

Titre : Korpusiñ asabl (KorA) : objectifs, résultats, outil, questions et limites

Auteurs : David ar Rouz, enseignant-chercheur en traductologie (Université Rennes 2) ; Ronan Stéphan, formateur de breton (Skol an Emsav)

Résumé :

Le projet KorA, formalisé dès 2018, partait de l'hypothèse qu'un outil de transcription collaboratif serait particulièrement utile pour le breton en vue 1) de myriadiser le travail à effectuer sur les milliers d'heures de collectage qui existent mais ne sont pas valorisées ; 2) d'exposer des apprenants à une langue spontanée, qui n'est presque plus entendue dans la société bretonne, pour un apprentissage éventuellement plus « affectif » (ar Rouz *et al.*, 2025) à partir de données (*data driven learning*). Si cette deuxième utilité était démontrée, en effet, elle ouvrirait des perspectives de financement pérenne d'un outil conçu dans cet objectif.

L'expérimentation menée dans le cadre de ce projet ([co-auteur], 2024), en tenant compte de plusieurs contraintes (absence de prototype, difficulté de la transcription, peu de temps disponible avec les groupes d'apprenants), n'a pas abouti à des résultats soutenant cette deuxième hypothèse, essentiellement en raison du faible nombre d'apprenants et de l'hétérogénéité des niveaux, surtout dans le groupe témoin.

En revanche, le travail a abouti à deux résultats importants :

1. une proposition didactique innovante tenant compte de la variation comme difficulté particulière pour les apprenants, ainsi que de la situation sociolinguistique ne permettant pas (plus) la constitution d'un corpus mental (Taylor, 2012) ;
2. une description précise de la forme que pourrait prendre un outil de transcription collaboratif, adaptable à bon nombre de langues.

La faisabilité de la proposition didactique initiale sera étudiée au regard des recherches en apprentissage à partir de données (Boulton & Cobb, 2017), de l'expérimentation menée et des contextes d'apprentissage existants. L'hypothèse trouve d'ailleurs, dans des études plus générales sur l'apprentissage, confirmation chez les enfants des besoins d'autonomie, de sens, de défis stimulants, de contribution réelle à la communauté... (Alvarez, 2016 ; Doucleff, 2021). Mais elle invite à une révolution éducative que l'outillage seul ne permettra sans doute pas de provoquer. L'outil imaginé sera ensuite présenté du point de vue des fonctionnalités attendues : transcription phonétique, transcription en différentes orthographe, traduction, collaboration en ligne avec organisation des projets et des rôles (enseignant/référent, par ex.). On mentionnera également les difficultés et questions que soulève le traitement de données audio(visuelles), notamment d'ordre juridique, mais aussi du point de vue des exploitations possibles, souhaitables, à autoriser ou interdire.

ALVAREZ Céline, 2016, *Les lois naturelles de l'enfant*, Paris, Éditions des Arènes.

AR ROUZ David, Marie BAIZE-VARIN, Ronan STÉPHAN, 2025, « Préparer l'outillage d'une transmission affective du breton et de l'arabe », dans Alexandra FILHON, Carole LE HÉNAFF & Hugues PENTECOUTEAU (dir.), *Diversité linguistique, éducation et transmission en Bretagne*, Rennes, Presses universitaires de Rennes (PUR), p. 161-174.

BOULTON Alex, Tom COBB, 2017, « Corpus use in language learning: A meta-analysis », *Language Learning*, vol. 67, n° 2, p. 348-393.

DOUCLEFF Michaeleen, 2021, *Chasseurs cueilleurs parents*, traduit de l'anglais par Élise Peylet, Paris, Leduc Éditions.

STÉPHAN Ronan, 2024, *Korpusiñ asabl. Application de traitement collaboratif de corpus. Intérêt et applications linguistico-didactiques, linguistiques et traductologiques*, thèse de doctorat inachevée préparée sous la direction de Marie Baize-Varin et de David ar Rouz, Rennes, LIDILE, Université Rennes 2.

TAYLOR John R., 2012, *The mental corpus: How language is represented in the mind*, Oxford, Oxford University Press.