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

Utilisation des modèles de parole pour la recherche lexicale par audio dans les dictionnaires : application au Teochew

Siman Chen Ilaine Wang Maxime Fily Pierre Magistry
Équipe de Recherche Textes, Informatique, Multilinguisme (ERTIM), 2 rue de Lille, 75007 Paris, France
siman.chen, ilaine.wang, maxime.fily, pierre.magistry@inalco.fr

RÉSUMÉ

Cet article porte sur l'utilisation des méthodes la détection de termes parlés par l'exemple (Query by Example - Spoken Term Detection) à une application réelle et dans un contexte de langue peu dotée : l'implémentation d'une fonctionnalité de requête orale pour un dictionnaire de teochew créé par sa diaspora. Cette fonctionnalité permet de faire des recherches dans un dictionnaire sans connaissance préalable des systèmes d'écriture utilisés, rendant ainsi le dictionnaire accessible aux locuteurs non-lettrés d'une langue. Nous comparons deux approches : la première repose sur la reconnaissance de la parole (ASR) avec un appariement *text-to-text*, la seconde sur la déformation temporelle dynamique (DTW) avec un appariement *audio-to-audio*. Nos résultats montrent que l'approche ASR fonctionne mieux. Par ailleurs, celle-ci a pu bénéficier d'optimisations supplémentaires, qui ne sont pas aussi directement applicables à l'approche DTW.

ABSTRACT

Leveraging Speech Models for Audio-based Lexical Retrieval in Dictionaries : the Case of the Teochew Language

This study presents our application of Query by Example - Spoken Term Detection methodologies to a real-world, low-resource scenario : building an audio-based query functionality for a Teochew dictionary. This functionality enables users to retrieve dictionary entries without prior knowledge of the writing systems, thereby enhancing the accessibility of the dictionary. To address this task, we investigate two approaches : (i) an Automatic Speech Recognition (ASR)-based approach using text-to-text matching (automatic romanization of a spoken query against the gold romanization from the dictionary), and (ii) a Dynamic Time Warping (DTW)-based acoustic framework for audio-to-audio retrieval (spoken query against audio recordings from the dictionary). Results show that text-to-text matching outperforms audio-to-audio matching ; however, the two approaches were not optimized under fully comparable conditions, as the ASR-based approach benefited from additional optimization, which was not equally available for the DTW method.

MOTS-CLÉS : Modèles de parole, langues peu dotées, reconnaissance automatique de la parole, Dynamic Time Warping, teochew, requête par l'exemple, détection de termes parlés.

KEYWORDS: Speech Models, Low-Resource Languages, Automatic Speech Recognition, Dynamic Time Warping, Teochew Language, Query-by-Example, Spoken Term Detection.

ARTICLE ACCEPTÉ À : SPEAKABLE@LREC 2026 : Workshop on Speech Language Models in Low-Resource Settings : Performance, Evaluation, and Bias Analysis, Palma, Mallorca, Spain, May 11-16, 2026.
URL : https://speakable-2026.github.io/

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 58–58.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.