# Reconnaissance vocale du discours spontané pour le domaine médical

Lucía Ormaechea Grijalba[1, 2] Pierrette Bouillon[1] Johanna Gerlach[1]
Benjamin Lecouteux[2] Didier Schwab[2] Hervé Spechbach[3]

(1) FTI/TIM, Université de Genève, Genève, Suisse
(2) Laboratoire d'Informatique de Grenoble, Saint-Martin-d'Hères, France
(3) Hôpitaux Universitaires de Genève, Genève, Suisse
{lucia.ormaecheagrijalba, pierrette.bouillon, johanna.gerlach}@unige.ch
{benjamin.lecouteux, didier.schwab}@univ-grenoble-alpes.fr
herve.spechbach@hcuge.ch

## RESUME

**Contexte** – Dans le domaine médical, et plus particulièrement dans les services d'urgence, les barrières linguistiques constituent un problème important. Une mauvaise communication entre un médecin et un patient qui ne partagent aucune langue peut mettre en danger la santé et la sécurité du patient (Hacke*r et* al., 2015). Pour faire face à cette situation, nous avons développé, pour le triage aux Hôpitaux Universitaires de Genève (HUG), le système de traduction BabelDr (FR vers ES, AR, TI, FA, PRS et LSF). Celui-ci repose sur un ensemble fini de phrases pré-traduites (mémoire de traduction, MT, environ 10 000 phrases canoniques), mais permet au spécialiste de s'exprimer oralement, pour en améliorer l'ergonomie (Bouj*on et* al., 2018). Le système lie le résultat de la reconnaissance vocale avec l'une des phrases pré-traduites au moyen de techniques neuronales (Muta*l et* al., 2020).

**Objectifs** – Actuellement, BabelDr repose sur le système de reconnaissance commercial NTE (*Nuance Transcription Engine*), spécialisé avec des données artificielles écrites. Celles-ci sont générées à partir de la grammaire BabelDr (*Synchronized Context-free Grammar, SCFG*) (Rayne*r et* al., 2018), qui met en correspondance les phrases canoniques de la MT avec des variations syntaxiques et sémantiques possibles (p. ex. *avez-vous de la fièvre ?* avec *êtes-vous fiévreux ?*, *est-ce que vous avez de la température ?*, etc.). Le but principal de cette étude est de voir quelle performance peut être atteinte pour ce type de discours oral spécialisé spontané, en utilisant la boîte à outils *open source* Kaldi (Povey *et* al., 2011).

**Méthodologie** – Pour développer le système de reconnaissance de la parole, nous avons privilégié une approche temps réel (basée sur la version online de Kaldi). Par ailleurs, nous avons eu recours à des modèles acoustiques hybrides HMM-DNN et une modélisation linguistique appuyée sur un modèle de langage générique interpolé avec une grammaire adaptée au discours médical et se basant sur des données générées avec la grammaire SCFG BabelDr. À l'aide d'un corpus oral de questions d'anamnèses et d'instructions médicales collecté aux HUG avec des médecins via BabelDr, nous avons évalué la performance du prototype Kaldi. Cela nous a permis, en outre, de mettre en regard les résultats obtenus avec les deux technologies.

**Résultats** – À la lumière des résultats globaux observés, une amélioration significative du système basé sur Kaldi est observée par rapport à NTE en termes de WER (14,37% vs 22,93% pour le corpus de test, cf. Table 1). Compte tenu du contexte spécialisé visé, où une erreur de traduction n'est pas acceptable, il est nécessaire d'avoir recours à des évaluations complémentaires, le WER

étant une mesure d'évaluation globale. Les résultats en termes de SemER (*Semantic Error Rate*, pourcentage de phrases orales incorrectement liées à la phrase canonique de la MT) montrent ainsi, sur le test, une meilleure précision sémantique des transcriptions effectuées par le système Kaldi. Les deux systèmes atteignent des taux d'erreur similaires sur le corpus de dev, mais nous observons une grande différence entre les résultats de Kaldi et ceux de NTE dans le corpus de test (14,73% vs 35,52%).

| | | **WER(%)** | | **SemER(%)**[1] | |
|---|---|---|---|---|---|
| ***Corpus*** | ***Phrases*** | **Nuance** | **Kaldi** | **Nuance** | **Kaldi** |
| **Dev** | **2864** | 20,99 | 14,15 | 17,59 | 21,11 |
| **Test** | **2708** | 22,93 | 14,37 | 35,52 | 14,73 |

TABLE 1 : Résultats en termes de Word Error Rate (WER) et Semantic Error Rate (SemER) obtenus avec les systèmes de reconnaissance vocale Nuance et Kaldi.

**Conclusion et perspectives** – Ces premières expériences montrent qu'un système spécialisé de reconnaissance automatique de la parole peut être compétitif en termes de performance par rapport à des systèmes plus généralistes. L'ajout de grammaires spécialisées au sein du décodeur permet ainsi d'atteindre des performances exploitables en production. La mise en production est prévue dans le courant de l'année 2021. À plus long terme nous souhaiterions développer un système qui adapte dynamiquement sa grammaire en fonction des évolutions de son utilisation. Une autre piste serait de générer non pas une transcription, mais directement la forme canonique. Nous envisageons également d'exploiter ce système dans le cadre d'un système de traduction automatique de la parole vers des pictogrammes. En effet, cette étude s'inscrit dans le projet FNS-ANR[2] PROPICTO (*PRojection du langage Oral vers des unités PICTOgraphiques*), qui vise à développer des ressources et des outils pour la transcription automatique de la parole française et sa traduction en pictogrammes.

MOTS-CLES : reconnaissance automatique de la parole – modélisation acoustique – modélisation linguistique – Kaldi – BabelDr – discours médical

# Références

BOUJON, V., BOUILLON, P., SPECHBACH, H., GERLACH, J., & STRASLY, I. (2018). Can speech-enabled phraselators improve healthcare accessibility ? A case study comparing BabelDr with MediBabble for anamnesis in emergency settings. *Proceedings of the 1st Swiss Conference on Barrier-Free Communication*, 50-65. https://doi.org/10.21256/zhaw-2018

HACKER, K., ANIES, M. E., FOLB, B., & ZALLMAN, L. (2015). Barriers to health care for undocumented immigrants : A literature review. *Risk Management and Healthcare Policy*, 175-183. https://doi.org/10.2147/RMHP.S70173

[1] Notons que l'évaluation en termes de SemER n'est effectuée que sur un sous-ensemble du corpus de développement, à savoir, 1222 phrases.

[2] 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).

MUTAL, J., GERLACH, J., BOUILLON, P., & SPECHBACH, H. (2020). Ellipsis Translation for a Medical Speech to Speech Translation System. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 281-290. https://www.aclweb.org/anthology/2020.eamt-1.30/

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. http://publications.idiap.ch/index.php/publications/show/2265

RAYNER, M., BOUILLON, P., TSOURAKIS, N., SPECHBACH, H., & GERLACH, J. (2018). Handling Ellipsis in a Spoken Medical Phraselator. In *Statistical Language and Speech Processing* (Vol. 11171, p. 140-152). Springer International Publishing. https://doi.org/10.1007/978-3-030-00810-9_13