

# 新聞學外語單字系統

## 前言-校學士結合領域介紹

國際與社會科學學院-應用華語文學系

文學院-韓國學與語文學分學程

通識教育中心-教育大數據微學程

選擇以「應用華語文學系」、「韓國學與語文學分學程」及「教育大數據微學程」作為主軸進行學習。三者間雖屬不同領域，但從語言教育、數據分析與跨文化溝通的整合應用中可發現，可以展現出一定的連結性與互補性，而同時能夠發展我未來的職涯乃至於學術發展目標。

- 應用華語文學系：**以語言學作為核心修習目的，像是學習華語語法、修辭學、語音學等課程，可以從結構層面深入理解華語或是其他語言特性。同時也在分析文本與語料時，具備更扎實的理論基礎，能進一步以語言學視角進行精確的語言分析與判讀。
- 韓國學與語文學分學程：**進一步拓展了語言比較與跨文化溝通能力。如觀光韓語、職場韓語等應用課程，方可實際感受韓語在實際語境中的語用特色與文化背景，也能從對比語言學的角度，探討韓語與華語在語法、語序，甚至是敬語系統上的差異，方可助於了解韓語母語者學習中文時可能面臨的語言遷移與學習困難。
- 教育大數據微學程：**科技的輔助方可有效的將學習效率提升，因此為我提供了能將語言行為進行量化與分析的能力。透過資料處理如學習分析工具實務應用、大數據程式設計等課程訓練，學會如何運用AI語料庫與自然語言處理技術，並進一步來分析語言學習者在語言使用上的錯誤模式與學習軌跡。最後進一步結合大數據分析找出其常見偏誤或是學習的盲點，最後找出符合其語言背景與文化脈絡的學習支援策略。



## 1 動機-為何選擇真實新聞語料？

- 學習理論：**Krashen的輸入假說可以知道學習者需要接觸到略高於其當前語言水平的可理解輸入  $i+1$ (該專題就是讓學習者自行尋找自己的 $+1$ )，因此想尋找比較適合學生學習語言的媒介。同時參考維高斯基的近側發展區理論，提供給學習者的「暫時性、可調整的」協助，幫助學生跨越獨立學習的門檻。期望將這個鷹架變成從單一語言教師擴展為「語言教師+AI」的混成體系。
- 縮短母語者差距：**傳統的教材常側重翻譯或死記硬背，可能在建立外語直覺會需要花費一段時間。因此藉由引入即時外語新聞，除了讓學習者在閱讀時事中同步解析陌生詞彙，提升國際媒體理解力。
- 遊戲化動機：**純粹的閱讀可能容易產生疲乏。系統將收藏的單字無縫銜接至「配對遊戲」、「打字練習」與「聽力測驗」等等，可透過即時練習，將認知的負荷轉化為持續投入遊戲的體驗。

感謝專題指導老師：應用華語文教學系 邱詩雯 副教授、通識教育中心兼任教授 陳俊豪 助理教授

感謝自主學習：自主探究授課老師：國文學系 曾曄傑 副教授

參考文獻：

Krashen, S. (1985) *The Input Hypothesis: Issues and Implications*. Longman, London.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

