

## 從語料庫統計的觀點分析 美國之音新聞英文被動句中文譯文

洪千惠

譯者英翻中時，常常會不自覺受到英文語法的影響，譯出的文字往往反應英文語法、不符合中文的表達習慣，成為西化中文。本文以「被動句型」為例，以科學的方法，來討論美國之音新聞英文「被動句」的中文譯文。本論文使用語料庫研究法，透過中文非譯文語料庫裡的例句，有系統地分析「被動句」中文譯文。中文非譯文語料來自英國「Sketch Engine (SKE)」繁體中文語料庫，中文譯文語料來自「NTU Chinese-English Parallel Corpora」。本文計算譯文語料在 SKE 語料庫裡的「出現頻率」及「詞彙搭配力 (t 值及 MI 值)」，並利用「文本索引」及「詞彙特性素描」功能討論詞語使用型態。本論文提出利用語料庫來檢視譯文是否為道地中文的方法，幫助譯者自我改善英到中譯文的品質。本文的另一個貢獻在於以統計量化方式闡述翻譯學裡語言規範的概念。

關鍵詞：被動句、語料庫、出現頻率、詞彙搭配力 (t 值及 MI 值)、翻譯品質、語言規範

收件：2011 年 11 月 5 日；修改：2011 年 7 月 9 日；接受：2011 年 7 月 11 日

## Corpus-Based Analysis for English-Chinese Translation of Passives in VOA News

Chien-hui Hung

When translating English into Chinese, translators are often unconsciously influenced by English syntax, which often results in the westernization of translated Chinese. This study addresses this issue by employing scientific methods to look at the English-Chinese translation of passives in Voice of America News. Based on the “Sketch Engine Traditional Chinese Corpus” and the “NTU Chinese-English Parallel Corpora,” we examine the quantitative characteristics of westernized Chinese in terms of word frequency and two statistical measures for collocations—Mutual Information and t score. Corpus tools such as “concordance” and “word sketch” in Sketch Engine were employed to observe and analyze collocation patterns in authentic Chinese. The goal is to help translators avoid using non-idiomatic and inauthentic Chinese and to advocate the use of both monolingual and bilingual corpora as important resources for improving the quality of translation, as well as the use of statistics to present the norm of the language in translation studies.

Keywords: CTS, passive construction, corpus, word frequency, collocability, norm

Received: November 5, 2010; Revised: July 9, 2011; Accepted: July 11, 2011

## 壹、前言

譯者英翻中時，常常會不自覺受到英文語法的影響，譯出的文字往往反應英文語法、不符合中文的表達習慣，成為西化中文。舉例<sup>1</sup>來說：

英文原文：He has been considered a front-runner in the contest.

中文譯文：他一直被認為在這次大選中處於領先地位。

此句中文譯文意思明白，可以讀得懂，只是這句中文譯文不是道地中文。相較於中文，被動語態廣泛出現在英語句子裡，但中文不一定要用「被」字表達。若翻譯成：「一般認為他在這次大選中處於領先地位」會較像中文。譯者完全按照英文語法，直譯這句原文。這種表達方式就是因英漢語法結構不同而造成的「西化譯法」（葉子南，2000）。而這句中文譯文充滿翻譯腔（translationese）<sup>2</sup>，可說是「歐化中文（Europeanized Chinese）」或是「英化中文（Anglicized Chinese）」（張振玉，1993）<sup>3</sup>。但並不是中文譯文有「被字句」就視之為「翻譯腔」。中文裡有被動語態，但是較少使用，按照傳統漢語語法，「被」字主要用在表達不幸的事件（黃宣範譯，2007；陳定安，1997）。英語被動的句子，譯成中文時，大部分的都可以用主動句型來表達（張培基、喻雲根、李宗杰、彭謨禹，1993）。除此之外，中文裡還有許多字可以用來代表被動，如「為」、「受」、「由」、「讓」等<sup>4</sup>。

自有翻譯活動以來，譯者應該採用何種翻譯原則、遵循何種翻譯標準，一直是翻譯界討論不休的議題。現代翻譯大師中，余光中（2002）認為意思精確，但讀起來不通順，在文法和修辭上是一種病態。他舉出「歐化分子」的三個毛病<sup>5</sup>，並強調亦步亦趨地模仿英文語法、逐字譯或死板的直譯是「假精確」。思果認為譯文的好壞可分三種：「不是中文」、「還可以讀得懂」、「像中文」（思果，2003，頁26）。而他同樣也是認定「譯文就要像中文」，他主張翻譯要譯意，中文譯文要能表達原文的意思。黃

邦傑（2006）認為譯者將英語語法結構硬搬過去，變成中文語法結構，是偷懶的行為。他非常反對「勉強看得懂但句法十分歐化的直譯」<sup>6</sup>。傅一勤（2004）對西化譯法採取比較樂觀的看法。他認為每一種語言的語法系統都具有彈性，相同的意思可以有許多不同的講法，而不同的用語造就不同的風格。他將翻譯的語言稱為「中介語言（interlanguage）」<sup>7</sup>，是譯者將一種語言翻成另一種語言時，取得兩種語言都可以接受的妥協說法。葉子南（2000）指出有些西化表達方式更能反映原語文化的特徵及原著的精神，可以幫助讀者了解原語符號體系，所以西化譯法還是有其存在的價值。王克非（2009）則表示西化譯法的產生有兩個原因：「一方面是不自覺的人用亦用，另一方面是譯者特意吸收外語的表現法，以豐富漢語的表現力」（頁 410）。

幾乎所有討論西化譯法的書都會談到「被動譯法」的問題。如余光中（2002）、思果（2001，2003）、傅一勤（2004）、葉子南（2000、2001）、潘穎薇（1997）、陳定安（1997）、劉宓慶（1997）、黃邦傑（2006）、黃宣範譯（2007）、吳潛誠（1995）及張培基、喻雲根、李宗杰、彭謨禹（1993）。對於中文譯文充斥「被動句」這現象，翻譯界的大師個個筆誅口伐。但這些學者所提出的心得看法及舉例大部分都只是翻譯多年的經驗談。鑑於強調客觀、用科學方法來研究翻譯學相關議題的趨勢正慢慢成形（Olohan, 2004）。因此本文以科學的方法——「語料庫研究法」來討論這個問題。

世界第一個語料庫為 Kucera 與 Francis 在西元 1967 年所創建的 Brown Corpus（楊惠中，2002）。那時的語料庫（corpora/corpus）指的僅只是語言資料的匯集。後來，隨著電腦科技的加入，現在所稱的語料庫有幾點特徵<sup>8</sup>：首先，這些語言資料的整合是根據某項原則或是規定，使得此資料庫具代表性；還有，這些資料通常存在電腦裡；最後，我們可以利用這資料庫作量化及質性的分析（Biber, Conrad, & Reppen, 1998; O'keeffe, McCarthy, & Carter, 2007）。而「語料庫語言學（corpus

linguistics)」所談的就是如何利用這些語料庫，輔以統計量化方式或語料庫文本索引功能，探討語言的分佈及使用型態。至於用那種語料庫研究方法較合適，或是要如何設計出可以代表自己研究目的的語料庫，文獻裡也已有諸多討論，如 Biber (1993)；Biber、Conrad 與 Reppen (1998)；McEnergy、Xiao 與 Tono (2006) 等。而英國翻譯家 Mona Baker 也在 1990 年後期結合「語料庫語言學」以及「翻譯學」成為一門新學科——「語料庫翻譯學 (CTS)」。Baker (1993, 1995, 1996, 2000) 提出用電腦語料及語言學統計方法來研究翻譯議題，包含描述翻譯文體特徵、個別譯者的翻譯風格等，將翻譯學研究帶入另一個新境界。

