

A Study on a Low-Resource Speech Recognition System for Taiwan Hakka Based on Whisper and LoRA (基於 Whisper 與 LoRA 的低資源模型之臺灣客家語語音辨識系統研究)

劉政廷

National Yunlin University of Science
and Technology Department of IM
(Program in Artificial Intelligence) / 雲
林縣斗六市大學路三段 123 號
zhengtingliu0104@gmail.com

廖宜翔

National Yunlin University of Science
and Technology Department of IM
(Program in Artificial Intelligence) / 雲
林縣斗六市大學路三段 123 號
sean0908699966@gmail.com

王恆佑

National Yunlin University of Science
and Technology Department of IM
(Program in Artificial Intelligence) / 雲林
縣斗六市大學路三段 123 號
yoyo1234870@gmail.com

邱重袁

National Yunlin University of Science
and Technology Department of IM
(Program in Artificial Intelligence) / 雲林
縣斗六市大學路三段 123 號
alan1109135@gmail.com

黃昭義

National Yunlin University of Science and Technology Department of IM (Program
in Artificial Intelligence) / 雲林縣斗六市大學路三段 123 號
joyh@yuntech.edu.tw

摘要

本研究旨在開發一套高效能的臺灣客家語自動語音辨識 (ASR) 系統，以應對客語作為低資源語言所面臨的保存與數位化挑戰。本研究採用 whisper-large-v3-taiwanese-hakka 作為基礎模型，此模型基於先進的 Transformer 編碼器-解碼器架構。為達成參數高效且能適應新語言的目標，我們採用了 LoRA (Low-Rank Adaptation) 微調策略，並特別針對模型中 q_proj、k_proj、v_proj、out_proj、fc1、fc2 等多個關鍵模組進行適配。實驗結果表明，相較於基礎模型在 FSR-2025 HAT-Vol2 測試集上 23.07% 的字元錯誤率 (CER)，經過 LoRA 微調後的模型表現出色，最終將 CER 顯著降低至 7.07%。訓練過程監控顯示，模型的驗證集損失與錯誤率皆穩定下降並收斂，證明 LoRA 策略能在不發生毀滅性遺忘的前提下，成功地將大型模型的知識遷移至客語辨識任務，提供了一個高效的解決方案。

Abstract

This study aims to develop a high-performance Automatic Speech Recognition (ASR) system for Taiwan Hakka, addressing the preservation and digitalization challenges it faces as a low-resource language. We utilized whisper-large-v3-

taiwanese-hakka as the base model, which is built upon an advanced Transformer encoder-decoder architecture. To achieve parameter-efficient adaptation for the new language, we employed the Low-Rank Adaptation (LoRA) fine-tuning strategy, specifically adapting key modules within the model, including q_proj, k_proj, v_proj, out_proj, fc1, and fc2. The experimental results demonstrate outstanding performance. Compared to the base model's Character Error Rate (CER) of 23.07% on the FSR-2025 HAT-Vol2 test set, the LoRA-tuned model achieved a significant reduction, bringing the final CER down to 7.07%. Monitoring of the training process showed that the model's validation loss and error rate both steadily decreased and converged. This confirms that the LoRA strategy can successfully transfer knowledge from the large model to the Hakka recognition task without suffering from catastrophic forgetting, providing an efficient solution.

關鍵字：客家語、自動語音辨識

Keywords: Hakka, Automatic Speech Recognition

1 前言

自動語音辨識 (Automatic Speech Recognition, ASR) 技術在高資源語言 (如英語、漢語普通話) 的應用已臻成熟，並廣泛

部署於智慧助理、語音輸入及智慧家居等場域。近年來，此領域的突破主要歸功於深度學習，特別是 Vaswani et al. (2017) 提出的 Transformer 架構，其顯著提升了語音辨識的準確率與泛化能力。然而，對於全球眾多資源匱乏的低資源語言而言，ASR 系統的發展因語料不足而依然面臨嚴峻挑戰。

