

行政院客家委員會獎助客家學術研究計畫

客語語料庫之系統化建構與量化分析  
The Construction of Hakka Corpus:  
Systemization and Quantitative Analysis

研究計畫主持人：賴惠玲

所屬單位：國立政治大學英國語文學系

中華民國九十九年十一月二十日

# 目 次

致謝扉頁	i
研究者簡介	ii
提要	01
1. 研究動機及目的	01
2. 語料庫介紹	04
2.1 語料庫為本相關研究之文獻回顧	04
2.2 臺灣客語方言語料庫之系統化架構	09
2.2.1 後端架設資料庫簡介	10
2.2.2 前端使用者介面簡介	11
3. 應用客語語料庫之語言學研究分析	17
3.1 研究之理論背景	17
3.2 臺灣客語方言完成體與持續體	20
3.2.1 臺灣客語方言完成體、持續體之本字考究	23
3.2.2 臺灣客語方言完成體、持續體之句法語意關係	24
3.2.3 臺灣客語方言完成體、持續體之語料分佈情形	28
3.3 臺灣客語「怕」構式之語法化與詞彙化	31
3.3.1 臺灣客語「怕」構式之句法及語意分析	34
3.3.2 臺灣客語「怕」構式之量化分析	38
4. 結論	44
參考文獻	48
附錄一	54
附錄二	55
附錄三	56
附錄四	57

## 表 次

〈表 1〉臺灣客語方言語料庫相關資訊	10
〈表 2〉臺灣客語方言語料庫系統權限分類	10
〈表 3〉客語「怕」字構式書面語料比例	40
〈表 4〉客語「怕」字構式口語語料比例	42
〈表 5〉臺灣客語方言語料庫典藏語料之體裁規劃	44
〈表 6〉「國立政治大學客語口語語料庫」網頁典藏語料之內容	46

## 圖 次

〈圖 1〉文章標題表單	11
〈圖 2〉文章內容表單	11
〈圖 3〉以 Python 程式語言撰寫之前端使用者介面	12
〈圖 4〉以 PHP 程式語言撰寫之前端使用者介面	13
〈圖 5〉查詢介面	14
〈圖 6〉文本查詢結果(文章表單列表)	15
〈圖 7〉關鍵字查詢結果	15
〈圖 8〉客語「怕」字構式書面語料之心理動詞分佈圖	39
〈圖 9〉客語「怕」字構式書面語料之情態詞分佈圖	39
〈圖 10〉客語「怕」字構式書面語料分佈圖	40
〈圖 11〉客語「怕」字構式口語語料之情態詞分佈圖	41
〈圖 12〉客語「怕」字構式口語語料分佈圖	42
〈圖 13〉客語「怕」構式圖	43
〈圖 14〉「國立政治大學漢語口語語料庫」語料呈現畫面	45

本研究計畫係接受行政院客家委員會獎助完成

## 研究者簡介

賴惠玲，台灣苗栗人，曾任國立政治大學英國語文學系系主任，現任國立政治大學英國語文學系專任教授並榮獲特聘教授，美國德州大學奧斯汀分校語言學博士。主要以客語語意語法為研究重點，研究領域包括語法化、詞彙語意學、構式語法及隱喻與轉喻等等。

## 個人著作一覽表

### (一) 期刊論文/專書論文(Journal articles)

1. **Lai, Huei-ling**, and Shih-min Li. 2010. Hakka corpus: Construction and Application. Post-conference Proceedings of the Eighth International Conference of Hakka Studies, 300-312. Taipei: Crane.
2. **Lai, Huei-ling**, and Wan-jyun Luo. 2010. Verbal polysemy of Hakka *piong3* 'to put': Frames and constructions. *Journal of Chinese Linguistics* 38.2: 224-260. **NSC 95-2411-H-004-026-MY2 ; NSC 97-2411-H-004-110 MY2** (AHCI; SSCI)
3. Chui, Kawai, and **Huei-ling Lai**. 2008. The NCCU Corpus of Spoken Chinese: Mandarin, Hakka, and Southern Min. *Taiwan Journal of Linguistics* 6.2: 119-142. (THCI Core, LA, LLBA, MLA)
4. **Lai, Huei-ling**. 2008b. Understanding and classifying two-part allegorical sayings: Metonymy, metaphor, and cultural constraints. *Journal of Pragmatics* 40.3: 454-474. **NSC 93-2411-H-004-039 ; NSC 94-2411-H-004-007** (AHCI; SSCI)
5. **Lai, Huei-ling**. 2008a. Using constructions as information management devices: Analysis of Hakka *lien5...ya3/du3* constructions. *Bulletin of the Institute of History and Philology Academia Sinica*. 79.2: 343-376. **NSC 95-2411-H-004-026-MY2** (AHCI; TSSCI)
6. Liao, Pei-yun, and **Huei-ling Lai**. 2007. Hakka verbs of removal: Integration of verbal meanings and [VX] constructions. *Taiwan Journal of Linguistics* 5.2: 1-44. (LA, LLBA, MLA)

7. Chiang, Shu-mei, and **Huei-ling Lai**. 2007. Grammaticalization and beyond: The invisible space behind the body-part term *boi55* (背) 'back' in Hakka. *Tsing Hua Journal of Chinese Studies* 37.1: 275-304. (TAHCI)
8. **Lai, Huei-ling**. 2007. Morphosyntactic and semantic changes of Hakka *si2*: A multi-level analysis. *Language AND Linguistics* 8.2: 553-574. (AHCI; TSSCI)
9. **Lai, Huei-ling**. 2006. Iconic coding of conceptualization: Hakka reduplicative constructions. *Language AND Linguistics* 7.2: 483-499. (CSA, LLBA, MLA)
10. **Lai, Huei-ling**. 2004. The syntactic grounding and conceptualization of Hakka BUN and LAU. *Concentric: Studies in Linguistics* 30.1: 87-105.
11. **Lai, Huei-ling** and Shu-mei Chiang. 2003. Intrapyschological force-dynamic interaction: Verbs of refraining in Hakka. *Taiwan Journal of Linguistics* 1.2: 35-64. (LA, LLBA, MLA)
12. **Lai, Huei-ling**. 2003c. The lexicalization patterns of verbs of hitting in Hakka. *Journal of Chinese Linguistics* 31.2: 340-361. (AHCI)
13. **Lai, Huei-ling**. 2003b. The semantic extension of Hakka LAU. *Language AND Linguistics* 4.3: 533-561. (CSA, LLBA, MLA)
14. **Lai, Huei-ling**. 2003a. Hakka LAU constructions: A constructional approach. *Language AND Linguistics* 4.2: 357-378. (CSA, LLBA, MLA)
15. **Lai, Huei-ling**. 2002b. From verb to verbal complement: Evidence from the verb DO in Hakka. *Form and function: Linguistic studies in honor of Shuanfan Huang*, ed. by Lily I-wen Su, Chinfa Lien, and Kawai Chui, 225-247. Taipei: Crane.
16. **Lai, Huei-ling**. 2002a. The grammaticalization of the verb DO in Hakka. *Journal of Chinese Linguistics* 30.2: 370-391. (AHCI)
17. **Lai, Huei-ling**. 2001. On Hakka BUN: A case of polygrammticalization. *Language AND Linguistics* 2.2: 137-153. (CSA, LLBA, MLA)
18. **Lai, Huei-ling**. 1999. Rejected expectations: two scalar particles CAI and JIU in Mandarin Chinese. *Linguistics* 37.4: 625-661. (SSCI)

## (二) 研討會/會議論文 (Conference papers / Proceedings)

1. Yeh, Chiou-shing, and **Huei-ling Lai**. 2010. A Study on the development of grammatical functions and semantic investigation of the morpheme *gam2* in Hakka (客語「敢」字之語法功能演變與語意探究). Paper to be presented at

- the 60<sup>th</sup> annual conference of the Chinese Linguistic Society of Japan, November 13-14, Kanagawa University, Yokohama, Japan.
2. Chiang, Shu-mei, and **Huei-ling Lai**. 2010. (從詞彙與構式看客語「好」字之多義情態功能). Paper presented at the 60<sup>th</sup> annual conference of the Chinese Linguistic Society of Japan. November 13-14. Kanagawa University, Japan.
  3. Li, Shi-min, and **Huei-ling Lai**. 2010. On (im)perfectivity of *do2/den2* (著) in Taiwan Hakka: A corpus analysis (台灣客語方言完成體與持續體之關係探索：語料庫觀點). Paper presented at the ninth International Conference of Hakka Studies(第九屆國際客家方言研討會). October 19-22. Chinese Academy of Social Sciences, Beijing, PRC.
  4. Chiang, Shu-mei, and **Huei-ling Lai**. 2010. Argument mismatches of Hakka potentiality-charactering constructions: Frames, constructions and information structure. Paper presented at the 6th International Conference of Construction Grammar (ICCG-6), September 3-5, Charles University, Prague, Czech Republic. **NSC 99-2411-H-004-185**
  5. Li Shih-min, and **Huei-ling Lai**. 2010. Understanding and classifying the causal relations of the causative BUN construction in Hakka. Paper presented at the 6th International Conference of Construction Grammar (ICCG-6), September 3-5, Charles University, Prague, Czech Republic. **NSC 99-2411-H-004-185**
  6. Li, Shih-min, and **Huei-ling Lai**. 2010. An analysis of (im)perfectivity of *do2/den2* (著) in Hakka: Interaction of lexical semantics and construction (客語兼表完成貌與持續貌「著」之探討：詞彙語意與構式互動的觀點). Paper presented at the 18th Annual Conference of the International Association of Chinese Linguistics (IACL-18) and the 22nd North American Conference on Chinese Linguistics (NACCL-22). May 20-22. Harvard University, Boston, USA. **NSC 97-2411-H-004-110-MY2**
  7. **Lai, Huei-ling**, and Yi-chen Chen. 2010. The effects of EFL learners' application of metaphor and metonymy in learning and retaining emotion expressions. Paper presented at 2010 International Conference on Applied Linguistics & Language Teaching (ALLT). April 15-17. National Taiwan University of Science & Technology, Taipei, Taiwan.
  8. **Lai, Huei-ling**. 2009. The interaction of lexical semantics and constructions: Present and Future (詞彙語意與構式的互動：現況與展望). Paper presented at Mini-workshop on Interaction of Lexical Semantics and Constructions (「詞彙語意與構式的互動」小型工作坊). November 14. National Chengchi University.

9. **Lai, Hwei-ling**, and Shih-min Li. 2009. The syntactic and semantic extension of Hakka GUO: A cross-categorical perspective (客語「過」構式句法及語意之衍變：跨範疇之觀點). Paper presented at the 59<sup>th</sup> annual conference of the Chinese Linguistic Society of Japan. October 24-25. Hokkaido University, Japan.
10. Chiang, Shu-mei, and **Hwei-ling Lai**. 2009. Form-function mismatches in non-referential interrogatives: A case study of *iu1 ma4ge3 ho2 V* constructions in Hakka (無指疑問句之「形式—功能」誤配：客語「有麼介好 V」構式研究). Paper presented at the 59<sup>th</sup> annual conference of the Chinese Linguistic Society of Japan. October 24-25. Hokkaido University, Japan.
11. Yeh, Chiou-shing and **Hwei-ling Lai**. 2009. Argument realization and transitivity of a family of *da2* constructions in Hakka: A Construction Grammar account (以構式語法觀點談客語「打」字群之論元體現與及物性). Paper presented at the 59<sup>th</sup> annual conference of the Chinese Linguistic Society of Japan. October 24-25. Hokkaido University, Japan.
12. Li, Shih-min, and **Hwei-ling Lai**. 2009. Interaction of aspect and construction: Evidence from the *ted4* construction in Hakka. Paper presented at the 17th Annual Conference of the International Association of Chinese Linguistics (IACL-17). July 2-4. EHESS, Paris, France.
13. **Lai, Hwei-ling**. 2009. The NCCU Corpus of Spoken Chinese: Hakka Corpus. Multi-Development and Application of Languages and Linguistics. May 15-16. National Cheng Kung University, Taiwan. (Invited speech)
14. **Lai, Hwei-ling**, and Shih-min Li. 2008. Hakka Corpus: Construction and Application (客語語料庫的建構及應用). Paper presented at the Eighth International Conference of Hakka Dialects. November 22-24. National Central University, Taiwan.
15. **Lai, Hwei-ling**, and Chiou-shing Yeh. 2008. Causative constructions with or without causative markers: Cases from Hakka. Paper presented at the 5th International Conference on Construction Grammar (ICCG-5). September 26-28. The University of Texas at Austin, USA. **NSC**  
**97-2411-H-004-110-MY2**
16. **Lai, Hwei-ling**, and Wan-jyun Luo. 2008. The implicit theme construction in Hakka: Interaction of structure, meaning and discourse. Paper presented at the 5th International Conference on Construction Grammar (ICCG-5). September 26-28. The University of Texas at Austin, USA. **NSC**  
**97-2411-H-004-110-MY2**
17. Chiang, Shu-mei, and **Hwei-ling Lai**. 2008. Argument Realization in the Family of Potential Constructions in Hakka. Paper presented at the 18<sup>th</sup>