Xiao (2010) 用語料庫研究法比較「中文譯文」跟「中文非譯文」的被動語態 (passive constructions)。Xiao 想要了解 Baker 等人所討論的翻譯文本的普遍特徵<sup>9</sup>，是否也可以在中文譯文裡觀察到。Xiao 發現因受到英文語法的影響，與中文非譯文的文章比起來，中文譯文裡有較多的被字句。而翻譯文本的普遍特徵裡的「正規化 (normalisation/conservatism)」指的是譯者受到譯入語 (target language) 的影響，過度使用譯入語的文法結構。這很明顯的與 Xiao 的研究結果相抵觸。換句話說，以被動句而言，翻譯文本的普遍特徵並不存在英到中的翻譯文本裡。本論文也是使用「語料庫研究法」，利用語料庫裡充足的語料證據，以「被動句」為例，有系統地討論英譯中出現的西化翻譯，也就是譯者採用英語的表達方式，因語法結構不同所造成的西化。與 Xiao 不同的是，本論文採用的語料包含較多的語料，而且分別用量化及質性的方法來討論。

解志強 (2010) 則提出以網路為本的翻譯理論模式。他認為網路語料最容易取得且數量龐大，建議譯者可以藉由語料庫裡的語料，譯出符合當地語言規範的譯文。解志強利用 Google 檢索到的網頁次數與部分摘要來闡述他的論點，並示範了幾種 Google 查詢的技巧。然而，就誠如解志強文中所總結，網路語料庫有其限制，而且若真要用語料庫裡的文章，當做翻譯的參考，語料庫的選擇非常重要。以這點看來，會誇大搜尋

筆數的 Google (頁 196) 不是一個好選擇，因為其不實的數據，無法呈現客觀的量化資料。另外，Google 的查詢也有其限制及缺點 (Kilgarriff, 2007)。此外若缺乏有系統的分析及歸納，僅靠搜尋筆數的數據，譯者不容易從眾多網路語料裡找到語言規範 (norm)<sup>10</sup>。因此，本文以語料庫語言學統計的方式來呈現語言規範的量化、客觀基礎，也就是藉由語料庫語言學統計來闡釋翻譯學裡面有關語言規範的概念。

為了了解被動語態在中文譯文 (target language) 裡的出現型態，是否受到英文原文 (source language) 的影響，本文利用中文語料庫，比較中文譯文及中文非譯文語料<sup>11</sup>，檢驗中文譯文裡詞彙的搭配關係是否在中文非譯文的文章裡經常出現。如果中文非譯文語料庫裡常可以找得到這樣的詞組搭配關係，表示這樣的中文譯文應該可以為中文為母語的大眾所接受，而無歐化中文之嫌。本文不僅利用量化的數據，還輔以質性語料例子的討論，提出利用語料庫來檢視譯文是否為道地中文的方法，幫助譯者自我改善英到中譯文的品質。

## 貳、研究方法

### 一、NTU Chinese-English Parallel Corpora<sup>12</sup>

本論文所使用的中文譯文語料來自「NTU Chinese-English Parallel Corpora」，為高照明所整合的英漢平行語料庫 (Gao, 2011)。原文 (source language) 是中文的中文語料有 14085307 個字，對應的英文譯文有 7312431 個英文詞；原文是英文的語料有 7275332 個詞，對應的中文譯文有 16056353 個中文字。此語料庫涵蓋許多主題，本文所採用的「美國之音」<sup>13</sup> 為新聞體裁，英文原文語料有 80299 個詞，中文譯文則有 157410 個中文字。本文所採用的語料，英文為原文，中文為譯文。這些語料分別以英文及中文斷句呈現，方便對照查看。

## 二、英國 Sketch Engine(SKE) 語料庫<sup>14</sup>

本論文所使用的中文非譯文語料來自「Sketch Engine (SKE)」<sup>15</sup> (Kilgariff, Rychlý, Smrž, & Tugwell, 2004)。它是由英國 Lexical Computing Ltd 機構所研發而成的詞彙素描系統，可用來研究詞彙的結構及與其他詞的搭配。SKE 的語料由好幾種語言組成，如中文、英文、德文、義大利文、日文等。本文所使用的語料為有七億詞的繁體中文語料庫<sup>16</sup>。語料來源擷取自二十世紀九十年代及二十一世紀初間，台灣中央通訊社的報導文章。

## 三、西化譯文檢驗方法

為了檢驗譯文，本文擬從「詞語的出現頻率 (word frequency)」以及「詞彙的搭配 (collocation)」兩大方面，利用統計方法提出量化的證據，並利用「文本索引 (concordance)」功能討論詞語使用型態。

### (一) 統計量化方式

#### 1. 詞語的出現頻率

當我們想要了解某個詞在某個語料庫的分佈情形，我們可以看這個詞所出現的次數。但是語料庫的大小 (size) 會影響字詞出現的次數，因此比較不同語料庫的原始次數是沒有意義的。為解決這問題，本文應用三種計算頻率的方法，比較字詞在不同語料庫的分析情形及差別。

#### (1) 標準化頻率 (Normalization) (Biber, Conrad, & Reppen, 1998)

「標準化」就是將個別的原始次數除以語料庫的大小後再乘上一個基準數。基準數的數值由語料庫的大小決定。

標準化頻率 = (詞在語料庫出現的次數 ÷ 語料庫的全部總詞數) × 基準數值

例如本文想了解「被」這一個字在兩個語料庫 (中文譯文、中

文非譯文)的分佈情形。假設「被」在譯文語料庫(12,000)出現「10」次,在非譯文語料庫(18,000)出現「12」次,基準數值為「1,000,000」。其標準化頻率如下:

譯文語料庫:  $(10/12000)*1000000=830$

非譯文語料庫:  $(12/18000)*1000000=660$

由上可知,「被」在譯文語料庫裡,每一百萬個字,會出現「830」次,但在非譯文語料庫,只出現「660」次。由此數據,我們的結論為「被」較常出現在中文譯文的文章裡。若沒有考慮語料庫的大小,只比較原始次數,會得到相反的結論。

(2) 差異係數 (Difference Coefficient) (Kennedy, 1998)

我們可以利用頻率差異係數來比較某個詞在兩種語料庫的分佈及差異度。

差異係數 = (A 語料庫頻率 - B 語料庫頻率) ÷ (A 語料庫頻率 + B 語料庫頻率)

