

台語口語及書面語體的多面向分析¹

張學謙

台東師範學院

這篇論文是以語料庫為本的社會語言學研究。本文嘗試辨認影響台語語體變異的深層言談面向，同時刻畫語體的篇章關係。本研究採用 Biber 的「多面向方法」。本研究建立的語料庫包含口語語體（9 類，94 篇）及書面語體（8 類，91 篇），總計 144,942 詞。經過相關的統計分析後，總共得出五個深層言談面向：人際關係 vs. 表達資訊；個人情感的表達；說服：邏輯 vs. 空間的連接；敘述；具親和力的說明 vs. 詳細報導。本文利用語體的面向分數來刻劃語體的篇章關係。研究結果顯示語體的變異需要透過多面向的分析來掌握，因為任何單一面向都不能絕對的區分口語及書面語體的差異。

關鍵詞：口語，書面語，多面向分析，台語

1. 前言

語體變異 (register variation) 是很普遍的語言現象。使用語言的時候，我們總是隨著言談的場合、目的、對象等情境因素的變化，而有不同的語言表現，也就是說，採用各種不同的語體。² 語體變異的研究至為重要，Hymes (1984) 認為語言項目 (verbal repertoire) 或是說語體變異的描述是社會語言學最主要的工作之一。

本文嘗試對台語的口語和書面語體作全面性的分析。研究的目的是：辨認影響台語語體變化的深層面向，同時也利用這些言談面向來刻劃語體的篇章關係 (textual relations)。傳統的語體分析在研究方法上有兩大的限制：(一)它們通常只研究少數特定的語言特徵、篇章、語體。研究的結果，由於取樣的差異，常常互

¹ 筆者感謝鄭良偉教授的指導，感謝胡萬川、江永進、楊允言、陳明仁先生提供語料，謝信一、王旭、林宗宏諸位教授提供寶貴意見，特此鳴謝。

² 本文將因情境或使用而異的語言變體，稱為語體 (register)。文類 (genre)，風格 (style) 和篇章類型 (text type) 等術語亦常被用來指稱類似的概念。

相抵觸或難以比較；(二)限於單變項的統計方法 (univariate statistics)，無法發現多面向的變異 (Biber 1986)。本研究採用的多面向方法 (Biber 1988, 1995) 可以克服上述的困難。首先，多面向方法將語料庫語言學 (corpus linguistics) 的方法運用在大規模調查語體變異上，包括電腦化的語料庫以及相關的電腦分析工具，如自動斷詞、語法標記、頻率統計等，讓我們可以在極短的時間內完成語料的處理。多面向方法又以多變項的統計方法得出大量的語言特徵在各種語體的共現關係，揭示語體差異的基本面向，因而得以進一步探求語體間彼此的篇章關係。

本文的組織如下：前言之後，第二節簡介 Biber 的多面向方法，本研究使用的語料、語言特徵及電腦工具，以及相關的統計結果；第三節是因素的功能性解釋，辨認出影響台語語體變異的五個基本面向；第四節探討各語體間在各個基本面向的篇章關係；第五節是結論。

2. 研究方法及統計分析

本研究採用 Biber (1988, 1995) 的「多面向方法」(Multi-dimensional approach)。「多面向方法」在方法論上有四個特點：(一)運用標準化、多樣化、電腦化的語料庫；(二)使用電腦程式來計算在大量篇章出現的眾多語言特徵 (linguistic features) 的頻率；(三)使用多變異統計分析，特別是因素分析 (factor analysis)，來決定語言特徵彼此之間的共現關係；(四)藉助微觀分析來判斷經由量化分析導出的共現模式背後的功能參數 (Biber 1988:63)。

2.1 語料庫、語言特徵及語料處理

根據上述原則，本研究建立包含不同溝通目的和言談情景的現代台語語體，包括口語及書面的語料，總共 185 篇文本。³口語語料庫有 9 類語體：私人交談、公開演講（實錄）、公開演講（編輯）、電視劇、新聞報導、民間故事、叩應、訪談（實錄）以及訪談（編輯）；書面語料庫有 8 類語體：社論、大眾散文、雜文、勸說文、電子郵件、私人書信、小說以及學術論文。總詞數約十四萬五千詞 (144,942 詞)。表 1 及 2 分別列出口語與書面語料庫的文本數，語體詞數以及平均詞數。

³ 本研究的口語及書面語料都標準化為漢羅書面語的形式。

語料庫建立之後，每篇文章都經過斷詞及語法標記的處理，同時計算語言特徵在個別文本出現頻率。這些工作都是透過鄭良偉教授研發的台華語料自動分析工具 (PCTMD) 所進行的。本文總共調查 55 個語言特徵，分成 17 大類：時態及時段標記；地方及時間副詞；代名詞；問句；介詞、形容詞、副詞；名詞形式；詞的類別 (lexical classes)；句法結構；從屬結構；連接詞；靜態形式 (stative form)；情態；否定；動詞類別；動詞補語；約簡形式；詞的選用 (lexical choice)。⁴ 語言特徵的頻率經過標準化後，進行因素分析。⁵

表 1. 口語語料庫

語 體	文本數	語體詞數	平均詞數
1. 私人交談	12	9,764	814
2. 公開演講 (實錄)	12	10,538	878
3. 公開演講 (編輯)	10	9,247	925
4. 電視劇	9	6,860	762
5. 新聞報導	8	5,949	754
6. 民間故事	12	8,639	719
7. 叩應	10	6,859	686
8. 訪談 (實錄)	10	9,206	921
9. 訪談 (編輯)	11	9,800	891
口語語庫總數	94	76,862	818

⁴ 本研究使用的語言特徵列於附錄。

⁵ 文本的篇幅長短不一，本研究將文本的長度標準化為 1000 詞，以利分析比較。

表 2. 書面語料庫

語 體	文本數	語體詞數	平均詞數
1. 社論	11	8,872	807
2. 大眾散文	13	8,632	664
3. 雜文	12	10,229	852
4. 勸說文	10	6,909	691
5. 電子郵件	10	6,681	668
6. 私人書信	13	8,432	649
7. 小說	10	9,234	923
8. 學術論文	12	9,091	758
書面語庫總數	91	68,080	748

2.2 因素分析

因素分析是以多面向方法研究語體變異的主要分析工具。因素分析的目的是把大量的變數（語言特徵）簡化為幾組共現的特徵（因素）。每組由統計分析得出的共現語言特徵，就稱為因素。因素分析有四個基本步驟：(1)建立相關矩陣；(2)萃取因素；(3)選擇及旋轉因素；(4)解釋因素。本研究使用 SPSS 統計套裝軟體（第七版）進行因素分析的統計工作。在因素分析之前，我們把語言特徵由原來的 55 個減為 36 個，刪除標準差及平均出現頻率低的語言特徵。這樣作可以增加統計的可信度。⁶ 建立相關矩陣是因素分析的第一個步驟，相關的大小表示兩個語言特徵之間共變的程度。本研究使用主成分法 (principal components analysis) 萃取因素。同時考慮特徵值 (eigenvalue) 及陡坡圖 (scree plot) 得出四因素解，五因素解，六因素解為可能的因素個數。分別考慮四因素解，五因素解，六因素解後，決定以五因素為最佳因素解。因素解釋之前，又用 Varimax 方法旋轉因素，以簡化因素結構，方便因素的解釋。