- International Congress of Linguistics (CIL-18). July 21-26. Korea University, Seoul, Korea. **NSC 95-2411-H-004-026-MY2**
18. **Lai, Hwei-ling**. 2008. 'Putting' and 'Removing' in Hakka: Argument Structure, Constructions and Information Management. Paper presented at the 18<sup>th</sup> International Congress of Linguistics (CIL-18). July 21-26. Korea University, Seoul, Korea. **NSC 95-2411-H-004-026-MY2**
19. **Lai, Hwei-ling**. 2008. A Corpus-based Exploration of Research Topics on Hakka Grammar. 行政院客家委員會 96 年度補助大學暨獎助學術研究計畫成果發表會. June 6. National Central University, Taoyuan.
20. Chui, Kawai, **Hwei-ling Lai**, and Hui-chen Chan. 2008. The NCCU Corpus of Spoken Chinese: Mandarin, Hakka, and Southern Min. Proceedings of the 7th Annual Wenshan International Symposium (Theme: Spoken Corpus: Language Documentation and Applications), 68-100. May 19. National Chengchi University, Taipei.
21. Chiang, Shu-mei, and **Hwei-ling Lai**. 2007. Evaluating and comparing in an implicit way: A case study of NP1 V NP2 *m5* X constructions in Hakka. Paper presented at the 10<sup>th</sup> International Cognitive Linguistics Conference. July 15-20. Jagiellonian University, Krakow, Poland. **NSC 95-2411-H-004-026-MY2**
22. **Lai, Hwei-ling**, and Shu-mei Chiang. 2007. The family of Hakka idiomatic *gin2* X *gin2* Y constructions: Paratactic constructions with subordinate meanings. Paper presented at the 10<sup>th</sup> International Cognitive Linguistics Conference. July 15-20. Jagiellonian University, Krakow, Poland. **NSC 95-2411-H-004-026-MY2**
23. Luo, Wan-jyun, and **Hwei-ling Lai**. 2007. "piong3" 'put' and its variants in Hakka: Frames and Constructions. Paper presented at the 8<sup>th</sup> Chinese Lexical Semantics Workshop. May 21-23. The Hong Kong Polytechnic University, Hong Kong.
24. Chiang, Shu-mei, and **Hwei-ling Lai**. 2006. A Family of NP1 V NP2 *m5* X constructions in Hakka. Paper presented at The Fourth International Conference on Construction Grammar (ICCG 4). September 1-3. The University of Tokyo, Tokyo, Japan. **NSC 95-2411-H-004-026-MY2**
25. **Lai, Hwei-ling**. 2006. Lien...ya/du constructions in Hakka. Paper presented at The Fourth International Conference on Construction Grammar (ICCG 4). September 1-3. The University of Tokyo, Tokyo, Japan. **NSC 95-2411-H-004-026-MY2**
26. **Lai, Hwei-ling**. 2006. The interaction patterns of two-part allegorical sayings: A typological perspective. Paper presented at IACL 14 & IsCLL 10 Joint

- Conference. May 25-29. Academia Sinica, Taipei, Taiwan. **NSC 94-2411-H-004-007**
27. **Lai, Hwei-ling**, and Hui-shan Gao. 2005. Linguistic manifestation of human thinking process: Evidence from Mandarin Chinese. Proceedings of 2005 International Symposium on Body & Cognition: A Multidisciplinary Perspective, 13-1~12-20. June 4-5. National Taiwan University, Taipei, Taiwan.
28. **Lai, Hwei-ling**. 2005. A taxonomic framework for two-part allegorical sayings. Paper presented at 9th International Cognitive Linguistics conference (ICLC 2005). July 17-22. Yonsei University, Seoul, Korea. **NSC 93-2411-H-004-039**
29. **Lai, Hwei-ling**. 2005. The semantic division of labor of BUN, LAU, TUNG in Hakka. Symposium on 90-92 NSC Projects. May 28-29. National Normal University, Taipei, Taiwan. **NSC 91-2411-H-004-020**
30. Chiang, Shu-mei, and **Hwei-ling Lai**. 2004. Verbs of urging in Hakka: A perspective from force-dynamics. Proceedings of the Fifth Chinese Lexical Semantics Workshop, ed. by Ji Dong Hong, Lua Kim Teng, and Wang Hui, 316-320. June 14-16. National University of Singapore, Singapore.
31. **Lai, Hwei-ling**. 2004. Iconic coding of conceptualization: Hakka reduplicative constructions. Proceedings of the Second Workshop of Chinese Dialects (第二屆漢語方言小型研討會—客家語研究), 33-47. May 22-23. Academia Sinica, Taipei, Taiwan.
32. **Lai, Hwei-ling**. 2004. Mental space links between expectations: Hakka SI as an evaluative indicator. Proceedings of the Ninth International Symposium on Chinese Languages and Linguistics (IsCLL-9), 333-350. November 19-21. National Taiwan University, Taipei, Taiwan.
33. **Lai, Hwei-ling**. 2004. Same constructions but different grammaticalization paths of Hakka LAU and TUNG. Proceedings of the Fifth International symposium on Taiwanese Languages and Teaching (第五屆台灣語言及教學國際學術研討會), 195-210. May 22-23. Prudence University, Taichung, Taiwan.
34. Lin, Su-chu, and **Hwei-ling Lai**. 2002. On semantic relatedness of the Mandarin polysemous word *CHU*. Proceedings of the First Cognitive Linguistics Conference, 97-113. January 12-13. National Chengchi University, Taipei, Taiwan.
35. **Lai, Hwei-ling**. 2002. The semantic extension of Hakka LAU: Metonymy through event frames. Paper presented at IACL-11. August 20-22. Aichi Prefectural University, Nagakute, Japan.

36. **Lai, Huei-ling.** 2002. The metonymic process of Hakka LAU. Proceedings of the Third Workshop on Chinese Lexical Semantics, 41-56. May 1-3. Academia Sinica, Taipei, Taiwan.
37. **Lai, Huei-ling.** 2002. Hakka LAU constructions: A constructional approach. Proceedings of the First Cognitive Linguistics Conference, 66-81. January 12-13. National Chengchi University, Taipei, Taiwan.
38. **Lai, Huei-ling.** 2001. The grammaticalization of Hakka BUN. Proceedings of the Symposium on Selected NSC Projects in General Linguistics from 1998-2000 (八十九年度語言學門 <一般語言學> 研究成果發表會), 431-449. National Taiwan University, Taipei, Taiwan.
39. **Lai, Huei-ling.** 2000. The polygrammaticalization of Hakka BUN. Proceedings of the Seventh International Symposium on Chinese Languages and Linguistics (IsCLL-7), 213-231. national Chung Cheng University, Chiayi, Taiwan.
40. **Lai, Huei-ling.** 2000. The lexicalization patterns of verbs of hitting in Hakka. Proceedings of the 2000 NCCU Teachers' Conference on Linguistics Research, 200-211. National Chengchi University, Taipei, Taiwan.
41. **Lai, Huei-ling.** 1999. From verb to verbal complement: Evidence from the verb DO in Hakka. Paper presented at the International Symposium on Grammaticalization at Potsdam University, Germany.
42. **Lai, Huei-ling.** 1996. The conditionals vs. CAI and JIU in Mandarin Chinese. Paper presented at ICCL-5, National Tsing Hua University in Taiwan.
43. **Lai, Huei-ling.** 1995. Rejected expectations: two time-related scalar particles CAI and JIU in Mandarin Chinese. Proceedings of NACCL-7/ICCL-4, 151-168. Madison, Wisconsin.
44. **Lai, Huei-ling.** 1995. The *lian...ye/dou* construction in Mandarin. Paper presented at the 65th annual meeting of Linguistics Society of America, New Orleans, USA.
45. **Lai, Huei-ling.** 1994. The interaction between topic-comment and focus-background structures: a study of the *lian...ye/dou* construction. Proceedings of a conference in celebration of the 10th anniversary of the Journal of Semantics, 513-522. Heidelberg, Germany.

### (三) 專書 (Books)

1. Lai, Huei-ling 2002. Hakka LAU constructions: A semantic analysis. Taipei: Crane.

2. Lai, Huei-ling. 1995. Rejected expectations: two scalar particles CAI and JIU in Mandarin Chinese, Ph.D. dissertation, The University of Texas at Austin.

#### (四) 研究計畫 (Research projects)

執行期間	擔任職務	計畫名稱	補助單位
2010.08~ 2011.07	主持人 Project Investigator	客語中形式與功能之錯配：詞彙語意、構式語意及事件結構 Form and function mismatches in Hakka: Lexical semantics, constructions and event structure	國科會 NSC 99-2410-H-004-185
2010.01~ 2010.12	主持人 Project Investigator	客語語料庫之系統化建構與量化分析 The Construction of Hakka Corpus: Systemization and Quantitative Analysis	行政院客家事務委員會 Council for Hakka Affairs, Executive Yuan 99-C07
2010.01~ 2010.12	主持人 Project Investigator	詞彙語意、構式及事件結構：以語言使用為基礎的觀點 (2/2) Lexical semantics, constructions, and event structure: Usage-based perspective (2/2)	99 年度政大發展國際一流大學及頂尖研究中心計畫 99H443
2009.08~ 2010.07	主持人 Project Investigator	以變式為訊息管理之策略：以客語為例 (2/2) Using construction alternations as information-packaging devices: Evidence from Hakka (2/2)	國科會 NSC 97-2410-H-004-110-MY2
2009.01~ 2009.12	主持人 Project Investigator	詞彙語意、構式及事件結構：以語言使用為基礎的觀點 (1/2) Lexical semantics, constructions, and event structure: Usage-based perspective (1/2)	98 年度政大發展國際一流大學及頂尖研究中心計畫 98H443A
2008.08~ 2009.07	主持人 Project Investigator	以變式為訊息管理之策略：以客語為例(1/2) Using construction alternations as information-packaging devices: Evidence from Hakka	國科會 NSC 97-2410-H-004-110-MY2

2008.02~ 2008.09	主持人 Project Investigator	2008 語言學卓越營：功能語言學及其應用 Taiwan Summer Institute of Linguistics: Functional Linguistics and its Application	國科會 NSC 97-2420-H-004-004
2007.08~ 2008.07	主持人 Project Investigator	客語特殊比較句式家族之研究：語意網絡及格式 (2/2) Families of Hakka idiomatic comparative constructions: Semantic networks and constructions (2/2)	國科會 NSC 95-2411-H-004-026-MY2
2007.01~ 2007.12	主持人 Project Investigator	政大外語學院、選研中心 2-6 學術資料庫：客語口語語料庫 NCCU Corpus of Spoken Chinese: Corpus of Spoken Hakka	政大選研中心、政大頂尖大學計畫 Election Study Center, National Chengchi University
2007.01~ 2007.12	主持人 Project Investigator	客語語法研究議題的開發：以語料庫為本 A corpus-based exploration of research topics on Hakka grammar	行政院客家事務委員會 Council for Hakka Affairs, Executive Yuan
2006.08~ 2007.07	主持人 Project Investigator	客語特殊比較句式家族之研究：語意網絡及格式 (1/2) Families of Hakka idiomatic comparative constructions: Semantic networks and constructions (1/2)	國科會 NSC 95-2411-H-004-026-MY2
2006.01~ 2006.12	主持人 Project Investigator	身體、認知與空間：以客語的身體部位詞為例 Body, cognition, and space: Cases of body parts in Hakka	行政院客家事務委員會 Council for Hakka Affairs, Executive Yuan
2006.03~ 2006.10	共同主持人 Co-Project Investigator	2006 語言學卓越營：意義與語法 Taiwan Summer Institute of Linguistics: Meaning and Grammar	國科會 NSC 95-2420-H-002-006
2005.01~ 2005.12	主持人 Project Investigator	客語歇後語之語意分類架構 The taxonomic framework of Hakka two-part allegorical sayings	行政院客家事務委員會 Council for Hakka Affairs, Executive Yuan
2005.08~ 2006.07	主持人 Project Investigator	客語歇後語語意之機制：隱喻與轉喻的觀點 2/2 The semantic mechanisms of Hakka two-part allegorical sayings: a metaphoric and metonymic	國科會 NSC 94-2411-H-004-007

		perspective	
2004.08~ 2005.07	主持人 Project Investigator	客語歇後語語意之機制：隱喻與轉喻 的觀點 1/2 The semantic mechanisms of Hakka two-part allegorical sayings: a metaphoric and metonymic perspective	國科會 NSC 93-2411-H-004-039
2004.01~ 2004.12	主持人 Project Investigator	客語重疊結構之語意研究 Semantic analysis on Hakka reduplication constructions	行政院客家事務委員會 Council for Hakka Affairs, Executive Yuan
2003.08~ 2004.07	主持人 Project Investigator	客語「死」字之研究：詞類轉換與語 意延伸 On Hakka SI: Categorical shift and semantic extension	國科會 NSC 92-2411-H-004-022
2002.08~ 2003.07	主持人 Project Investigator	客語「分」、「搵」、「同」語意功能 分工 The semantic division of labor of BUN, LAU, TUNG in Hakka	國科會 NSC 91-2411-H-004-020
2001.08~ 2002.07	主持人 Project Investigator	客語近義動詞之詞彙化類型 Lexicalization patterns of Hakka near synonymous verbs	國科會 NSC 90-2411-H-004-013
2000.08~ 2001.07	主持人 Project Investigator	客語 LAU 字句的分析：時貌的觀點 On Hakka LAU constructions: an aspectual viewpoint	國科會 NSC 89-2411-H-004-041
2000.01~ 2000.12	主持人 Project Investigator	客語「分」字語法化的研究 The grammaticalization of BUN in Hakka	國科會 NSC 89-2411-H-004-003
1999.01~ 1999.12	主持人 Project Investigator	客語近義動詞的語意分析 Semantic analysis on Hakka near synonymous verbs	國科會 NSC 88-2411-H-004-022