差異係數數值介於「1」與「-1」間,正值代表此詞在 A 語料庫出現較多,負值代表在 A 語料庫出現較少。數值<sup>17</sup>愈大,表示差異愈大。比「0.5」大或比「-0.5」小的數值表示兩個語料庫的頻率差異大。

(3) Log-Likelihood (LL) 值<sup>18</sup> (Biber, 1993; Rayson & Garside, 2000)

LL 值可以比較某個詞在兩個語料庫(語料庫 A 及語料庫 B)裡出現的頻率差異,是否達到統計顯著。正的 LL 值表示:與語料庫 B 比較,語料庫 A 裡這個詞的出現頻率相對較多。負的 LL 值則表示語料庫 A 裡這個詞的出現頻率相對較少。LL 值愈大(不管正或負),兩個頻率差異也愈大。LL 值  $\geq 3.8$  時即達到統計顯著水準 ( $p < 0.05$ )。LL 檢定簡單快速,網路上甚至有 LL 值計算器 (Log-Likelihood Calculator)。不過,

如果詞彙出現頻率過低，那計算出來的 LL 值將不那麼可靠 (McEnergy, Xiao, & Tono, 2006)。

## 2. 詞彙的搭配

詞彙的搭配意指「兩個或兩個以上的詞在文本中很短距離內的共同出現」<sup>19</sup> 詞跟詞共同出現的情況很多，有的時候是出於偶然，有的時候卻是常用的慣用語。例如，我們會說「風雨」「交加」，而不會說「風雷」「交加」。研究詞彙的搭配及計算其搭配力 (collocability)，可以讓我們了解兩詞共同出現是因為語言的慣用習慣，還是只是隨機出現。t 值及 MI 值可以闡述兩詞的搭配力 (Church & Hanks, 1990; Church, Gale, Hanks, & Hindle, 1991)。

以下為 MI 值及 t 值計算方式的介紹 (以「被發現」為例)：

### (1) Mutual Information (MI) 值：

$$I(x, y) = \log_2 \frac{P(x, y)}{P(x)P(y)}$$

$P(x)$ ：「被」在非譯文語料庫出現的機率

$P(y)$ ：「發現」在非譯文語料庫出現的機率

$P(x, y)$ ：「被」「發現」同時在非譯文語料庫出現的機率

$I(x, y)$ ：MI 值

MI 值所代表的是兩詞搭配程度的強弱。MI 值至少要大於零。MI 值愈大，表示「被」與「發現」兩詞的關聯，強度愈大，愈有可能是真正的搭配語。但 MI 值要多大，這兩詞才是真正的搭配語，則要視語料庫的大小來定。語料庫越大，即使兩詞一同出現的次數很少，或只是碰巧出現一次，也會使得兩詞的 MI 值遠大於零。因此，要了解兩詞的搭配力有多大，不能只看 MI 值，還要考慮語料庫大小 (McEnergy, Xiao, & Tono,

2006)。

(2) t 值 (t-scores) <sup>20</sup>

$$t \approx \frac{f(x,y) - \frac{f(x)f(y)}{N}}{\sqrt{f(x,y)}}$$

f(x)：「被」在非譯文語料庫出現的次數

f(y)：「發現」在非譯文語料庫出現的次數

f(x,y)：「被」「發現」同時在非譯文語料庫出現的次數

N：非譯文語料庫總字數。

這個檢定將語料庫大小放入考量。t 值愈大，表示兩者搭配力愈大，其共同出現的次數也愈高。當 t 值  $\geq 1.645$ ，就已經達到統計顯著水準 ( $p < 0.05$ )。

MI 值與 t 值雖然同為闡述兩詞搭配力的統計度量方法，兩者所代表的意義卻不盡相同。t 值的高低取決於兩詞共同出現的次數；t 值愈大表兩個詞共同出現的次數愈多。MI 值偵測的是兩詞搭配程度的強弱；MI 值愈高，表兩詞共同出現的機會就愈大，但是並不表示這兩詞共同出現的次數就愈高。要了解兩詞的搭配能力，要由這兩項指標共同決定，不能只看單一指標的數值。

本文研究新聞英語的中文譯文，檢驗中文譯文裡詞彙的搭配關係是否在中文非譯文的文章裡經常出現。這兩個搭配語的檢驗 (MI 值及 t 值) 可以幫助本文，藉由中文非譯文語料，釐清中文譯文裡的詞組搭配關係是否為歐化中文。

## (二) 文本索引

提高譯文品質的方法已廣受多方討論，也已有許多專家學者提出由多年的翻譯經驗歸納出的翻譯技巧或是基於不同理論檢驗得出的翻譯方法，如思果 (2001, 2003) 等人。本文提出利用語料庫改善譯文品質的方法。利用語料庫的「文本索引 (concordance)」功能，可以觀察特定

字詞在句子裡的使用方法。查詢的字詞會置於中間且用不同的顏色標示出來，易於研究左、右兩邊的前後文。除此之外，「Sketch Engine」的「詞彙特性素描 (word sketch)」功能可以查詢特定字詞當不同詞性使用的使用型態。本文利用這些功能，透過例子的討論，介紹利用語料庫來改善譯文品質的方法。

## 參、研究發現

### 一、「被動」字頻率比較

吳潛誠 (1995) 指出，除了「被」字之外，其他像是「挨、遭、遇、教、讓、給、受、獲、承、蒙」等字也可以表達被動之意。林杏光 (1990) 指出的五種現代漢語可用的被動結構中，也表示「挨、受、遭、叫、讓、給、為」等字有表達被動意義。本文作者遂調查這些字在這兩個語料庫裡出現的頻率，其統計如表 1。

表 1 「被動」字頻率比較

	美國之音 中文譯文		Sketch Engine 中文非譯文	
	157410	%	706428333	%
被	404	0.257%	859559	0.122%
受	245	0.156%	245305	0.035%
教	117	0.074%	13521	0.002%
獲	63	0.040%	163583	0.023%
承	58	0.037%	3192	0.0005%
給	55	0.035%	231614	0.033%
遭	46	0.029%	156111	0.022%
讓	38	0.024%	412343	0.058%
遇	24	0.015%	10864	0.002%
蒙	17	0.011%	6344	0.0009%
挨	2	0.001%	1295	0.0002%

表 1 分別檢驗「NTU Chinese-English Parallel Corpora」中文譯文語料庫及「Sketch Engine」繁體中文語料庫，比較數個可以表達被動之意的詞在這些語料庫裡出現的次數。語料庫名稱下方的數字代表總詞數，例如「Sketch Engine」繁體中文語料庫裡共有 7 億詞。由表 1 可得知，在眾多被動字中，「被」字在各語料庫出現的次數最多。「被」字在「美國之音」中文譯文語料庫出現「404」次，在「Sketch Engine」中文非譯文語料庫則出現了「859559」次，占語料庫的比率分別為「0.257%」及「0.122%」。