⁶ Gorsuch (1983) 建議因素分析時，變項數（在此為語言特徵）和樣本數（文本數）的比率至少要 1：5。本研究有 185 篇文本，因此至多可以使用 37 個語言特徵。

表 3 列出分布在五個因素的語言特徵及其因素負荷量 (factor loading)。因素負荷量，不管正負值，表示因素和語言特徵關聯的強度。每一語言特徵在各個因素都有因素符合量，不過它們並不是同等重要，得分愈高的愈能代表該因素。表 3 僅列出因素符合量高於 $|\cdot 30|$ 的語言特徵。因素解釋時，僅考慮絕對值超過 $\cdot 35$ 的語言特徵（顯著特徵）；因素負荷量介於 $|\cdot 35|$ 和 $|\cdot 30|$ 的語言特徵僅作為輔助說明之用（參考特徵）。因素如同時有正、負因素負荷量的語言特徵，表示正負二極形成互補分配的關係。如表 3 所示，部分的語言特徵出現在不同的因素，譬如第一人稱單數代名詞出現在因素 1 及因素 2。這可能反映出部分的語言特徵有多重的溝通功能 (Grabe 1984:48-9)。因素解釋時，需要探求這些語言特徵與該因素的其他語言特徵共享的言談功能。不過，計算面向分數時，同一語言特徵只計算其得分最高的面向，其餘不計。

表 3. 五個因素的語言特徵

因素 1

言談詞	.80
句尾語氣詞	.80
台語特別詞	.79
合音詞	.73
指示代名詞	.69
問句	.69
第二人稱單數代名詞	.57
完成貌	.55
言談動詞	.54
志願情態	.54
存在貌	.44
(進行貌	.40) ⁷
(第一人稱單數代名詞	.40)
(第三人稱單數代名詞	.37)
(方向補語	.34)
(副詞	.33)

名詞	-.86

⁷ 括弧的語言特徵表示：(1)這個語言特徵在別的因素有較高的得分，或(2)參考特徵。不列入面向分數的計算。

修飾標記	-.83
屬性形容詞	-.71
type/token 比率	-.71
並列連接詞	-.70
介詞	-.65

因素 2

強調詞	.71
謂語形容詞	.70
私有動詞	.65
第一人稱單數代名詞	.53
(第二人稱單數代名詞	.33)
(可能情態	.33)
(時間副詞	-.31)

(沒有負成分)

因素 3

連繫詞	.67
表讓步的從屬子句	.51
表條件的從屬子句	.46
可能情態	.39
必須情態	.32

進行貌	-.55
(時間副詞	-.45)
(動作動詞	-.35)

因素 4

時段標記	.66
地方副詞	.61
方向補語	.58
副詞	.57
第三人稱單數代名詞	.44
(時間副詞	.33)

(沒有負成分)

因素 5

第一人稱多數代名詞（包含聽者）	.63
表原因的從屬子句	.56
（言談動詞	.43）
（指示代名詞	.38）

動作動詞	-.58
（type/token 比率	-.43）

3. 因素的功能性解釋：台語的 5 個言談面向

因素是經由因素分析得出在統計上有顯著意義的共現模式 (co-occurrence pattern)。面向 (dimension)則是因素經過功能性解釋的結果。因素的解釋主要是基於共現的語言特徵之間有共通的溝通功能的假設 (cf. Ervin-Tripp 1972)。這些共通的功能可能是情境的、社會的或認知的功能。因素解釋時，除了需要考慮各極語言特徵共同出現的理由外，也得探索互補分配的原因。以下是五個因素的功能性解釋。

3.1 第一面向：人際關係 vs. 表達資訊的面向

如表 3 所示，共同出現在因素 1 負極的語言特徵包括（重要性由高到低）：名詞、修飾標記、屬性形容詞、type/token 比率、並列連接詞、以及介詞。這些語言特徵的言談功能很明顯的和資訊的傳遞有關。名詞是傳達資訊內容的主要手段。台語的修飾標記 *e* 可作為屬格助詞或名物化助詞 (nominalizer)，有整合訊息的功用。屬性形容詞修飾名詞組。高比率的 type/token 比率顯示用詞的多樣化以及精確的用字，因此需要較長的準備時間，通常出現在書面語。並列連接詞可以用來擴充意念單位 (idea unit)。介詞顯示名詞組間的不同關係。簡言之，在因素 1 負極的語言特徵和訊息焦點、細心準備以及訊息整合等言談功能息息相關。

共同出現在因素 1 正極的語言特徵包括：言談詞、句尾語氣詞、台語特別詞、合音詞、指示代名詞、問句、第二人稱單數代名詞、完成貌、言談動詞、志願情態、存在貌、進行貌、第一人稱單數代名詞、第三人稱單數代名詞、方向補語以及副詞。基本上，這些語言特徵的言談功能跟結構散漫、親切的人際互動 (fragmented, involved interaction) 有關。言談詞通常在口語使用，極少出現在書面語。言談詞有承接先前言談的功能，反映交談者積極投入的互動關係。句尾語氣

詞同樣很少在書面語中使用，主要用來表示情態或者是說話者的態度：疑問、警告、勸告、驚訝等 (Cheng 1977, Lien 1988)。台語特別詞較常出現在隨意性的談話。一般而言，正式的情景或者是書面語使用較多的台華共通詞 (Cheng 1987)。因此，特別詞可以當作非正式情景的標記之一。合音詞常常出現在隨意或者是快速談話的場合。問句除了可以提問外，也有鼓勵聽者積極參與交談的功能，因而是參與感 (involvement) 或人際互動的標記 (Chafe 1985)。出現在因素 1 的代名詞有：指示代名詞、第二人稱單數代名詞、第一人稱單數代名詞以及第三人稱單數代名詞。指示代名詞得依賴適當的情景來辨認指涉物，是較為散漫的資訊傳達方式。人稱代名詞則標記參與者的社會關係。第一人稱代名詞、第二人稱代名詞分別標記說話者本身及聽話者的參與 (Chafe 1985)。第三人稱代名詞常被視為敘述的標記。這個面向有三種時態標記：完成貌、存在貌、進行貌。時態可視為觀察情境內部時間成分的不同方式 (Comrie 1976:3)。這三種時態標記表示說話者對情境的主觀看法 (Hsieh 1992, 鄭良偉 1997a:29)。因此可以當作個人與情境的關聯 (cf. Chafe 1985)。志願情態表示主觀意圖也有類似的功能。因素 1 的部分語言特徵似乎和敘述有關，比如：第三人稱、時態標記外，言談動詞、方向補語、副詞。整體而言，因素 1 的正極傾向表現人際互動的語言功能，注重談話雙方積極參與及個人主觀意念，比較依賴言談的情境來傳達訊息，由於受到言談時間限制，結構較為零散。

