

# RECUEIL DE DONNÉES ET ANALYSE DE CORPUS EN STRUCTURE INFORMATIONNELLE LA PHRASE COMPLEXE EN BERBÈRE, PERSPECTIVE COMPARATIVE JOURNÉES D'ÉTUDES TEDONNÉES, UNIVERSITÉ DE FRANCHE-COMTÉ, UFR TAL, 12 AVRIL 2024

Ourida Manseri

INaLCO — USPC, EA 4092, LaCNAD

## Introduction

Branche des langues afroasiatiques, la langue berbère (tamaziɣt) est constituée de parlers ou langues parlées en Afrique du Nord, du Maroc en passant par l'Algérie, la Tunisie, Libye, jusqu'en Égypte, et de la Mauritanie, au Mali, Niger jusqu'au nord du Burkina Faso (fig.1). Malgré un fond structurel commun (VSO), il existe entre ces parlers des disparités de diverses natures. Les premiers résultats de mon travail sur la comparaison des structures de la phrase complexe confirment cette affirmation. Dans ce poster nous parlerons du traitement des données issues de cette recherche, les enjeux, la réalité qui s'y reflète et quelques préconisations.

## La question de recherche

Malgré une structure interne globalement similaire, les premiers résultats de mon étude sur la phrase complexe et la structure informationnelle (SI) dans les parlers berbères indiquent des disparités. Partie intégrante de ma recherche, le recueil, la transcription, annotation, traduction des corpus oraux et écrits est un processus long, car la tâche est manuelle et ralentit l'avancée des travaux. En berbère il n'existe pas encore de système de reconnaissance ou d'annotation automatique de la langue, alors qu'il est urgent de recueillir des données auprès des derniers locuteurs des parlers menacés. Quelles stratégies et quelles actions entreprendre pour préserver le patrimoine linguistique berbère en voie de disparition?

## Le travail sur le corpus

J'ai recueilli mon corpus en Kabylie pendant deux missions sur le terrain au village Aït-Ouanèche, au printemps-été 2022 et été 2023 auprès de deux informatrices monolingues de 93 et 79 ans. Nées dans le village limitrophe, elles maîtrisent la langue littéraire orale. J'ai principalement enregistré, au dictaphone Sony ICD-UX570, 4 Go (LPCM 44.1kHz/16 bit, .wav), l'informatrice de 79 ans dans sa maison. J'ai recueilli des conversations, contes et légendes, récits de vie, poésie, y compris des poèmes de sa composition. J'ai utilisé des données linguistiques publiées pour mon corpus écrit représentant des parlers de différentes régions d'Afrique du Nord. J'ai transcrit l'audio avec ELAN (voir fig.2) et l'écris avec FLEx. Je me suis basée sur les conventions de *Leipzig Glossing Rules* et suivi la méthode utilisée dans le cadre du projet CorpAfroas (Mettouchi, 2007).

e-mail [ourida.manseri@hotmail.fr](mailto:ourida.manseri@hotmail.fr)



Fig. 1: Les parlers menacés ou éteints (points orange, rouge)

## La langue berbère et le TAL

Elan est un outil puissant pour les étapes (segmentation, transcription, annotation, traduction, etc.) de travail. Contrairement aux langues dotées, on ne dispose pas d'une telle technologie, malgré des tentatives (Outahajala *et al.* 2014, 2020). L'automatisation de ces processus est une urgence pour la préservation des parlers où il ne subsiste plus que quelques locuteurs.

**En conclusion**, la documentation ou autodocumentation (Mettouchi 2018) des parlers menacés doit être prise en charge par les États en menant une politique de recueil systématique, création d'un système de TAL pour l'automatisation de la reconnaissance et la mise au point d'un corpus berbère interopérable, afin de préserver les parlers qui sont encore bien vivants.

## Références bibliographiques

- Baude, O. *et al.* Corpus oraux, guide des bonnes pratiques 2006. CNRS Editions, Presses Universitaires Orléans, 2006.  
Ehret, C. "Berber Peoples in the Sahara and North Africa: Linguistic Historical Proposals". *Burials, Migration and Identity in the Ancient Sahara and Beyond*, ed M. C. Gatto et al., CUP, 2019, p. 464-94.  
Galand, Lionel. Études de linguistique berbère. Collection linguistique 83. Leuven: Peeters, 2002.  
Meakins, F. *et al.* *Understanding linguistic fieldwork*. First edition, 2018.  
Mettouchi, A. et C. Chanard (eds) *The CorpAfroAs Corpus of Spoken Afroasiatic Languages*, 2014. <http://dx.doi.org/10.1075/scl.68.website>.

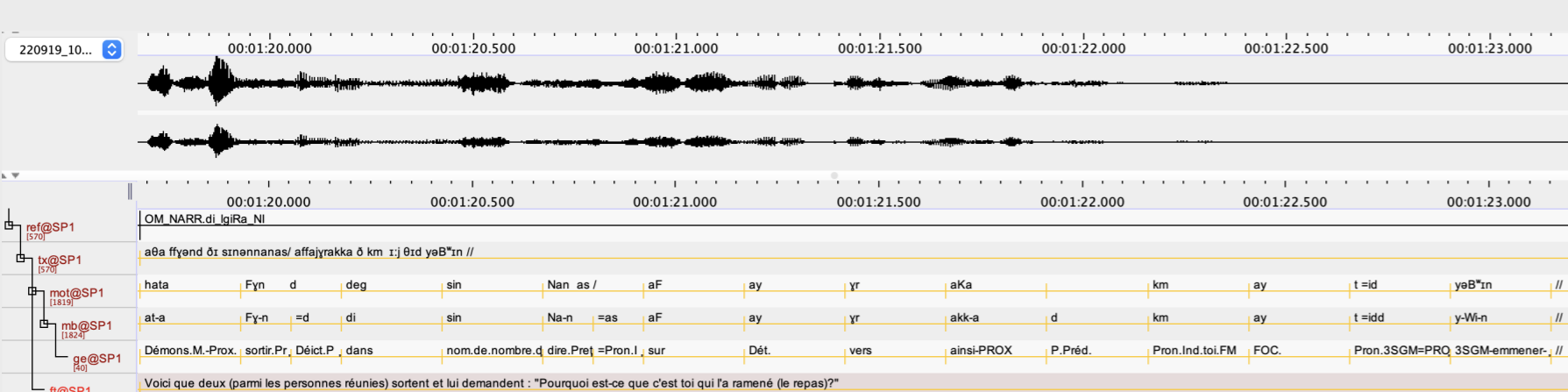


Fig. 2: Transcription phonétique morphophonologique, glose et TL d'un énoncé