Actes de CORIA-TALN 2026
Nantes, France, 29 juin au 3 juillet 2026

Exploiter les similitudes linguistiques pour la transcription vocale à faibles ressources

Valentina Fedchenko[1] Eric Jordan[1, 2]
(1) ERTIM, INALCO, 65 rue des Grands Moulins, 75013 Paris, France
(2) LACITO, CNRS, 7 rue Guy Môquet, 75017 Paris, France
valentina.fedchenko@inalco.fr, eric.jordan@inalco.fr

RÉSUMÉ

Cette étude examine comment les modèles acoustiques auto-supervisés à grande échelle (tels que XLSR et MMS) représentent la similarité linguistique et si cette information peut être exploitée pour optimiser la reconnaissance automatique de la parole (ASR) pour les langues à faibles ressources et présentant une forte variation dialectale. Bien que ces modèles excellent dans l'apprentissage par transfert interlinguistique, leurs représentations internes des variations dialectales fines restent encore peu comprises. Nous nous concentrons sur le yiddish, une langue présentant un continuum dialectal complexe, afin de vérifier si une mesure interne de similarité acoustique du modèle, appelée Acoustic Token Distribution Similarity (ATDS), permet de prédire les performances de l'ASR. Notre méthodologie consiste à affiner les modèles sur des données issues de différents dialectes du yiddish et à mesurer l'ATDS entre le yiddish et des langues apparentées. Les résultats confirment que l'ATDS constitue un prédicteur significatif : une similarité acoustique plus élevée dans l'espace latent du modèle est corrélée à des taux d'erreur de caractères (CER) plus faibles après affinage. Cette relation est particulièrement forte dans les couches moyennes à supérieures du modèle MMS et pour les données intra-domaine. L'ATDS capture une similarité acoustique dépendante du modèle, qui ne correspond pas toujours aux relations linguistiques généalogiques, mais qui demeure un indicateur pratique du potentiel d'apprentissage par transfert. Nous concluons que l'ATDS constitue un outil précieux pour sélectionner les langues sources lors du développement de systèmes ASR plus efficaces et sensibles aux variations dialectales dans des contextes de documentation linguistique. Néanmoins, ses valeurs absolues doivent être interprétées avec prudence et mises en perspective avec les connaissances linguistiques existantes.

ABSTRACT

Leveraging Linguistic Similarity for Low-Resource Speech Transcription.

This study investigates how large-scale, self-supervised acoustic models (like XLSR and MMS) represent linguistic similarity and whether this can optimize Automatic Speech Recognition (ASR) for low-resource and dialectally diverse languages. While these models excel at cross-lingual transfer learning, their internal representations of fine-grained dialectal variation remain opaque. We focus on Yiddish, a language with a complex dialect continuum, to test if a model's internal acoustic similarity metric—Acoustic Token Distribution Similarity (ATDS)—predicts ASR performance. Our methodology involved fine-tuning models on Yiddish dialects and measuring ATDS between Yiddish and related languages. Results confirm that ATDS is a meaningful predictor : higher acoustic similarity in the model's latent space correlates with lower character error rates (CER) after fine-tuning. This relationship is strongest in mid-to-upper layers of the MMS model and for in-domain data. Crucially, ATDS captures model-dependent acoustic similarity, which does not always align with genealogical

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 20–21.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

linguistic relationships but remains a practical indicator of transfer learning potential. We conclude that ATDS is a valuable tool for selecting donor languages to develop more efficient, dialect-sensitive ASR systems for language documentation, even if its absolute values require careful interpretation against linguistic knowledge.

MOTS-CLÉS : apprentissage par transfert, similarité linguistique, diversité dialectale, contexte à faibles ressources, reconnaissance automatique de la parole, dialectes yiddish.

KEYWORDS: transfer-learning, linguistic similarity, dialectal diversity, low-resource context, automatic speech recognition, Yiddish dialects.

ARTICLE ACCEPTÉ À : The Fifteenth biennial Language Resources and Evaluation Conference : LREC 2026.
URL : https://lrec.elra.info/lrec2026-main-288