因素 1 的正負二極顯示語言產生環境 (production circumstances)、語言結構和言談目的息息相關。⁸ 因素 1 的正極，表示參與感以及互動性的語言特徵通常也是結構零散的標記。負極表示訊息整合跟訊息修飾的語言特徵傾向反映細心準備或者是編輯的語言產生環境。正負二極同時分別跟口語模式 (mode) 和書面模式相關。這可能是因為書面模式提供了口語通常所欠缺的編輯和準備的機會。就言談目的、語言產生環境、語言模式而言，正負二極的差異可由以下的對立項表述：互動的 vs. 非互動的，即席的 (on-line) vs. 有準備的 (planned)，口語的 (oral) vs. 書面的 (literate)。⁹ 因此，我們將第一個面向標示為：親切的人際互動焦點 (involved interactional focus) vs. 結構緊密的資訊焦點 (integrated informational focus)，簡稱為：人際互動 vs. 表達資訊的面向。

⁸ Chafe (1982) 指出造成口說與書寫差異的二個基本面向。首先，書寫是緩慢、刻意、可編輯的過程，而口語則迅速完成，結果造成書面語的結構緊密 (integrated)，口語的結構零散 (fragmented)。其次，書寫是個人單獨的行為，而口語通常在社會互動的情形下發生，因而書面語客觀中立 (detached)，而口語則顯現出積極參與 (involved) 的性質。

⁹ 在此，oral 指涉典型的口語，即日常對話，而 literate 指涉典型的書面語，即強調資訊的說明文 (Biber 1995)。

3.2 第二面向：表達個人情感

表 3 顯示共同出現在因素 2 正極的語言特徵有：強調詞、謂語形容詞、私有動詞、第一人稱單數代名詞以及第二人稱單數代名詞、可能情態、以及時間副詞。最後三個語言特徵的因素負荷量都沒有超過 .35，是參考特徵。強調詞、謂語形容詞和私有動詞都跟表達立場或者是表示對談話主題的關注相關。強調詞通常用在非正式的場合，表示說話者或作者對主題的情感投入 (Chafe 1985)。謂語形容詞常出現在結構鬆散的言談，主要的功能是表明想法或者立場 (stance)。私有動詞指表現心理活動的動詞（如：想講，感覺），可以用來表示個人的態度、思想或感情 (Biber 1988:105)。第一人稱代名詞是自我投入 (ego-involvement) 的標記，第二人稱代名詞的使用顯示說話者對聽眾的關注 (Chafe 1985)。上述的五個語言特徵和個人情感的表達有關。因此，我們將因素 2 稱呼為表達個人情感的面向。

3.3 第三面向：說服的面向：邏輯 vs. 時間的聯接

如表 3 所示，因素 3 正極共同出現的語言特徵有：連繫詞、表讓步的從屬子句、表條件的從屬子句、可能情態以及必須情態。這些語言特徵都跟說服或者辯論有關。連繫詞 *si* 是靜態、信息的風格的標記。連繫詞同時也有表示對照或者是強調的功能 (Chen 1995)。表讓步和表條件的從屬子句跟勸說或辯論有關。條件句有表示假設情況、引導背景知識、請求或者對照的功用 (Wang 1996)。表讓步的從屬子句主要是對先前言談作對照性的說明。可能情態標示陳述的可信度程度。必須情態表示邏輯的必然性和責任義務。正極的這些語言特徵常常在勸說文或者辯論性的言談共同出現，主要的功能是說明抽象的理念，而不是具體事件的敘述。

出現在負極的語言特徵有三個：進行貌、時間副詞和動作動詞。最後兩個語言特徵在其他的因素上都有較高的因素負荷量，因此不計算他們在這個因素的面向分數。這三個語言特徵和強調時間聯接的敘述相關。時間副詞用來標示敘事的時間定位 (temporal orientation)，顯示和具體現實的關聯。動作動詞紀錄主體的行為動作。進行貌顯示動作正在進行當中。這二個語言特徵是動作取向的 (action-oriented)，跟正極的靜態說明相反。因此，正負二極呈現互補分配的關係。基於正負二極的各別共通功能以及互補分配的理由，第三個面向可以說是區別以邏輯關聯為主的勸說文和以時間聯接的敘述文。我們將第三個面向命名為：說服的面向：邏輯 vs. 時間的聯接。

3.4 第四面向：敘述的面向

因素 4 只有正極沒有負極。共同出現在正極的語言特徵有：時段標記、地方副詞、方向補語、副詞、第三人稱單數代名詞和時間副詞。這些語言特徵都和敘述面向有關。時段標記用來標記事件或情境的時間段落（開始、繼續、完成等）。地方副詞和方向補語跟空間的定位 (spatial orientation) 有關聯。方向補語指示動作的方向。地方副詞和時間副詞表明人物、事件的時空關係。副詞則說明動作的方式。第三人稱代名詞是敘述的標記。

根據上述說明，第四個面向很明顯的跟強調空間關係的敘述有關。這個面向區別敘述和非敘述的言談（如：說明文、勸說文）。因此，第四個面向就稱呼為：敘述的面向。

3.5 第五面向：具親和力的說明 vs. 詳細報導的面向

如表 3 所示，因素 5 由六個有顯著因素負荷量的語言特徵構成，其中三個在別的因素有較高的負荷量。因為因素 5 只有三個顯著的語言特徵，因素解釋並不容易，下列的解說，只是暫時性的。¹⁰

正極的語言特徵有：第一人稱多數代名詞（包含聽者）*lan*（咱們）、表原因的從屬子句、言談動詞以及指示代名詞。這些語言特徵似乎和親切的說明文體相關。*Lan* 在人際溝通上有多種的語用功能：表示親近、關懷、拉近彼此的距離等（鄭良偉 1997b:5）。因此，*lan* 可以表達談話雙方的親切的關係，建立『我群』彼此一致的認同感 (in-group solidarity)。表原因的從屬子句除了說明理由、提供資訊以外，還有表達情意的功能 (Biber 1988:107)。言談動詞通常是敘述的標記，與因素 5 其他語言特徵共同出現的原因不明。指示代名詞常用在未事先準備、或者有言談時間壓力的情景。簡單的說，上述的語言特徵主要是和親切的（口語）說明體相關。

負極只有動作動詞以及 type/token 比率。這二者同時出現在負極令人訝異，因為他們分屬於動詞和名詞的風格 (verbal and nominal style)。動作動詞主要用在

¹⁰ Gorsuch (1983) 提出一個經驗法則，認為一個因素至少需要五個變項（語言特徵），才有辦法給予適當的解釋。

敘事或者是報導文體。詞的多元化主要是書面現象，因為選詞用字需要較多的準備時間和編輯。或許負極的語言特徵和有計劃的報導 (planned reportage) 有關。