臺灣客語作為重要的本土語言之一，便面臨著語料稀缺且使用人口逐年下降的雙重困境。為此，發展高效能的客語 ASR 系統，不僅是技術層面的探索，更對語言的數位保存、教育推廣與文化傳承具有深遠的學術與社會意義。

在此背景下，「2025 Formosa 語音辨識挑戰賽」提供了標準化的 HAT-Vol2 客語語料庫，為本研究奠定了關鍵的實驗基礎。同時，OpenAI 的 Whisper 模型 (Radford et al., 2023) 透過在海量多語言資料上的預訓練，展現了在低資源情境下卓越的遷移潛力。因此，本研究的核心動機便是結合前述契機，將先進的大型預訓練語音模型應用於客語辨識任務，並透過微調策略探索其效能極限，以期為客語數位化工程提供一個穩固的技術方案。

2 文獻回顧

為奠定本研究之理論基礎與確立研究定位，本章將回顧自動語音辨識技術的發展脈絡、探討其在低資源語言上面臨的挑戰，並檢視臺灣本土語言相關研究之現況。首先，我們將追溯 ASR 技術的演進，從早期的統計模型（如隱馬可夫模型 (Hidden Markov Models, HMMs)）發展至現今由深度學習主導的端到端架構，特別是 Transformer 模型的崛起如何革新了整個領域。

2.1 自動語音辨識技術的演進

ASR 的發展可追溯至 20 世紀 50 年代早期，當時的系統主要依賴基於動態時間規劃 (Dynamic Time Warping, DTW) 的方法來進行模式比對。隨著統計學與機器學習的進步，HMMs 成為主流，並在 1980 至 1990 年代主導了語音辨識的研究方向。HMMs 能夠有效地建模語音信號的時間序列特性，並與高斯混合模型 (Gaussian Mixture Models, GMMs)

結合，形成了經典的 GMM-HMM 框架，廣泛應用於語音辨識系統中。

進入 2010 年後，深度學習技術的興起帶來了革命性的改變。深度神經網路 (Deep Neural Networks, DNNs) 被引入到聲學建模中，逐漸取代傳統的 GMM。接著，卷積神經網路 (Convolutional Neural Networks, CNNs) 與長短期記憶網路 (Long Short-Term Memory, LSTM) 更進一步提升了模型的表現，使得 ASR 系統在大型語料下取得顯著突破。2017 年，Vaswani 等人提出的 Transformer 架構，首次完全拋棄循環結構，透過自注意力機制 (Self-Attention Mechanism) 同時捕捉長距依賴與上下文資訊，顯著提升了序列建模能力。此後，基於 Transformer 的模型（如 Conformer、wav2vec 2.0）逐漸成為主流，並奠定了現今端到端語音辨識系統的基礎。

2.2 低資源語言的 ASR 挑戰

儘管深度學習推動了語音辨識的快速進展，但這些成功主要集中於高資源語言，如英語與漢語普通話。低資源語言面臨兩大挑戰：

1. 語料不足：缺乏大規模且標註完善的語音-文字對齊資料，使得模型難以進行有效訓練。
2. 語言特性複雜：許多低資源語言存在方言差異、口語化強烈或文字系統尚未標準化等問題。

為克服這些困境，研究者提出了多種方法：

- 跨語言遷移學習 (Cross-lingual Transfer Learning)：先在高資源語言進行預訓練，再將模型微調於低資源語言。
- 多語言訓練 (Multilingual Training)：同時使用多種語言進行訓練，以共享跨語言特徵。