# 客語語料庫之系統化建構與量化分析\*

計畫代號：99-C07

計畫主持人 賴惠玲 教授  
國立政治大學英國語文學系

## 提要

本研究旨在建構一系統化典藏臺灣客語方言書面及口語之客語語料庫，並運用此客語語料庫進行客語語言研究之量化分析。本研究建構之客語語料庫為一個基於後端架設的資料庫及前端使用者介面所架構成的語料庫平臺，後設資料庫目前收錄典藏約 670 篇文章，每篇文章均系統化紀錄其出處、作者、文體等相關資訊，前端使用者之網頁版介面已具文本輸入、文本查詢、關鍵字查詢及語料和使用管理等功能，以提供研究者使用此線上語料庫進行客語語意、句法、文本分析、對比分析、應用教學等學術研究。再者，本研究以語料庫為本之觀點，應用此客語語料庫之頻率分佈等量化分析佐證，用以檢視西方語言學以使用為本的相關學說與理論，本文以客語「著」構式及「怕」構式為例，示範如何運用語料庫進行語言學相關研究，並闡述此研究方法之重要性。

**關鍵詞：**臺灣客語方言語料庫、語料庫平臺、後設語言、介面功能、量化分析、構式語法、體貌、語法化、詞彙化

## 1. 研究動機及目的

根據客委會網站提供的 93 年度台灣客家民眾客語使用狀況，30 歲以下年青客家人有 3 成能聽解客語，僅 1 成可流利使用；在家庭語言方面，30 歲以下約有 60%使用台灣華語、20%使用台灣閩南語、未滿 10%使用台灣客語。曹逢甫、

---

\* 本計畫之參與人員包括：協同主持人真理大學英美語文學系助理教授強舒嫻；研究助理交通大學資訊科學與工程研究所碩士生聶家祺，政治大學語言學研究所博士生葉秋杏、李詩敏及碩士畢業生羅婉君、劉容瑜。本計畫之部分研究成果於「第九屆國際客家方言研討會」發表三篇論文（論文詳細資訊請參閱參考文獻），與會學者對本語料庫之建構及應用提出許多寶貴意見，特此銘謝。

黃雅榆（2002）以 S 型曲線點出客語語言流失的臨界點在 16-25 歲，語言流失一旦達到這個臨界點，在這之後的流失率就會產生滾雪球效應。再者，根據 2002 年德國科隆瀕危語言學會議所制訂的語言瀕危性鑑定的統一方法，客語瀕危性已屬瀕危語言的遭到侵蝕的語言（團體內部的一部分成員已轉用其他語言，而另一部份成員，包括兒童仍在學習使用的語言）。

爲了保存臺灣客語目前使用之形式，建立書面語或口語語料庫爲有效方法之一。語料庫的建構除了數位化典藏語言、保存語言文化資料之外，應用在語言學上可做詞彙、句法、語意的研究或者文本流暢及文體類型比較等分析研究，更可運用至其他相關學門的研究，如語文教學、跨語言比較的比較語言學、計算機輔助教學、自然語言處理、電腦語言辨識等。口語語料庫除了擁有以上羅列的語料庫優點，更可反映人們真實使用語言的行爲與習慣，不僅保存當代用語與口音，亦可幫助需要使用真實語料的學科直接使用口語語料庫的資源進行教學或研究。此外，語料庫還可以和統計資訊知識結合，以「數字會說話」的量化統計反映或預估語言真實使用的走向。

語料庫的用處在世界各語言已漸受重視，以漢語體系的語言爲例，在中文語料庫方興未艾、蓬勃發展之際，相形之下，不論是書面語或口語語料，客語語料庫的缺乏更爲明顯。因爲缺乏充足的語料，而增加客語研究的困難度，無法對客語進行更深入的探討，導致促進客語研究的能量不足，是以客語語料庫的建立更是刻不容緩。本研究建構奠基於本人過去研究的基礎上，包括本人於



2007 年執行完成之 96 年度客委會計畫「客語語法研究議題的開發：以語料庫為本」以及本人於 2007 年負責執行之政治大學外語學院、選研中心 2-6 學術資料庫「國立政治大學漢語口語語料庫」(<http://140.119.172.200/>)之「客語口語語料庫」。過去研究所建構的客語語料庫僅能單機檢索，電腦必須安裝 Apache 程式，利用 PHP、MySQL 等程式語言方能檢索語料；利用此方式建立的文本有以下缺點，例；需要手動斷行、語料分類不易、無法比較不同文體或類型之差異、無法進階檢索、無法條件限制、無法批次取代、無法搜尋取代；這些的語料處理上的困難及限制，一方面限制目前語料庫使用的方便性，一方面也限制了許多可開發的研究議題。針對以上階段所建構的客語語料庫缺失，本研究主要目的即在以系統化方式建構客語語料庫，改以 Python 和 PHP 程式語言撰寫客語語料庫介面，可建立新增文本，並整合過去建立之舊文本以放進資料庫，同時還有搜尋語料、進階搜尋、排序、批次修改等功能，並可透過線上連線、搜尋檢索，以避免前次語料更新繁複之步驟。此外，也希望透過客語語料庫的建構，使用者可利用大量語料開發更多客語研究議題，本研究就示範以客語「著」構式及「怕」構式為例之兩個構式的量化研究議題，結合語料庫語言學與統計學之研究方法，一方面使客語語料庫更系統化，一方面提升客語研究的量化分析之精準化及科學化。

## 2. 語料庫之介紹

以下將先回顧以語料庫為本相關研究之文獻，再簡介臺灣客語方言之系統架構。

### 2.1 語料庫為本相關研究之文獻回顧

以漢語語料庫為本的研究，目前以利用中研院平衡語料庫做現代漢語研究分析最多。如蔡美智等 (Tsai et al. 1998)、張麗麗等 (Chang et al. 2000)、劉美君等 (Liu et al. 2000)、黃居仁 & 洪嘉馥 (Huang and Hong 2005) 等研究由語料庫為本比較近義詞之表現。藉由詞彙語意研究分析，瞭解近義詞的相似和差異、詞彙的多義性、語意的呈現機制和延伸方式，進而整理漢語詞彙、編纂字典，以進行理論及計算機之應用研究，倡導漢語語料庫語言學研究；例如：黃居仁等 (1996, 1997) 以語料庫為本編纂國語日報量詞詞典；黃居仁等 (Huang et al. 2000) 及安可思等 (Ahrens et al. 2003) 根據語料庫發展出動詞語意表達模式 (Module-Attribute Representation of Verbal Semantics; MARVS) 理論架構；劉美君 (2002) 建構中文動詞詞網 (Chinese VerbNet)，以框架語義 (frame semantics) 為理論，以語料庫為本的方法研究分析漢語動詞；洪嘉馥 & 黃居仁 (Hong and Huang 2008) 利用兩岸不同單位所建構的語料庫，探索兩岸詞彙使用的差異現象，並藉此分析建立從語料庫中抽取兩岸對應詞彙的研究方法；鍾曉芳 (Chung 2007)、洪嘉馥 (2010) 分別利用語料庫界定隱喻源域、詞義預測。同時，語言

學分析應用於計算機資訊處理，使電腦可分析語言、學習新的語言知識、自動辨識分析未知詞；例如：黃居仁等（Huang et al. 1998）利用統計方式、電子辭典資源去分析名詞語意成分，以計質與計量方法支持語言學分析；陳克健&陳超然（Chen and Chen 1998）從帶有分詞及詞類標記的語料庫抽取出未知詞，以語法語意及統計機率關係表達開放型未知詞組合成分的構詞律模型，用以測度和辨識未知詞的複合詞合理程度；曾慧馨等（2002）則以合併兩種方式去計算構詞規律，以預測未知動詞的詞類。綜觀以上漢語研究，均是建構在語料庫為本所開發出的研究面向。

國外語料庫起步較早，規模完備，語體富多樣性，因此結合人工智慧、統計演算法、語言學等學門的計算語言學或語料庫語言學之應用研究已頗具成果。例如：Verspoor & Lowie (2003) 設計不同的教學法，並利用統計學的分析法以瞭解多義詞學習在何者教學情境可獲得最佳效益。Stefanowitch & Gries (2003) 採用費雪精確度檢定(Fisher exact test)證明在語言理論中語法結構的確可被視為標誌，並為心理語言學方面之語言習得提供一解釋分析方式。再如 Gries (2010) 以語料庫為本，採用統計方法證明英文的時態分佈（過去式、現在式）和英國或美國語料庫、語言使用時間（早期、晚期）有相關性；此外，Gries 運用統計學各種運算去檢驗、測試語言學分析應用（如：語意學的多義詞分析、歷史語言學的語言變遷、語言習得的先後與頻率關係），證明統計學確實可廣泛運用於語言學學門，但前提是需要有語料庫的奧援。

相較於國外的量化分析研究的成熟性，漢語語言學的量化統計囿於語料庫建構和取得的不易，大多相關文獻在處理台灣華語，如王珏瑄 & 黃美金 (Wang and Huang 2006) 探討台灣華語三個連接詞「因為」、「所以」和「然後」在口語語料庫的言談功能與分佈百分比關係，印證 Traugott (1995b) 提出的「副詞語法化趨向」。鍾曉芳等 (Chung et al. 2005) 採用建議上層共用知識本體 SUMO (Suggested-Upper-Merged-Ontology) 的定義隱喻詞彙，以協助使用概念領域分類詞彙、建立詞彙之間關係。鍾曉芳 & 安可思 (Chung and Ahrens 2008) 以量化比較和分析近義詞在語料庫的詞義，並以互見訊息 (Mutual Information) 的運算，比較不同近義詞的搭配詞，增加 MARVS 在語義分析的可檢驗性、運用在自動擷取詞彙語意上的基礎。

再觀以客語語料庫為本的研究，因為客語語料庫的欠缺、建構系統不完整及尚未對外公開等因素，而無法如台灣華語般蓬勃發展。台灣目前的客語語料庫相關建構資訊，於國立交通大學申請客委會 92 年度補助大學校院發展客家學術機構的計畫「國立交通大學客家文化學院籌設暨客家文史工作推動計畫案」

(申請內容，見客委會網頁

<http://www.hakka.gov.tw/ct.asp?xItem=8048&ctNode=1669&mp=298>)，計畫內容

顯示台灣客語自然口語語料庫為當初計畫預期目標，然而目前並未見到相關研究成果發佈於網路上。由中研院鄭錦全負責執行的「閩南語典藏-歷史語言與分布變遷資料庫」(<http://southernmin.sinica.edu.tw/index.htm>)，執行期間

2003/01~2006/12)，計畫目標之一為建立客家語語料庫，然目前並未見到客家語語料庫的上線。謝杰雄（2006）碩士論文，即在透過客語語料分析，以語料庫加工處理為目標導向，期建立客語分詞標準、詞類分類標準，建置一單機型客語語料庫為主，做為嘗試規劃和發展 WEB 語料庫的基礎。近期，客語研究學者紛紛體認客語語料庫的建構是刻不容緩的，語料庫的建構除了最基本的資料典藏數位化之外，可應用在其他學科的應用研究，因此近年來各學術單位或研究學者紛紛提出建構語料庫的計畫與構想；包括前面所提到的，本人於 2007 年執行完成 96 年度客委會計畫「客語語法研究議題的開發：以語料庫為本」、本人於 2007 年負責執行的「國立政治大學客語口語語料庫」，後續相關研究的迴響與出版包括：受邀於政大舉辦之第七屆文山國際學術研討會介紹政大客語口語語料庫（Chui et al. 2008）、於中央大學發表客委會 96 年度補助大學暨獎助學術研究計畫成果（Lai 2008）、於第八屆國際客方言學術研討會發表「客語語料庫的建構及應用」（Lai and Li 2008）、受國科會邀請演講「面對瀕臨死亡的語言我們能做些什麼？以客語為例」（賴惠玲 2008）、與政大國語口語語料庫主持人徐嘉慧一同撰寫介紹政治大學漢語口語語料庫之期刊（Chui and Lai 2008）、受國立成功大學邀請於語言多元化發展與應用國際研討會（MDALL）演講客語語料庫議題（Lai 2009）、經邀稿和審核後刊登於第八屆國際客方言學術研討會論文集之「客語語料庫的建構及應用」一文（賴惠玲 & 李詩敏 2010），以上均顯示客語語料庫的重要性為學界各領域接受與肯定。此外，謝杰雄（2008）延續碩士

論文研究，實際演示一線上客語語料庫「山哈客家語語料庫」

(<http://210.240.1.36/hakka/index.php>)，語料有斷詞、詞類加工，然語料庫註冊會員有使用權限的限制，需提供語料才能長久使用，並非全然免費對外開放；目前語料大多為書面語，缺乏可顯示真實使用情境的口語語料。中研院語言典藏計畫預計自 2007/01 至 2012/12 執行「閩客語典藏」計畫

