

AESOP-ILAS 語料庫

一、語料庫簡述

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. 收錄語言: 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: 分為 AESOP-ILAS 1 及 AESOP-ILAS 2 二部分。(詳見三)
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, 詳見附錄 A)
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)、一部份經人工調整過之音節檔 (*.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: 5.42 GB, 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 (詳見四)
 - (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 (詳見四)

二、 錄音設定

1. 語音格式:

- (1) 音檔格式: *.wav
- (2) 取樣頻率: 16000 赫茲 (Hz)
- (3) 位元率: 16 位元
- (4) 聲道: 單聲道

2. 錄音軟體: TWNAESOP Recording Tool (以香港中文大學與中國科學院深圳先進技術研究院共同開發之錄音軟體 CUHK-SIAT recording tool.exe 為雛型, 依本錄音實驗目的修訂而成)

3. 錄音地點:

- (1) AESOP-ILAS 1: 安靜的教室、辦公室、實驗室
- (2) AESOP-ILAS 2: 中央研究院語言學研究所語音實驗室

三、 語料庫說明

AESOP-ILAS (Asian English Speech cOrpus Project - Institute of Linguistics, Academia Sinica) 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。

AESOP 國際聯盟成立宗旨, 為共同建立一個代表亞洲口說英語多樣性的亞洲英語語音資料庫及平台, 由來自日本、台灣、香港、中國、泰國、菲律賓、越南、印尼、蒙古的語言學家、語音科學家與教育學者組成。主要目標包括: 以共同的核心語料及標註系統、相同的錄音設備, 蒐集並分享、比較上述地區的亞洲英語語料。每一位 AESOP 合作成員需獨立尋求研究經費的補助, 完成分派的工作, 同時配合各語言的特性及各自的研究需求, 自行設計並蒐集核心語料以外的語料。所有 AESOP 蒐集建構的語料庫, 也將成為公開資源。AESOP 台灣地區語料庫 (後定名: AESOP-ILAS 語料庫) 建置主持人為中央研究院語言學研究所特聘研究員兼所長鄭秋豫, 以台灣代表身份應邀參與, 負責 AESOP 資料庫共同核心內容之設計, 並蒐集核心語料中台灣地區英語的語料。

AESOP-ILAS 語料庫藉由「亞洲口音英語跨國語音資料庫——建構台灣口音英語語音資料庫 TWNAESOP (TaiWaN Asian English Speech cOrpus Project)」此一國際合作資料蒐集研究計畫建置, 計畫資助單位為蔣經國國際學術交流基金會 (計畫編號: DB002-D-08, 執行期間: 2009/07/01~2012/12/31)。該計畫共有四位學者參與: 主持人鄭秋豫為中央研究院語言學研究所語音實驗室召集人, 研究專長為語音學跨學科研究、語料庫語音學、計算語言學及口語韻律之量化分析及建模。三位計畫協同主持人為: 國立陽明大學 Tanya Visceglia 助理教授, 其母語為美式英語, 研究專長為台灣國語與台灣口音英語的語音特性分析; 香港中文大學蒙美玲 Helen Meng 教授, 專長為粵語、國話與英

語音成分的研究，並將其研究成果直接應用於自動語音辨識（ASR）及相關語音科技的開發；日本早稻田大學 Yoshinori Sagisaka 教授，同時也是 AESOP 跨國研究計畫及資料庫建置的總召集人，其研究專長為計算語言學及跨語言韻律分析與模型。

此語料庫內容分為二部分：第一部分為 AESOP-ILAS 1，第二部分為 AESOP-ILAS 2。語料共 14 GB，約 812 小時。

AESOP-ILAS 1，即 AESOP 國際語音聯盟核心語料台灣部分，於 2009 年 7 月至 2010 年 9 月錄製，共 8.64GB，約 500 小時，共有 500 人次口語資料：包括 L1 英語語料 12 筆，L2 台灣口音英語語料 488 筆，每筆語料錄音時間平均約 1 小時。語料內容包括：朗讀目標詞嵌入短句文本、朗讀短語篇、誘發性半自發人機對話，及看圖描述獲取之自發性語篇（詳見四）。本語料庫的文本設計與以往台灣口音英語語料庫最大不同處在於：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。

