

中華民國計算語言學學會

平衡語料庫 申請說明

中央研究院漢語平衡語料庫(簡稱Sinica Corpus)第3.0版，為一包含五百萬目詞的帶標記平衡語料庫。本語料庫中每個文句都依詞斷開，並標示詞類標記。語料的蒐集也盡量做到平衡分配在不同的主題和語式上，是現代漢語無窮多的語句中一個代表性的樣本。語料庫得中央研究院及中華民國行政院國家科學委員會補助，由中央研究院中文詞知識庫小組執行、研究，並授權中華民國計算語言學學會發行。申請人需向學會提出申請，簽妥授權使用協議書，並同意確實遵守協議書上之約定條款。

申請文件：

由服務機關開立身份證明書,內文需言明本資料庫僅用於個人研究，絕不用於商業用途。

備妥一式兩份使用協議書（本會與申請人各執乙份）。

請備妥上列文件寄至：

中華民國計算語言學學會
台北市南港區研究院路二段128號
中央研究院資訊所（轉）

付款方式：

劃撥:帳號：19166251，戶名：中華民國計算語言學學會

支票：請加註「中華民國計算語言學學會」抬頭

信用卡：請填寫信用卡資料

工本費用：

國內個人：

會員 NT 1,000

非會員加收處理費NT 500 NT 1,500

國外個人：

會員 US\$ 50.

非會員加收處理費 US\$ 50 US\$ 100.

國內團體 (1-10 人 使用)

會員 NT 10,000

非會員加收處理費NT 1,000 NT 11,000

國外團體 (1-10 人使用)

會員 US \$ 500.

非會員加收處理費 US \$ 100. US \$ 600.

中華民國計算語言學學會 平衡語料庫授權使用協議書

立協議書人（以下簡稱本人）現經中華民國計算語言學學會授權，取得中央研究院漢語平衡語料庫之使用權。本人同意切實遵守下列條款：

- 一. 平衡語料庫之電子型式，組成內容型式及詞類與文類標記，著作權屬中央研究院詞庫小組。智慧財產權屬中央研究院及國科會共有，並授權中華民國計算語言學學會進行非營利性研究授權。
- 二. 平衡語料庫文件內容之著作權屬原始著作人所有。
- 三. 本協議書係針對使用上述第一項之著作內容進行研究之授權。
- 四. 牽涉上述第一項及第二項著作內容之任何引用之商業行為應與著作權所有人另定約規定之。
- 五. 本語料庫內之部份或全部內容，及其直接衍生資料（即詞彙庫，字/詞頻統計，詞類統計）不得包含/引用於任何商業產品中。
- 六. 本人不得將此語料庫之全部或部份移轉給第三人。
- 七. 本人應用此語料庫從事學術研究而獲致成果時，應於本人所發表的論文或其他公開文件中申明誌謝 (1)中華民國計算語言學會，(2) 中央研究院詞庫小組，及 (3) 中華民國國家科學委員會。
- 八. 本人如對此語料庫內容或附屬程式有所改進或擴充，願意無條件回饋計算語言學學會，以共同促進中文計算語言學之發展。
- 九. 本人如有侵犯此語料庫著作權或智慧財產權之行為，應自行負擔法律責任。

立協議書人： (蓋章)

服務單位： (蓋章)

職稱：

E - m a i l：

中華民國 年 月 日

審核單位： (由本會填寫)

中華民國計算語言學學會 平衡語料庫授權使用協議書

立協議書人（以下簡稱本人）現經中華民國計算語言學學會授權，取得中央研究院漢語平衡語料庫之使用權。本人同意切實遵守下列條款：

- 一. 平衡語料庫之電子型式，組成內容型式及詞類與文類標記，著作權屬中央研究院詞庫小組。智慧財產權屬中央研究院及國科會共有，並授權中華民國計算語言學學會進行非營利性研究授權。
- 二. 平衡語料庫文件內容之著作權屬原始著作人所有。
- 三. 本協議書係針對使用上述第一項之著作內容進行研究之授權。
- 四. 牽涉上述第一項及第二項著作內容之任何引用之商業行為應與著作權所有人另定約規定之。
- 五. 本語料庫內之部份或全部內容，及其直接衍生資料（即詞彙庫，字/詞頻統計，詞類統計）不得包含/引用於任何商業產品中。
- 六. 本人不得將此語料庫之全部或部份移轉給第三人。
- 七. 本人應用此語料庫從事學術研究而獲致成果時，應於本人所發表的論文或其他公開文件中申明誌謝 (1)中華民國計算語言學會，(2) 中央研究院詞庫小組，及 (3) 中華民國國家科學委員會。
- 八. 本人如對此語料庫內容或附屬程式有所改進或擴充，願意無條件回饋計算語言學學會，以共同促進中文計算語言學之發展。
- 九. 本人如有侵犯此語料庫著作權或智慧財產權之行為，應自行負擔法律責任。

立 協 議 書 人： (蓋章)

服 務 單 位： (蓋章)

職 稱：

E - m a i l：

中 華 民 國 年 月 日

審核單位： (由本會填寫)