- 參數高效微調 (PEFT)：如 LoRA (Hu et al., 2022)，僅調整部分權重即可快速適應新語言，降低計算成本與記憶體需求。

這些方法已在多種低資源語言上展現成效，例如 Meta 的 BABEL 計畫

(Harper, 2014) 針對 26 種語言建立了語音資源，wav2vec 2.0 也在非洲語言與南亞語言上獲得成功應用。

2.3 臺灣本土語言的語音研究

臺灣的語言多樣性極為豐富，包括閩南語、客家語、原住民族語言等。然而，相關的語音辨識研究仍處於起步階段。過去研究多集中於閩南語，例如透過 Kaldi 工具建立 GMM-HMM 與 DNN-HMM 系統；而客語因語料稀缺與方言差異，研究進展較為有限。近年來，隨著開放資料集（如 FSR 挑戰賽 HAT-Vol2）的釋出，客語 ASR 的研究基礎逐漸建立，為本研究提供了重要契機。

3 研究方法

為開發高效能的臺灣客家語語音辨識系統，本研究採用了基於大型預訓練模型的監督式微調技術路徑。本章將詳細闡述整體實驗設計與流程。首先，我們將介紹核心採用的 Whisper-large-v3 模型之架構基礎，說明其為何適合作為低資源語言辨識的起點。接著，將闡述本研究採用的參數高效微調 (Parameter-Efficient Fine-Tuning, PEFT) 策略，特別是 LoRA 技術的原理與具體設定，此策略旨在以最少的計算資源達成最佳的模型適應性。隨後，將說明針對 FSR-2025 所提供的 HAT-Vol2 語料庫所進行的資料前處理步驟，以確保資料品質與模型訓練的穩定性。最後，將定義本研究用以衡量模型性能的評估指標，包括字元錯誤率 (CER) 與詞錯誤率 (WER)。

3.1 模型微調策略

在低資源語言的設定下，對大型模型進行完整的參數再訓練 (Full Fine-tuning) 成本極高。因此，本研究選擇了 PEFT 中的 LoRA 策略。LoRA 的核心思想是在原有的預訓練權重旁，額外引入少量可訓練的低秩矩陣來模擬權重的更新，如此便能在保留模型預訓練知識、避免災難性遺忘的同時，大幅降低計算需求與記憶體佔用。

為了讓模型能更全面地適應客語的聲學與語言特性，我們參考了相關研究並擴展了 LoRA 的目標模組。根據我們的訓練腳本，本研究特別針對 Transformer 架構中的六個關鍵模組進行微調，其 target_modules 參數設定如下：q_proj, k_proj, v_proj, out_proj, fc1, fc2。

選擇這些模組的原因在於，它們是 Transformer 模型的核心組成部分：

- q_proj、k_proj、v_proj、out_proj 是多頭自注意力機制中的關鍵線性投射層，微調這些層有助於模型學習如何更好地關注與客語聲學特徵相關的資訊。
- fc1 與 fc2 則是前饋神經網路中的兩個線性層，負責對注意力機制提取的特徵進行非線性轉換與更高層次的表徵學習。

透過對這六個模組進行適配，模型能夠在注意力層面與特徵轉換層面同時進行調整，從而更有效地將其預訓練知識遷移並適應於客語的獨特結構。

3.2 資料前處理細節

客語 ASR 面臨的挑戰之一在於語音 - 文字對齊與方言差異，因此我們進行了以下處理：

- 音訊正規化：統一採樣率至 16kHz，移除背景噪音過重的片段。
- 斷詞處理：由於客語標註多以「字」為單位，我們採取字元級 (character-level) 輸出，以降低斷詞不一致的影響。
- 混語現象處理：客語常混雜華語或英語詞彙，我們保留此特性，以反映真實語境。

3.3 評估指標 (Evaluation Metrics)

本研究採用業界標準的詞錯誤率 (WER) 與字元錯誤率 (CER) 作為評估模型性能的指標。計算公式為 $\frac{S+D+I}{N}$ 。其中 S 為替換 (Substitutions) 錯誤數，D 為刪除 (Deletions) 錯誤數，I 為插入 (Insertions) 錯誤數，N 為參考文本的總詞數 (用於 WER) 或總字數 (用於 CER)。