由於希望盡可能蒐集到最多人次語料，涵蓋最大語者變異，以獲取最具代表性的台灣英語語料樣本，AESOP-ILAS 1 的文本設計著重口語英語音段與超音段語音層面，每位語者錄音時間控制在 1 小時內，但也因控制因素有限的情況下，如欲使用此批語料建立計算語言學模型時，將面臨語料變異性不足（data sparsity）的問題。換言之，AESOP-ILAS 1 語料庫並未包含 (1) 足夠的音位配列結構變異（phonotactic variation）以及 (2) 二語英語的語篇韻律與語用層面因素之考慮。

針對以上不足，研究團隊於 2010 年完成 AESOP-ILAS 1 語料庫建置後，開始考察多個主要英語語料庫的文本設計內容（詳見四），準備錄製 AESOP-ILAS 2。計畫主持人鄭秋豫並多次與合作主持人 Yoshinori Sagisaka 教授等人，深入討論工作進度安排與內容設計，最後決定錄製語料的文本內容及錄音方式。

AESOP-ILAS 2 於 2011 年 11 月至 2012 年 7 月錄製，共 5.42GB，約 312 小時，共有 40 人次口語資料：包括 L1 英語語料 10 筆，L2 台灣口音英語語料 30 筆，每筆語料錄音時間平均約 7.8 小時。語料內容包括：朗讀單詞、朗讀窄焦點短句、朗讀長篇語篇段落語料、誘發性半自發人機對話，及中文短語篇段落語料（詳見四）。

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。除原始音檔外，同時收錄對照文本，以及經 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，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。

四、 語料內容及來源

A. AESOP-ILAS 1

每位發音人共完成8項錄音作業：作業(1)至(6)為朗讀語句或短語篇，作業(7)為誘發性半自發人機對話（computer-prompted dialogue），作業(8)則為自發性看圖描述（picture description task），回答5個書面問題。平均錄音時間約1小時。

- (1) 目標字 X 嵌於承載句句重音位置（Target words embedded in carrier sentences）：目標字 20 個（1~4 音節的單字，選自美國 CMU（Carnegie Mellon University）電子常用詞辭典，主重音平均分佈於第 1 到第 4 音節），嵌入承載句 “I said X five/ten times.” 句型中，共 20 句。
- (2) 同（1）目標字嵌於簡單句韻律邊界（prosodic boundaries）位置：將（1）的 20 個目標字嵌入 3 種韻律邊界中，分別為 Yes-No 問句尾、Wh 問句尾、2-phrase 短語（句）尾，每種 5 句，共 15 句。
- (3) 同（1）目標字嵌入強調句語句對比重音（contrastive stress or narrow focus）位置：共 20 句。
- (4) 介係詞嵌入強調句語句對比重音（contrastive stress）位置：共 5 句。
- (5) 字面相同但有一個意義以上的短句：每組語句字面相同或十分相似，但因出現的情況（context）不同，意義不同，說法也不同。發音人依不同的情況，將不同的意義讀出。共 5 組，10 句。
- (6) 短篇段落「北風與太陽」：IPA（International Phonetic Association）口語文本實例 “The North Wind and the Sun”，包含絕大多數英語音位對比（phonemic contrast）的短語篇，共計 144 個音節、113 個詞、5 個句子（8 個主要子句、5 個從屬子句）、3 個段落。
- (7) 誘發性人機對話：發音人與電腦各自扮演不同的角色對話，發音人扮演航空公司訂位人員，電腦扮演訂票顧客。發音人根據電腦所發出的客戶語音與螢幕上呈現的回答文本提示，與電腦對話，共 17 組話輪（turn）。
- (8) 看圖描述：發音人根據螢幕上顯示的圖片，回答 5 個問題。【*此部分不含轉寫文本】

B. AESOP-ILAS 2

1. 語料考察：

為蒐集更多語料，補全 AESOP-ILAS 1 不足部分，研究團隊首先考察相關英語語料庫的語料（文本）內容，包括：TIMIT, IViE, UME-ERJ 與 CASS-AESOP，做為 AESOP-ILAS 2 文本設計之參考。各語料庫語料（文本）內容摘要如表 1 所示：

