

## Titre de la thèse

*Korpusiñ asambI* : vers un apprentissage de langues minorées à partir de données authentiques

## Résumé

Les langues bretonne et arabe partagent un trait commun : aucun groupe social n'a adopté leurs formes standardisées écrites en tant que variété orale exclusive. Il en résulte que les apprenants sont confrontés à un certain nombre de variétés dès le début de leur apprentissage, même lors d'interactions avec des locuteurs lettrés. Si cette variation polymorphique constitue avant tout un élément de surface, elle peut retenir toute l'attention de l'apprenant, voire désorienter ce dernier, occultant de ce fait la profonde unicité de la langue. Par ailleurs, les locuteurs de langue première n'ont pas uniquement recours aux formes normées afin d'assurer l'intercompréhension. Grâce à un ensemble de stratégies, ils adaptent et complètent la forme orale de leur parler d'origine, selon l'interlocuteur, le besoin et le contexte. Ce travail s'adresse aux enseignants et formateurs, parfois déstabilisés face à la variation : elle est décrite et associée à des pistes d'enseignement, selon le niveau.

Cette thèse propose la création d'un nouvel outil, KorA, permettant d'effectuer de la transcription en mode collaboratif à partir de son propre corpus de documents sonores. L'intérêt de cet exercice est qu'il conduit à ne pas se limiter à une simple compréhension globale de l'intrant : l'apprenant est amené à tout observer, du moins s'il est aidé par d'autres apprenants ou son formateur. L'apprenant pourra créer son propre corpus de documents sonores afin, par exemple, d'observer les réalisations orales du même locuteur selon le contexte ou de s'approprier un parler précis, par sa description formelle. De plus, une étude pilote effectuée auprès d'apprenants adultes du breton et visant à tester une autre fonctionnalité proposée par l'outil KorA, la description grammaticale collaborative à partir de transcriptions, est ici exposée. Enfin, un tel outil ouvrirait en outre des perspectives de recherche à partir de la possibilité de constituer collaborativement des corpus, pour différentes langues, qui trouveraient des usages bien au-delà de l'enseignement-apprentissage, par exemple en linguistique ou en traduction et rédaction.

## Nature du financement

Contrat ARED (co-financement région Bretagne/Université Rennes 2). Obtenu le : 28/09/2020