綜合上述的討論，第五個面向的言談功能可能是區別親近的說明文體和有計劃的報導。因此，第五個面向就暫時命名為：具親和力的說明 vs. 詳細報導的面向。

4. 語體之間的篇章關係

第三節已經以溝通功能的觀點，將因素分析的結果解釋為台語的言談面向。這些功能性的解釋需要進一步的檢驗。本節將以語體之間的篇章關係來考驗上節的分析。我們將各別語體在各面向的面向分數 (dimension score)，由高到低，排列成行，透過言談面向來解釋語體的分佈關係。面向分數的計算方式是先計算各別文本在各面向的得分（加上正極，減去負極的顯著特徵的規範化分數），然後計算各語體的面向分數的平均數。因素負荷量界低於 $|.35|$ 的語言特徵，不列入面向分數的計算。相同的語言特徵如果出現在不同的面向，則只計算得分最高的面向。得出語體在面向得分之後，我們以 General linear models (GLM) 的統計方法測試語體在各面向的得分是否有顯著的差異。由表 4 可知，這五個面向都能有效的區分語體差異。¹¹

表 4. GLM 的結果：語體差異

面向	F value	Probability (p)	R ²
1	53.7	p < .0001*	84%
2	11.8	p < .0001*	53%
3	9.8	p < .0001*	48%
4	10.8	p < .0001*	51%
5	8.8	p < .0001*	46%

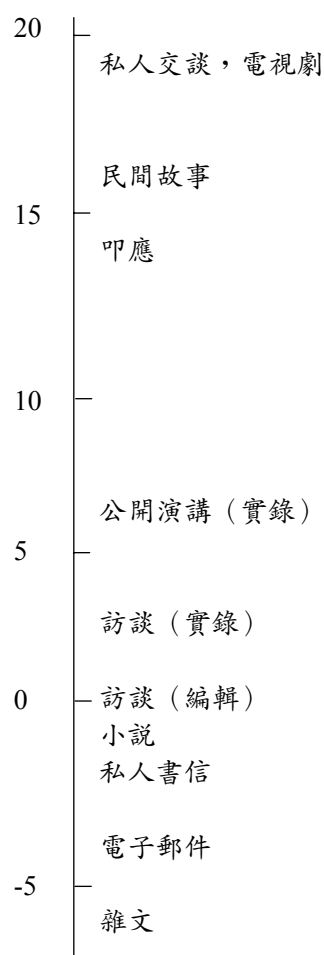
¹¹ F 值顯示某面向能否顯著的區分語體。p 值如小於 .05，則表示達到統計的顯著程度。R² 顯示面向區別語體的重要性。

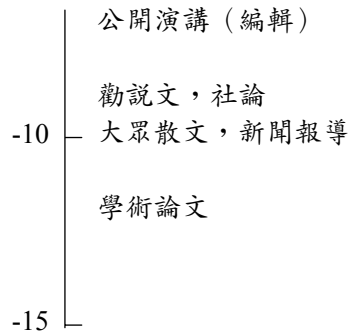
4.1 語體在第一面向的關係

圖 1 列出台語 17 種語體在面向 1 的分佈情形。由圖 1 可知，私人交談(19.4)，電視劇(18.8)的面向分數最高，而大眾散文(-10.0)，新聞報導(-10.9)及學術論文(-11.9)分數最低。民間故事(16.4)、叩應(14.6)也有相當高的得分；雜文(-6.1)、經編輯的公開演講(-7.0)、勸說文(-8.2)、社論(-8.6)的得分則是相當低。這個分佈情形符合我們對面向 1 的解釋：親切的人際互動焦點

圖 1. 十七種台語語體在第一面向的分布情形

面向 1：人際關係 vs. 表達資訊的面向





vs. 結構緊密的資訊焦點。面向 1 得分較高的都是強調人際互動關係的語體，通常又是依賴言談情境 (situated) 的即席談話，由於言談時間的壓力，結構較為零散。由取自私人交談的例文 1，可以看出強調人際互動關係的語體常使用的語言特徵：

例文 1：私人交談 (s008) ¹²

A: 我 今仔日 chong 到 XX 寺去。

B: 是按怎？

A: 我 就 m 知影按怎騎按呢。一直騎感覺 m 著，a

B: a 『小姐』也 toe 你 去 o？伊也騎一台，抑是你 ka 載？

A: 我 ka 載 lah. A chong 到 XX 寺去，你 知無？

B: 我 ka 你 講，你 ka 人 oat 了傷早，你 ka 人 oat，你 ka 人 oat，

C: Oat 了傷早，猶未到你 著 oat leh 走。

A: 我 想講奇怪，a thah 2,3 個月，

C: 猶未到你 著 oat leh 走 a。

B: 你 到這個 hoⁿ，

A: 二、三個月無來按呢 hoⁿ，he 路 soah 感覺無啥著。

例文 1 是表現人際互動的典型文本。例中運用面向一表示互動的語言特徵（劃底線的語詞）：言談詞 *a*，句尾語氣詞 *o*，*lah*，*a*，*hoⁿ*，合音詞，*he*，*saⁿ*，問句，第一，第二人稱。例文 1 同時使用極少使用負極的語言特徵。而得分低的語體則是強調信息焦點的說明文。它們通常是書寫的語體，有較多的準備時間或編輯的機會，結構較為嚴謹。取自學術論文的例文 2 則顯示以資訊傳達為重點的語體所具有的語言特徵（劃線的部分）：

¹² s 代表口語文本，w 代表書寫文本。數字是文本的編號。

例文 2：學術論文 (w005)

教育部決定 1996 年，開始 ti 中小學實施本土語言教育，雖然也開始進行一寡基本的研究，譬如，委託人文及社會科學指導委員會，針對台語、客語、原住民語的語言，分別建議一套標音系統，但是猶是傷少，做為教育內容的一部分來講，咱至少猶需要文字的標準化、常用詞的調查、師資的訓練、教材的編纂、補助教材的充足，等等…

例文 2 顯示以信息為主的文本經常使用與信息整合，信息說明相關的語言特徵，例如：名詞、屬性形容詞、連接詞、修飾標記和介詞。例文 2 和例文 1 剛好成為對比，例文 2 很少使用例文 1 的語言特徵，反之亦然。

乍看之下，面向 1 似乎可以以語言模式的差異將語體一分為二：口語語體出現在正極，書面語體則出現在負極。然而，仔細的檢閱圖 1 將發現單靠模式本身並無法解釋所有語體變異，例如，新聞報導（-10.9）是以口語呈現，卻出現在負極，面向分數遠低於書面表達的私人書信（-1.9）。由此可見，面向 1 的語體變異尚需考慮言談目的或語言產生環境。就言談目的而言，私人書信強調感情投入和人際互動，而新聞報導則重視資訊的客觀呈現，這兩者產生環境亦不同，寫信時較為隨性，新聞報導卻是透過詳細編輯、準備的文稿以口語呈現的。由編輯和實錄語體的對比最容易看出語言產生環境對語體的影響，實錄的演講和訪談（6.3; 3.4）面向分數都遠高於經編輯的演講和訪談（-7.0; 0.7）。由例文和語體在圖 1 的分布可知面向 1 雖然和語言模式有關，然而面向 1 並沒有將口語與書面語截然二分。面向 1 的語體變異需要結合語言模式、言談目的和語言產生環境才能得到解釋。由上討論可知語體在面向 1 的篇章關係符合我們對面向 1 的解釋。

