

(意傳科技有限公司) 推行 文化部推行語言多樣性友善環境補助作業 實施計畫書

一、計畫名稱：

臺灣嬌聲 2.0

二、目的：

近年來華語逐漸成為強勢語言，也使得華語語音服務普及化，像是火車、客運等運輸業者的廣播、公共場所的提醒聲、或是大型活動的展場資訊，如此巨量的服務需求也加速成熟了華語語音合成技術。但相較於華語，本土母語的語音服務以及相關語音技術開發則寥寥無幾，然而各家單位要招募或是任聘短期的專業語音服務人才並不容易，此時語音合成就是個簡便省力的方式，藉由電腦程式將輸入的文字自動轉換並輸出語音檔，提高產出的效率。同時也希望由臺語語音合成拋磚引玉，提昇公共場域或交通運輸上的臺語語音普及度。

臺灣嬌聲 2.0 是臺灣嬌聲的優化版，推廣臺語、幫助語言學習的開源臺語語音合成計劃。電腦語音合成是由程式將文字變成聲音，比起傳統錄音方式更為輕鬆省力，使語音應用也更為彈性。根據國家語言發展法草案第十條說明，政府機關應提供「播音」等公共服務，像是各類機關的導覽、服務說明，就能用語音合成，讓每個人的母語都能在公共場所出現，推廣母語更有可行性。

語言合成也能用在教材、有聲書等語言輔助學習，儘管目前市面上已有許多文字教材，但少有新穎詞彙的參考資料，此時語音合成便可輔助開發設計數位多媒體教材，更加豐富多媒體情境的學習需求。目前臺灣嬌聲已經用在許多開源的公益專案，像是 iTaigi、台語兒、鬥拍字、寫啥物、傷倚矣。臺灣嬌聲當初因為僅一人開發，在苦無相關人力資源的協助下，未能完善全部功能，以致於影響語音的效果。

於此，為了增強社會上的母語語音能聽度，以及豐富資訊界對臺語語音合成的開源技術，希望藉由這次的機會推出更完善的臺灣聲 2.0，讓臺語語音合成的聲音能更正確、更自然，增加臺語在公共場所的能見度，達到語言平等的目標並回饋社會大眾。

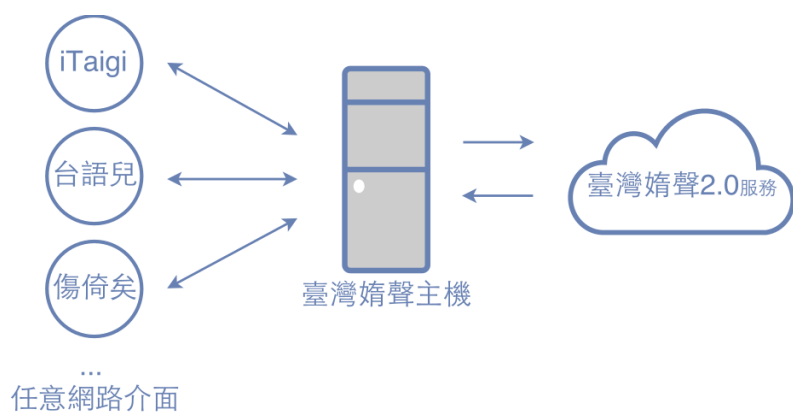
三、辦理期程：

[illegible]

實作口語調標記人機介面												1	五(一)7
人工標注口語調標記												2	五(一)8
實作變調系統												2	五(一)9
詩文讀寫工作坊切音、標注羅馬字												1	五(二)1
設計錄音文稿												3	五(二)2
錄語料												3	五(二)3
語料口語調標注												2	五(二)4
訓練語音合成模型												2	五(三)1
系統整合												1	五(三)2
相關資料公開發表												2	五(四)

四、計畫內容：

臺灣嬌聲是推廣台語、幫助語言學習的開源台語語音合成系統，將文字變成聲音，比起傳統錄音方式更為輕鬆省力，使語音應用也更為彈性。語音合成系統也能結合網路服務，提供各機關團體使用。本計畫推出臺灣嬌聲 2.0，提昇整體語音服務品質。