(<http://minhakka.ling.sinica.edu.tw/bkg/index.php>)，該計畫，內容包括建立客語的語言典藏（包括長篇語料文本、辭彙與口語語音），目標在於建立以研究為主要目標的閩客語語音、詞彙、文本及地理資訊的資料庫，並提供完善的檢索。江俊龍執行之國科會 97 年研究計畫「台灣客家語語料庫之建置及應用」，該計畫以客語「口語和文本」為收集對象，進行用字規範、標音規範、錄製音檔、音檔轉寫文字、建立分詞體系與規範、建立詞類體系與規範、進行分詞與統計等項目，最終目的在建置 200 萬字的漢字客語語料庫，研發電腦斷詞程式，設計與建置語料庫 WEB 網站，架設伺服器。

從以上文獻回顧可發現，客語語料庫建構已引起各界的重視，除了保存客語語言、文化等最基本素材之外，應用在語言學上可做詞彙、句法、語意等分析研究，為語言學研究提供更為完整、縝密的分析基礎；更可再運用已建構的語料庫發展出各種與其他學門相關的研究，諸如語言教學、比較語言學、計算語言學、自然語言處理等。本研究計畫架構在本人過去客語語料庫所蒐集的語料素材、處理語料的經驗、客語研究的能量上，針對過去客語語料庫單機版的

缺點，期能以系統化方式整合過去的文本語料與現行政大客語口語語料庫之口語語料，書面語與口語語料並重，期能規劃建構出一可供連線作業的客語語料庫，除供檢索查詢之外，尚包括以各種條件，如作者、文體、出處、類型，篩選搜尋結果，以及進階設定共現詞彙功能、頻率統計功能、批次修改功能，希望藉由此功能周詳之客語語料庫，提供語言學研究者在考量詞彙語意之外，更可觀察詞彙語意與句法構式的互動，在更大的語境下勾勒出語言使用的實際情形，並輔以統計學門之專業知識，將語料庫應用在量化分析上，以「數字會說話」來讓客語研究更科學化、多元化及精密化。

## 2.2 臺灣客語方言語料庫之系統化架構

本語料庫之建構目的在於建置一個穩定、編輯者編輯與查詢介面親切 (user-friendly) 之系統。語料庫的穩定，其意義在於當系統文章量增多時，仍能維持一定之查詢速度，並能定期備份，保持資料之正確性。語料庫目前所使用的後設資料庫系統為 MySQL 3.2.0.1，操作介面為 python tkinter GUI，網頁程式為 php 5.3.0 及 html 語法，網站架設於 Apache 2.2.11 版本。目前語料庫中所記載的文章總數和相關訊息如〈表 1〉所示：

〈表 1〉臺灣客語方言語料庫相關資訊

項目	值
文章數	673
總句數	9822
最長文章所含長句數	211
最短文章所含長句數	1
長句平均字數	35.54

本系統開放的使用者權限分三個階層，其權限如〈表 2〉所示：

〈表 2〉臺灣客語方言語料庫系統權限分類

語料庫使用者	語料庫使用者為一般使用者，可以新增語料及查詢語料。
語料庫管理者	語料庫管理者除了新增與查詢語料外，負責維護語料的正確性，可以修改語料內容及刪除語料內容。
系統管理者	系統管理者負責維護系統運行，新增使用者及修改使用者權限。

### 2.2.1 後端架設資料庫簡介

語料庫後端架設資料庫共有兩個表單，分別為：文章標題 (Booktitle)、文章內容 (Bookcontent)。文章標題表單記載與文章出處相關的資訊，如〈圖 1〉所示，內容包括書名、出版社名、作者名、編者名、頁碼、文章名、出版日期、語料類別、建立者及建立日期，以利書本的反查及勘誤。文章內容表單為長句內容，如〈圖 2〉所示，利用標點符號「。」、「？」及「！」切分文章，將文章切成長句，以長句為儲存單位，儲存該句的字數。利用「。」、「？」及「！」



為切分單位，原因在於以此三個標點符號結尾的句子通常可做為一個語意段落完整的標記，故以此做為語料查詢的最小單位。

〈圖 1〉文章標題表單

bookPK	author	editor	bookname	secname	type	time	page	pub	builder	createtime
0	涂敏松	蔣宜琪	苗栗縣客語故事集 (二)	白頭公	Story	87.3	2		0	2010-01-21 13:30:15
1	涂敏松	蔣宜琪	苗栗縣客語故事集 (二)	田螺殼	Story	87.3	14		0	2010-01-21 13:30:15
2	張月英	黃碧蓮、江淑珍	苗栗縣客語故事集 (二)	猴仔	Story	87.4	26		0	2010-01-21 13:30:15
3	張水妹	林麗美、江淑珍	苗栗縣客語故事集 (二)	狗兄	Story	87.5	30		0	2010-01-21 13:30:15

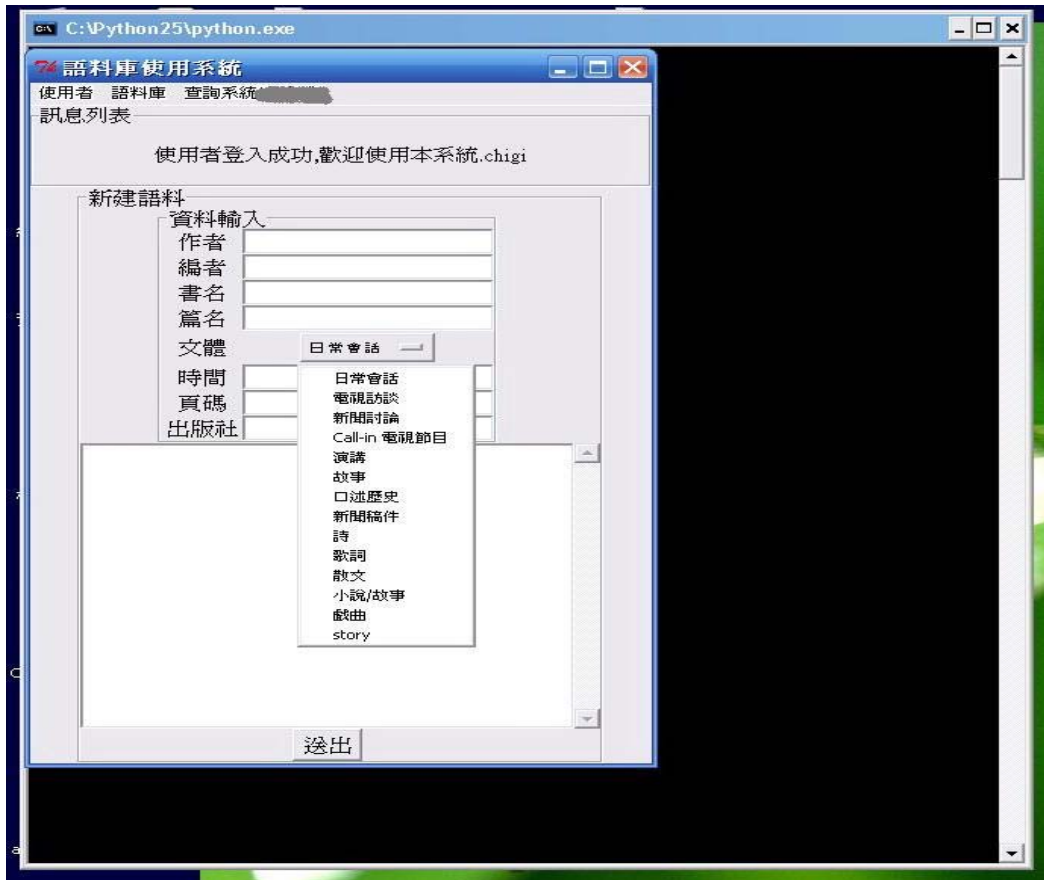
〈圖 2〉文章內容表單

bookPK	order	content	wordNum
0	0	在一个大樹林肚項，歇有一群白頭公仔，嚟燥就飛去河灘項食水，生活過來真快樂。	38
0	1	有一日，有一隻白頭公仔，異大隻又異聰明，佢自家喊自己係老大。	30
0	2	佢同大家講：「前幾日ngai5為到愛看天頂个天宮，飛過一條大河灘，發現河灘背个田莊，種有一大片个禾仔...	81
0	3	所有个白頭公仔聽佢怎樣講，全部就講好啊！	20
0	4	老大講：「en1這兜做毋得食歸串仔个穀子，農夫種田當辛苦，愛拈跌到地泥下个穀來食正做得，大家做得到無...	57

## 2.2.2 前端使用者介面簡介


語料庫前端使用者介面，目前以網頁形式呈現。計畫初期規劃擬以 Python 程式語言撰寫，利用 PHP 與 MySQL 作為系統及後設資料庫的開發工具，其使用者介面如〈圖 3〉所示：

〈圖 3〉以 Python 程式語言撰寫之前端使用者介面



而後考量使用者之便利性與親切性，改以網頁形式呈現，使用者可省去下載軟體安裝及面對陌生視窗之不便，故而後改以 php 及 html 語法撰寫網頁程式，網站架設於 Apache 2.2.11 版本，以網頁版介面形式呈現，其使用者介面如〈圖 4〉所示。因語言程式改寫之關係，故目前網頁版已完成改寫之功能包含：文本輸入、文本查詢、關鍵字查詢、語料和使用管理等功能，搭配詞查詢、批次修改、頻率統計等功能仍在改寫中。

〈圖 4〉以 PHP 程式語言撰寫之前端使用者介面

文章名	<input type="text"/>
作者名	<input type="text"/>
編者	<input type="text"/>
書名	<input type="text"/>
頁碼	<input type="text"/> 例如 p. 20-p. 25
出版社	<input type="text"/>
出版時間	<input type="text"/> 例如 87.5
語料類別	<input type="text" value="日常會話"/> 
	<ul style="list-style-type: none"><li>口語對話</li><li>日常會話</li><li>電視訪談</li><li>新聞討論</li><li>Call 語音轉錄</li></ul>

關於系統運作方面，文本輸入功能<sup>1</sup>需由熟稔客語用字之人員新增及輸入文本，因此輸入人員必須熟悉輸入格式，確保輸入格式正確，也必須瞭解哪些用字已有教育部公布之標準用字，哪些用字目前尚未有標準用字，且輸入者在遇到雖已有標準用字但一般輸入法卻無法輸入的難字時，需知道如何輸入相對應的拼音及調號。為預防輸入者輸入錯誤，在其新增文本時，系統會紀錄新增者的資訊及新增時間，方便語料庫管理者校對及修正。

關於客語用字之處理，需特別提出說明的是，雖然目前教育部已頒佈部分規定用字，但部分用字無法以一般輸入法輸出，進而有所謂的「難字」的產生。針對這些難字，目前處理的方式是改以輸入其相對應的拼音加調號；例如「𑄎」以「ngai5」標記，並以空白區隔連續的難字。針對難字部分，為了在新的難字

<sup>1</sup> 文本輸入與新增介面如〈圖 4〉所示。

出現時，輸入者能正確且快速地修改語料庫中相同的字，因此程式設計師特地建立一反查用表格，即「相關詞彙」(Relatedword)。「相關詞彙」為一記載相關詞用法的紀錄，例如「涯」與「ngai5」為相關詞彙，表格紀錄此相對應資訊，若是有新的用字時，使用者可利用此表格進行批次取代或代換搜尋。

修改文章的功能包括修改標題及修改內容，但僅有文章輸入者本人及語料庫管理員以上權限的使用者可修改標題，如此可避免語料混亂。如欲修改標題，可藉由搜尋方式，修改者可根據搜尋結果來判定文章是否要修改標題。修改文章內容的功能，可以藉由單篇文章內容修訂以及多篇文章用字的批次修改之方式進行修改。

查詢功能主要分文本查詢、關鍵字查詢兩種方式，細目如〈圖 5〉所示。文本查詢功能，可按文章名、作者名、書名等方式查詢，查詢結果為文章表單列表，如〈圖 6〉所示；如欲進一步查詢每一篇文章的內容，可再點擊查詢結果。關鍵字查詢功能，可輸入關鍵字查詢，查詢結果為含有該關鍵字之句子，如〈圖 7〉所示，可點擊文章名，進一步瞭解全篇文章內容。

〈圖 5〉查詢介面

搜尋文章名	<input type="text"/>	<input type="radio"/> 模糊比對	<input type="radio"/> 完全比對
搜尋作者名	<input type="text"/>	<input type="radio"/> 模糊比對	<input type="radio"/> 完全比對
搜尋書名	<input type="text"/>	<input type="radio"/> 模糊比對	<input type="radio"/> 完全比對
搜尋關鍵字	<input type="text"/>	<input type="radio"/> 模糊比對	<input type="radio"/> 句首 <input type="radio"/> 句尾
<input type="button" value="新查詢"/>			