當分析其他的被動字，作者發現在「Sketch Engine」語料庫裡，「讓」、出現的頻率最高，為「412343」次；「獲」、「受」、「遭」、「給」出現的頻率也不低，分別為「163583」次、「245305」次、「156111」次及「231614」次。而在中文譯文語料庫裡，「美國之音」語料庫，「受」的出現頻率僅次於「被」，共出現「245」次，其次為「獲」、「給」、「遭」及「讓」，其出現次數分別為「63」、「55」、「46」及「38」。本文繼續比較這兩個語料庫裡出現頻率較高的幾個被動字。結果列於下表，表 2。

表 2 「美國之音」與「SKE」常出現被動字與「被」頻率比較

	美國之音		SKE	
	被	與「被」的比率	被	與「被」的比率
受	245	61%	245305	29%
獲	63	16%	163583	19%
給	55	14%	231614	27%
遭	46	11%	156111	18%
讓	38	9%	412343	48%

表 2 比較同是新聞體裁的兩個語料庫，目的在於了解是否「被」字在中文譯文語料庫裡的使用次數高於中文非譯文語料庫。第三欄及第五欄內的數字代表此被動字與「被」字出現頻率的比率。以「Sketch Engine」

語料庫出現頻率第二名的「讓」字，其出現頻率為「被」字的一半（48%），但在「美國之音」中文譯文語料庫裡，「被」字出現了「404」次，「讓」字才出現「38」次，其比率為「9%」，不到「被」字的十分之一。「遭」、「給」及「獲」情況與「讓」相同，它們與「被」在中文譯文語料庫的比率皆低於中文非譯文語料庫。「遭」的比率分別為「11%」及「18%」；「給」的比率則分別為「14%」及「27%」；「獲」的比率則分別為「16%」及「19%」。唯一不同的是「受」，此字與「被」的比率似乎在中文譯文語料庫裡較高。但當我們仔細研究此中文譯文語料庫時，我們發現「受」在非當被動之意的句子很多，光是「受傷」跟「接受」的句子就高達「141」句。若我們將這兩種句子扣掉，那「受」與「被」的比率只剩下「25%」，低於中文非譯文語料庫的「29%」。這些發現證明了中文譯文語料庫裡，「被」字明顯多用，其出現次數高於其他被動字，可能的解釋為當譯者英翻中時，一看到英文被動句（BE 動詞 + 過去分詞）時，就翻成中文的「被…」，而較少用其他的字。

## 二、中文譯文與 SKE 中文非譯文語料庫「被」字頻率比較

由表 1 可分別得知，「被」字在中文譯文及中文非譯文語料庫出現的頻率。但完全只看出現的次數，而不考慮語料庫的大小，這樣的比較結果不是很正確。表 3 考慮語料庫的大小，列出四種計算頻率的方法，分別比較「被」字在「SKE」語料庫與三個不同體裁中文譯文語料庫裡的出現次數。

表 3 「被」字頻率比較

	美國之音 中文譯文	Sketch Engine 中文非譯文
總詞數	157410	706428333
「被」字出現次數	404	859559
標準化頻率 (每一百萬詞)	2566	1217
差異係數	0.36	
Log-Likelihood (LL)	178.16	

由表 3 可得知，當我們比較同是新聞體裁的「美國之音」與「SKE」語料，「被」字在中文譯文及「SKE」語料庫的標準化頻率分別為「2566」及「1217」。這也表示在中文譯文語料庫裡，每一百萬詞，「被」出現「2566」次；而在 SKE 語料庫，每一百萬詞，「被」才出現「1217」次。這兩個數據表示，「被」字在中文譯文語料庫裡出現的頻率較 SKE 語料庫高。「被」字在譯文及 SKE 語料庫的差異係數為「0.36」。此值大於零，表示「被」字在中文譯文語料庫裡出現的次數較多。LL 值為「178.16」，表示「被」字在中文譯文語料庫裡明顯多用。此 LL 值大於「3.18」，代表「被」字在中文譯文及中文非譯文語料庫兩個頻率的差異已達統計顯著。

### 三、中文譯文語料庫用「被」字表達的動詞

我們發現，在中文譯文語料庫裡的所有被字句裡，譯者使用被動句型最多的三個動詞，分別為「打」、「殺」及「關」。其中，「被打」出現頻率最高，共出現「116」次；「被殺」出現「59」次，「被關」則出現了「50」次。但在 SKE 語料庫裡，卻是「被殺」出現頻率最高。「被殺」出現了「2879」次，「被打」出現了「1071」次，「被關」則出現了「1287」次。以「被」與這三個動詞搭配程度來看，「被」與「殺」兩詞 t 值為

「53.656」，MI 值為「8.279」；「被」與「關」兩詞 t 值為「35.873」，MI 值為「5.053」；「被」與「打」兩詞 t 值為「32.723」，MI 值為「3.510」。由於這三組詞 t 值皆大於「1.645」，這表示這三個動詞與「被」字搭配力皆已達統計顯著。至於搭配程度，「被」與「殺」兩詞的 MI 值比另兩組詞高，這表示「被」與「殺」兩詞的關聯及搭配程度較「被」與「關」及「被」與「打」來得強。「被打」、「被關」以及「被殺」都是不幸的事。之前也已提及，按照傳統漢語語法，「被」字主要表達不幸的事。但當本文作者研究中文譯文語料庫裡的句子，卻發現語料庫裡有許多動詞，並不是不好的事，卻也用被動句型表達。本文作者整理十八個在譯文語料裡與「被」一同出現的動詞，再用 SKE 語料庫檢驗這些被動動詞及中文譯文，其結果列於表 4。

表 4 中文譯文語料庫用「被」字表達的動詞在 SKE 語料庫出現頻率

	出現次數	t 值	MI 值
被「認為」	12003	109.458	3.930
被「視為」	11477	107.126	8.235
被「發現」	7710	87.771	4.478
被「殺」	2879	53.656	8.279
被「釋放」	1403	37.455	5.369
被「關」	1287	35.873	5.053
被「打」	1071	32.723	3.510
被「解除」	1037	32.201	4.573
被「感染」	877	29.612	3.834
被「證實」	630	25.098	6.693
被「任命」	403	20.074	3.640
被「證明」	357	18.893	2.295
被「選中」	279	16.703	7.099
被「隔離」	209	14.457	4.173
被「封閉」	196	14	3.576
被「注意」	39	6.244	-1.95
被「吃」	63	7.928	-0.051
被「移開」	10	3.162	5.204
被「創造」	8	2.828	-3.710

表 4 記錄出現在中文譯文語料庫裡，十八個被動動詞在「Sketch Engine」中文非譯文語料庫出現的頻率、 $t$  值及 MI 值。這十八個詞依出現次數的多寡順序排列。由於  $t$  值  $\geq 1.645$  就已達到統計顯著，由表 4 可得知，這十八個動詞與「被」字的搭配力皆已達統計顯著的程度，而這些動詞中，「認為」與「被」共同出現的次數最多， $t$  值最大。但在搭配程度方面，「被」「注意」、「被」「吃」以及「被」「創造」的 MI 值小於零，這表示「注意」、「吃」及「創造」這三詞與「被」不可能是搭配語，而其他 MI 值大於零的詞中，「殺」及「視為」與「被」的搭配程度最強。