圖一、提供網路服務，供各專案以及各種機關服務導覽使用

目前臺灣嬌聲已經用在許多開源的公益專案，像是 iTaigi (圖二)、傷倚矣 (圖三)、台語兒、鬥拍字、寫啥物。其中 iTaigi 是月瀏覽量高達 11 萬以上的線上大眾臺語辭典，利用臺灣嬌聲取得即時的語音合成資料，不論詞彙、例句新舊都能播放出來，同時利於一般民衆透過聲音學習臺語讀寫。

寶可夢

台語怎麼講

哇！原來「寶可夢」是這樣子講的！快分享給你的朋友知道吧 [分享 32](#)

橐袋仔怪獸

▶ ☰

lak-tê-á-kuài-siū
出處：沒有人

華語：寶可夢

35

7

♥ 按呢講好

🔍 按呢怪怪

寶可夢

▶ ☰

Pô-khó-bōng
出處：沒有人

華語：寶可夢

26

5

♥ 按呢講好

🔍 按呢怪怪

抱起毛

▶ ☰

phô-khí-moo
出處：Ôo Siok-tsing

華語：寶可夢

27

9

♥ 按呢講好

🔍 按呢怪怪

抱去摸

▶ ☰

phô khi bong
出處：沒有人

華語：

神奇寶貝 寶可夢 Pokémon GO

14

7

♥ 按呢講好

🔍 按呢怪怪

抱去捐

▶ ☰

phôo khi mooh
出處：沒有人

華語：寶可夢

14

12

♥ 按呢講好

🔍 按呢怪怪

神奇寶貝

▶ ☰

Sîn-kî-pó-puè
出處：沒有人

華語：寶可夢

3

6

♥ 按呢講好

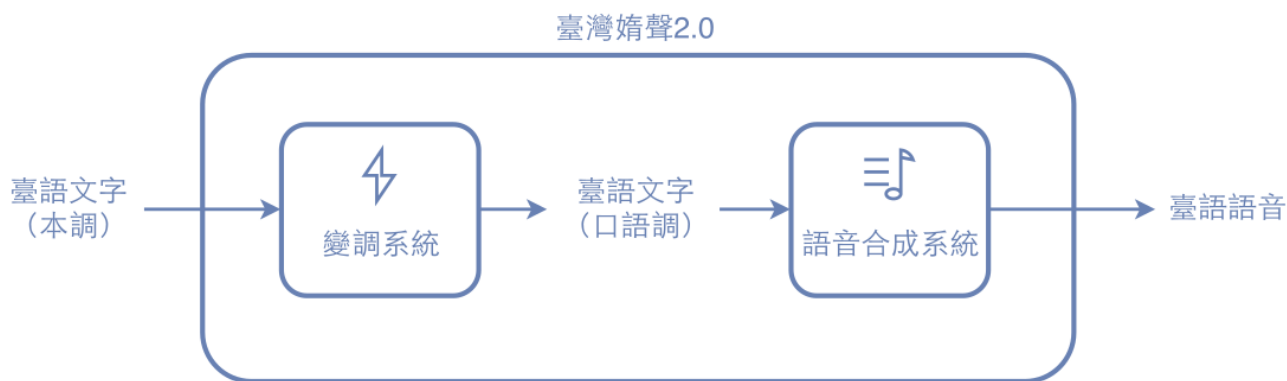
🔍 按呢怪怪

圖二、iTaigi 自動產生語音



圖三、傷倚矣做影片

一套完整的臺語語音合成服務，需要充分善加利用臺語的語言特性，因此臺灣嬌聲原團隊針對前版缺漏的臺語語音特性，撰寫更完善的變調系統和語音合成系統，整合出優化版的臺灣嬌聲 2.0（圖四）。