〈圖 6〉文本查詢結果（文章表單列表）

書名
<a href="#">苗栗縣客語故事集（二）</a>
<a href="#">頭前溪个故事</a>
<a href="#">臺灣桃園客家方言</a>
<a href="#">東勢鎮客語故事集</a>
<a href="#">東勢鎮客語故事集七</a>
<a href="#">東勢鎮客語故事集二</a>
<a href="#">東勢鎮客語故事集三</a>
<a href="#">東勢鎮客語故事集五</a>
<a href="#">東勢鎮客語故事集六</a>
<a href="#">東勢鎮客語故事集四</a>
<a href="#">油桐花下个思念</a>
<a href="#">客家雜誌第65期</a>
<a href="#">客家雜誌第66期</a>
<a href="#">客家雜誌第67期</a>
<a href="#">客家雜誌第68期</a>
<a href="#">客家雜誌第69期</a>

〈圖 7〉關鍵字查詢結果

文章名	頁碼	
<a href="#">白頭公</a>	2	在一个大樹林肚項，歌有一群白頭公仔，喺燥就
<a href="#">白頭公</a>	2	有一日，有一隻白頭公仔，異大隻又異聰明，佢
<a href="#">白頭公</a>	2	所有个白頭公仔聽佢怎樣講，全部就講好啊！
<a href="#">白頭公</a>	2	一大群个白頭公仔開始飛啲，無幾久仔就飛過一快樂，飛等過去。
<a href="#">白頭公</a>	2	等這群白頭公仔飛到田項，愛拈跌到地泥下个剝高个地方，向四面看，發現係舊去佢重擡架一個

目前尚未改寫完成的搭配詞查詢、批次修改、頻率統計等功能仍在改寫中。

其中的搭配詞查詢功能，在於提供使用者輸入兩個詞彙，查詢結果為兩個詞彙

的共現長句，並可進一步設定兩個詞彙之間的距離，以設定詞彙關係的緊密度；以客語「緊...緊...」結構查詢為例，輸入的兩個關鍵字即為「緊」、「緊」，查詢結果即包含「緊走緊快」此類詞組。批次修改功能，在於提供一次性大量修改用字或錯誤，可節省逐個修改的時間。頻率統計功能，在於提供關鍵字詞出現的個數，以便量化分析使用。

為協助語料庫管理者管理語料，程式設計提供語料庫管理者可查詢最新輸入的文章，亦可查詢特定輸入者的文章，如此可用以檢驗文本輸入之正確性與否，並可進行後續之文章勘誤。

目前語料庫所含文本主要為書面語之故事集、客家雜誌，未來將收錄政治大學口語語料庫之口語語料；同時在資源整合上，包括國立新竹教育大學台灣語言與語文教育研究所葉瑞娟教授、國立新竹教育大學英語教學系黃漢君教授與吳睿純教授等諸位教授與所屬研究團隊所建構之客語經典歷時文本、客語戲劇節目對白、客語對答問卷內容等，均將一同彙整至此客語語料庫內。目前，已有不少國內外專家學者詢問此客語語料庫之使用方式，然而因書面語料、網路語料因版權問題尚未解決，故語料庫仍無法公開對外開放檢索，此問題仍待各單位之努力與合作，方能俾使此客語語料庫對外開放使用。同時，因不同文本之差異性大，涉及語料儲存方式之細節，仍需要更多時間撰寫和微調程式，俾使語料庫語料更具多樣性。

本研究目前已完成客語語料庫架構及重要的資訊系統部分，計畫成果之

貢獻大致如原預期，且也吸引客語研究者的高度興趣與詢問，足見客語語料庫系統化建構和量化分析以及本研究計畫之重要性。未來研究目標，希望朝向以語料角度為主的研究，並希望能夠結合語言學與其他學門，如計算語言學、資料庫研究等之合作，希望發展重疊詞搜尋功能、客語自動斷詞程式、客語自動標音程式、量化方面的頻率計算或字數統計、並進而系統化建立電子化之客語辭典及客語和其他語言之多語辭典，期冀客語語料庫之檢索系統與應用更加完善。

### 3. 應用客語語料庫之語言學研究分析

以下將 3.1 節將簡介語言學研究之相關理論背景，以及根據此學術理論和應用本計畫之客語語料庫所呈現之量化分析部分研究成果，3.2 節介紹客語「著」構式，3.3 節介紹客語「怕」構式，藉此示範如何運用語料庫進行語言學相關研究，並闡述此研究方法之重要性。

#### 3.1 研究之理論背景

本研究計畫採語料庫為本之方式檢視語言在真實語境中使用之形式與功能，故採用以語言使用為本之觀點的構式語法 (Construction Grammar) 為理論架構。構式語法堅信語意是由語境而生，認為形式和語意均屬句法的一部份；由於構式語法具使用者為主之特質，並在跨語言中尋求通性，故構式語法研究

範圍包括語言的核心及邊緣現象 (Kay 1995)。構式語法的最基本成分是同時承載形式、語意的結構，可小至詞素、大至篇章，包括詞彙固定的成語、部分詞彙固定的片語、全部是抽象模組的結構，並具特殊語意、語用功能或言談語境；也因此即便是衍生語法置於詞彙儲藏室處理的成語，構式語法也將其看成是有固定形式和規律性，因而構式語法認為句法和詞彙並非是獨立分開的模組，而是一連續性的概念 (Kay and Fillmore 1999; Goldberg 2006)。由於結構是形式和語意配對而成，不但具普遍性，亦有其獨特性，兩者不可一切為二，故有構式群或構式家族的概念產生 (Goldberg & Jackendoff 2004)；基於簡易性以及規則性，語言中最普遍的結構多半是最典型的形式，為了溝通的目的，新結構會因不同的語用或言談功能需求而由原本的結構認可產生，因此，不論是規則性或獨特性結構，都具共通特性。

對於動詞語意和構式的映射處理，語言現象顯示個別動詞的細微語意差異與構式語意密切互動，包括語意相近、句法行為相似的溝通動詞 *talk*、*speak*、*whisper* 不能同時合法出現在結果構式；同屬同一語意類別的攝取動詞 *eat*、*devour*、*swallow*，句法替換和論元結構卻不同，因而許多學者呼籲，在處理構式語意時必須同步考慮動詞詞彙語意的內涵 (Boas 2003, 2008a, 2008b; Croft 2003)。因此，構式語意和詞彙語意具高度相關性，每個構式語意的基模都需規範可一起出現的動詞分類為何；是故，詞彙和構式的關係並非僅是單純的二分法，詞彙語意、時貌等構式成分均在構式中扮演重要角色，



且彼此間的互動需參照詞彙所引出的世界文化等框架知識、由構式而形成的語意和句法映射的句法訊息，如此才能解釋詞彙語意和構式之間的互相搭配或強制轉換情形。這也是本研究計畫認為構式語法能適切解釋不同體貌的體貌表徵，及其與共現詞彙的搭配互動關係。

對於論元體現的處理，傳統觀點認為句子是由兩個部分，即主語和賓語所組成。然而，主語和賓語這兩個句法論元並非全然和語意論元的施事者和受事者相呼應，如以下例子所示 (Fillmore 1970)：

(1a) John hit the tree.

(1b) A rock hit the tree.

在(1a)，主語 *John* 為施事格，但(1b)主語 *a rock* 卻為工具格，而其它論旨角色，如處所、終點、刺激等都可出現在主語位置；因此，構式語法學家將重點放在形式及語意功能之間的關係。

除了共時的語言現象之外，語言的使用常常與歷時的改變息息相關，經常有詞彙化 (lexicalization) 或語法化 (grammaticalization) 的情形。Traugott (1995a, 2005) 認為語法化是一種語言過程，可以是歷時性也可以是共時性。Bybee, Perkins, & Pagliuca (1994) 對於語法化的詮釋則為，語法化是指詞彙語素 (lexical morphemes) 逐漸發展成為功能語素 (grammatical morphemes)，或者是詞彙語素結合另一個詞彙或功能語素後，發展為功能語素的一種過程。例如，英文的 *be*

*going to* 即是由原先表達移動的動詞發展而成表示時態的助詞語法功能。

在情態意義方面，Lyons (1977) 根據情態邏輯將情態意義分成兩類，第一類為「知識性情態」(epistemic modality)，表說話者在談論命題之可能性和必要性時，參照自己的信念或意見所表現出來的態度；第二類為「義務性情態」(deontic modality)，表說話者的行為在情境中表達要求、許可、規定、義務、禁止等意義。Palmer (1979) 則將情態意義區分為三類，除知識性情態及義務性情態外，還有「動力性情態」(dynamic modality)，用以表達施事的能力以及自主性。英文情態詞之重新分析 (reanalysis)，更加佐證根源義務性情態與知識性情態之動因關係以及在語言表現上之多義性形式，而情態詞功能的改變和其前後詞彙之改變與共現亦是密不可分，故本研究計畫採用構式語法之理論分析客語「怕」構式所出現的語境，以檢視「怕」與其前後詞彙之關係及「怕」之功能的改變。

### 3.2 臺灣客語方言完成體與持續體

誠如蔣紹愚 (2005) 所言，儘管古漢語和方言的一些語法形式現在雖然不一致，但卻有關係。從共時或歷時的文獻可看出，古漢語和部分漢語方言使用同一個標記表示「完成」與「持續」，例如古漢語和閩南語「著」、吳語「仔」、湘語「達」(梅祖麟 1989；楊秀芳 1992；孫朝奮 1997；吳福祥 2004；蔣紹愚 2005，2006；陳前瑞 2009)。例如，(2a)和(2b)的長沙話「達」(邢公畹 1979) 分別標誌完成體、持續體，(3a)和(3b)的吳語「仔」(Mei 1979) 分

別為完成體詞尾、持續體詞尾，(4)的閩南語「著」(楊秀芳 1992)表持續和先行兩種歧義<sup>2</sup>。

- (2a) 他吃達飯達
- (2b) 坐達吃
- (3a) 吃仔飯哉
- (3b) 騎仔馬尋馬
- (4) 阿英穿著

關於以上所提，古漢語和漢語方言使用同一形式兼表完成體和持續體之探討，雖過去文獻均認為兩種體貌之間具相關性，但對於其來源、孰先孰後之語法化演變發展途徑，仍莫衷一是(Mei 1979; 劉丹青 1995; 錢乃榮 2002; 吳福祥 2004; 蔣紹愚 2005, 2006; 陳前瑞 2009)。

臺灣客語方言的完成體念 do31，一般採用的漢字為「著」，而持續體念 den31，一般採用的漢字為「等」，如以下例子所示。出自《苗栗縣客語故事集》的例句(5a)、(6a)，出自《東勢鎮客語故事集》的(5b)、(6b)，出自《客家話小王子》的例句(5c)、(6c)，以及出自《臺灣客家語常用詞辭典》的(5d)、(6d)，雖各典籍的作者不同，但均同時使用「著」標誌完成體、「等」標誌持續體。

---

<sup>2</sup> 「著」的語音形式會因方言或其在句中位置而有不同讀法(楊秀芳 1992)，**錯誤! 找不到參照來源。**之「著」應讀作 le<sup>0</sup>。

- (5a) 你這兜做鄰長、村長較大人，印仔打落去，講清楚題錢係爲到造橋，人正毋會恗**著**佢騙錢。《苗栗縣客語故事集，頁 90》
- (5b) 啊就，逐日該家僮滿哪去河壩，去尋，尋**著**一個釣魚个。《東勢鎮客語故事集，頁 62》
- (5c) 佢拿吾傑作分大人看，問佢兜會分這張圖嚇**著**麼？《客家話小王子，頁 3》
- (5d) 阿哥利用暗晡頭去打工，逐隻月做得賺**著**五千個銀。《臺灣客家語常用詞辭典》
- (6a) 等到暗晡頭，時間差毋多了，就摻飯菜拈到桌項，大家房間**等**。《苗栗縣客語故事集，頁 6》
- (6b) 啊就企緊煞毋會死，無死又無想落來，企**等**恬恬。《東勢鎮客語故事集，頁 20》
- (6c) 這張圖係畫一尾蛇哥擘**等**嘴愛將一條動物吞落肚屎去。這就係該張圖个複製。《客家話小王子，頁 3》
- (6d) 朝晨當多學生仔背**等**書包仔去上課。《臺灣客家語常用詞辭典》

從漢語體系以及語言類型學觀點來看，和北方話、吳語、閩語、湘語等語系同屬於漢語體系的臺灣客語方言應也存有使用同一詞彙兼表完成體和持續體之用法。然而，臺灣客語方言採用不同的漢字詞彙標誌完成體、持續體，究竟此兩種體貌的本字是否同源？完成體和持續體在臺灣客家方言中的關係爲何？本文將藉由語料庫觀點分析臺灣客家方言完成體和持續體在實際語料中之分佈情形，並採構式語法( 諸如 Kay 1995; Kay and Fillmore 1999; Goldberg 1995, 2006; Croft 2003; Boas 2008a, 2008b )之理論架構探討完成體及持續體構式的句法特徵及和這兩種體貌共現的動詞語意類型，用以分析此兩種體貌之本字及其句法和語意關係。以下分別將就臺灣客語方言完成體、持續體的本字、句法語意關係、分佈情形逐步探討。

### 3.2.1 臺灣客語方言完成體、持續體之本字考究

孫朝奮 (1997) 指出只有南方方言還保留「著」字完成體用法，梅祖麟 (1989) 則認為保留「著」字完成體用法是南方話的特點之一，且臺灣客語方言均一致地採用「著」標誌完成體，故臺灣客語方言完成體標記之本字採「著」，拼音為 do31，此較無爭議。