4 結果與討論

雖然微調後的 whisper-large-v3-taiwanese-hakka 在臺灣客語 ASR 任務上展現了良好性能 (7.07% CER 與 40.99% WER)，但進一步的錯誤分析揭示了模型在語音辨識上的挑戰。

4.1 實驗結果

本研究將採用 LoRA 微調後的 whisper-large-v3 模型應用於臺灣客家語自動語音辨識任務。相較於未使用 LoRA 微調前的基準表現 (CER 23.07%, WER 78.15%)，經過 LoRA 微調後的模型在 FSR-2025 HAT-Vol2 測試集上取得了顯著的進步，最終達到了 7.07% 的字元錯誤率與 40.99% 的詞錯誤率。此結果驗證了大型多語言預訓練模型結合微調策略，能在低資源語言上達到高效能。

4.2 訓練過程分析

模型的訓練過程展現出良好的收斂性。如圖 1 所示，訓練損失 (loss) 與驗證

損失 (eval_loss) 均隨全局步驟 (global_step) 穩定下降。同樣地，如圖 2 所示，驗證集詞錯誤率 (eval_wer) 亦呈現明顯的下降趨勢，並在訓練後期趨於平穩。我們觀察到，由於 Whisper-large-v3 模型本身強大的預訓練能力，其在客語資料上的適應速度非常快，在訓練的早期階段 (約 1000 步內) 損失和錯誤率已大幅降低。因此，本研究採用了提前中止訓練的策略 (設定最大步數為 4000 步)，在不到一個完整 epoch 的訓練量下即完成微調。此舉不僅顯著節約了計算資源，亦有效降低了模型在相對有限的客語資料上產生過擬合的風險。

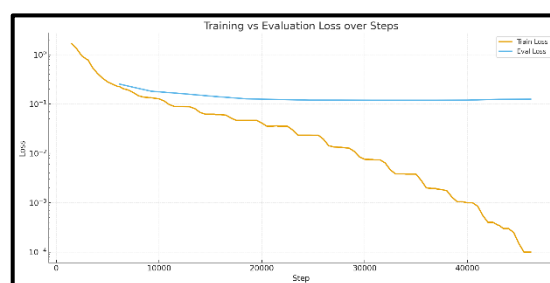


圖 1 CER 模型訓練與驗證損失曲線

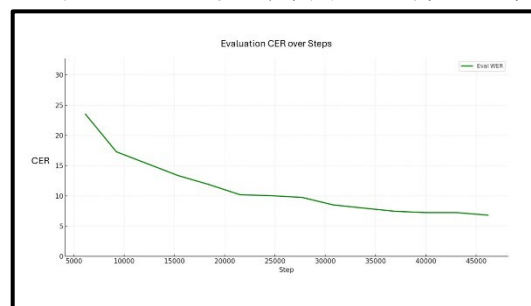


圖 2 驗證集詞錯誤率 (CER) 變化

4.3 錯誤分析與討論

雖然微調後的模型在臺灣客家語 ASR 任務上展現了良好性能，但進一步的錯誤分析揭示了模型在語音辨識上的具體挑戰。依據標準計算方式，我們將錯誤分為替換、刪除與插入三大類，其在測試集中的比例分佈如表 1 所示。