圖四、臺灣嬌聲 2.0 運作流程

我們將會：

1. 變調規則加強
2. 語音合成系統補足必要語音
3. 模型訓練整合
4. 相關資料公開發表

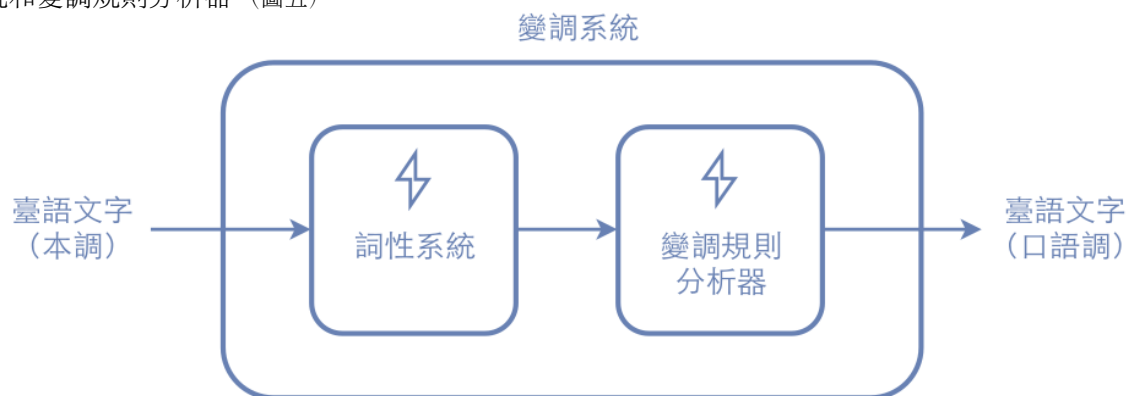
完成後，除了將臺灣嬌聲 2.0 提供給既已使用臺灣嬌聲的公益專案以外，計劃中所撰寫的程式皆以 CPAL 或 MIT 公開授權，語料及數學模型以 CC BY-SA 公開。

五、 實施方法：

(一)、變調規則加強

電腦能夠唸台語之前，得讓電腦先學會台語文怎麼唸。台語文的唸法是指平時說話發出的口語調，而非書寫習慣的本調。以「火」舉例，火車、煙火的「火」是一樣的書寫拼音，但是用唸的就可發現聲調不同。說話時，每個字根據其所在位置而適時改變聲調，形成一套巧妙的變調規則。

臺灣嬌聲 2.0 計劃以楊允言老師「台語口語調系統實作研究」為基礎，實作出一套完整且穩定的台語語句的變調系統。該研究提出 20 條口語調規則，其中有 6 條牽涉到詞性，因此變調系統有兩個關鍵：詞性系統和變調規則分析器（圖五）。



圖五、變調系統運作流程

1. 準備詞性標注語料

從團隊現有的開放語料庫當中，尋取 10,000 句文字語料作為詞性語料庫的基礎。

2. 實作輔助人工初步詞性對應系統

在沒有現成的臺語詞性標注工具之下，將參考句型相似且發展成熟的華語詞性工具，作為臺語詞性的標注參考，將第一步準備好的臺語翻譯成華語，再標注華語的詞性，最後再用華語詞的詞性對應回去臺語詞。

3. 實作輔助人工詞性標注人機介面

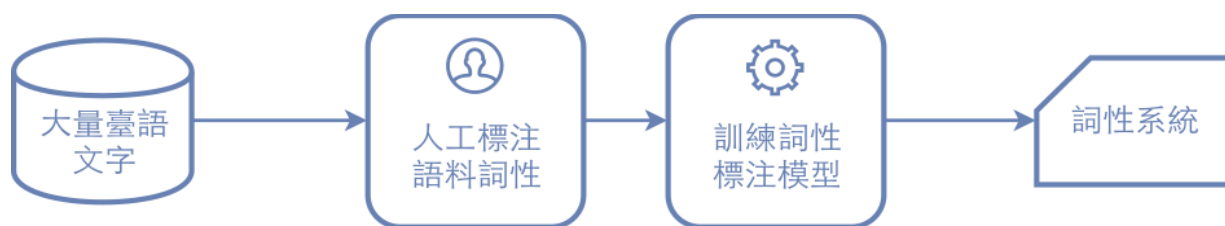
台語詞性需要人工標注過，所以需要設計一個人機介面，方便檢查。因為臺語有漢字、羅馬字等兩種主流書寫法，若要標注詞性，並不適合以一般文書軟體編輯。一來程式難以解析，二來有錯誤無法即時偵測。為了能更有效率校對臺語詞性，需要撰寫一個網站介面，迅速完成目標。

4. 標注語料詞性

聘請臺語文專業人士，使用第 3 步的人機介面，配合第 2 步的輔助工具，逐詞編纂第 1 步語料庫 10,000 句的詞性。詞性將參考採用 CKIP 的分類原理，然而 CKIP 詞性是以華語為目標設計的，並不一定完全適合臺語，因此必須調整詞性分類，建構出一套適合臺語的詞性分類。