臺灣客語方言持續體的漢字使用，一般採「等」，其拼音為 den31。然而，徐兆泉所編之《臺灣四縣腔海陸腔客家話辭典》，有關「等」之說明如下：客語借用字；表進行式的助詞，意同北京話的「著」。因此，採用「等」字標誌持續體，可能是因為臺灣客語方言持續體的讀法接近臺灣華語「等」字，故借「等」的音來標誌持續體，為音近而借自臺灣華語之借用字。而我們認為臺灣客語方言表持續體的本字應該是「著」，和表完成體之本字相同，則臺灣客語方言也有一字兼表完成體與持續體的用法<sup>3</sup>，其考量有以下幾點原因。第一，從語言類型學觀點來看，和北方話、吳語、閩語、湘語等語系同屬於漢語體系的臺灣客語方言應也存有以同一形式「著」兼表完成體和持續體之用法。第二，雖有其他學者認為臺灣客語方言表完成體和持續體標記的讀音不同故兩者並無關係，然而現代漢語「著」完成體標記讀音 zhao2、持續體標記讀音 zhe5，古漢語卻是同源。第三，經私下請教臺灣客語本字研究專家—前台北市立教育大學應用語文研究所教授—古國順教授，古教授認為標誌持續

---

3 本文所探討的臺灣客語方言以「著」表持續貌之用法適用於臺灣使用人口最多的兩大客語次方言，即：四縣客語、海陸客語；並不適用於大埔客語，因大埔客語以「緊」標誌持續貌（江敏華 2007）。

體「等」之本字應為「著」，「等」乃音近之借字；此觀點可從梅祖麟 (1989) 所提的避免功能負擔 (function load) 太重之說法來解釋。第四，持續體標記 den31 可能經歷了語音演變，《1994 汕頭版客語聖經》即採用「著」字標記羅馬字「ten」。因此，從語言類型學觀點、漢語體系實證、古教授看法，以及語音演變及客語聖經記載，均可以大膽推測，臺灣客語方言持續體標記的本字應為「著」，故臺灣客語方言和漢語體系相同，也有採用同一詞彙，即「著」字，兼表完成體和持續體之用法。

然而，語音推演上，除內部構擬之外，若能藉由和其他同語系之語言相互比較建構之比較法，則更具可信度。故若能採錄得到其他客語方言之第一手語料，藉由比較構擬出客語方言持續體標記的原音，將有助於其本字考究。

### 3.2.2 臺灣客語方言完成體、持續體之句法語意關係

本節分別探討「著」之完成體及持續體的句法語意特徵。句法上，完成體標記「著」常與標誌已然體的情態助詞「有」(鄭綮 2005) 共現，但完成體「著」不能與持續體標記「緊」(Lai and Chiang 2007; 江敏華 2007) 共現，如下例(7)：

(7) 佢兜頭下有 (\*緊) 略略仔驚著喔

推究其原因，(7)的「有」標誌已然，表示動作的完成或情況發生變化，為一完整體（張愛民 1992），與「著」標誌完成體、「著」與其前面動詞表瞬成述語之體貌特性相符；而「緊」標誌持續體，和「著」表完成體、「著」與其前面動詞表瞬成述語之體貌特性相悖，故無法共現。

此外，「V-著」表完成體時，「著」為法相標記 (phase marker)<sup>4</sup>，為一動補結構，允許能願標記「得」（或其變體「以」、「仔」）和「毋」插入，如例(8a)及(8b)所示。因此，能願標記的插入為完成貌「著」構式之特徵之一。

- (8a) 恁細个字佢還看 **得/以仔 著**，目珠實在當利  
(8b) 歸大群仔鳥仔，打紙炮仔嚇**毋**走，**毒咩毒毋著**

語意上，完成體標記「著」對其前面的共現動詞有選擇性限制，可和認知類動詞、會面類動詞、情緒類動詞、投擲類動詞合用，例子分別如下所示：

- (9a) 隔壁班个老鼠妹知**著**，一定會毋搭佢俚啦  
(9b) 今朝晨在路項睹**著**塞車  
(9c) 頭擺人講細人仔煞**著**，愛去收驚  
(9d) 頭先無細義打**著**你，實在當敗勢

此外，感知類動詞也可以和「著」合用，且語意和情緒類動詞一樣均偏向負面，獲得類動詞也可以和「著」合用，但語意偏向正面，如例(10a)及例(10b)：

---

4 法相標記及體貌標記之判定方式參見 Lien (2005)。

(10a) 寒天愛著加兜仔衫，正毋會冷著

(10b) 老妹這擺演講比賽得著一面金牌

即便是同類型動詞，仍要考慮動詞的細微語意以及動詞語意和構式其他成分的搭配。參看以下(11a)至(11c)的例子，同屬姿態類動詞的「坐」、「企」，例(11a)可接受、(11b)卻不能接受，因為凳仔的本質本就是分人坐，而不是\*分人企，故而(11a)合法、(11b)不合法。倘若將句子的其他成分代換，「企著」可以是合法的，如下例(11c)表示以站姿取得位子。因此，同屬一類型之動詞是否能與「著」共現而表完成貌，仍應考慮動詞語意細微差異及其與構式成分的互動關係，動詞語意與構式語意的融合仍有其約制，以免過度概化

(over-generalizations)。

(11a) 該凳仔分佢坐著

(11b) \*該凳仔分佢企著

(11c) 僇等暗久正企著這隻位仔，麼人都做毋得喊僇走

句法上，與持續體「著」共現的「有」標誌習慣，不可插入能願標記但可和「緊」共現。持續體「著」標記可與標誌持續體的「緊」共現，但其共現有所限制，「緊」只能和表動作動詞的持續體「V-著」合用，表示其動態之變化，如例(12a)，但不能和以處所倒裝構式體現的表狀態動詞的持續體「V-著」合用，如例(12b)。



(12a) 隨著時代個進步，社會也（有）/（緊）踴著變化

(12b) 聽講該間屋（有）/（\*緊）匠著當多金銀財寶

在以上探討曾提到標誌已然體的情態助詞「有」可和完成體「著」共現，而非持續體「著」也可以和「有」共現，如上例(12a)及(12 b)所示。此處的「有」並非是標誌已然體，而是標誌習慣。和表完成體的「著」不同的是，表持續體的「著」已發展為一體貌標記，故「V-著」表持續體時，無法插入能願標記。

語意上，出現在處所倒裝構式，可和持續體標記「著」共現的動詞類型有：放置類動詞、印記類動詞、存在類動詞、捧提類動詞、穿戴類動詞。前三類動詞的主語為典型處所，後兩類動詞的主語為擴大意之處所，即以捧提或穿戴之身體部位表處所，如下例(13)之「頭那」。

(13) 頭那時時戴著笠嘛，熱天出門正無會熱著

因持續體「著」表示狀態或事件的持續，故其要求共現的動詞語意具持續伴隨或接續，如前面所提處所倒裝構式的五類動詞本身都可表示動作開始施行後的狀態持續；除存在類動詞外，其他四類動詞可以不必藉由處所倒裝結構之語序就可指涉持續貌。其他類動詞，如伴隨類動詞、接續類動詞、投擲類動詞，均可與持續體「著」共現，指涉動作之持續，請參閱以下的例子

(14a)至(14c)：

- (14a) 山狗太陪著鄉間个孩童成長，都市叢林係尋毋著其个蹤跡  
(14b) 新竹市關東橋連著下員山、竹中塹頭重埔一帶，還係良田百頃，  
    風光當好  
(14c) 擲著球仔擩石頭，盡好搞

此外，如果「V-著」出現在連動式結構，在訊息結構裡擔任背景功能

(background)，請看例子(15)：

- (15) 第三遽个係畫眉仔，其對該兩儕个頭那頂飛著過

動作動詞「飛」所形成的「V-著」構式均擔任背景事件功能，表達前景動作施行的方式，「飛著」表達前景動作（即畫眉鳥在那兩人頭上經過）施行的方式。<sup>5</sup>

### 3.2.3 臺灣客語方言完成體、持續體之語料分佈情形

本文所採用的臺灣客語語料，其來源為「客語語料庫」（含國立政治大學客語口語語料庫，網址 <http://140.119.172.200/>，及客語書面文本）。

出現在完成體標誌「著」前的動詞類型，不論是書面語或是口語語料，

---

5 陳前瑞（2009）也提到「V-著」在晚唐詩句做背景性事件用法。

皆包含以下六大類：認知類、會面類、情緒類、感知類、獲得類及其他類等。各類動詞在書面語語料中的分佈百分比如下：感知類 38.5%、獲得類 24.4%、認知類 18.9%、其他類 9.4%、會面類 5.5%、情緒類 3.3%；在口語語料中的分佈百分比如下：感知類 34.5%、獲得類 29.8%、認知類 26.2%、其他類 7.1%、會面類及情緒類均各 1.2%。比較其分佈，動詞類型的分佈排序相同，惟口語的認知類比例較書面語高。分析此兩者的分佈差異，其可能之原因在於目前收錄的口語語料體裁大多為日常生活會話類型，說話者大多在陳述或閒聊自己所遭遇與接觸的經驗，故在口語語料中，認知類動詞比例較高。

出現在持續體標記「著」前的動詞類型，書面語包含以下九大類：放置類、印記類、存在類、捧提類、穿戴類、伴隨類、接續類、身體姿態類、其他類；口語則僅包含七大類，其前面動詞類型，目前在口語語料庫中並未出現印記類、穿戴類動詞。各類動詞在書面語語料中的分佈百分比如下：身體姿態類 24.3%、伴隨類 23.7%、接續類 13.9%、其他類 13.6%、捧提類 12.3%、存在類 6.0%、放置類 3.0%、穿戴類 2.2%、印記類 1.1%；在口語語料中的分佈百分比如下：其他類 67%、身體姿態類 15%、放置類 7%、捧提類 5%、伴隨類 4%、接續類及存在類均各 1%。分析此兩者的分佈差異，在口語語料中的其他類比例最高，乃在於口語語料的語料比數較少，故樣本數分佈目前無法呈現常態性分佈，因此其他類比例較高。而在書面語料部分，雖然捧提類也運用到身體部位，然而捧提類動詞有時會出現在處所倒裝構式裡，而處所

倒裝構式的主語並非最典型的施事者；<sup>6</sup> 語料呈現亦顯示，但凡借用處所倒裝構式的動詞類型，其分佈比例較低；此點，可能與認知語言學的典型和非典型有關。而伴隨類和接續類動詞的比例均不低的原因乃在於他們和表持續體的「著」共現時，不但可以指涉空間，亦可以指涉時間（如：*踰著*；*續著*）。

值得注意的是，前面我們所造出的投擲類動詞，需視後面所接為目標或客體而表不同體貌（如：*擲著*細阿妹；*擲著*石頭），但目前語料庫裡尚未存在此類動詞。但凡大多數動作動詞，出現在「著」前，皆可兼表完成體和持續體（如：*看著*；*捉著*），但真實語料常有語境可提供各種訊息，使我們得以判定「著」表完成體或持續體。但會面類、獲得類動詞只出現在完成體標記「著」之前（如：*睹著*；*贏著*），存在、放置、印記、伴隨、接續類等動詞只出現在持續體「著」之前，且存在類動詞一定要以處所倒裝結構體現（如：*寫著*；*劫著*）。換言之，「著」雖具雙重功能，但完成體標記「著」對其共存動詞語意類型的選擇性較持續體標記「著」為窄。這也符合了標誌持續體的「著」之時貌標記的語法化程度較標誌完成體標記的「著」之法相標記為高之語法化現象。

語料顯示，若無適當語境，大多動詞類型出現在「著」之前，均可能指涉完成體和持續體；採用構式語法之理論分析，以語言使用為出發點，語境往往可提供充分的訊息。未來若口語語料更多時，可再繼續探究其前面動詞

---

6 典型施事者與典型受事者可參見 Dowty (1991) 所提之「論元選擇原則」(Argument Selection Principle)。

的分佈是否以感知類等動詞為多，可再深究口語語料和主觀化 (subjectivity) 的關係，亦可再探究處所倒裝構式等違反論元映射關係之語料的分佈情形。

### 3.3 臺灣客語「怕」構式之語法化與詞彙化

心理動詞的兩個論元之語意角色，分別為經驗者 (experiencer) 及刺激物 (stimulus)。然而，因心理動詞之語意不同，主語與賓語所扮演的語意角色在句法的位置上會隨之調整，如以下二例之比較：

- (16a) 佢當驚阿伯。
- (16b) 佢愛去嚇阿伯。

在(16a)，「佢」為經驗者，「阿伯」為刺激物；而在(16b)則恰好相反，「佢」扮演刺激物角色，「阿伯」則是扮演被嚇之經驗者角色。因此，心理動詞之動詞語意及其論元表現一直是句法與語意互動研究的重要議題。本文將針對臺灣客語心理動詞「怕」所表現之句法、語意、語用等特徵，進行細部剖析。客語主要以以下兩個心理動詞表達「害怕」的語意：「驚」與「怕」。「驚」的語意表示害怕或擔心，如(17a)與(17b)：

- (17a) 你毋好緊講驚鬼，鬼毋會煩人...。
- (17b) 大家驚你个叫聲會分飛行機聽到。

而「怕」也和「驚」相同，同樣有表達害怕、擔心之語意，如(18a)與(18b)：

(18a) 毋怕啦！毋怕啦！姨婆你驚麼个！

(18b) 毋怕啦，一下轉，你做你去啦！

觀察客語語料後發現，客語多用「驚」表示擔心、害怕意，在語言上的體現為極豐富的句式，而「怕」大多只有否定祈使句式「毋怕」，此對比呈現一有趣且不對稱現象。此外，被「驚」搶奪語意功能之「怕」，尚有一「驚」所欠缺之語意表徵為「恐怕」，如下例(19)：

(19) 這擺你怕保俵毋得囉...