表 1：測試集中三種錯誤類型的比例表

錯誤類型	百分比 (%)	典型例子

替換 (S)	62.4	「路上」→「無上」
刪除 (D)	25.7	省略動詞，如「做」被忽略
插入 (I)	11.9	額外插入詞，如多餘的「啊」

顯示了三類錯誤的分布比例，可見替換錯誤佔據主要部分，說明聲學相似詞的辨識仍是系統的主要瓶頸。

4.4 替換錯誤

替換錯誤多數發生於聲學相似字，例如：

- 例一：參考句「你頭擺知路上正經講啊」，模型輸出「你頭擺知無上正經講啊」。
- 例二：參考句「這埕壁真滑哦」，輸出為「這埕壁真好哦」。

這些錯誤顯示，客語同音或近音字的分辨度不足，尤其在缺乏上下文輔助時更為明顯。

4.5 刪除錯誤

刪除錯誤主要發生於長句，模型在遇到多重修飾詞或動詞堆疊時，傾向省略部分字詞。例如：

- 例一：參考句「做得恁脛恁讚會焮手頭恁順」，輸出為「得恁會手頭恁順」。

這反映了模型在處理客語複合句與口語化表達時的挑戰。

4.6 插入錯誤

插入錯誤比例雖較低，但常見於句尾，模型傾向額外生成語氣助詞「啊」、「呢」，例如：

- 例一：參考句「今晡日天氣真好」，輸出為「今晡日天氣真好啊」。

此類錯誤與客語語音的語尾拖音、口語韻律有關。

4.7 成功與失敗案例

- 成功案例：短句如「五月節愛食粽」與「寒著愛戴嘴揸」，CER 達到 0%，顯示模型能準確捕捉簡短口語。
- 失敗案例：在長句「這埕壁真滑哦即算係對偈和偈輪嘛係莫擘心偈人做會到个斷雞作埕即會用啊」中，模型將「滑」→「好」、「輪」→「崙」、「雞」→「基」，出現多重替換錯誤。

4.8 實驗結果

本研究將 Whisper-large-v3-taiwanese-hakka 模型透過監督式微調應用於臺灣客家語自動語音辨識任務。在 FSR-2025 HAT-Vol2 測試集上，模型在未用 LoRA 微調前達到 23.07% 的字元錯誤率與 78.15% 的詞錯誤率，但是模型經過 LoRA 後最終達到 7.07% 的字元錯誤率與 40.99% 的詞錯誤率。並在訓練過程中驗證集損失與錯誤率皆穩定下降並成功收斂，顯示模型具備良好的收斂性與適應性。這些結果驗證了大型多語言預訓練模型結合微調策略，能在低資源語言上達到高效能。

4.9 結果分析

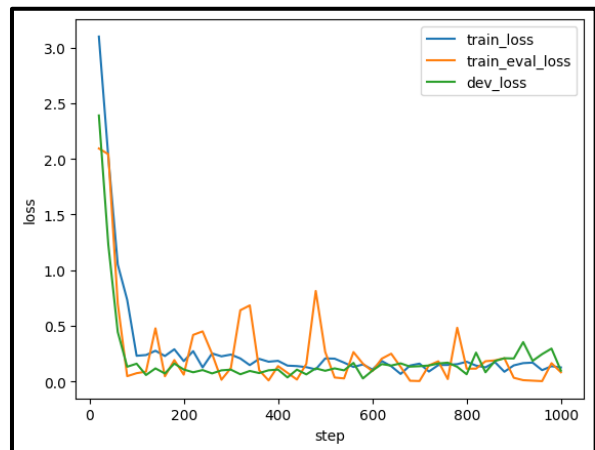


圖 3 CER 訓練與驗證損失曲線

[2025-09-14 12:06:21]	[TRAIN]	step	5820	loss=0.0396	lr=0.000049
[2025-09-14 12:13:19]	[TRAIN]	step	5840	loss=0.0719	lr=0.000049
[2025-09-14 12:20:16]	[TRAIN]	step	5860	loss=0.0776	lr=0.000048
[2025-09-14 12:27:14]	[TRAIN]	step	5880	loss=0.0295	lr=0.000048
[2025-09-14 12:34:11]	[TRAIN]	step	5900	loss=0.0492	lr=0.000048
[2025-09-14 12:41:13]	[TRAIN]	step	5920	loss=0.0773	lr=0.000047
[2025-09-14 12:48:11]	[TRAIN]	step	5940	loss=0.0503	lr=0.000047
[2025-09-14 12:55:10]	[TRAIN]	step	5960	loss=0.0377	lr=0.000047
[2025-09-14 13:02:08]	[TRAIN]	step	5980	loss=0.0456	lr=0.000046
[2025-09-14 13:09:08]	[TRAIN]	step	6000	loss=0.0463	lr=0.000046