5. 整合詞性系統

在完成詞性標注後，我們可以用詞性語料庫，經過機器學習的技術，訓練出一個臺語詞性系統(圖六)。



圖六、詞性系統取得流程

6. 準備口語調語料

從團隊現有的開放語料庫當中，尋取 2,000 句文字語料作為口語調語料庫的基礎。一般語料只有漢字和本調臺羅，並未有口語調語料。需要尋找包含口語調資料的語料庫，挑選適合的 2,000 句臺語語料，並準備用其口語調，自動判斷每一字的變調類型。

變調類型將採用「臺語口語調系統實作研究」的分類標準，分別為規則變調、本調、隨前變調、輕聲、再變調、仔前變調、三連音變調、升調等八種。

7. 實作口語調標記人機介面

第 6 步的變調類型需要人工標注，確保資料的正確性。而臺語有漢字、羅馬字等兩種主流書寫法，若要標注變調類型，並不適合以一般文書軟體編輯。一來程式難以解析，二來有打字錯誤無法即時偵測。為了能更有效率標注口語調，需要撰寫一個網站介面，迅速完成目標。

8. 人工標注口語調標記

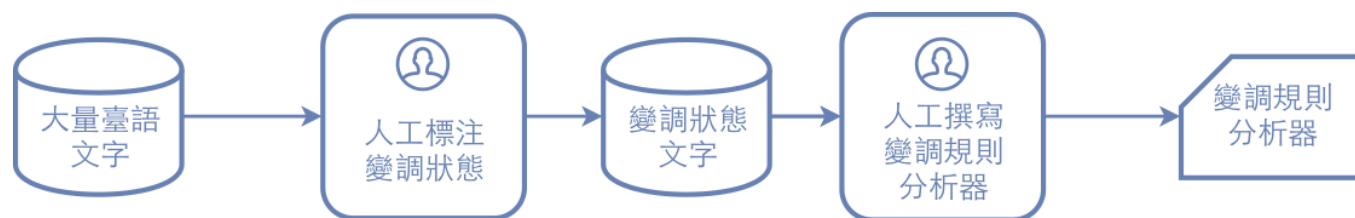
變調規則非常複雜，無法隨時用人工檢查系統錯誤。口語調標記的目的是用來自動追蹤變調系統的完成度，因此正確度相當重要。而口語調語料的自動判斷變調，仍然需要人工標注，確保語料正確性。

我們將用第 7 步的人機介面，標注第 6 步準備的 2,000 句台語語句。口語調的標記是可以重覆利用的資產，只要有標記，就可以實作出效果一樣的程式。因此口語調的標記一旦完成，往後若有別的計劃，這個標記就是大家的基石。

9. 實作變調規則分析器

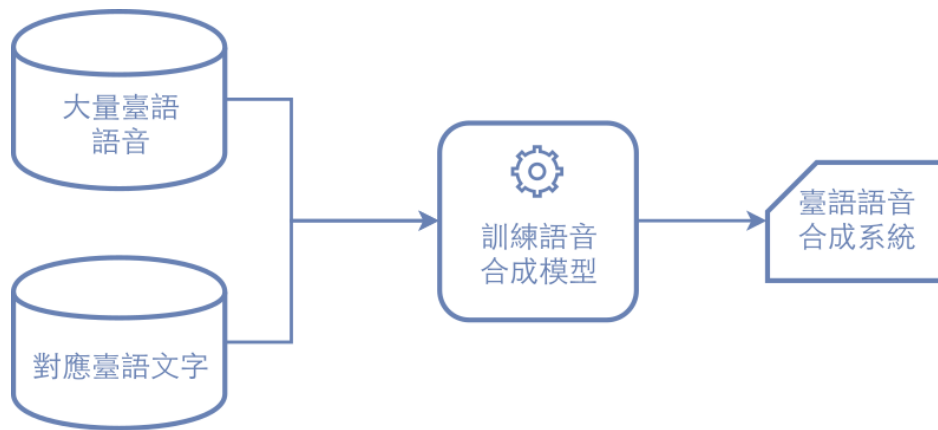
在臺語口語調系統實作研究中，標注 1,576 個音節約 158 句口語調標記，而觀察出 20 條變調規則，正確率可高達 89%。臺灣嬌聲 2.0 完成 20 條變調規則（註 1）後，我們將利用第 8 步標注的 2,000 句語口語調標記，進行觀察，會再針對不足的地方加強變調規則分析器，接近可商用的水準。（圖七）

