

AESOP-ILAS 語料庫簡介

1. 名稱: AESOP-ILAS (Asian English Speech cOrpus Project - Institute of Linguistics, Academia Sinica) 亞洲口音英語跨國語音資料庫—中研院語言所台灣二語英語語料庫
2. 語言: 台灣地區國語母語者的二語英語口語。亦收錄對應的美式英語母語, 及少量台灣地區國語母語口語。
3. 目的: 提供台灣二語英語的韻律及聲學特徵研究用。
4. 內容: 分 AESOP-ILAS 1 及 AESOP-ILAS 2 二部分。
5. 錄製時間:
 - (1) AESOP-ILAS 1: 2009 年 7 月~2010 年 9 月
 - (2) AESOP-ILAS 2: 2011 年 11 月~2012 年 7 月
6. 語音種類: 麥克風語料 (頭戴式 Sennheiser PC155 麥克風)
7. 發音人: 總計 540 人次
 - (1) AESOP-ILAS 1: 美式英語母語者 (L1) 12 人 (男性 6 人, 女性 6 人), 台灣地區國語母語者 (L2) 488 人 (男性 231 人, 女性 257 人), 共 500 人次
 - (2) AESOP-ILAS 2: 美式英語母語者 (L1) 10 人 (男性 5 人, 女性 5 人), 台灣地區國語母語者 (L2) 30 人 (男性 15 人, 女性 15 人), 共 40 人次
8. 語料庫大小: 總計 14 GB, 約 812 小時; 包含原始音檔 (*.wav)、對照文本 (*.corpus)、經 HTK 分割校準音段處理之檔案 (*.phn)、一部份經人工調整之音節檔 (*.adjusted)
 - (1) AESOP-ILAS 1: 8.64 GB, 約 500 小時
每筆語料錄音時間平均 1 小時
 - (2) AESOP-ILAS 2: 5.42GB, 約 312 小時
每筆語料錄音時間平均 7.8 小時 (L1: 5.25 小時; L2: 8.7 小時)
9. 語料設計:
 - (1) AESOP-ILAS 1: 朗讀短句與短語篇語料共 6 種, 誘發性半自發人機對話 1 種, 自發性看圖描述 1 種, 共 8 種
 - (2) AESOP-ILAS 2: 朗讀單詞、窄焦點短句、長語篇段落語料各 1 種, 誘發性半自發人機對話 1 種, 中文短語篇段落語料 1 種, 共 5 種

10. 簡介:

AESOP-ILAS 語料庫為「亞洲口音英語跨國語音資料庫 AESOP (Asian English Speech cOrpus Project) 國際聯盟」的台灣二語英語部分，由蔣經國國際學術交流基金會同名專題計畫資助，計畫編號 DB002-D-08，執行期間民 98.07.01～101.12.31，計畫主持人為中研院語言所特聘研究員兼所長鄭秋豫。

本語料庫共 14 GB，約 812 小時。分為二部分：AESOP-ILAS 1 共 8.64 GB，約 500 小時，共有 500 人次口語資料：計 12 位美式英語母語者、488 位台灣國語母語者，每位錄音時間約 1 小時。AESOP-ILAS 2 共 5.42GB，約 312 小時，共有 40 人次口語資料：計 10 位美式英語母語者，每位錄音時間約 5.25 小時、30 位台灣國語母語者，每位錄音時間約 8.7 小時。除原始音檔外，同時收錄對照文本、以及 HTK 分割校準音段處理之檔案 (*.phn)，與一部份經人工調整之音節檔 (*.adjusted)，標記對應語音（子音與母音）之邊界時間碼。

本語料庫不以音段或某特定或個別語音特徵為目標，而以音節、詞組、片語、短句與多短語語段等各級不同大小的語言層次及單位，強調較全面性、具溝通訊息的韻律現象及語音特徵，以期提供更豐富的語音分析與評量指標。

基於學術資源公開共享之理念，於 2015 年 4 月釋出 AESOP-ILAS 語料庫，提供國內外學術研究單位使用，可供英語教學、語音研究、語音建模、語音辨識與合成等多方面學術應用。

本資料庫之智慧財產權屬中央研究院，授權中華民國計算語言學會發行，供國內學術機構非營利使用。申請人需向中華民國計算語言學會提出申請，簽妥授權使用協議書，並同意確實遵守協議書上之約定條款。