臺灣華語並無(19)的用法，而是藉由「恐怕」、「大概」等詞彙化詞彙表達，此為客語特殊之處。教育部重編國語辭典中，「怕」之動詞釋義有二者：一為「恐懼、害怕」，如「他表弟很怕老虎」；二為「擔心、疑慮」，如「我怕妳累壞了」。然而，副詞釋義之舉例皆為非現代用法，如下例：

(20a) 可能、或許、大概。

儒林外史·第二十六回：「他怕不也有二五六歲，他對人自說二十一歲。」

(20b) 豈、難道。表反問。

初刻拍案驚奇·卷三十八：「我怕不要周全？只是關著財利上事，連夫妻兩個，心肝不託著五臟的。」

(20c) 如果、倘若。

永樂大典戲文三種·小孫屠·第三出：「梅香，怕有賞春佳客來買酒，你與我安排了酒器，整頓則個。」

相較於現代華語「怕」之使用限制於動詞用法，客語「怕」字還可以用以表達可能、大概、難道等語意，如(20a)、(20b)所示。換言之，表達恐怕、大概意義的情態詞在華語中是由「恐怕」、「大概」等詞彙化詞彙來呈現，但在客語中，「怕」字除了有動詞用法，亦有情態詞用法。此外，客語「怕」動詞與副詞之頻率，亦有其差異；除了常見的否定句式「毋怕」之外，「怕」表「害怕」之語意多半只用於俗語、諺語，如(21)：

(21) 男驚投錯行，女怕嫁錯郎。

因此，客語「怕」絕大多數情況是用以表達說話者主觀性不確定之情態意義，如(22)：

(22) 佢个兩隻腳，怕有十公分長。

(21)與(22)「怕」之論元體現、句法位置及使用頻率，其表現均不同。首先，心理動詞「怕」所要求的兩個必要論元角色〈經驗者，刺激物〉，在例(21)中，分別體現為句中之主語「女」與賓語「嫁錯郎」，然而在(22)中則體現為句子的說

話者（隱匿不現）與命題「佢个兩隻腳有十公分長」。在句法方面，(21)的動詞「怕」支配句中整個動詞片語，而(22)的「怕」則是位於更高層的句法位置。(21)句用法多半限制於諺語型式，而(22)句表達「可能、恐怕」語意之「怕」構式具高度能產性。關於以上分析，將在 3.3.1 節進一步討論。遵循構式語法框架，本文試圖為客語「怕」字之句法分佈、語意類型及語用特徵進行探究剖析。

### 3.3.1 臺灣客語「怕」構式句法及語意之分析

本文採用語料庫研究，配合理論架構，主要將語料分為兩類。A 類如例子(23)，B 類如例子(24)：

- (23) 男驚投錯行，女怕嫁錯郎。
- (24) 佢个兩隻腳，怕有十公分長。

以構式語法觀點而言，構式為一形式與語意之配對。「怕」之論元語意角色為「經驗者」與「刺激物」。此二語意角色在 A 類中，分別投射在主語與賓語之位置，相關例子如下：

- (25a) 禾怕寒露風，人怕老年窮。
- (25b) 草怕斷根，禾怕枯心，蟲怕夜燈。
- (25c) 毋驚青龍高萬丈，只怕白虎回頭望。

除了在諺語中，「怕」多半用於否定祈使句式中，如：



- (26a) 毋怕啦，來去啦！  
 (26b) 毋怕，毋怕，天光日換 佢去回話。  
 (26c) 毋怕啊，你做你講，有麼个事情？  
 (26d) 毋怕啦，佢無阿爸無阿媽...。  
 (26e) 毋使驚，毋使怕，大家來看海浪花。

但語料中並無「怕」可出現在祈使句中，如：

- (27) \*怕啦，來去啦！

因此，「毋怕」在客語中如同熟語一樣產生固化 (fossilization) 現象。總之，句法方面，此構式為典型 (canonical) 形式；如同前述所提，此用法多用於諺語中，即使是一般用法，也只觀察到否定祈使句式，因此在語意方面則帶有非典型的熟語 (idiomatic) 特徵。

相較之下，在 B 類中，「經驗者」與「刺激物」則是投射在說話者與命題之位置，此類型為一高度能產之構式 (highly productive)，在許多資料中都可找到相關例子如下：

- (28a) 這兩種鳥仔，怕無一定係“害鳥”。  
 (28b) 這擺你怕保佢毋得囉...。  
 (28c) 吾俵仔今晡日怕愛來打佢囉！

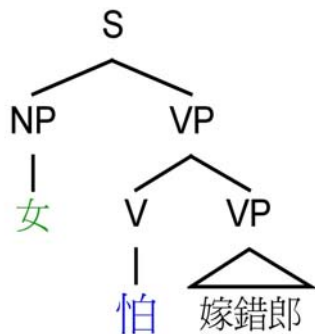
在以上例子中，主語「這兩種鳥仔」、「你」、「吾俵仔」並非經驗者，真正的經

驗者是投射在說話者本身身上，而刺激物則是句中的命題內容。此外，因為說話者對於命題內容之猜測性或不肯定性，說話者根據其本身之經驗選用情態詞「怕」表達出其認知情態 (epistemic modality) 中的不確定性 (uncertainty)。

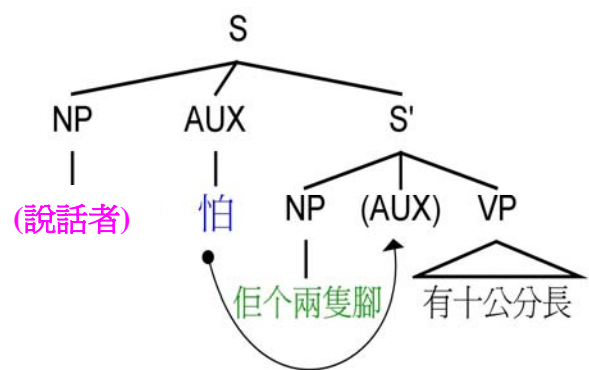
以上兩類的構式在句法呈現方面之不同比較如下：

- (29) 男驚投錯行，女怕嫁錯郎。
- (30) 佢个兩隻腳，怕有十公分長。

**A類**



**B類**



「女」和「佢个兩隻腳」在(29)與(30)中同樣佔據主語位置，但在句法表現，「女」是主要子句中的主語，但「佢个兩隻腳」則是附屬子句的主語；(30)中主要子句的主語是說話者本身，在句中隱而不現。

若將上例(29)及(30)改為否定構式，亦有不同呈現，如：

(29') 男驚投錯行，女毋怕嫁錯郎。

(30') 佢个兩隻腳，怕無十公分長。

當「怕」擔任一心理動詞時，否定詞「毋」「不」直接接在動詞前；若「怕」為一情態詞時，否定詞則是在附屬子句中運作，接在附屬子句中的動詞前<sup>7</sup>。

總之，「怕」字構式由原本單純指涉一人物、物體或事件，變成指涉說話者的態度或看法，「怕」字也從心理動詞變成情態詞；因此，從 A 類到 B 類之變化充分展現語法化之現象。

除了語法化之外，「怕」字構式亦呈現詞彙化。「怕」與「驚」或「敢」此二同義、反義心理動詞亦詞彙化成情態詞「驚怕」與「敢怕」，皆表達說話者對命題內容之主觀不確定性。換言之，表情態除可透過論元角度：經驗者投射到說話者、刺激物投射到命題內容之外，亦可藉由詞彙化機制來表達出情態概念。以「驚怕」為例，「驚」與「怕」為近義關係 (synonymy)，除了可單純表達害怕之語意，如(31)；亦可表達「恐怕」之意，表達出語境中的讓步 (concessive) 語氣，並顯現出知識性情態及主觀性特徵，如(31b)：

(31a) 阿八呢盡賢孝，驚怕阿媽愁，毋敢講正經。

(31b) 佢看佢，氣色不對，……，驚怕不久人世囉。

再以「敢怕」為例，「敢」與「怕」為反義關係 (antonymy)，然而兩者結合

<sup>7</sup> 原句為：「佢个兩隻腳，怕+ NEG +有十公分長。」而「NEG +有」在客語中則以「無」字顯現。

之語意為「恐怕」之意，同樣表達語境中的讓步語氣及顯現出知識性情態及主觀性特徵，如(32)：

(32) 敢怕今晡日會落水哦！‘恐怕今天會下雨哦！’

表達否定語意時，否定詞「毋」(不)並非接在「敢」字之前，而是在命題子句之前，表達說話者認為情境不可能發生的主觀判斷，試比較(33a)與(33b)之異同：

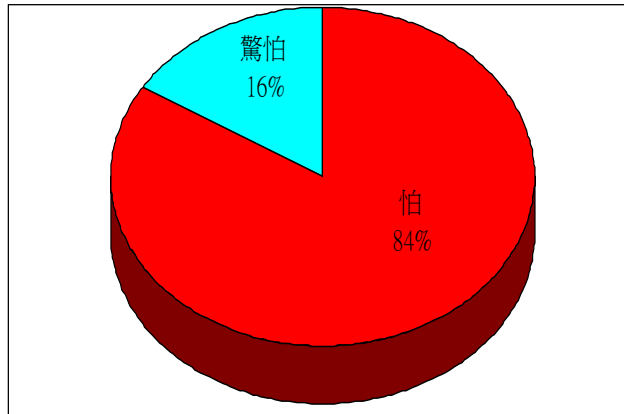
(33a) 敢怕今晡日毋會落水哦！

(33b) \*毋敢怕今晡日會落水哦！

### 3.3.2 量化分析

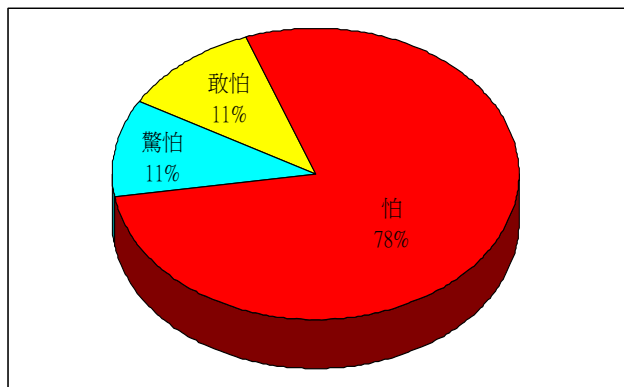
在量化分析方面，書面與口語各自有不同呈現。在書面語料部分，共搜尋得到 106 筆「怕」之語料。本研究將語料分為兩大類：表擔心、害怕之動詞義及表恐怕、大概之情態詞義。A 類共有 61 筆，可再細分為「怕」及「驚怕」；「怕」例子共有 51 筆，佔 84%，值得注意的是，否定祈使句式便有 29 筆，其他則多為俚語用法；「驚怕」表擔心、害怕的例子則有 10 筆，佔 16%。相關圖示如〈圖 8〉：

〈圖 8〉客語「怕」字構式書面語料之心理動詞分佈圖



B 類表恐怕、大概義則有 45 筆，其中「怕」一字表達情態意義共有 35 筆，佔 B 類的 78%；以「驚怕」及「敢怕」形式呈現的各為 5 筆，各佔 B 類中的 11%，如〈圖 9〉：

〈圖 9〉客語「怕」字構式書面語料之情態詞分佈圖

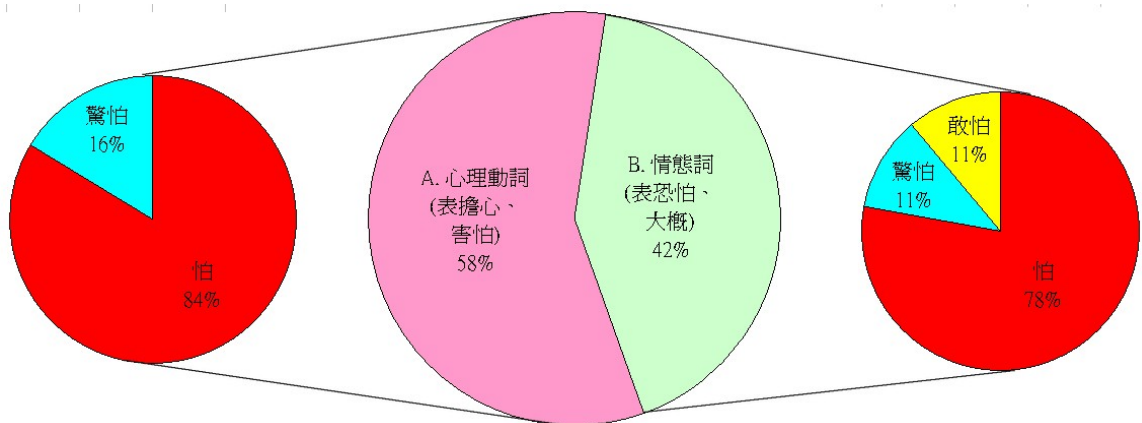


A 類佔書面語料總比例之 58%，B 類則佔 42%，兩者差距並不大。比較書面語料之 A 類及 B 類，可得到以下〈表 3〉及其相對應圖示〈圖 10〉：

〈表 3〉客語「怕」字構式書面語料比例

語意表徵	構式	筆數	總數比例	總數
A. 心理動詞 (表擔心、害怕)	怕	51	58%	61
	驚怕	10		
B. 情態詞 (表恐怕、大概)	怕	35	42%	45
	驚怕	5		
	敢怕	5		
書面語料筆數				106