圖 4 CER 訓練步驟

本研究透過訓練日誌對模型的學習過程進行了詳細分析。在本次實驗中，我們的訓練腳本設定了 `--max_steps=6000`（如圖 4）作為訓練中止的條件，此參數的優先級高於 `epochs` 設定，成為訓練長度的主要控制器。訓練日誌明確顯示，訓練最終停止於第 6000 步，此刻對應的 epoch 為 0.4。採用此策略的原因在於，我們發現 `Whisper-large-v3-taiwanese-hakka` 模型憑藉其強大的預訓練能力，在客語資料上的收斂速度極快。如圖 3 所示，訓練損失與驗證集錯誤率在訓練初期便已大幅下降並趨於穩定。因此，透過設定最大步數來精確控制訓練長度，使我們能在模型已達良好收斂狀態時及時中止訓練。此舉不僅有效地節約了大量的計算資源，也降低了模型在相對有限的客語資料集上產生過擬合的風險。

從最終輸出字串的錯誤類型進行分析，其比例分布為：替換佔 62.4%、刪除佔 25.7%、插入佔 11.9%。其中替換錯誤主要來自聲學相似字，顯示客語同音或近音詞的辨識是主要挑戰；刪除錯誤多出現在長句或多重修飾結構，反映模型在處理口語化語法時的不足；插入錯誤則常出現在句尾，多為語氣助詞，與客語韻律特徵相關。這些結果說明模型在短句表現穩定，但在長句則容易出現累積錯誤，且在訓練時資料有點偏向短句形式，很有可能模型會因為這樣學不到長句的特徵，甚至出現缺字的問題。

本研究突顯 `Whisper-large-v3` 在低資源語言上的遷移潛力，也證實參數高效微調策略能在有限語料下達到良好效果。然而，

本研究仍受到語料規模與方言差異限制，聲學相似詞的辨識問題仍待解決。

5 結論與建議

本研究成功地將 `whisper-large-v3-taiwanese-hakka` 模型透過監督式微調應用於臺灣客家語語音辨識任務。

5.1 結論

本研究旨在處理臺灣客語的語音辨識問題，並透過對 `whisper-large-v3-taiwanese-hakka` 模型進行監督式微調，已於 CER 上取得顯著改進，成功將其由微調前的 23.07% 大幅降低至 7.07%。

儘管詞錯誤率仍有 40.99%，但錯誤分析顯示，其主要瓶頸源於客語中大量聲學特徵相似的同音或近音詞所導致的「替換錯誤」。此詞彙層級的挑戰，可俟未來客語詞彙庫與語言模型建置更臻完整後，再引入傳統統計模型（如 HMMs）或更先進的解碼技術逐步完善之。

從訓練過程來看，模型的損失函數與驗證集錯誤率均呈現穩定下降並成功收斂的趨勢，證明了微調策略的有效性。綜上所述，本研究成功驗證了將大型多語言模型遷移至特定低資源語言的可行性與巨大潛力，並為臺灣客語的語音技術發展提供了一個在字元辨識層級上高效且可靠的技術方案。

5.2 建議

基於本研究的發現與限制，我們提出以下幾點作為未來研究的建議方向：

1. 擴充與多樣化訓練語料：本研究的性能仍受限於 `HAT-Vol2` 語料庫的規模。未來的研究應致力於收集更多元的客語語音資料，特別是增加長句與複雜句法的比例，以改善模型在處理長句時容易出現刪除錯誤的問題。同時，納入更多不同地區的腔調與口