雖然這些動詞的  $t$  值皆達統計顯著，承如傅一勤（2004，頁 136）所言，「在強大外語勢力的衝擊下，把中性動詞也用被動句表達，是很自然的事」。只是，傅一勤也表示這些「被」字句可能都不是上好的中文，有修改的空間。因此，本小節繼續討論使用這些動詞時，不要用「被」句型表達，而改用其他句型的可能性。以下為這些譯文<sup>21</sup>的分析：

#### <例一>

**英文原文：Ten people are confirmed dead.**

**中文譯文：已經有 10 人被證實死亡。**

在 Sketch Engine 語料庫裡，含有「證實死亡」的句子共有「157」句，用被動句型來表達，句子內含「被證實死亡」的句子只有「20」句，而「已證實死亡」或「已經證實死亡」的句子則有「50」句，句子如「重傷的犯人在送到山頂醫院後，已證實死亡」。其餘的句子皆用主動句型來表達，句子如「掩埋的二十人中，有一人獲救，三人證實死亡」，或是「自週四以來，已有一百六十七人證實死亡」。

利用 SKE 語料庫裡「詞彙特性素描」功能，查閱常與「證實」一同出現的搭配詞，也不見「被」字的記錄。緊鄰且出現在「證實」前、與「證實」搭配力強的修飾語，除了之前提過的「已／已經」，還有「無法」、「間接」、「也」、「首度」、「初步」、「未／尚未」、「親口」、「加以／予以／予」、「不能」、「進一步」、「無從」、「稍後」、「首次」、「再次」等字。

**<例二>**

**英文原文：He has been considered a front-runner in the contest.**

**中文譯文：他一直被認為在這次大選中處於領先地位。**

「被認為」在 SKE 語料庫出現了「12003」次，是表 4 眾多詞中出現次數最多的。但本文作者發現，在「Sketch Engine」語料庫裡，用主動方式表達「認為」這動詞的句子高達「634991」句。這表示被動句型還是占了極少數。另外，含有「一直認為」的句子有「1193」句，但語料庫裡卻沒有「一直被認為」的句子。同樣的例子還有如「More bodies are found」。若將此英文譯成「發現更多屍體」會比翻譯成「更多屍體被發現」更像中文。英文用被動句，有時候是為了不想把主詞表示出來，有表達客觀的意味。由於中文常常可以省略主詞，因此雖用主動，但可以達到英文用被動句表現客觀的效果，也能避免使用被動句型，減少英式中文的發生。

**<例三>**

**英文原文：Many have been released.**

**中文譯文：許多人已經被釋放。**

此句譯文「許多人已經被釋放」是按照英文文法逐字翻過來。在 SKE 語料庫裡，大多數的句子皆是描述甲方「釋放」乙方，如「以色列釋放阿拉伯人」。但也有用乙方為主詞的句子，如「被控涉嫌內亂的政治犯已陸續釋放」。若將此句改成「被控涉嫌內亂的政治犯已陸續被釋放」，念起來會很奇怪，不是好的中文。

在利用「詞彙特性素描」功能，查閱常與「釋放」同出現的搭配詞，並不見「被」字的記錄。在這些「釋放」的修飾詞裡，「予以」或是「予」皆有「被」的意思。雖然「予以釋放」或「予釋放」在 SKE 語料庫出現的次數並不高：「予以釋放」出現了「171」次；「予釋放」出現了「37」次。但「予以」與「釋放」的搭配強度顯著，MI 值達「6.15」；「予」與「釋放」的搭配程度也不弱，MI 值為「5.136」；相較之下，雖然「被」與

「釋放」共同出現了「1403」次，兩詞的 MI 值也才「5.369」，這表示「予」或是「予以」與「釋放」的搭配程度較「被」強。

再查閱「獲釋」在 SKE 語料庫裡的分佈情況，我們發現兩詞一共出現了「10125」次，兩詞的 MI 值也高達「14.136」。因此，若將此句翻譯成「許多人已獲釋」會較「許多人已獲釋放」更像中文說法。

<例四>

**英文原文：Elder statesman Shimon Peres was named a deputy prime minister.**

**中文譯文：政壇元老西蒙·佩雷斯被任命為副總理。**

查閱 SKE 語料庫，本文作者發現「任命」可當動詞也可當名詞使用，若當動詞使用，則大致可分為兩大類，第一類為甲方「任命」乙方的句子，如「亞賽拜然國會今天重新任命穆塔里波夫為總統」，第二類句子為本例題六，是以乙方為主詞的句子。即使如此，SKE 語料庫裡的句子，也少見用被動句型表達。句子如「葉爾勤再度任命柴諾米丁組閣」、「吳敦義是留任，並不是重新任命」等。因此，此句的譯文「被任命」的「被」是多餘的，可以省略。「被」字可以省略的例子還如下：

<例五>

**英文原文：The course was being cleared.**

**中文譯文：跑道已被封閉了。**

「Sketch Engine」中文非譯文語料庫裡，雖然也可以找得到「被封閉」或是「被移開」的句子，但為數不多。幾乎所有的句子都省略「被」字。因此「被封閉」或是「被移開」中的「被」字都是多餘的。

<例六>

**英文原文：More than 57-hundred people in 11 Indian states have been affected.**

**中文譯文：印度 11 個邦已經有 5 千 7 百多人被感染。**

譯者可考慮用中文裡其他有被動意思的字，如表 4 的「受」字。語料

庫的資料裡，「被感染」出現了「877」次，「受感染」則出現了「1052」次，其 MI 值為「5.905」，也大於「被感染」的 MI 值「3.834」。因此，若將此句譯文修改成「印度 11 個邦已經有 5 千 7 百多人受感染」，即可以免除譯文英文腔的困擾。相同的例子如下面這組翻譯：

#### <例七>

**英文原文：It was hardly noticed.**

**中文譯文：很難被注意到。**

在 SKE 語料庫裡，「被」與「注意」同時出現的次數與機率都不大。「被注意」才出現「39」次，其 MI 值也小於零。但是「受注意」出現頻率卻高達「180」次，「受」與「注意」共同出現的 MI 值比零大，為「2.064」，有可能為真正的搭配語。因此本文作者建議譯者將此句英文原文譯成「幾乎不受注意」。

#### <例八>

**英文原文：It might be eaten also.**

**中文譯文：所有蘿卜的葉子都已經被吃掉了。**

查閱 SKE 語料庫，當然可以發現「…被吃」的句子，但為數不多。計算「被吃」的 MI 值，也發現其值小於零，這表示「被」與「吃」不可能是搭配語。而中文裡卻常可見用「非受格句 (unaccusative construction)」來代替「被」字的使用，句子如「藥吃完了」、「泡麵吃光了」等。若將此句譯成「所有蘿卜的葉子都已經吃掉了」或是「所有蘿卜的葉子都已經吃光了」會較「所有蘿卜的葉子都被吃掉了」更像中文。