語料庫名稱	單位	年份	語料內容摘要
TIMIT	麻省理工學院	1993	A. 460 phonetically-compact sentences B. 2 calibration sentences C. 1890 randomly selected sentences
IViE	牛津大學	1998	A. 22 phonetically-controlled sentences with different grammatical structures sentences B. A passage of “The Cinderella Fairy Tale” C. A retold version of the Cinderella fairy tale D. A map task E. Discussion on a given topic
UME-ERJ	東京大學	2000	A. 8 sentence sets (S1-S8) 1. 460 TIMIT-based phonetically-balanced sentences 2. 32 sentences including phoneme sequences difficult for Japanese to pronounce correctly 3. 100 sentences designed for test set 4. 94 sentences with various intonation patterns 5. 120 sentences with various accent and rhythm patterns B. 5 words sets (W1-W5) 1. 302 minimal-pair words 2. 300 phonemically balanced words 3. 109 words with various accent patterns
CASSAESOP	中國社科院	2010	A. AESOP core sentences: the same as AESOP-ILAS 1 B. The text pieces from UME-ERJ

			<p>C. A passage of “The Cinderella Fairy Tale” from IViE</p> <p>D. CASS-E: 415 English sentences to cover a large variety of intonation variations</p> <p>E. CASS-C: 124 Chinese tokens of various syllables, words, sentences, dialogues and paragraphs</p>
--	--	--	--

表 1 考察之英語語料庫語料（文本）內容摘要

2. 語料設計及說明：

為誘發：(1) 全面性涵蓋音位配列結構變異（**phonotactic variation**）；(2) 寬／窄語意焦點（**broad/narrow focus**）；(3) 大範圍語篇及語用層面二語英語的口語語料，以及考量與 AESOP 國際語音聯盟成員分享各自錄製語料的協議，AESOP-ILAS 2 文本設計納入了 5 項錄音作業：作業(1)至(3)為英語朗讀語料，作業(4)為英語誘發性半自發人機對話（**computer- prompted dialogue**），作業(5)為台灣國語朗讀語料。平均錄音時間約 7.8 小時。

- (1) 音位配列結構變異單詞（**Phonotactic variations from phonetically-rich isolated words**）：參考 Pitrelli 等人（1995）Phonebook 作法，簡化人工篩選程序，參考 (1) Syllabified CMU Pronouncing Dictionary (v. 0.6), (2) The Word Frequency List，產生最後字表的來源（輸入）材料（**source material**），第一階段從近 13 萬單字中初步自動篩選產生 32,950 候選字表，第二階段再以程式篩選出 6,122 個單字，最後人工刪除程式未剔除的不適當字、專有名詞或縮寫等單字產生 5,400 個單字字表。
- (2) 寬／窄語意焦點語句（**Broad and narrow focus sentences**）：176 句誘發寬／窄語意焦點語句，取自 CASSASEOP 中 CASS-E 部份語句內容。
- (3) 長篇段落「灰姑娘」：取自 IViE 長篇段落 “The Cinderella Fairy Tale”，其段落長度遠長於 AESOP-ILAS 1 語料中的短篇「北風與太陽」（“The north wind and the sun”），提供足夠的句調變異與語篇韻律訊息，文本長度共計 1,000 個音節、759 個詞、93 個主要子句、49 個從屬子句、14 個段落。
- (4) 特定情境設計誘發性對話（**DCT Situational Dialogue**）：Discourse Completion Tasks（以下稱 DCT）原本是用以瞭解二語學習的言語行為（**speech act**）的研究方法（Blum-Kulka, 1989），但其形式是透過特定情境的對話設計，以誘發涵蓋對話雙方適當的提問與回應情節，因此研究團隊認為透過 DCT 亦可獲取各種互動式語篇形式的語料，符合研究需要。

AESOP-ILAS 2 所使用的文本，原始文本為早稻田大學 Michiko Nakano 教授研究團隊所設計，Michiko Nakano 教授是國際知名的語言評測專家，文本內容依二語英語能力的精熟程度分成 5 級，從 A2 基礎級到 C2 進階級，每一級共 19 個單元，包含 15 至 17 組對話，全部共有 86 組 2 人 A-B 對話與 3 組 3 人 A-B-C 對話。AESOP-ILAS 2 採用其中的 86 組 2 人 A-B 對話內容，文本中部份情境或內容係依台灣地區的文化民情與中文特性適度修改而成。發音人與電腦依照情境，各自扮演不同的對話者角色 A 或 B，發音人根據電腦所發出的對話語音內容與螢幕上呈現的回答文本提示，與電腦對話。

- (5) 中文短篇段落「北風與太陽」：2011 年，AESOP 國際聯盟決議採用 IPA (International Phonetic Association) 口語文本實例「北風與太陽」文本，做為共同的 L1 文本，以利 language transfer 或 tone borrowing 的語音分析比較，釐清母語對於二語英語的影響為何。