註 1. 在原本臺灣嬌聲中，實作了臺語口語調系統實作 20 條中的 10 條（第 1、2、3、11、12、13、14、17、18、20 條），其中臺灣嬌聲 2.0 要實作的第 4、5、6、7、8、9 條與詞性相關，第 10、15、16、19 條則是各別詞的例外處理。



圖七、變調規則分析器取得流程

（二）、語音合成系統補足必要語音



圖八、語音合成系統取得流程

語音合成（圖八）需要同一個發音人的聲音。目前團隊擁有的開放語料當中，質量最佳的莫過於王秀容老師的教育部閩南語辭典 18726 詞，以及在新北市一零四學年度九百例句工作坊裡 150 句錄音。因此將以擴增王秀容老師（教育部閩南語辭典發音人）的語音為目標，蒐集王秀容老師的新北市一零三年詩文讀寫工作坊的 5 篇錄音，並額外錄製 2 小時的語音。

1. 詩文讀寫工作坊切音、標注羅馬字

新北市一零三年詩文讀寫工作坊裡有 5 篇王秀容老師的詩文錄音，3 篇童謠的錄音。將選用無背景音樂的 5 篇詩文錄音，總計約略 28 分半。由於訓練語音合成的語音需要以句為單位，會將原本整篇的音檔切做約略 540 句的音檔。而詩文語料有提供臺語漢字，仍然需要標注羅馬字，供程式訓練數學模型。

2. 設計錄音文稿

針對目前語音合成的結果進行分析，找出目前的弱點例如較罕見的第九調聲調，長句語音，並設計相對應的文稿，讓發音人補足對應的語料。若以錄製 2 小時的新語料為目標，以一句 10 字 3 秒計算，需要準備 2400 句語句。

3. 錄語料

邀請王秀容老師前來錄音，場地以大學語言相關系所錄音室為優先以減少花費，並聘請音控人員協助錄音。語音錄製後，請專業切音人士剪輯成一句一句的音檔。

4. 語料口語調標注

由撰稿者人工逐句標注變調標記，確保語音合成的語音品質。

因些需要將全部語料的變調標記標注過一次，包含教育部辭典 18726 詞、例句工作坊 150 句、詩文讀寫 5 篇、新錄音 2400 句等語料。以詞性標注、口語調標記兩個人工介面來進行，維持工作效率，並確保資料格式完整。

(三)、模型訓練整合

目前使用 HTS (HMM-based speech synthesis system) 語音合成工具。原理是將聲音分析成頻譜參數，分門別類到口語標記 p, ph, b, ... 各種音素上。在使用時，再將頻譜參數轉換成語音訊號。

1. 訓練語音合成模型

將語料轉換成 HTS 格式，並經由統計、機器學習等過程，取得語音合成的數學模型。彙整教育部辭典、例句工作坊、詩文讀寫、新錄制的語料並統一資料格式之後，再利用原先臺灣嬌聲的程式，轉換成 HTS 格式，並經由統計、機器學習等過程，取得語音合成的數學模型。

2. 系統整合

整合原先臺灣嬌聲的程式，變調系統以及新版的語音合成數學模型到網路主機上，並維持穩定運作。為了減輕網路主機成本以及讓服務更穩定，將選擇租賃網路虛擬主機。

最後，上線檢驗並測試在 iTaigi 和其他網站上能正常運作，確保服務穩定性。本計劃也會完全開放此語音合成的網路服務，提供給有需要的專案。

(四)、相關資料公開發表

在網路上公開發表程式、語料，以及妥善整理過後的詞類分類原則、詞類標記語句、口語調標注和實作方法。這些都是可重複利用的成果，也期盼做為他人日後技術發展的參考。

六、經費概算：(說明計畫總經費、自籌經費、申請補助經費)

以下以句為單位，一句以 10 字計算

項目	單價	數量說明	數量	金額	說明
工程師工作費	300	每月 172 小時*11 月	1892	567,600	工作項目：五(一) 1-3,5-7,9、五(二) 1、五(三) 1-2、五(四)。