4.2 語體在第二面向的關係

在 3.2 節我們將因素 2 解釋為：表達個人情感的面向。因此，這個面向可以區分表主觀情感和表客觀資訊的言談。個別語體在圖 2 的分布支持我們的分析。私人書信（4.6）在這個面向的得分最高；電視劇（2.7），訪談（2.6）亦有相當高的得分。因為這些語體重視個人主觀情感的表達。表達感情是語言的重要功能。Jakobson (1990:73) 將語言表達情感的功能 (emotive function) 或是表達功能 (expressive function) 定義為：“焦點放在說話者 (addresser)，目的在直接的表達說話者對所說的話的態度。”由於新聞報導（-3.7）和學術論文（-2.5）強調客觀的信息傳播，因此在這個面向的得分相當低。例 3 顯示私人書信如何運用面向 2

的語言特徵表達個人的情感（劃線的部分）：

例文 3：私人書信 (w031)

今仔日早起我看 X teh 寫批 ho 你。伊寫 kah 真歡喜。我 tu ti 邊仔 phah 電腦聽伊直直 leh o-lo 家己講伊的台語批寫了真好。寫了 theh ho 我看。我對頭笑到尾...伊感覺家己的批寫了真好, 真 “creative”, 閣 theh 去影印, 收集起來, 因為是伊的第一張台語批。致到我 是感覺真好笑。

例文 3 使用很多的表示情感的語言特徵, 例如: 第一, 第二人稱代名詞(我, 你), 強調詞(真), 私有動詞(感覺, 歡喜)等。私人書信可以引起讀者強烈的情緒反應, 如例文 4:

例文 4：私人書信 (w030)

慈愛的賢妹及賢妹婿查某孫全家收信平安：

我 ti 26 暗, 有收著賢妹的慈音深深感激, 真多謝你的安慰。

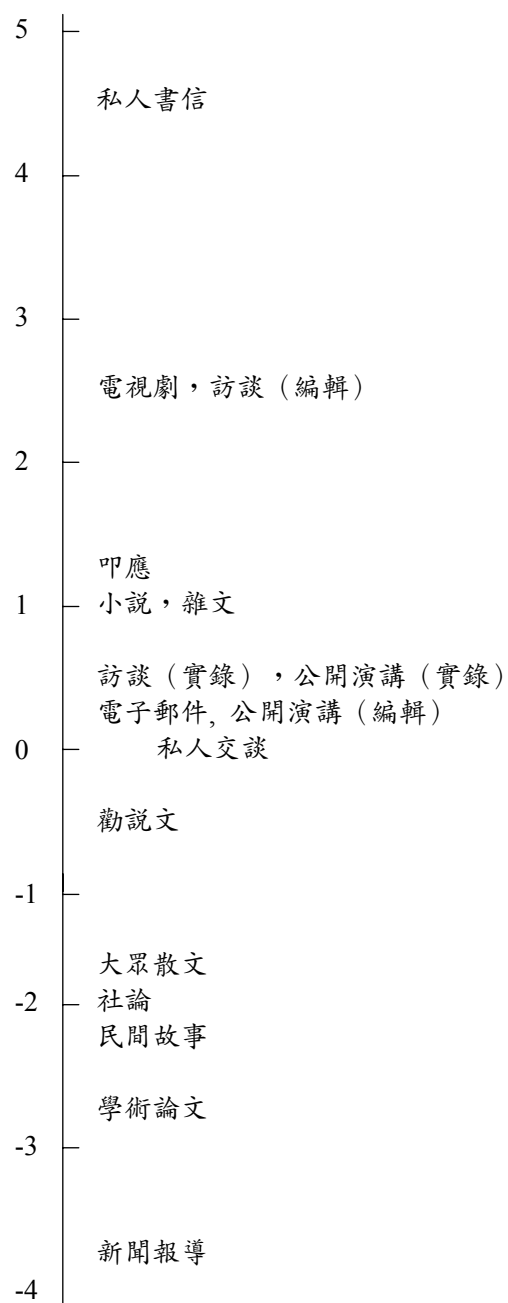
我 na 讀隴喉管 na tiⁿ...我 na 讀目屎 na 流, 想著 m 知到甚物時陣咱姊妹才有機會相逢 neh。

一般認為口語較書面語容易流露情感, 並沒有得到本研究的支持。事實上, 本研究顯示台語的私人書信是最常透露個人情感的語體, 而私人談話(0.0)在這個面向的得分不高, 甚至低於電子郵件(0.4)。因此, 語言媒介並不是影響表達情感這個面向的主要因素, 社會文化對語言使用的規範才是最重要的因素 (cf. Besnier 1988)。言談目的也是影響情感表達的因素。一般來說, 說明文和辯論文比較不重視情感的表達, 因此, 學術論文(-2.5)、社論(-1.9)、大眾散文(-1.5)和勸說文(-0.6)的分數都很低。敘事語體(小說、雜文)和強調人際互動的語體(訪談、演講)則比上述語體為高。¹³

整體而言, 個別語體在圖 2 的分布支持我們將面向 2 解釋為表達個人情感的面向。語言模式和情感表達沒什麼關聯, 情感的表達主要是決定於言談目的及言談的社會文化規範。

¹³ 民間故事的得分較其他敘事文體為低, 可能是因為民間故事較少將情感以言詞表示, 而是透過手勢、表情、動作等非語言的管道 (cf. Tannen 1982:42)。

圖 2. 十七種台語語體在第二面向的分布情形
面向 2：表達個人情感的面向



4.3 語體在第三面向的關係

在 3.3 節我們將第三面向標記為：說服的面向：邏輯 vs. 時間的连接。也就是說，這個面向區分以邏輯關聯為主的勸說文和以時間连接的敘述文。圖 3 顯示各語體在第三個面向的分布情形。由圖 3 可知，辯論性質的語體得分較高，如電子郵件（4.8）、叩應（4.0）、勸說文（3.4）、社論（2.1）。本文所收集的電子郵件都是辯論語體的討論信函。它在這個面向的得分最高，例文 5 顯示說服性質的文本運用邏輯连接來達成說服的言談目的：

例文 5：電子郵件（討論信函 w064）

如果我採用一套簡單閣好學的系統我想我應該真緊著學好勢 a。
閣來才是問題，如果我用這套系統無作品、無教材、無字典。
簡單講愛啥無啥。這陣我可能會放棄這個系統...

