「西英漢三語平行語料庫」(Corpus Paralelo de Español, Inglés y Chino, CPEIC)

「西英漢三語平行語料庫」(CPEIC, 2007-2017 年）透過不同語言分析主題的計畫申請，在逐年累積語料量、增加語料類型和改善檢索功能的模式下持續語料庫的建置工作，歷經由書面語到口語、由雙語到三語、由單機初階到網路進階等方面，進行功能與技術性之提升。建構程序主要包括 1.語料收集與彙整及 2.系統開發兩部分。

1. 語料收集與彙整分四部分：聖經語料（書面語）、童話故事語料（書面語）、聯合國大會語料（書面語和口語）和電影字幕語料（口語）。藉由不同來源、主題與文體來豐富語料的多樣性。

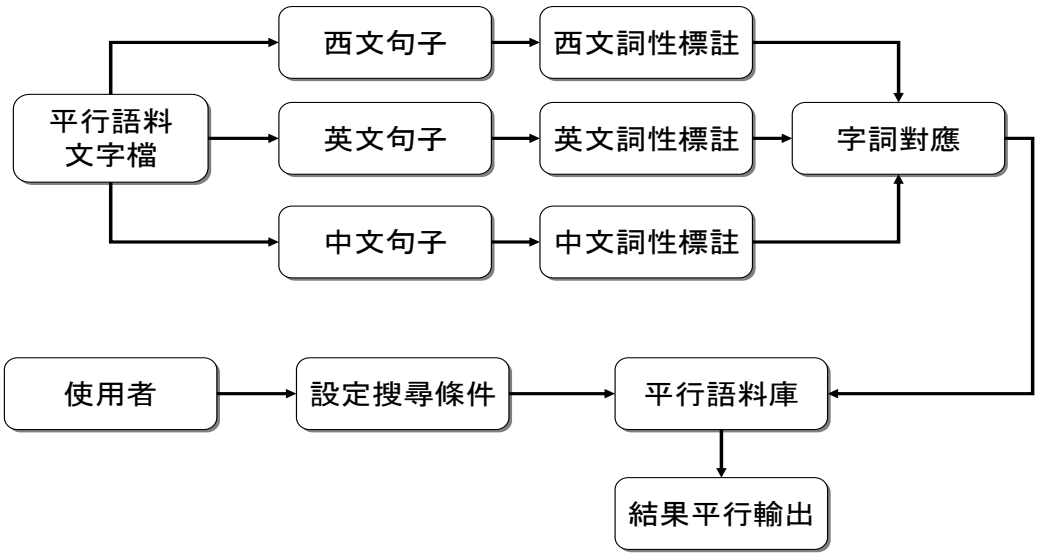
2. 系統開發分成以下四部份。

(1) 詞類註記：西、英文採用TreeTagger，中文則是採用中研院的CKIP進行詞類的標註。

(2) 三語對應：透過Giza++進行三語字詞的平行對應。

(3) 資料處理：處理過的資料被匯入資料庫(MySQL)中，可方便資料的管理與備份，而後續的使用者查詢也可以透過我們所開發的程式，產生SQL語法快速的抓取所要資料並呈現給使用者。

(4) 使用者介面開發：使用者所查詢的介面是由HTML搭配JQuery和Perl做整體的開發。在使用者介面和檢索結果的呈現上，採用網頁方式[1]，如下圖所示；語料庫建構的相關訊息請參見下表。



「西英漢三語平行語料庫」建構流程圖

[1] 公開檢索的連結網址為 http://corpora.flld.ncku.edu.tw/，點選「西語」(Spanish)之「西英漢三語平行語料庫」(CPEIC)。

「西英漢三語平行語料庫」(CPEIC) 之建構

語料庫[2]	西英漢三語平行語料庫 (2007-2017)
來源	聖經[3]、童話故事[4]、聯合國大會文件[5]和電影字幕[6]
文字搜尋	√
複合條件搜尋	√
詞類搜尋	√
文字對應	√
可搜尋語言數	3
三語言字數量	西 1,245,004 字、英 1,213，840 字和中 1,564,945 字

[2] 協助之老師與助理：盧文祥、鄭安中、林紀蓮老師，簡君聿、沈敬濠、邱國豪、唐子軒、周洵、陳麒琛、呂昭儀、蔡昆育、王廷軒、張耀升、張展毓等語料庫技術助理和洪聖允、朱玉馨、歐宜亭、黃湘云、黃靖珈、宋采育、葉孟欣、林佳錤、陳亭潔、吳柏姍、倪晨綾等語料處理助理。

[3] https://www.biblegateway.com/。包含完整的新約和舊約共 66 卷。

[4] http://en.childrenslibrary.org/, http://itunes.apple.com/hk/app/id440153337?mt=8。含「三隻小豬、三隻熊」等童話故事共 13 則。

[5] http://www.un.org/zh/ga/documents/index.shtml。含口語「會議逐字記錄」和書面語第六十五屆「大會決議」文件，包括行政和預算、法律、社會人道主義和文化、特別政治和非殖民化、經濟和金融等主題。

[6] 電影包含「Hable con ella/Talk to her/悄悄告訴她、Truman/Truman/特魯曼和 Volver/Volver/玩美女人」。西語來源依序爲 http://www.opensubtitles.org/es/subtitles/74969/talk-to-her-es，
https://www.opensubtitles.org/en/subtitles/6516419/truman-es，
https://www.opensubtitles.org/zh/subtitles/5868290/volver-es。英語來源依序爲:
http://www.opensubtitles.org/es/subtitles/3106260/talk-to-her-en， http://subhd.com/ar0/338333，
http://subsmax.com/subtitles-movie/volver-2006/ni。中文來源依序爲:http://subhd.com/a/31244，
http://subhd.com/ar0/339222， https://www.opensubtitles.org/zh/subtitles/5960301/volver-zt。