項目	單價	數量說明	數量	金額	說明
詞性標注	16.3	10000 句	10000	163,000	工作項目：五(一)4。單價參考「中央政府各機關學校稿費支給基準數額表」特別稿件撰稿費用
變調標記標注	10	2000 句	2000	20,000	工作項目：五(一)8。單價參考「中央政府各機關學校稿費支給基準數額表」一般稿件撰稿費用
詩文讀寫切音、 標注羅馬字	10	540 句	540	5,400	工作項目：五(二)1。單價參考「中央政府各機關學校稿費支給基準數額表」一般稿件撰稿費用
設計錄音稿	10	2400 句	2400	24,000	工作項目：五(二)2。單價參考「中央政府各機關學校稿費支給基準數額表」一般稿件撰稿費用
發音人錄音費	2000	4 工作天	4	8,000	工作項目：五(二)3。不含事先準備
錄音室場地費	2000	4 工作天	4	8,000	工作項目：五(二)3。優先找大學語言相關系所錄音室，酌估場地費
音控費	2000	4 工作天	4	8,000	工作項目：五(二)3。單價參考「中央政府各機關學校稿費支給基準數額表」一般稿件撰稿費用
音檔剪輯費	2000	4 工作天	4	8,000	工作項目：五(二)3。每個音檔需後製處理，分句剪輯。
語料口語調標注-新語料	10	2400 句	2400	24,000	工作項目：五(二)4。單價參考「中央政府各機關學校稿費支給基準數額表」一般稿件撰稿費用
語料口語調標注-教育部辭典	1	18726 詞	18726	18,726	工作項目：五(二)4。單價參考「中央政府各機關學校稿費支給基準數額表」一般稿件撰稿費用
語料口語調標注-例句工作坊	10	150 句	150	1,500	工作項目：五(二)4。單價參考「中央政府各機關學校稿費支給基準數額表」一般稿件撰稿費用
語料口語調標注-詩文讀寫	10	540 句	540	5,400	工作項目：五(二)4。單價參考「中央政府各機關學校稿費支給基準數額表」一般稿件撰稿費用
虛擬主機費	630	11 個月	11	6,930	工作項目：五(三)2。1inode 虛擬主機，每月 20 美金。
總計				868,556	

- 總經費：868,556
- 自籌經費：568,556
- 申請補助：300,000

除了工程師工作費為經常性支出外，補助經費將用於：詞性標注、變調規則標注、詩文讀寫切音標注羅馬字、設計錄音稿、發音人錄音費、錄音室場地費、音控費、音檔剪輯費、語料口語調標注、虛擬主機費。

七、人力編制：（說明人力配置運用方式）

工作項目說明參考第五節實施方法的内容，不在此贅述。

人力	工作項目說明
工程師	五(一) 1-3,5-7,9、五(二) 1、五(三) 1-2、五(四)
標注稿費	五(一) 4,8、五(二) 1, 2, 4
發音人錄音費	五(二) 3
音檔剪輯	五(二) 3

八、預期效益：（對語言多樣性友善環境推動之正面影響）

	臺灣嬌聲	工研院 文字轉語音 Web 服務	Cyberon TTS 賽微文字轉語音 Web 服務	臺灣嬌聲 2.0 (本計劃預定目標)
支援輸入教育部臺語漢字	0	部份支援(註 1)	X	0
支援輸入教育部羅馬字	0	X	部份支援(註 2)	0
支援輸入白話字	0	0	X	0
支援輸入漢羅	0	0	X	0
支援輸入連字符	0	X	0	0
支援空白斷開羅馬字	0	0	X	0
華語詞翻譯	X (註 3)	0	0	X (註 3)
變調規則實作數量	10 (註 4)	5 (註 5)	2 (註 6)	20 (註 7)
支援第九調高升調	X	X	X	0
開源程式	0	X	X	0
開放語料	0	X	X	0
開放模型	0	X	X	0

附註

1. 支援：佇、佗。無支援：佻。
2. 不支援正式版，只支援數字版。
3. 考慮語言學習用途，避免臺語華語互相影響，華語翻譯以另一計劃處理。
4. 以臺語口語調系統實作研究的 20 條來看，第 1、2、3、11、12、13、14、17、18、20 條。
5. 以臺語口語調系統實作研究的 20 條來看，第 1、2、3、11、14 條規則。
6. 以臺語口語調系統實作研究的 20 條來看，第 1、2 條規則。
7. 將實作的第 4、5、6、7、8、9 條與詞性相關，第 10、15、16、19 條則是各別詞的例外處理。

