

# Measurements of turn-taking and linguistic behaviors in clinical settings

Rachid Riad<sup>1, 2, 4</sup> Lucas Gautheron<sup>1</sup> Emmanuel Dupoux<sup>1, 3</sup> Anne-Catherine Bachoud-Lévi<sup>4</sup> Alejandrina Cristia<sup>1</sup>

(1) Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'Etudes cognitives, ENS, EHESS, CNRS, PSL University, Paris, France

(2) CoML Team, INRIA, Paris, France

(3) Facebook Artificial Intelligence Research, Paris, France

(4) Laboratoire de Neuropsychologie Interventionnelle, Département d'Etudes cognitives, ENS, INSERM, UPEC, PSL University, Paris, France

riadrachid3@gmail.com, alecristia@gmail.com\*

## RÉSUMÉ

---

Les conversations impliquant des patients, dans des conditions naturelles, sont de précieuses sources d'informations pour le suivi médical. Il est devenu plus facile de collecter de grands corpus de données vocales. Cependant, obtenir des indicateurs pertinents à partir d'enregistrements audio est actuellement hors de portée pour la plupart des cliniciens et des chercheurs. *Alors, comment pouvons-nous obtenir des mesures du tour de paroles et des marqueurs linguistiques pour des applications cliniques ?* Les méthodes envisagées doivent être *fiabiles, peu coûteuses* et garantir la *vie privée*. Nous discuterons des avantages et des inconvénients de plusieurs approches qui sont actuellement développées et examinées dans notre laboratoire : 1) Développement de nouveaux outils pour l'annotation par des experts de fichiers audio (Titeux\* *et al.*, 2020), 2) Crowdsourcing avec des citoyens qui veulent aider dans l'avancée des recherches scientifiques (Semenzin *et al.*, 2020), 3) Développement d'algorithmes pour détecter et identifier les tours de parole durant les entretiens cliniques (Riad *et al.*, 2020).

## ABSTRACT

---

### Measurements of turn-taking and linguistic behaviors in clinical settings

Conversations involving patients, in natural conditions, are valuable sources of information for the medical follow-up. Thanks to recent technological advances, it became easier to collect large naturalistic corpora of speech data. However, getting meaningful insights from these long naturalistic datasets is currently out of reach for most clinicians and researchers. *Then, how do we obtain measurements of naturalistic turn-taking and linguistic behaviors for clinical applications ?* Especially, for scientific endeavor and clinical practice, the methods should be *reliable, cost-effective* and guarantee the *privacy* of the subjects. Indeed, these long recordings can vary greatly by their duration, by their recording conditions and by the variability of speaker traits (that we might want to capture and measure for as clinical outcomes). In this presentation, we will discuss advantages and drawbacks of several approaches that are currently being developed and examined in our laboratory to process audio recordings in clinical settings. On the one hand, researchers and clinicians can rely on expert annotators of speech and language data. We developed standardized ways of assigning annotation

---

\*. We acknowledge ANR-16-DATA-0004 ACLEW, ANR-17-EURE-0017; and the J. S. McDonnell Foundation, ANR-10-IDEX-0001-02 PSL\*, ANR-19-P3IA-0001 PRAIRIE 3IA Institute, Neuratris, Facebook AI Research (Research Gift), Google (Faculty Research Award), Microsoft Research (Azure Credits and Grant), and Amazon Web Service (AWS Research Credits).

to expert annotators (Titeux\* *et al.*, 2020), with the evaluation of their agreements. After training the annotators, the quality of annotations is high but it becomes quickly infeasible to annotate huge corpora. On the other hand, to overcome this scaling issue, we also investigated the capabilities of the crowd-sourcing project on Zooniverse to leverage citizens' help to solve simple classification tasks on short audio chunks drawn randomly from the daylong recordings (Semenzin *et al.*, 2020) of children with and without Angelman syndrome. Finally, we witnessed an increase in performance of the speech processing algorithms thanks to neural networks trained on large open datasets. Yet, it is not known how these algorithms transfer and perform on speech productions of individuals with speech and language deficits. We evaluated the state-of-the-art pipelines to detect and identify speaker turns for conversational clinical interviews between neuropsychologists and patients with Huntington's Disease (Riad *et al.*, 2020). We found that these algorithms required a sufficient amount of well annotated data to reach significant performance. Finally, these algorithms provided halfway satisfactory results concerning speech features relevant for clinical practice.

---

MOTS-CLÉS : tours de parole ; daylong recordings ; annotations ; traitement de la parole pathologique.

---

## Références

RIAD R., TITEUX H., LEMOINE L., MONTILLOT J., SLIWINSKI A., BAGNOU J. H., CAO X. N., DUPOUX E. & BACHOUD-LÉVI A.-C. (2020). Comparison of speaker role recognition and speaker enrollment protocol for conversational clinical interviews. *arXiv preprint arXiv :2010.16131*.

SEMENZIN C., HAMRICK L., SEIDL A., KELLEHER B. & CRISTIA A. (2020). Towards large-scale data annotation of audio from wearables : Validating zooniverse annotations of infant vocalization types. In *SLT*.

TITEUX\* H., RIAD\* R., CAO X.-N., HAMILAKIS N., MADDEN K., CRISTIA A., BACHOUD-LÉVI A.-C. & DUPOUX E. (2020). Seshat : A tool for managing and verifying annotation campaigns of audio data. In *LREC*, Marseille. \* Equal contribution.