoriées par le Ministère de la culture français, depuis 1999, puisqu'il est parlé dans la pointe de Givet (département des Ardennes). Il s'agit d'une langue d'oïl, comme le français, forte de plus de 300 000 locuteurs estimés et de ressources numérisées (site Wikipédia en wallon, dictionnaires en ligne, etc). Il n'existe cependant pas en wallon de système de synthèse de la parole à partir du texte donnant accès de façon sonore à toute information disponible sous forme écrite.

La synthèse vocale, désormais présente sur la plupart des téléphones et dans les transports en commun, trouve de nouvelles applications didactiques auprès du grand public, notamment des plus jeunes, qui n'ont pas hérité oralement de la langue. Des systèmes de synthèse vocale ont récemment été développés pour des langues minorisées comme le breton (Lolive 2017) et l'occitan (Corral et al. 2020). Si une dizaine voire une vingtaine d'heures de parole était encore nécessaire, il y a quelques années, les avancées dans le domaine des réseaux de neurones permettent de constituer des systèmes à partir de deux ou trois heures d'enregistrements (Strickland et al. 2023). C'est à peu près la durée de lecture du *Petit Prince*, deuxième livre le plus traduit après la Bible : il a été traduit dans quelque 400 langues dont le wallon namurois. Nous en avons également fait lire une version (ré)écrite dans un wallon normalisé (disponible sur Wikisource), auprès d'un locuteur natif, un auteur qui pourra enrichir ce corpus de nouveaux enregistrements de ses propres livres.

La version enregistrée servira de base au développement d'un système de conversion orthographique-phonétique, lequel devra se montrer robuste à d'autres systèmes d'écriture. Un tel système (graphème-phonème, GP) est en cours de développement au LISN. Écrit sous forme de règles qui pourront être paramétrées pour s'adapter à différentes régions, ce système sera utilisé pour segmenter en phrases et en phonèmes le corpus audio dont nous disposons.

Les données ainsi traitées permettront à un modèle de synthèse automatique de générer n'importe quelle suite de phonèmes et par conséquent — interfacé avec le convertisseur GP — n'importe quel texte. Les systèmes actuels suivent un processus en deux étapes : le modèle acoustique et le vocodeur. Le modèle acoustique convertit le texte en un spectrogramme-mel correspondant — un spectrogramme modifié qui se rapproche de la perception humaine. Le vocodeur utilise ensuite ce spectrogramme pour générer l'onde sonore. Les approches d'apprentissage profond dominent les deux étapes, utilisant notamment des avancées récentes telles que les transformateurs (modèles auto-attentionnels) et les réseaux antagonistes génératifs (GAN). Un des modèles acoustiques les plus répandus est Tacotron 2 (Shen et al. 2018), il est fondé sur une architecture encodeur-décodeur et produit des spectrogrammes-mel à partir du texte ; ensuite, WaveGlow (Prenger, Valle, et Catanzaro 2019), un réseau de neurones

Sous la tutelle de :

www.lisn.upsaclay.fr

Twitter [@LisnLab](https://twitter.com/LisnLab)

LinkedIn [LisnLab](https://www.linkedin.com/company/lisnlab)



université
PARIS-SACLAY



CentraleSupélec

Inria

Site Belvédère : Campus Universitaire Bâtiment 507
Rue du Belvédère – 91405 Orsay Cedex

Site Plaine : Campus Universitaire bâtiment 650
Rue Raimond Castaing – 91190 Gif-sur-Yvette

génératif par *flow*, peut être utilisé comme vocodeur pour générer la parole à partir des spectrogrammes-mel.

Pour réaliser le projet, le ou la stagiaire devra manipuler différents outils de traitement automatique de la langue écrite et de la parole. La connaissance du wallon n'est évidemment pas obligatoire, mais un intérêt pour les langues peu dotées et ce que peuvent apporter les nouvelles technologies constituera un plus : des allers-retours avec des linguistes seront à prévoir. Un co-encadrement expert en *machine learning* facilitera la tâche de l'étudiant ou de l'étudiante. Ce travail représentera une expérience enrichissante pour des candidats intéressés par les outils au niveau de l'état de l'art. Pour candidater, merci de nous envoyer un curriculum vitæ ainsi qu'une brève lettre de motivation.

Durée : 5–6 mois (à partir de février 2024)

Gratifications : environ 650 €/mois

Localisation : LISN, 508 rue John von Neumann, Campus F-91405 Orsay CEDEX

Contact : Philippe Boula de Mareüil mareuil@lisen.fr

Co-encadrant : Marc Evrard marc.evrard@lisen.upsaclay.fr

Références

- Corral, Ander, Igor Leturia, Aure Séguier, Michael Barret, Benaset Dazéas, Philippe Boula de Mareüil, et Nicolas Quint. 2020. « Neural Text-to-Speech Synthesis for an Under-Resourced Language in a Diglossic Environment: the Case of Gascon Occitan ». In *SLTU-CCURL*, 53-60.
- Lolive, Damien. 2017. « Vers plus de contrôle pour la synthèse de parole expressive ». Thèse d'habilitation à diriger des recherches, Université de Rennes 1.
- Prenger, Ryan, Rafael Valle, et Bryan Catanzaro. 2019. « Waveglow: A flow-based generative network for speech synthesis ». In *IEEE-ICASSP*, 3617-21.
- Shen, Jonathan, Ruoming Pang, Ron J Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, et al. 2018. « Natural TTS synthesis by conditioning wavenet on mel spectrogram predictions ». In *IEEE-ICASSP*, 4779-83.
- Strickland, Emmett, Dana Aubakirova, Dorin Doncenco, Diego Torres, et Marc Evrard. 2023. « NaijaTTS: A pitch-controllable TTS model for Nigerian Pidgin ». In *ISCA Speech Synthesis Workshop*.

Sous la tutelle de :

www.lisen.upsaclay.fr

Twitter [@LisnLab](https://twitter.com/LisnLab)

LinkedIn [LisnLab](https://www.linkedin.com/company/lisen-lab)



Site Belvédère : Campus Universitaire Bâtiment 507
Rue du Belvédère – 91405 Orsay Cedex

Site Plaine : Campus Universitaire bâtiment 650
Rue Raimond Castaing – 91190 Gif-sur-Yvette