

摘要

現代漢語連續口語對話語音語料庫」(Mandarin Conversational Dialogue Corpus) 語料收集時間為2001年。發音人由中央研究院調查研究工作室自臺北市市民中按三大年齡層(16-25歲、26-35歲及36-45歲)，男女各半，隨機抽樣選出60位，共錄製30組對話，為25.6小時，每個對話平均長度約50分鐘。對話內容不限定特定主題。

Sinica MCDC8 為其中八個對話資料。文字轉記以語音軟體PRAAT進行，共計12萬字。以語者說話輪之語言內容為檔案切分之單位。

技術優勢

中研院漢語對話語音語料庫(Sinica MCDC8)包含八個對話之聲檔與文字轉記檔。語音檔以.WAV儲存。文字轉記檔以PRAAT格式儲存，具有段落之時間標記。

本院覽號

24T-1031223

公告日期

智財權狀態

know-how

應用範圍

語音辨識、語音訓練、語言學語料庫之應用。

創作人

曾淑娟