## 2 系統處理形式與架構流程

### 1 Input

新聞純文本

目前支援中文或韓文的新聞 URL 與純文本輸入，作為分析的源頭。

### 3 Graph

D3.js 力導向渲染

計算引力與排斥力，依據 TOCFL 難度動態調整節點大小，視覺化呈現。

### 2 Mapping

解析關聯圖譜

透過 AI 解析詞彙原型、詞性與語法結構，並自動對應中文/英文翻譯及例句。

### 4 Gamification

串接小遊戲

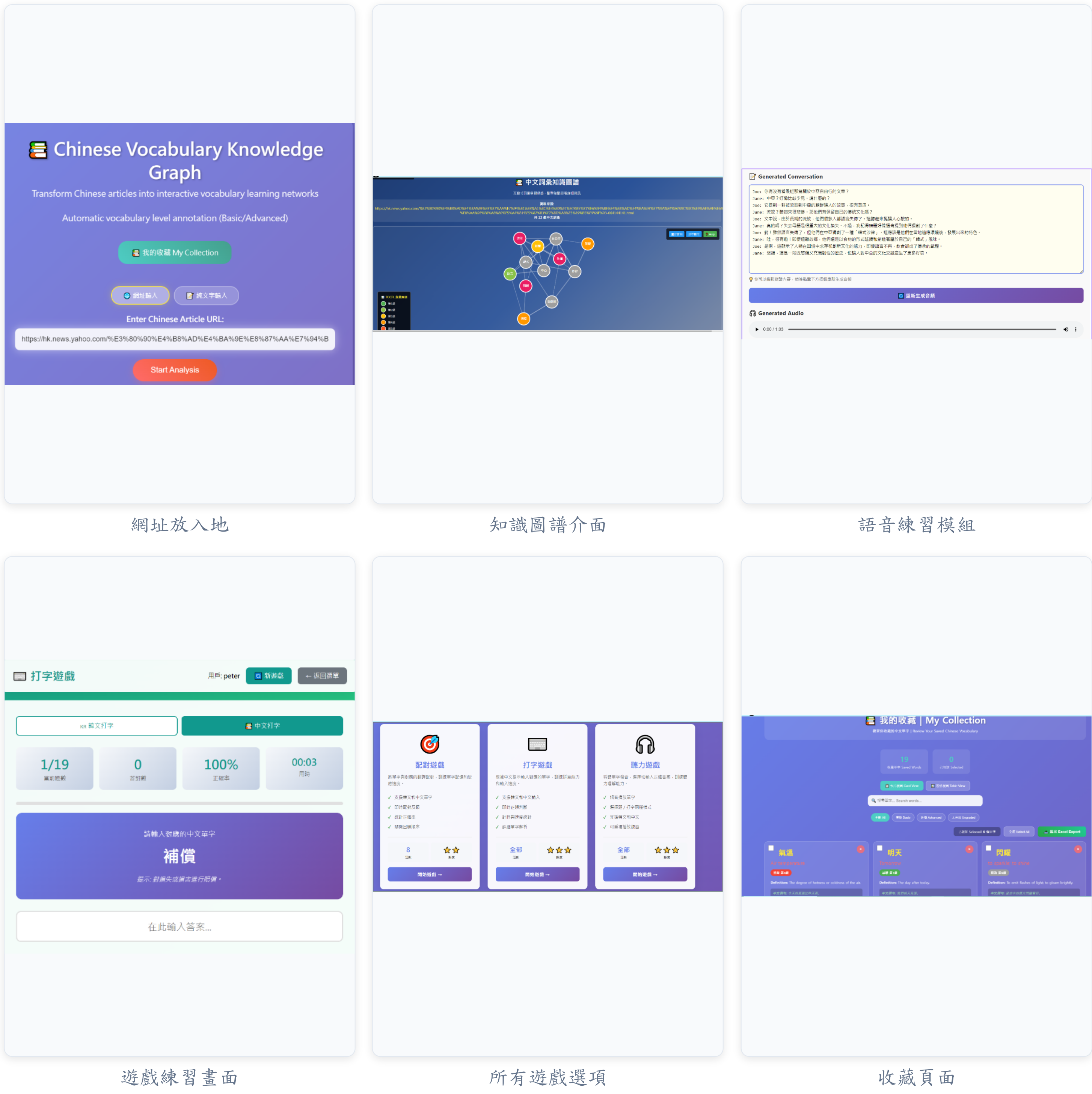
雙擊節點收藏單字後，資料同步至 Supabase，用於測驗與語音對話生成。

知識圖譜視覺化

動態 TTS 語音生成

遊戲化記憶機制

TOCFL 分級整合



## 3 如何呈現

- 工程效能：**內建國教院 14,452 字的 TOCFL 官方標準詞彙庫。為解決龐大文本比對的延遲，底層建構 Hash Table 結構，確保單字比對與遊戲配對的即時性。
- 穩定輸出：**在調用 API 時，定義 JSON Schema (response\_schema)，並透過設定 nodes 與 links 為 Required Array，並設定 Persona 與 Constraints (禁用生僻字等等)，來解決 LLM 輸出格式不穩定的問題。
- 雲端同步：**採用 Supabase (PostgreSQL)，即時同步使用者的學習紀錄與 TOCFL 等級，搭配 Vercel 與 Railway 雙軌部署，實現零停機更新。