例文 5 運用表條件的從屬子句“如果”、可能情態“可能”以及必須情態“應該”等語言特徵論證文字系統的優缺點。這些語言特徵都跟說服或者辯論有關。

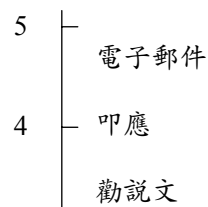
例文 6 顯示在面向 3 得分低的語體的語言特徵：

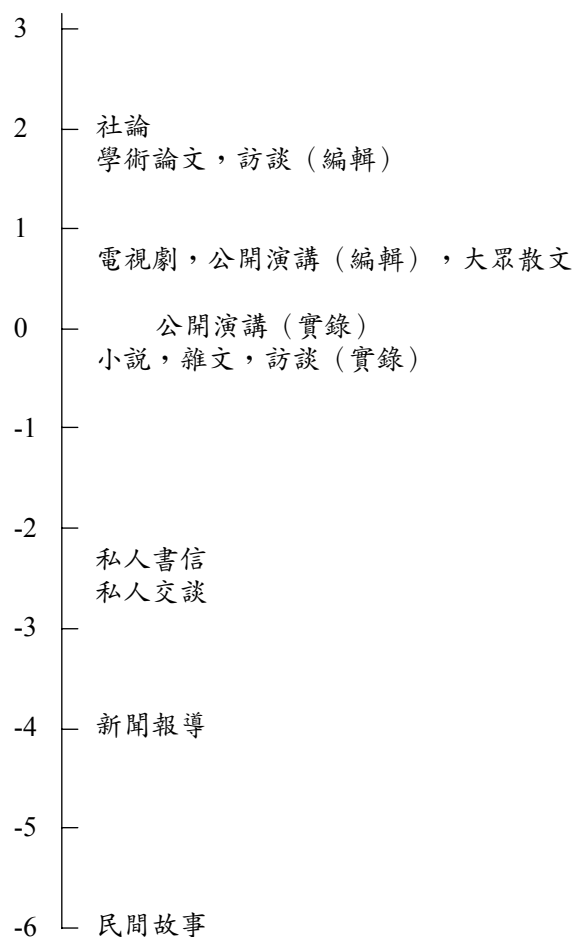
例文 6：民間故事（s024）

阿 in 兩個這個囡仔講講昨昏有這種代誌按呢。阿這個囡仔講。
M，我閣 m 捌聽我卡早 oan-a ti 遮睏講。我卡早暗時若歹天 houN。
我 oan-a 定定 ti 遮睏按呢。

例文 6 顯示敘述語體使用時間副詞（昨昏、卡早、暗時、定定）來標記事件的時間定位。由於敘述語體的重點在記錄事件的開展，而非邏輯論證或說服，因此，這類語體在面向 3 的得分較低，如小說（-0.7）、雜文（-0.9）、新聞報導（-3.8）、民間故事（-5.7）。

圖 3. 十七種台語語體在第三面向的分布情形
面向 3：說服的面向：邏輯 vs. 時間的连接





由上討論可知面向 3 區分以邏輯關聯為主的勸說文和以時間連接的敘述文，圖 3 各語體的分布情形及例文都支持我們的解釋。

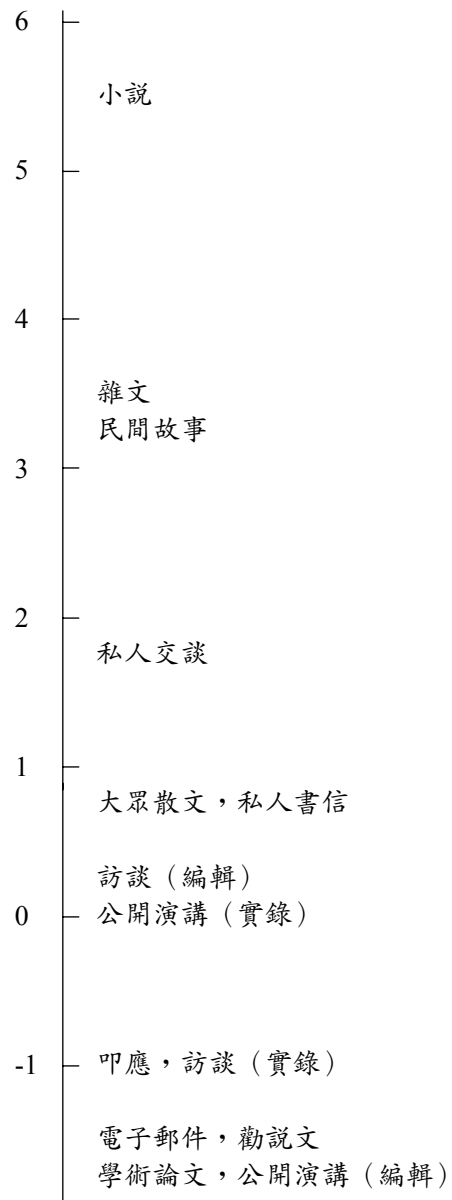
4.4 語體在第四面向的關係

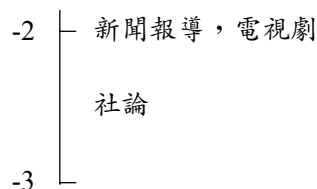
第四面向是強調空間關係的敘述面向。也就是說，這個面向區別敘述和非敘述的言談。由圖 4 可知，敘述語體有相當高的得分，小說（5.5）最高，接著是雜文（3.6）和民間故事（3.3）。而社論（-2.5）、勸說文（-1.5）、電子郵件（-1.5）和學術論文（-1.6）等非敘述語體的得分都很低。書寫的敘述語體（小說和雜文）比口說的敘述語體（民間故事）有較高的得分，可能是因為書寫提供較多的時間

營造故事的空間背景，而口語則受到言談時間的限制。

圖 4. 十七種台語語體在第四面向的分布情形

面向 4：敘述的面向





例文 6 顯示台語小說使用敘述語言特徵的情形：

例文 6：小說 (w086)

伊坐 ti 彼石階頂，彼石階仔四界攏是松樹落落來的松球，伊一面喘氣，
一面用手 ng 去拭面頂的汗彼爬起來的山路已經消失 ti 若波浪的山腳，
頭前山路的尾溜便是伊幾年前來每暗攏 teh 眠夢的清水寺

上引文顯示台語小說使用相當多的方位詞和地方副詞來指明故事的場景，例如：石階頂、四界、面前、山腳、頭前。同時也使用二個方向補語：落落來、爬起來。第三人稱“伊”共出現三次，指涉故事中的主角。

圖 4 顯示敘述是私人交談中常有的言談現象。大眾散文（0.7）、私人書信（0.7）、訪談（編輯，0.4）、演講（實錄，0.3）常常混合敘述和非敘述的目的，所以這些語體分布在中間地帶。新聞報導（-1.9）在這個面向的得分相當低，可能因為這類語體僅強調時間定位，而不注重空間的定位。因此，雖然同樣和敘述相關，第三和第四面向有不同的焦點，前者強調時間定位，後者則強調空間定位。