## 肆、結論

由於受到英文原文的影響，中文譯文並沒有達到像中文那樣流暢自然的境界。本文分析新聞英文中文譯文，透過 SKE 中文非譯文語料庫裡的例句，觀察道地中文如何表達被動語態，並用統計方法分析，提出量化的

證據，如經由計算出的 MI 值討論與「被」字共同出現字詞的搭配程度。其結論可用來當作檢測譯文是否為西化英文的準則。

本文作者發現不論是中文譯文或是中文非譯文語料庫，「被」字出現的次數遠高於其他被動字。本文在考慮語料庫大小的情況下，發現「被」字在中文譯文語料庫裡出現的頻率較 SKE 語料庫高，而且所求得之統計數據，如「標準化頻率」、「差異係數」值及「LL」等值皆達統計顯著。

「SKE 中文非譯文語料庫」裡有為數不少的「被」字句，而且也不完全僅表達不幸的事，但在中文譯文語料庫裡，「被」字句出現的頻率卻還是遠高於中文非譯文語料庫。原因可能是譯者翻譯時，受到英文文法的影響，一看到英文句裡「be 動詞+過去分詞」的句型，就翻成「被…」，而不考慮是否有其他較好的字可以代替「被」，也不考慮這樣的中文是否為道地的中文。本文的結論為：相較於中文，英文裡較常用被動句子表達；中文裡雖也有被動語態，但較少使用。當語意需要表被動語態時，除了是不幸的事情用「被」字句外，中文有以下幾種表達方式：一是用其他的字來代替「被」，不同的動詞搭配不同的字，如「不受注意」、「遭逮捕」、「獲釋」等。二是用主動的語氣，將要表達的動詞放在後面。三是翻譯時，遵循「能省則省」原則，將「被」字省略，例如省略「許多人已『被』釋放」的「被」，省略「跑道已『被』封閉的『被』」等。四則是用「非受格句 (unaccusative construction)」來代替「被」字的使用，如用「吃掉了」來代替「被吃掉」。

## 伍、研究限制與未來研究建議

基於近年來強調以客觀的方法來研究翻譯學的趨勢，本論文藉由語料庫語言學裡的統計量化方式來闡釋翻譯學裡有關語言規範的概念。本論文使用語料庫研究法，透過 SKE 中文非譯文語料庫裡的例句，觀察新聞英文「被動句」道地中文的表達方式。本文藉著比較中文譯文語料與

SKE 中文非譯文語料，利用量化研究方法，試圖找出譯文歐化的證據。

但一連串的語料分析後，我們發現有些譯文語料很明顯是西式英文，但其 MI 值卻還是大於「0」，如「被證實」、「被選中」等。這研究結果當然受到我們所使用語料庫的品質所影響。本文所使用的「Sketch Engine」繁體中文語料庫，雖有七億詞，卻有其限制。因白話文的語法彈性，容易受到西方語言的影響，因此 SKE 的中文語料難免有些呈現歐化現象，可是 SKE 繁體中文語料已是到目前為止，我們所能找到最大的中文非譯文語料庫了。

其次，SKE 收錄語料全來自台灣中央通訊社的報導。體裁雖以新聞為主，但裡頭又可細分為社論、國內新聞、國外新聞等。各類組成比例與美國之音不盡相同，但本文並沒考慮到這個因素，因此研究結果的差異有可能反應了組成語料庫比例的不同。除此之外，譯文語料庫完全是美國之音的英文譯文，若能包含更多不同來源的新聞譯文，研究結論將較能推論到所有的新聞譯文。

另外，美國之音的新聞很有可能是由大陸譯者所翻。因此，本文所發現的差異有可能是台灣及大陸兩地中文使用的差異，而不是中文譯文與中文非譯文的差異。為排除這個變數，建議未來研究考量譯者身份。

至於本文所利用的統計量化方法中，在分析「被動字」的頻率上，本文並沒有考慮斷詞，導致「受傷」或「接受」也會包含在「受」的頻率內。建議後續考量中文斷詞，在分析詞頻前，先處理斷詞問題。

而計算詞語搭配能力的  $t$  值及 MI 值，只能計算兩個緊鄰詞彙的搭配程度，如「幸運」與「多」兩個詞，而無法計算「幸運」「多」「了」三個緊鄰詞彙的搭配程度。後續研究可以考慮用其他的統計方法，如「連鎖詞串 (n-gram)」<sup>22</sup> 來研究兩個以上的詞連續一起出現的次數，討論這樣的詞串是不是中文裡的慣用語。

本研究雖有以上所提種種限制，譯者（尤其是中文非母語的譯者）還是可以參考本論文的分析方式，運用語料庫的「文本索引 (concor-

dance)」<sup>23</sup>及「詞彙特性素描 (word sketch)」<sup>24</sup>等功能，自我改善英到中的譯文品質。本文的另一個貢獻在於以統計量化方式闡述翻譯學裡語言規範 (norm) 的概念。語言是一群人在文字上的使用的共識，本文利用語料庫語言學裡的統計來解釋、歸納翻譯文本裡的語言習慣。

未來研究除了以上所述，考慮語料庫文體比例，台灣大陸中文使用習慣、使用中文斷詞及其他統計方法外，也可以朝其他幫助譯者改善譯文品質的工具或軟體著手。例如，研究發明會自動挑錯的「譯文編輯器」，指出如「被字句」在內的常見錯誤；開發「常見翻譯錯誤練習系統」，讓翻譯系學生可以針對常見的錯誤，加強練習，並培養、提升對這些錯誤的認知。

## 註解

1. 例句來源：NTU Chinese-English Parallel Corpora。
2. 這個詞「translationese」最早是由 Nida (1959) 所提出。
3. 「西化」或「歐化」語法是中文語法受西洋語法影響，所產生的一種語法，又因英語比其他語言更為普遍，所以也有人稱作「英語化」或「英化中文」(吳恬綺，2005)。
4. 吳潛誠 (1995) 表示：「挨、遭、遇、教、讓、給、受、獲、承、蒙、得、邀、見」等字皆可表達被動之意。
5. 三個毛病分別為：「第一、見字而不見句；第二、以為任何英文句都可以在中文裡找到同義詞；第三、以為把英文句子的每一部分都譯過來後，就等於把那句子譯過來了」(余光中，2002，頁 61)。
6. 見黃邦傑 (2006，頁 224)。
7. Selinker (1972) 是最早談到「interlanguage (IL)」這概念。
8. 本文採用的定義，原文如下：「A corpus is a principled collection of texts. A corpus is a collection of electronic texts usually stored on a computer. A corpus is available for qualitative and quantitative analysis (O'keeffe, McCarthy, & Carter, 2007, p. 1-2)」。
9. 分別為「簡易化 (simplification)」、「明朗化 (explicitation)」、「正規化 (normalisation/conservatism)」及「中性化 (levelling out)」。以上普遍特徵，本文採用陳瑞清 (2005) 的翻譯。
10. 根據 Hatim 的定義，語言規範 (norm) 表「The conventions (in the sense of implicitly agreed-upon standards) of acceptable content and rhetorical organization (2001, p. 231)」。