以上(2)、(3)、(4)語料蒐集，有助日後與 AESOP 團隊所蒐集之語料，以及 IViE 語料交叉比較研究。

五、錄製資料一覽

A. AESOP-ILAS 1

AESOP-ILAS 1 語料庫共錄製 500 筆語料，含 L1 (母語為美式英語者) 12 筆，及 L2 (母語為台灣地區國語，英語為第二外語者) 488 筆，錄製地點、人數、性別比例詳見表 2:

		錄製數量	男/女 (人數)	語料大小
L1 發音人		12	6/6	258MB
L2 發音人 語料蒐集單位 (錄製地點)	中研院語言所	35	9/26	8.33 GB
	中研院資訊所	14	10/4	
	台灣大學	74	29/45	
	清華大學	39	23/16	
	交大語言所	79	38/41	
	交大電信所	39	20/19	
	中正大學	61	31/30	
	師範大學	48	21/27	
	台北科技大學	58	30/28	
	銘傳大學	41	20/21	
合計		500	237/263	8.58 GB

表 2 AESOP-ILAS 1 發音人來源資料摘要

B. AESOP-ILAS 2

AESOP-ILAS 2 語料庫共錄製 40 筆語料，含 L1（母語為美式英語者）10 筆，及 L2（母語為台灣地區國語，英語為第二外語者）30 筆。

此語料庫是由 5,400 個英語音位配列結構變異單詞、176 句英語誘發寬／窄語意焦點語句、1 英語長篇段落「灰姑娘」“The Cinderella Fairy Tale”、86 組英語誘發性對話，及 1 中文短篇段落「北風與太陽」所組成。由於單一語者所需錄製語料數量遠較 AESOP-ILAS 1 龐大，考慮到錄音時間與疲勞等因素，錄音作業分多次進行，語料內容摘要、每人所需錄製的語料數量，以及錄製總數如表 3 所示：

作業	語料總數	L1/L2 發音人 (人數)	錄製數量 (L1/L2)	錄製總數 (L1/L2)
音位配列結構變異單詞	5,400	10/30	600/200 <small>(重疊錄製: 60/20)</small>	6,000/6,000
寬／窄語意焦點語句	176		176/88	1,760/2,640
長篇「灰姑娘」	1		1/1	10/30
特定情境設計誘發性對話	86 組, 492 句		492/492	4920/14760
中文短篇「北風與太陽」	1	0/30	0/1	0/30

表 3 AESOP-ILAS 2 語料錄製數量摘要

六、資料夾名稱與檔案名稱說明

1. 資料夾名稱說明

- (1) L1: 英語（母語）語料；L2: 台灣口音英語語料
- (2) 發音人代號編號規則：
性別（F: 女性；M: 男性）+ L1/L2（1/2）+ 發音人流水編號（001~999）
- (3) adjust: 經人工調整之音節檔（僅一部分檔案有）
- (4) corpus: 錄音文本（建議使用軟體：記事本、WordPad）
- (5) phn: 經 HTK 分割校準音段處理之檔案
- (6) wav: 音檔

2. 檔案名稱說明

- (1) AESOP-ILAS 1 檔名: Record #(of Recording session)_# (of files).wav
(corpus)
- (2) AESOP-ILAS 2 檔名: Recording session_# (of files).wav (corpus)

七、發音人基本資料人數統計與百分比

錄音作業開始前，每位發音人需在軟體登入視窗中填寫簡要個人基本資料，包括：性別、母語、教育程度、教育背景、（自評）英語能力精熟度等，基本資料可選擇之項目如表 4 所列：

性別	母語	教育程度	教育背景	自評英語能力
男	英語	博士	藝術學院	精通
女	台灣地區國語	碩士	商管學院	良好
	閩南語	學士	教育學院	普通
	客家語	高中	工學院	不良
	南島語	小學	人文學院	
	其他	其他	法學院	
			醫學院	
			理學院	
			社會科學院	
			其他	