總而言之，面向 4 以言談目的區別語體：敘述語體在這個面向有較高的得分，而非敘述語體的得分偏低。語言模式與這個面向的關係不大。

4.5 語體在第五面向的關係

在 3.5 節，我們將台語語體變異的第五面向暫時命名為：具親和力的說明 vs. 詳細報導的面向。這個面向的只有三個顯著的語言特徵，因此，我們只提出嘗試性的解釋。

第五面向可能是區別親近的說明文體和有計劃的報導，主要呈現說明語體及敘述語體的對立。如圖 5 所示，得分高的語體主要是具公開性質的談話，例如：訪談（實錄，3.2）、公開演講（實錄，2.4）、叩應（2.0）。¹⁴其中，以訪談的得

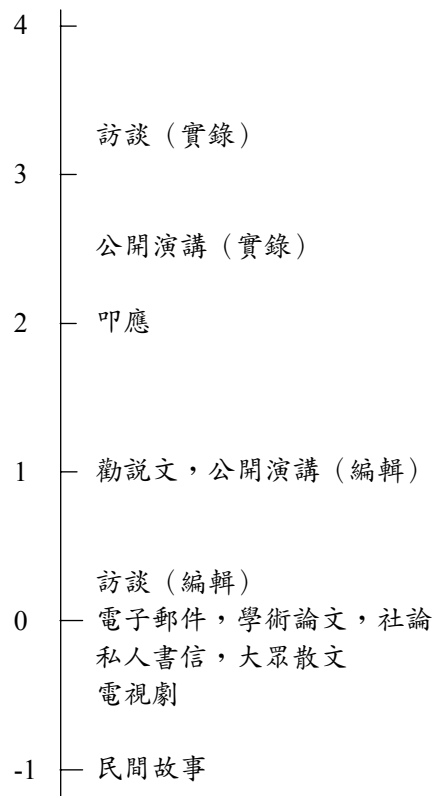
¹⁴ 經編輯的訪談（0.4）與演講（0.9）得分較低。

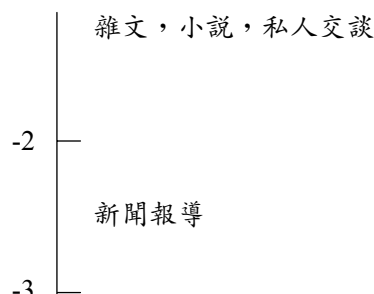
分最高。這類公開談話的主要目的通常是宣揚理念（競選演講，講道），說明事理（知識性的訪談），交換意見，聯絡感情（叩應），特別重視爭取聽眾的認同，拉近談話雙方的距離。另一個影響面向 5 語體分布的因素是計劃與否。得分高的語體主要是即席說明的言談 (on-line exposition)。書寫的說明語體的得分則較低，大多分布在中間地帶，例如：電子郵件（0.2）、學術論文（0.1）、社論（0.1）。例文 7 顯示公開演講運用正極的語言特徵達成彼此一致的認同感（劃線的部分）：

例文 7：公開演講 (s010)

使咱的人生無法度自在的是由於咱過去習慣的索仔 ka 咱縛 leh'ho 咱團
團轉，ho 咱無法度自由的是咱的情絲無法度斬斷。若是咱有法度斷了，
咱的腳下這就自由自在 a。

圖 5. 十七種台語語體在第五面向的分布情形
面向 5：具親和力的說明 vs 詳細報導的面向





一般而言，在負極出現的語體大多跟敘述有關：民間故事（-0.9），雜文（-1.3），小說（-1.6），私人交談（-1.6），新聞報導（-2.5）。這些與敘述有關的語體排序主要是根據計劃時間的多寡而定：愈有計劃時間的語體，得分愈低。新聞報導得分最低，是最能代表負極特質的語體。在負極共現的語言特徵只有動作動詞以及 type/token 比率。新聞主播通常是按照事先準備好的文稿播報新聞的。由於書面語的影響，新聞報導在選詞用字上較有變化，而有較高的 type/token 比率。又因為報導事件的關係，新聞也有較高比率的動作動詞，如例文 8：¹⁵

例文 8：新聞報導 (s052)

未來的民營電廠是賣電 ho 台電公司做整體的調度。對疏解電力不足有真大的幫助。近年來政府興建電廠面臨真濟的阻礙，所以決定開放民間來參與共同解決電力無夠的困境。

由上述語體關係的討論可以得知，第五個面向的主要是區別親切，即席的說明語體和有計劃的報導語體。這樣的語體分布關係大致上符合第五個面向的言談功能：具親和力的說明 vs. 詳細報導的面向。

5. 結論

本文以語料庫語言學為工具，運用多面向的方法，進行台語語體變異的研究。目的是辨認台語語體變異的基本面向，同時刻劃語體的篇章關係。本研究總共辨認出五個言談面向：(一)人際關係 vs. 表達資訊的面向；(二)表達個人情感的面向；(三)說服的面向：邏輯 vs. 空間的连接；(四)敘述的面向；(五)具親和力的說明面向

¹⁵ 劃底線的語詞是動作動詞。

vs. 詳細報導的面向。

就各語體間的篇章關係而言，第一個面向區分表達人際關係的語體（以私人交談，電視劇為代表性的語體）以及表達資訊的語體（以學術論文為代表性的語體）；第二個面向是表達個人情感的面向（以私人書信為代表性的語體）；第三個面向是說服的面向，區分邏輯的連接（討論性質的書信，叩應，勸說文，社論）和空間的聯接（新聞報導）；第四個面向是敘事的面向，代表性的語體是小說，雜文及民間故事；第五個面向區分具親和力的說明語體（訪談，公開演講，叩應）以及詳細報導的面向（新聞報導）。

不同的語體在各個面向的分布情形，證實我們對因素的解釋是適切的。各語體，如我們所預測的，因基本言談面向的差異，而有相應的分布。本文亦支持相關研究的結果 (Biber 1988, Besnier 1988, Kim 1990, Biber and Hared 1992)：任何一個面向都不能絕對的區分口語及書面語體的差異。因此，語體的變異需要透過多面向的分析來掌握。

附錄：本文使用的語言特徵及其類別¹⁶

A. 時態及時段標記

1. 存在貌 (existential aspect)
2. 經驗貌 (experiential aspect)
3. 完成貌 (perfect aspect)
4. 進行貌 (progressive aspect)
5. 時段標記 (phase marker)

B. 地方及時間副詞

6. 地方副詞 (place adverbial)
7. 時間副詞 (time adverbial)

¹⁶ *表刪除的語言特徵。

C. 代名詞

8. 第一人稱單數代名詞
9. 第一人稱多數代名詞（排除聽者）*
10. 第一人稱多數代名詞（包含聽者）
11. 第二人稱單數代名詞
12. 第二人稱多數代名詞*
13. 第三人稱單數代名詞
14. 第三人稱多數代名詞*
15. 指示代名詞
16. 反身代名詞*