11. 用統計分析比較兩個語料庫的研究並不只限於翻譯學。第二語言習得 (Second Language Acquisition, SLA) 及外語教學 (Foreign Language Teaching, FLT) 經由學習者電腦語料庫 (Computer Learner Corpus, CLC) 的建立, 利用語料統計分析 (如全部及平均字數, 高頻率字數, 連接詞等), 比較母語人士語料及非母語人士語料 (Granger, 2004)。
12. 見附錄一、附錄二。
13. <http://www.voafanti.com/gate/big5/www.voanews.com/chinese/>
14. 見附錄三。
15. <http://www.sketchengine.co.uk/>
16. 以下敘述一律以「SKE 中文非譯文語料庫」、「SKE 中文語料庫」或「SKE 語料庫」簡稱。
17. 指絕對值而言。
18. 「Chi-squared (X<sup>2</sup>) Test」也可以用來比較兩個語料庫裡出現的頻率差異, 但有鑑於「Chi-squared」統計方法的諸多限制, LL 檢定會比「Chi-squared Test」理想 (Rayson, Berridge, & Francis, 2004)。因此, 本文只用 LL 值來比較兩個語料庫的頻率差異。
19. 此定義取自楊惠中 (2002) 在文中 (頁 155) 所提辛克萊的定義。
20. Z 值也可用來偵測搭配力 (Hunston, 2002)。Z 值跟 t 值的計算很相似, 當樣本數很大時, 兩者的數值差異會很小。
21. 本小節的例子皆來自「NTU Chinese-English Parallel Corpora」。
22. 鄧敏君 (2005) 曾討論過這種方法, 本文引用其對 (n-gram) 的中文翻譯。
23. 見附錄四。
24. 見附錄五。

## 參考文獻

- 王克非 (2009)。近代翻譯對漢語的影響。**比較、鑑別、應用——英漢對比應用研究** (頁406-416)。上海：上海外語教育出版社。
- 余光中 (2002)。余光中談翻譯。北京：中國對外翻譯出版社。
- 吳恬綺 (2005)。國中翻譯試題歐化現象之研究。國立臺灣師範大學翻譯研究所碩士論文。未出版，台北。
- 吳潛誠 (1995)。中英翻譯：對比分析法。台北：文鶴出版社。
- 林杏光 (1990)。漢語句型。北京：中國國際廣播出版社。

- 思果 (2001)。翻譯新究。台北：大地出版社。
- 思果 (2003)。翻譯研究。台北：大地出版社。
- 張振玉 (1993)。翻譯散論。台北：東大圖書股份有限公司。
- 張培基、喻雲根、李宗杰、彭謨禹 (1993)。英漢翻譯教程。台北：書林出版社。
- 陳定安 (1997)。英漢比較與翻譯 (最新版)。台北：書林出版社。
- 陳瑞清 (2005)。漢譯文本的形合趨勢：以語料庫為本的翻譯研究。翻譯學研究集刊，9，161-196。
- 傅一勤 (2004)。翻譯的語言：一種中介語言。傅一勤中英語言學論文集 (頁133-144)。台北：輔仁大學出版社。
- 葉子南 (2000)。英漢翻譯理論與實踐。台北：書林出版社。
- 葉子南 (2001)。高級英漢翻譯理論與實踐。北京：清華大學出版社。
- 黃邦傑 (2006)。新編譯藝譚。台北：書林出版社。
- 黃宣範 (譯) (2007)。C. N. Li & S. A. Thompson 著。漢語語法。台北：文鶴出版社。
- 楊惠中 (2002)。語料庫語言學導論。上海：上海外語教育出版社。
- 解志強 (2010)。翻譯中的語言規範問題：談網路的語料庫功能。編譯論叢，3 (2)，163-202。
- 劉宓慶 (1997)。英漢翻譯訓練手冊。台北：書林出版社。
- 潘穎薇 (1997)。將錯就錯—從西化中文談英文教學。第二屆兩岸外語教學研討會論文集 (頁8.1-8.10)。台北：淡江大學外語學院。
- 鄧敏君 (2005)。語料庫輔助翻譯研究之方法論。翻譯學研究集刊，9，405-426。
- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honor of John Sinclair* (pp. 233-250). Amsterdam & Philadelphia: John Benjamins.
- Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*, 7(2), 223-243.
- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and translation* (pp. 175-186). Amsterdam & Philadelphia: John Benjamins.
- Baker, M. (2000). Towards a methodology for investigating the style of a literary translator. *Target*, 12(2), 241-266.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243-257.

- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Church, K.W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16, 22-29.
- Church, K.W., Gale, W. A., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical acquisition: Using on-line resources to build a lexicon* (pp. 115-164). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gao, Zhao-Ming. (2011). Exploring the effects and use of a Chinese-English parallel concordancer. *Computer-Assisted Language Learning*, 24(3), 255-275.
- Granger, S. (2004). Computer learner corpus research: Current status and future prospects. *Applied corpus linguistics: A multidimensional perspective* (pp. 123-145). Amsterdam & Atlanta: Rodopi.
- Hatim, B. (2001). *Teaching and researching translation*. Harlow: Pearson Education Limited.
- Hunston, S. (2002). *Corpora in applied linguistics*. England: Cambridge University Press.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). The Sketch Engine. *Proceedings of Euralex* (pp. 105-116). Lorient, France: publisher.
- Kilgarriff, A. (2007). Googleology is bad science. *Computational Linguistics*, 33(1), 145-151.
- McEnery, A., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London and New York: Routledge.
- Nida, E. A. (1959). Principles of translation as exemplified by Bible translating. In R. A. Brower (Ed.), *On translation* (pp. 11-31). Cambridge: Harvard University Press.
- Olohan, M. (2004). *Introducing corpora in translation studies*. London and New York: Routledge.
- O'keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. *In Proceedings of the workshop on Comparing Corpora*, held in conjunction with *The 38th Annual Meeting of the Association for Computational Linguistics*, 1-6.
- Rayson P., Berridge D., & Francis B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. In Volume II of Purnelle

- G., Fairon C., & Dister A. (Eds.) *Le poids des mots: Proceedings of the 7th International Conference on Statistical Analysis of Textual Data, Louvain-la-Neuve, Belgium, March 10-12, 2004*, Presses universitaires de Louvain, 926-936.
- Selinker, L. (1972). Interlanguage. *IRAL*, 10(3), 209-231.
- Xiao, R. (2010). How different is translated Chinese from native Chinese? A corpus-based study of translation universals. *International Journal of Corpus Linguistics*, 15(1), 5-35.