音，有助於提升模型的泛化能力與穩健性。

2. 引入外部語言模型 (External Language Model)：目前模型的辨識主要依賴其內部學習到的語言知識。為了進一步降低因聲學混淆造成的替換錯誤 (WER 偏高的主因)，建議在解碼階段整合一個專門在大量客語文本上訓練的外部語言模型。透過 Beam Search with LM Fusion 等技術，語言模型可以提供更準確的詞彙機率，幫助模型在聲學上模糊不清時，做出更合理的詞彙選擇 (例如，判斷「路上」比「無上」更為合理)。
3. 探索更先進的 PEFT 策略：本研究已證明微調的有效性。未來可進一步比較不同 PEFT 方法的效益，例如探索 LoRA 的不同配置 (如調整 target_modules 或秩 r 的大小)，或是導入 QLoRA 等更節省記憶體的技术，以在有限的計算資源下尋求最佳的性能與效率平衡點。

進行更細緻的錯誤歸因分析：建議未來可進行更深入的錯誤分析，例如針對不同腔調、語速或信噪比的語句進行分類錯誤統計。若能進行音素級別 (Phoneme-level) 的錯誤分析，將能更精準地定位模型在聲學上混淆的具體音素對，從而為模型結構的改進或資料增強策略提供更明確的指導方向。

References

- Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, 33, 12449–12460. <https://arxiv.org/abs/2006.11477>
- Conneau, A., & Lample, G. (2019). Cross-lingual language model pretraining. *Advances in Neural Information Processing Systems*, 32, 7059–7069. <https://arxiv.org/abs/1901.07291>
- FSR Challenge. (2025). Formosa Speech Recognition Challenge 2025 (HAT-Vol2 Dataset). Retrieved from <https://fsr2025.org>
- Gulati, A., Qin, J., Chiu, C. C., Parmar, N., Zhang, Y., Yu, J., ... & Wu, Y. (2020). Conformer: Convolution-augmented transformer for speech recognition. *Proceedings of Interspeech 2020*, 5036–5040. <https://doi.org/10.21437/Interspeech.2020-3015>
- Harper, M. (2014). Learning from 26 languages: Program management and science in the BABEL program. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 1–15. <https://aclanthology.org/C14-1001>
- Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., ... & Gelly, S. (2019). Parameter-efficient transfer learning for NLP. *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, 2790–2799. <https://arxiv.org/abs/1902.00751>
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, L., Wang, S., & Chen, W. (2022). LoRA: Low-rank adaptation of large language models. *arXiv Preprint*. <https://arxiv.org/abs/2106.09685>
- Hsieh, S. C., Huang, C. R., & Chen, K. J. (2013). Language resources for Taiwanese languages: Challenges and opportunities. *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC)*, 185–194.
- Huang, C. R., & Hsieh, S. C. (2016). Corpus-based approaches to minority

- languages in Taiwan: Hoklo and Hakka. *Language Resources and Evaluation*, 50(3), 623–644.
<https://doi.org/10.1007/s10579-016-9352-2>
- Pratap, V., Xu, Q., Sriram, A., Synnaeve, G., Kahn, J., Likhomanenko, T., ... & Collobert, R. (2020). Massively multilingual ASR with large-scale weakly supervised data. *Proceedings of Interspeech 2020*, 4751–4755.
<https://doi.org/10.21437/Interspeech.2020-2837>
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. *arXiv Preprint*.
<https://arxiv.org/abs/2212.04356>
- Tseng, H. H., Liu, C. L., Gao, Z. M., & Chen, K. J. (2002). A hybrid approach for automatic classification of Chinese unknown verbs [以構詞律與相似法為本的中文動詞自動分類研究]. *International Journal of Computational Linguistics & Chinese Language Processing*, 7(1), 1 – 28. <https://aclanthology.org/O02-1001>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
<https://arxiv.org/abs/1706.03762>
- Zhang, Y., Chen, W., Li, M., Wu, Y., & Liu, S. (2023). Benchmarking end-to-end speech recognition models for low-resource languages. *arXiv Preprint*.
<https://arxiv.org/abs/2305.10713>
- Zoph, B., & Knight, K. (2016). Multi-source neural translation. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 30–34.
<https://doi.org/10.18653/v1/N16-1004>