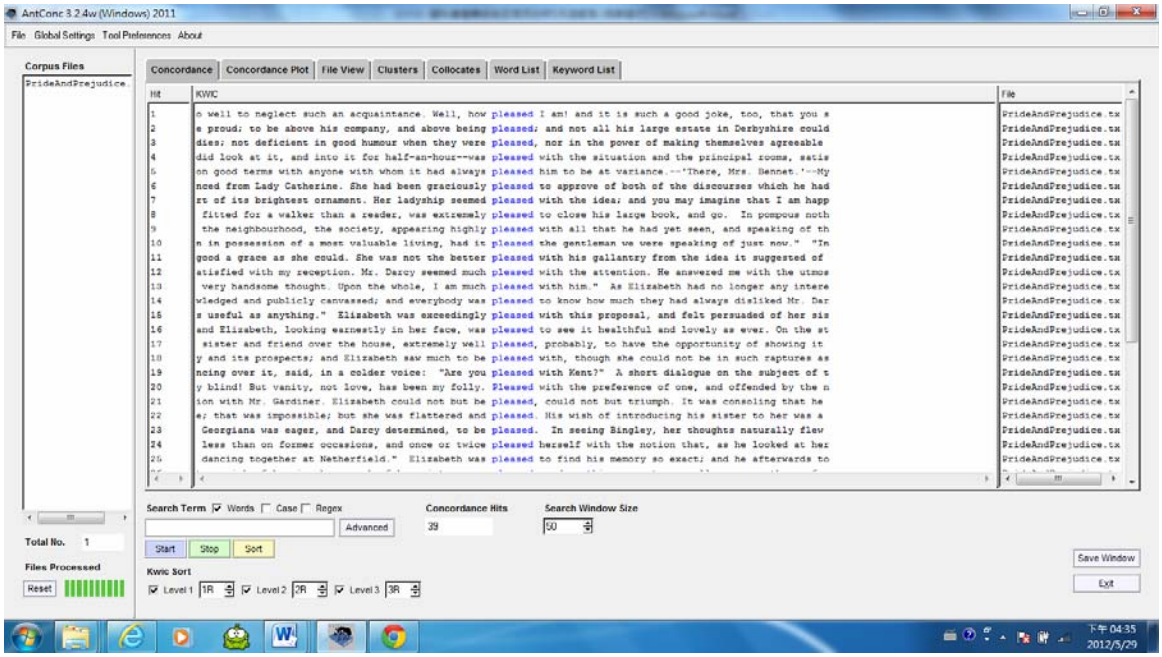
中文句法樹庫（Penn Chinese Treebank ）(http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2004T05)。兩者在語言，語料來源，語料庫大小，標記集，標記單位，標記訊息，及依據的語言學理論都不相同。

Sinica Treebank 與 Penn Chinese Treebank 最大的差別在於結構樹的語法單位不同。前者以標點符號作為分隔不同結構樹的單位，因此一個結構樹很多時候只是一個詞組（如 PP, NP）而不是一個完整的句子。而後者除小部分結構樹是句子的片段（以 FRAG 標示）大部分的結構樹是完整的句子(sentence)(以 IP 標示)。另外 Sinica Treebank 語法結構採取中心語主導原則 (Head-Driven Principle)，註明中心語(Head)和其他成分（如附加語）的語法和語意訊息，表達出句子中詞和詞之間的語法結構和語意角色關係，而 Penn Chinese Treebank 並沒有中心語與語意角色的訊息，而是在詞組上加註如主詞 SBJ 受詞 OBJ 等語法功能的方式來取代。

由於 Sinica Treebank 有未簡化標記，簡化標記及精簡標記三種標記集，相較於 Penn Treebank 只有一種標記集，Sinica Treebank 的三種不同的標記集可以作為不同的特徵。除此之外只有 Sinica Treebank 有標示語意角色的訊息，Penn Chinese Treebank。由 Linguistic Data Consortium (LDC)所發行，其中標示語意角色的 Penn Chinese Treebank 稱為 Chinese Proposition Bank。

參、語料庫語言學的工具

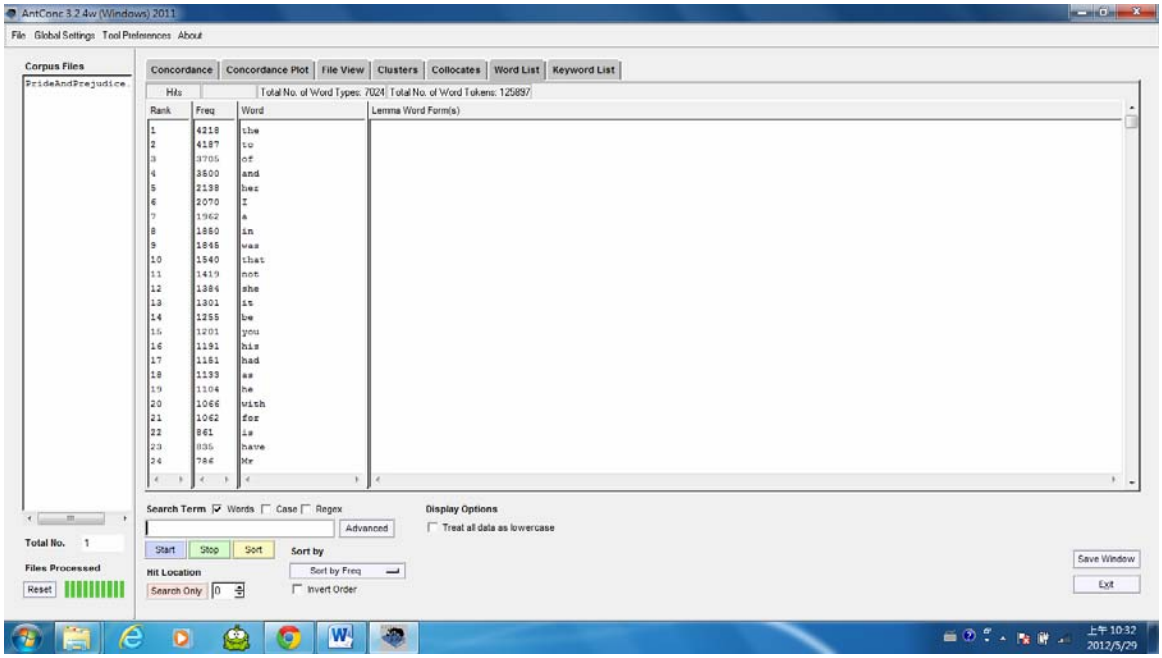
一、關鍵詞前後文程式(concordancer)：輸入一個關鍵詞或字串，程式自動將語料庫中所有包含這個詞或字串例子找出來置中並顯示前後語境。Antconc 是一個免費軟體，可以計算語料關鍵詞的頻率，並檢索關鍵詞以及搭配語。下面的畫面擷取自 Antconc 關鍵詞上下文檢索程式 concordancer 的功能。



圖一 AntConc 關鍵詞前後文排序程式

http://www.antlab.sci.waseda.ac.jp/antconc_index.html

二、詞頻程式：計算某個特定的字串或每個出現在語料庫中的詞的頻率。如上面 Antconc 內建 concordancer 功能，搜尋某一個關鍵詞時，下方 Concordance hits 會顯示這個關鍵詞在這個語料庫出現幾筆。如下圖，點選 Antconc 上方 Wordlist 即可計算每個出現在語料庫中的詞的頻率，且會依照頻率高低排序。



圖二 AntConc 詞頻排序程式