D. 問句

17. 問句

E. 介詞、形容詞、副詞

18. 介詞
19. 屬性形容詞
20. 謂語形容詞
21. 形容詞重疊
22. 副詞

F. 名詞形式

23. 名詞
24. 修飾標記 (modifier marker)

G. 詞的類別 (lexical classes)

25. 言談詞 (discourse particles)
26. 加強詞 (amplifiers)*
27. 強調詞 (emphasizers)
28. 和緩詞 (downtoners)*

H. 句法結構

- 29. 使役式 (causative construction)*
- 30. 處置式 (disposal construction)*
- 31. 被動式 (passive construction)*

I. 從屬關係 (subordination features)

- 32. 表原因 (causative subordinator)
- 33. 表條件 (conditional subordinator)
- 34. 表讓步 (concessive subordinator)
- 35. 表時間 (temporal subordinator)

J. 連接詞

- 36. 並列連接詞
- 37. 同位連接詞*
- 38. 轉折連接詞*

K. 靜態形式 (Stative form)

- 39. 連繫詞 (*si*)

L. 情態

- 40. 能力情態
- 41. 必須情態
- 42. 可能情態
- 43. 志願情態

M. 否定

- 44. 否定

N. 動詞類別 (verb types)

- 45. 動作動詞 (action verb)
- 46. 私有動詞 (private verb)

- 47. 言談動詞 (public verb)
- 48. 動詞重疊*

O. 動詞補語 (verb complement)

- 49. 方向補語
- 50. 結果補語*
- 51. 程度補語*

P. 約簡形式 (reduced forms)

- 52. 句尾語氣詞
- 53. 合音詞 (contraction)

Q. 詞的選用 (lexical choice)

- 54. 台語特別詞
- 55. type/token 比率 (前 400 詞中不同的詞數)

引用文獻

- Besnier, Niko. 1988. The linguistic relationships of spoken and written Nukulaelae registers. *Language* 64:707-36.
- Biber, Douglas. 1986. Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language* 62:238-414.
- _____. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- _____. 1995. *Dimensions of Registers Variation: A Cross-linguistic Comparison*. Cambridge: Cambridge University Press.
- _____, and Mohamed Hared. 1992. Dimensions of register variation in Somali. *Language Variation and Change* 4:41-75.
- Chafe, Wallace. 1982. Integration and involvement in speaking and writing, and oral literature. *Spoken and Written Language: Exploring Orality and Literacy*, ed. by D. Tannen, 35-54. Norwood, N.J.: Ablex.

- _____. 1985. Linguistic differences produced by differences between speaking and writing. *Literature, Language, and Learning: The Nature and Consequences of Reading and Writing*, ed. by D.R. Olson, N. Torrance and A. Hildyard, 105-23. Cambridge: Cambridge University Press.
- Chen, Lilly Lee. 1995. The verb *si* 'be' in Taiwanese. *Proceedings of the Second International Symposium on Languages in Taiwan*, 442-452. Taiwan.
- Cheng, R. L. 1977. Taiwanese question particles. *Journal of Chinese Linguistics* 5.2:153-185.
- _____. 1987. Borrowing and internal development in lexical change-A comparison of Taiwanese words and their Mandarin equivalents. *Journal of Chinese Linguistics* 15.1:105-132.
- Comrie, Bernard. 1976. Aspect: An Introduction to the Study of the Verbal Aspect and Related Problems. Cambridge: Cambridge University Press.
- Ervin-Tripp, Susan. 1972. On sociolinguistic rules: Alternation and co-occurrence. *Directions in Sociolinguistics*, ed. by John J. Gumperz and Dell Hymes, 213-250. NY: Holt.
- Gorsuch, Richard. 1983. *Factor Analysis*, 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Association.
- Grabe, William. 1984. *Towards Defining Expository Prose within a Theory of Text Construction*. Los Angeles: University of Southern California dissertation.
- Hsieh, Hsin-I. 1992. In search of a grammatical foundation for dialect subgrouping. *Chinese Languages and Linguistics I: Chinese Dialects*, ed. by Committee for Publications, 333-377. Taipei: The Institute of History and Philology, Academia Sinica.
- Hymes, Dell. 1984. Sociolinguistics: Stability and consolation. *International Journal of the Sociology of Language* 45:39-45.
- Jakobson, Roman. 1990. The speech event and the functions of language. *On Language*, ed. by Linda R. Waugh and Monique Monville-Burston. Cambridge: Harvard University Press.
- Jang, Shyue-chian (Tiuⁿ, Hak-khiam). 1998. *Dimensions of Spoken and Written Taiwanese: A Corpus-based Register Study*. University of Hawaii dissertation.

- Kim, Yong-Jin. 1990. *Register Variation in Korean: A Corpus-based Study*. Los Angeles: University of Southern California dissertation.
- Lien, Chin-fa. 1988. Taiwanese sentence-final particles. *The Structure of Taiwanese: A Modern Synthesis*, ed. by R.L. Cheng and Shuan-fan Huang, 209-234. Taipei: The Crane Book Co.
- Tannen, Deborah. 1982. The myth of orality and literacy. *Linguistics and Literacy*, ed. by William Frawley. New York: Plenum Press.
- Wang, Yu-fang. 1996. The Informational Sequences of Adverbial Clauses in Chinese Spoken and Written Discourse. Taipei: National Taiwan Normal University dissertation.
- 鄭良偉. 1997a. Tense interpretation of four Taiwanese modal verbs. 〈台、華語的時空、疑問與否定〉，《台語、華語的結構及動向》第3冊，19-36。台北：遠流出版社。
- . 1997b. 〈台、華語的代詞、焦點與範圍：從林瑜鏗教授“咱們英語中心”談起—北京話的『咱們』和台語的『咱』〉，《台語、華語的結構及動向》第4冊，3-10。台北：遠流出版社。

[Received 1 December 1998; accepted 15 October 1999]

台東師範學院語文教育學系
台東市 950 中華路一段 684 號
hakkhiam@ms11.hinet.net

A Multi-dimensional Analysis of Spoken and Written Taiwanese Register

Tiuⁿ, Hak-khiam

Taitung Teachers College

This paper is a corpus-based study of sociolinguistics. The aim of this study is twofold: (1) to identify underlying dimensions of register variation in Taiwanese, and (2) to characterize the textual relations among different registers. This study adopts Biber's (1988) multi-dimensional approach to register variation. The corpus comprises nine spoken registers (94 texts) and eight written registers (91 texts), totaling 144,942 words. Texts are segmented, tagged, and counted by computational tools. The distribution of 36 linguistic features in these texts is subjected to factor analysis. Five underlying dimensions are identified in this study: interpersonal vs. informational; personal expression of emotion; persuasion: logical vs. temporal linking; narrative; and involved exposition vs. precise reportage. Textual relations among these registers support the interpretation of dimensions. Since no single dimension is capable of explaining all of the register variation in Taiwanese, the results of this study show that register variation are best captured by multi-dimensional construct.

Key words: multi-dimension, spoken, written, register, Taiwanese