（一）、對社會之貢獻

所有目前合作的公益專案，都將能取得優化後的語音合成資料。iTaigi 截止 2017/10/29 總共收錄了 60996 詞（去掉同音詞）及 13836 條例句，其中高達 96.45% 的詞彙及 100% 的例句皆使用臺灣嬌聲網路服務，將獲得更優良清晰的合成語音品質。每月至少 11 萬瀏覽量將因為臺灣嬌聲 2.0 的推出而受惠。

公共播音服務只需要藉由網路連線至臺灣嬌聲 2.0，就能快速取得任意合成語音。紐西蘭的毛利語復興成果舉世聞名，背後成功的關鍵在於翻轉毛利語在毛利人自己心中的地位，甚至改變各界對毛利語的想法、肯定毛利語。當臺語使用者在公共場合廣泛地聽見自己的母語，「原來我的母語可以在公共場所被聽見」、「原來我們這麼潮，電腦也會說母語」，社會肯定的力量更是一種心靈支持。

（二）、對研究之貢獻

本計劃所產生的語料，皆以 CC BY-SA 的方式公開授權。因此錄音檔、標注詞性及變調語料，皆能供他人重覆利用。不要花費重覆的資源在同一件事上，讓資源能更有效率的利用。CC BY-SA 授權項目條列如下：

1. 詞性標注結果 10000 句
2. 變調標注 2000 句
3. 教育部辭典 18726 詞變調規則
4. 例句工作坊 150 句變調規則
5. 詩文讀寫 5 篇切音結果
6. 詩文讀寫 5 篇羅馬字
7. 詩文讀寫 5 篇變調規則
8. 新錄音 2400 句語音
9. 新錄音 2400 句錄音文稿
10. 新錄音 2400 句變調規則
11. 詞性標注模型
12. 語音合成模型

（三）、對資訊社群之貢獻

近年來開源程式的風氣逐漸茁壯，為了響應開源程式，本計劃所實作的程式，皆以 CPAL 或 MIT 授權，促進開源社群風氣，讓資訊技術進步更加快速。授權項目條列如下（註 1）：

1. 初步詞性對應系統
2. 詞性標注人機介面
3. 詞性系統
4. 口語調標記人機介面
5. 語音合成服務

註1. 以本團隊撰寫的部份為主。若有整合到第三方工具或函式庫，該部份以其原始授權為主。

(四)、對其他母語之貢獻

1. 客語

客語和臺語同樣是漢語語系的一支，若往後要精進客語的語音合成，可以直接共享本計劃的成果。因此，標注詞性及口語調的工作流程是相同的，像是工作內容中的實作初步詞性對應程式、實作詞性標注人機介面、實作詞性系統、實作口語調標記人機介面、實作變調系統，皆是可以直接拿來使用的成果。

2. 族語

台灣有豐富的民族與語言，這些文化也極需要資訊技術的幫忙。不同族語的重音規則也不同，有些重音會因為詞性不同，而出現在不同的音節。因此詞性的初步標注、和介面仍然是一個可以再利用的資源。

九、檢附文件：(其他有關本計畫之補充說明文件)

(一)、臺灣聲源開源程式碼

https://github.com/sih4sing5hong5/tai5-uan5_gian5-gi2_hok8-bu7/wiki

(二)、本團隊語言合成服務

1. 鬥拍字-臺語語音合成: <https://鬥拍字.意傳.台灣/>
2. 搵手-客語語音合成: <https://搵手.意傳.台灣/>
3. lama1-阿美語語音合成: <https://lama1.意傳.台灣/>
4. puniq-泰雅語語音合成: <https://puniq.意傳.台灣/>
5. sapuz-布農語語音合成: <https://sapuz.意傳.台灣/>
6. hapoy-賽夏語語音合成: <https://hapoy.意傳.台灣/>

(三)、使用本團隊語音合成服務的專案:

1. iTaigi: <https://itaigi.tw>
2. 台語兒: <https://baconbao.github.io/TaiwaneseWizard/>
3. 阿美語萌典: <https://amis.moedict.tw/>