## 附錄一、NTU Chinese-English Parallel Corpora



## 附錄二、NTU Chinese-English Parallel Corpora- 「VOA: 被」



### VOA

共有404項搜尋結果

- **line 3:** They say scores of people were injured.  
[上一句](#) [下一句](#)  
**line 3:** 有數十人被打傷。  
[上一句](#) [下一句](#)
- **line 4:** But the Associated Press quotes Kuang Biao as saying he has not been suspended.  
[上一句](#) [下一句](#)  
**line 4:** 他沒有被停職。  
[上一句](#) [下一句](#)
- **line 3:** More than 40 people were also injured in the explosion.  
[上一句](#) [下一句](#)  
**line 3:** 還有40多人被炸傷。  
[上一句](#) [下一句](#)
- **line 7:** The helicopter was not hit.  
[上一句](#) [下一句](#)  
**line 7:** 直升機沒有被擊中。  
[上一句](#) [下一句](#)
- **line 7:** Three British pilgrims were killed during a bus ambush south of the capital.  
[上一句](#) [下一句](#)  
**line 7:** 兩名英國朝聖者被打死。  
[上一句](#) [下一句](#)
- **line 7:** Some protesters were detained.  
[上一句](#) [下一句](#)  
**line 7:** 一些示威者被拘留。  
[上一句](#) [下一句](#)
- **line 3:** Ten people are confirmed dead.

## 附錄三、英國 Sketch Engine (SKE) 語料庫


檔案(E) 編輯(E) 檢視(V) 我的最愛(A) 工具(I) 說明(H)

[Home](#) [Concordance](#) [Word List](#) [Word Sketch](#) [Thesaurus](#) [Sketch-Diff](#)

Corpus: Chinese GigaWord 2 Corpus: Taiwan, traditional

Query:

Keyword(s)

Context

Text Types

## 附錄四、SKE「文本索引 (Concordance): 被」


檔案(E) 編輯(E) 檢視(V) 我的最愛(A) 工具(I) 說明(H)

[Home](#) [Concordance](#) [Word List](#) [Word Sketch](#) [Thesaurus](#) [Sketch-Diff](#)

[View options](#) [Sample](#) [Filter](#) [Sort](#) [Frequency](#) [Collocation](#) [Save](#)

Page  of 27805  [Next](#) | [Last](#)

CNA19910101.0022 台北一日電) <p> 新光集團吳東亮被綁架案發生後，已使多金  
 CNA19910101.0027 個時期的合唱音樂，以往被浪漫派光芒所掩，近年  
 CNA19910101.0027 演唱技巧與純淨的音色，已被譽為文藝復興時代合唱曲的  
 CNA19910101.0041 遭強制拆除、三百五十七家被處以罰款，累計實收罰金  
 CNA19910101.0045 、龍祥等地下投資公司相繼被法院裁定破產。八月底再  
 CNA19910101.0062 新腎上腺素於神經末梢得以被再吸收，因而可造成中樞  
 CNA19910101.0062 中心和抑制胃腸蠕動，也曾被用為減肥藥，賭徒為提振  
 CNA19910101.0066 三十三人，因涉及他項刑案，被其他單位接押，致實際出獄  
 CNA19910101.0077 人中，一人因未抵抗、未被殺害，另一人跳海逃生。  
 CNA19910101.0083 懸而未決的政治問題，諸如被黜「黨總書記」趙紫陽的  
 CNA19910101.0092 共有三十六艘船隻在本區域被洗劫，打破過去數年的紀錄  
 CNA19910101.0092 打破過去數年的紀錄。</p><p> 被搶劫的船隻有油輪、散裝  
 CNA19910101.0092 甲板，肆意洗劫。</p><p> 一艘被劫的貨櫃船船長說，這些  
 CNA19910101.0093 特稿) <p> 在共黨獨裁者西奧塞古被推翻一年之後，由於經濟  
 CNA19910101.0100 訣別書。方聲洞的赴義前黨父書被選錄在高中國文課本內。  
 CNA19910101.0101 一名最兇悍的菲籍船員則被殺傷後跳海逃生，但下落  
 CNA19910101.0114 大陸民主運動人士王軍濤可能被中共判以重刑的消息受到  
 CNA19910101.0114 現年二十七歲，去年王軍濤被捕之後，她也曾遭拘禁了  
 CNA19910101.0127 卡查多，於黎明前在自宅被捕。</p><p> 菲空軍官員說，  
 CNA19910101.0127 兵變，他在這次兵變救平後被捕。</p><p> 當時軍事法庭判

Page  of 27805  [Next](#) | [Last](#)

## 附錄五、SKE「詞彙特性素描 (Word Sketch): 證實」

← → ↻ ↺ ↻ 🔍 ☆ ↻ ↻ 檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(I) 說明(H)

Home Concordance Word List **Word Sketch** Thesaurus Sketch-Diff

Turn on clustering More data Less data Save

**證實** Chinese GigaWord 2 Corpus: Taiwan, traditional freq = 41143

PP 向	567	15.1	Modifier 10179	6.0	Subject 21878	5.0	Object	3256	0.7	Modifies 733	0.1
法新社	42	7.75	無法	1374	8.37	發言人	636	6.9	真實性	25	7.35
傳媒	11	5.54	間接	134	8.29	官方	263	6.54	死訊	15	6.94
新聞界	7	4.65	初步	260	7.57	官員	1343	6.41	法新社	28	6.66
世人	6	4.47	首度	174	7.45	實驗	126	6.39	正確性	13	6.48
媒體	79	3.3	加以	141	6.88	證據	128	6.23	據效	17	6.3
記者	141	1.12	尚未	187	6.78	盧仁發	52	6.08	虛驚一場	8	6.05
中央社	67	0.46	予以	143	6.61	結果	347	5.72	傳聞	38	5.74
中心	14	0.13	稍後	43	6.44	警方	532	5.67	說法	95	5.63
			無從	31	6.35	章孝嚴	88	5.58	功效	12	5.42
			不能	196	6.34	消息	222	5.53	死者	25	5.31
			已	1487	6.31	廖正豪	57	5.53	個案	30	5.27
			已經	343	6.28	軍方	130	5.52	死因	10	5.25
			首次	110	6.27	檢方	95	5.51	傳言	10	5.22
			也	1399	6.1	研究	464	5.47	消息	150	5.15
			未	537	6.08	蔣仲苓	41	5.46	病例	37	5.07
			似乎	36	6.04	傳聞	54	5.4	傳媒	8	4.65
			親口	21	6.02	本人	58	5.37	報導	153	4.47
			進一步	171	5.89	科學家	58	5.33	身份	66	4.3
			予	70	5.84	試驗	47	5.3	憂慮	9	4.02
			迄未	21	5.79	說法	90	5.18	疑慮	11	3.9
			不肯	22	5.77	顏慶章	37	5.17	事實	23	3.77
			再度	109	5.63	林義夫	34	5.16	細節	9	3.76
			立即	106	5.49	醫學	65	5.14	理論	9	3.72
			再次	41	5.48	王金平	55	5.09	案例	8	3.67
			終於	35	5.46	正面	53	5.08	話	23	3.56

SentObject of 2424 7.6

願	1521	8.67
拒絕	412	7.39
不予	64	7.15
足以	75	6.91
有待	65	6.04
無異	7	5.16
願意	66	4.61
拒	6	4.11
等候	6	3.89
等待	7	3.13
說法	9	3.11
敢	8	3.0
需要	12	1.02
要求	18	0.33
希望	22	0.13