表 4 基本資料可選項目一覽表

A. AESOP-ILAS 1

1. 發音人性別：

L1: 男: 6 人 (50%) ; 女: 6 人 (50%) 。

L2: 男: 231 人 (47%) ; 女: 257 人 (53%) 。

2. 發音人母語：

L1: 英語: 12 人 (2%) 。

L2: 台灣地區國語: 385 人 (77%) ; 閩南語: 93 人 (19%) ; 客語: 10 人 (2%) 。

3. 發音人教育程度：

L1: 博士: 3 人 (25%) ; 碩士: 6 人 (50%) ; 學士: 3 人 (25%) 。

L2: 博士: 14 人 (2%) ; 碩士: 237 人 (49%) ; 學士: 237 人 (49%) 。

4. 發音人教育背景：

L1: 商管學院: 1 人 (8%) ; 教育學院: 4 人 (34%) ; 人文學院: 4 人 (34%) ; 理學院: 2 人 (16%) ; 其他: 1 人 (8%) 。

L2: 藝術學院: 13 人 (3%) ; 商管學院: 44 人 (9%) ; 教育學院: 5 人 (1%) ; 工學院: 178 人 (36%) ; 人文學院: 114 人 (23%) ; 法學院: 3 人 (1%) ; 醫學院: 12 人 (2%) ; 理學院: 77 人 (16%) ; 社會科學院: 9 人 (2%) ; 其他: 33 人 (7%) 。

5. L2 發音人自評英語能力精熟度結果：
精通：11 人（2%）；良好：117 人（24%）；普通：293 人（60%）；不良：67 人（14%）。

B. AESOP-ILAS 2

1. 發音人性別：
L1: 男：5 人（50%）；女：5 人（50%）。
L2: 男：15 人（50%）；女：15 人（50%）。
2. 發音人母語：
L1: 英語：10 人（25%）。
L2: 台灣地區國語：29 人（73%）；台語：1 人（3%）。
3. 發音人教育程度：
L1: 博士：2 人（20%）；碩士：3 人（30%）；大學：5 人（50%）。
L2: 博士：2 人（7%）；碩士：21 人（70%）；大學：7 人（23%）。
4. 發音人教育背景：
L1: 藝術學院：1 人（10%）；商管學院：1 人（10%）；教育學院：2 人（20%）；工學院：1 人（10%）；人文學院：2 人（20%）；社會科學院：1 人（10%）；其他：2 人（20%）。
L2: 藝術學院：8 人（27%）；商管學院：1 人（3%）；教育學院：3 人（10%）；工學院：8 人（27%）；人文學院：2 人（7%）；理學院：4 人（13%）；社會科學院：4 人（13%）。
5. L2 發音人自評英語能力精熟度結果：
精通：11 人（27%）；良好：14 人（35%）；普通：15 人（38%）。

八、 附錄

* 附錄 A: 麥克風 **Sennheiser PC155** 產品特色與規格:

PC155 The ultimate gaming tool

- ▶ **XL, velour padded ear caps** block out surrounding noise and give you ultimate wearing comfort for heavy duty gaming
- ▶ With the unique **3D, 360-degree sound experience**, sound comes at you from all sides including above and below
- ▶ The **USB adapter with integrated sound card** lets you take your own personal extreme surround sound with you wherever you play



Our best pro-gamers headset, the PC 155 USB lets you play, watch, talk or listen almost anytime, almost anywhere. It will keep you one step ahead of any attackers and right in the middle of the action.

- ▶ Adjustable headband for the perfect fit
- ▶ Adjustable, bendable, pivotable, noise-canceling microphone improves the clarity of your communication, and gives you more comfort and flexibility
- ▶ In-line volume control and mic mute switch lets you decide volume intensity and have a side conversation in privacy
- ▶ Can be used with CD/DVD/MP3 players and for IP-based communications via a detachable USB adapter that reveals 2 x 3.5 mm plugs
- ▶ Single-sided cable keeps chin and neck free so you can concentrate on your game
- ▶ Cable wind-up clip keeps excess cable wire - between headphones and adapter - out of the way when you play.

Technical data

Headphones
Frequency response 18 – 22,000 Hz
Impedance..... 32 Ω
SPL (at 1 KHz, 1 Vrms)..... 114 dB
Cable length 3 m / 9.8 ft
Connector Plug USB adapter and 2 x 3.5mm jack

Microphone
Frequency response 80 Hz – 15,000 Hz
Pick-up pattern..... Uni-directional
Sensitivity -38 dB
Impedance..... -2 k Ω

Accessories

Included:

- ▶ Detachable USB adapter
- ▶ Wind-up cable clip

Optional:

- ▶ Replacement foam ring ear cushions
- ▶ Headband pad, also available in leatherette
- ▶ Microphone windscreen