〈圖 10〉客語「怕」字構式書面語料分佈圖

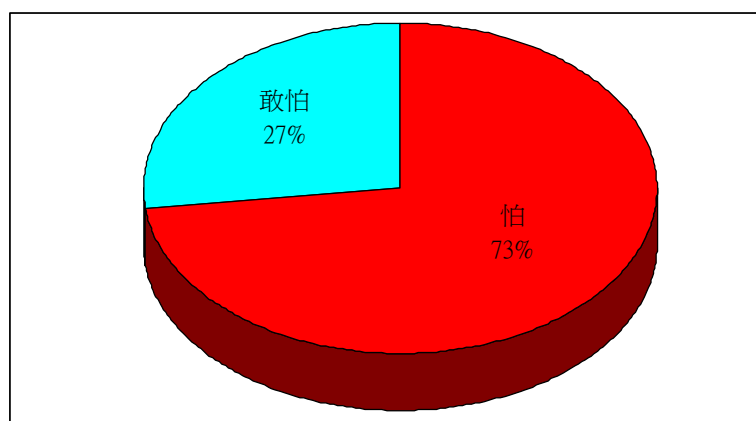


探究 A 類與 B 類在書面語的分佈差距不大的原因與書面語的語體有關，本語料庫的書面語料主要來源為臺灣各縣市出版的故事集佔大宗，其次為其他出

版社出版的故事或客家雜誌之客語故事，由於故事的敘述及性質常常會表達故事主角的遭遇及情緒，因此「怕」會傾向表達其基本語意表示擔心、害怕，再者故事或寓言大多以醒世或勸教為主要訴求，因此經常會以諺語或者熟語來結尾以勸誡世人行善或者保留美德，因此「怕」也經常以基礎語意出現在諺語中。然而，隨著語言的改變，在現代客語的使用，「怕」的情態用法則佔多數，例如在口語語料的分佈情形。<sup>8</sup>

口語語料部分，在政大客語口語語料庫中共搜尋得到 13 筆「怕」之語料：2 筆表擔心、害怕義，並皆以否定祈使句式表現；表恐怕、大概義有 11 筆，其中由「怕」一字表達此情態之例子有 8 筆，佔 B 類中的 73%，以「敢怕」形式出現則有 3 筆，佔 27%。在口語語料庫中無「驚怕」之例，請見〈圖 11〉。

〈圖 11〉客語「怕」字構式口語語料之情態詞分佈圖



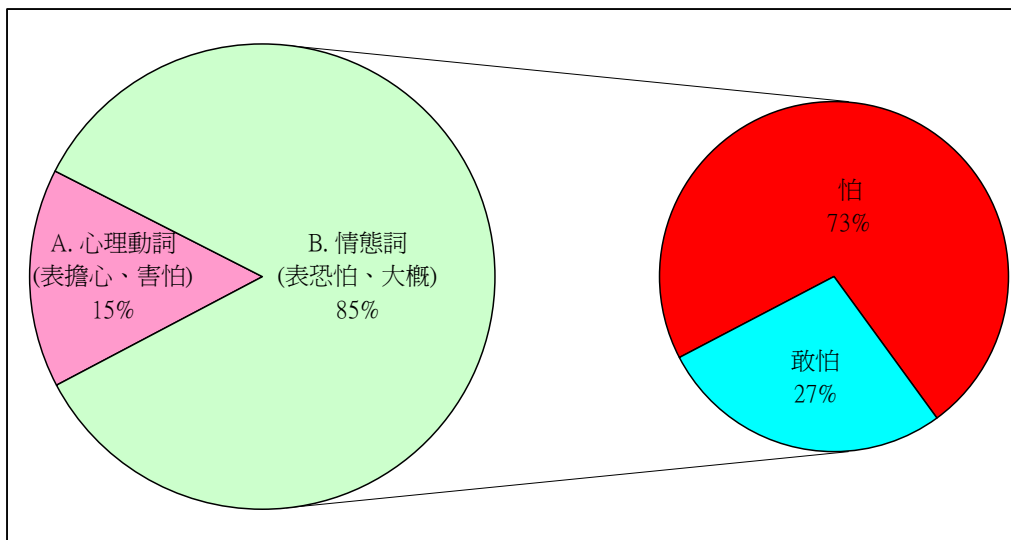
〈表 4〉以及〈圖 12〉即為口語語料中 A 類以及 B 類之比較呈現：

<sup>8</sup> 有關語體中之語言的句法及語意行為表現為值得探討的議題，有待未來研究。

〈表 4〉客語「怕」字構式口語語料比例

語意表徵	構式	筆數	總數比例	總數
A. 心理動詞 (表擔心、害怕)	怕	2	15%	2
B. 情態詞 (表恐怕、大概)	怕	8	85%	11
	敢怕	3		
口語語料筆數				13

〈圖 12〉客語「怕」字構式口語語料分佈圖



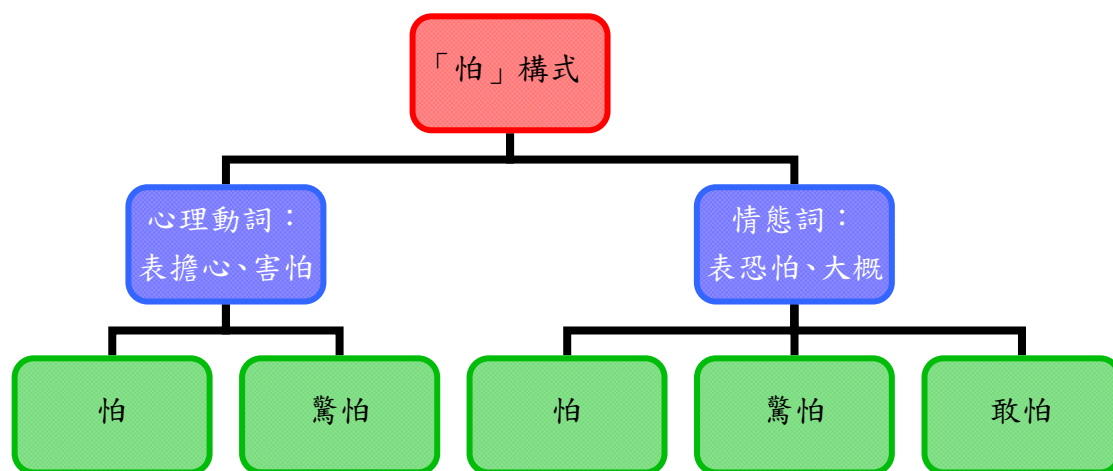
比較〈表 3〉與〈表 4〉，可發現「怕」字構式在書面及口語使用有所不同。在構式方面可發現，「驚怕」在口語語料中完全沒有出現。而在比例上，不同於



書面語料，口語語料較常使用「怕」字構式表恐怕、大概，比例高達 85%；而表擔心、害怕則只有 15%，可推測在口說中表擔心害怕之心理動詞可能多由「驚」形式呈現。此外，單用「怕」來表達情態意義之比例也高過詞彙化之詞彙「敢怕」，這點在書面、口語語料中皆同。

與台灣華語相比，客語「怕」字之語意及語法表現更為豐富。擔任心理動詞與情態詞之「怕」在論元體現方面以及語意表達方面各有不同。關於「怕」字構式之運作圖示於下：

〈圖 13〉客語「怕」構式圖



在量化分析方面，也分別針對語料類型之差異，依照使用頻率的比列提出相關圖示及表格，並比較兩者之異同。未來可針對「怕」構式的語意關連，在構式語法框架的呈現機制等相關議題再進一步研究。

## 4. 結論

本語料庫目前典藏書面語料，未來將典藏口語語料。尋找特定年齡群是未來建構語料庫的目標之一，蒐集不同體裁文本及語料為另一項目標，未來在體裁規劃上希望朝如〈表 5〉所示的分類進行。

〈表 5〉臺灣客語方言語料庫典藏語料之體裁規劃

對話	日常會話 電視訪談 新聞討論 Call-in 電視節目
獨白	演講 說故事 口述歷史 新聞稿件

截至目前為止，臺灣客語方言書面語料已典藏 670 篇文章，每篇文章均系統化紀錄其出處、作者、文體等相關資訊。未來將加入語料庫中典藏的口語語料，目前已收錄三十筆語料（十九筆日常會話，十筆獨白說故事，一筆電視訪談），數量仍在持續進行中。其中已有六筆各約二十分鐘的口語對話（五筆日常會話，一筆電視訪談）已公開於「國立政治大學漢語口語語料庫」網頁

（<http://140.119.172.200/>），已完成漢字轉寫、音檔切割，並標記羅馬拼音、詞語標註及英文翻譯等資訊，語料呈現於網頁之畫面如〈圖 14〉所示，詳細內容則

如〈表 6〉所示。<sup>9</sup>

〈圖 14〉「國立政治大學漢語口語語料庫」語料呈現畫面

84	M	<p>Yes, I see. Oh, wait, that's so good. That's a really good nut.</p> <p>[5 ei 5] ... &lt;L2 那麼 L2&gt; en1 就 dang .. 薯 點心 食 哩 [6 la 6]</p> <p>ei na4me5 en1 qiu3 dang dong1 diam3xim1 siid8 le1 la5</p> <p>PF then 1PL EMP REPAIR act-as snack eat PRT PRT</p> <p>Well. Then, we also consider pestled tea to be a snack.</p>
85	F1	<p>[6 哦 6] ... [7 這樣形 .. 到 山肚 一 7] 禮拜 hon .. 咩 毋 會 驚 梅死</p> <p>o1 an2ngiong2hin5 do3 san1du2 id4 li5bai3 hon me3 m5 voi3 giang1 iau1si2</p> <p>BC this-way stay mountain one week PRT also NEG will worry starve</p> <p>啦 [8 hon 8]</p> <p>la5 hon</p> <p>PRT PRT</p> <p>Oh. Thus, you needn't worry about the provisions even if you stay in the mountains for a week.</p>
86	M	<p>[7 肚餓 梅 食 个 啦 hon 7] ... (1.8) [8 &lt;L2 對 L2&gt; 8]</p> <p>du2sii5 iau1 siid8 ge3 la5 hon du4</p> <p>stomach hungry eat NOM PRT PRT right</p> <p>If you are hungry, you can eat it to relieve the pangs. Right.</p>
87	F2	<p>[8 毋 會 毋 會 8] [9 毋 會 毋 會 .. 係 9]</p> <p>m5 voi3 m5 voi3 m5 voi3 m5 voi3 he3</p> <p>NEG will NEG will NEG will NEG will yes</p> <p>No, no, no. That's right.</p>
88	F1	<p>[9 係 .. 還 有 旁人 e 9] ... 阿 .. 阿伯 你 愛 歇腳 一下 囉</p> <p>he3 han5 iu1 pong5ngin5 e a1 a1bag4 ngi5 oi3 hied4kun3 id4ha3 lo1</p> <p>yes EMP have others PRT REPAIR uncle 2SG need take-a-rest for-a-while PRT</p>

<sup>9</sup> 有關「國立政治大學漢語口語語料庫」建構步驟與詳細內容，請參見賴惠玲，李詩敏（2010）。

〈表 6〉「國立政治大學客語口語語料庫」網頁典藏語料之內容

	參與者		語料長度		
	女性	男性	分鐘	話輪	字數
H001-CN-C-MFF-OMO 客家擂茶製作	2	1	20	503	11848
H002-CN-C-FF-MM 搭乘公車經驗	2	0	20	396	10849
H003-CN-C-FFF-OOOOOO 婆媳相處	3	0	22	415	12460
H004-CN-C-MMM-OMM 東勢經濟發展	0	3	23	365	11129
H005-CN-C-FF-OOM 花蓮生活回憶	2	0	20	669	14931
H005-CN-C-MFF-MOOOO 客家家庭主婦的生活	2	1	20	453	14170
總計：	11	5	125	2801	75,387

應用系統化客語語料庫所呈現的量化語言分析方面，本計畫以臺灣客語方言表完成體和持續體「著」、涉及語法化之主觀性與詞彙化之固化等觀點之「怕」構式在語料庫中的分佈分析為例，展現語料庫之應用成果。

整體而論，本語料庫研究尚有以下未竟完成之處：其一，由於時間之限制與人員之不足，本語料庫暫無法進行斷詞與詞類標記。惟斷詞與詞類標記將使語料庫的功能更為強大，此兩項工程將留待未來繼續執行。其二，前面提及，在語料庫建置之初，尙未能考慮文體之平衡，且此語料庫目前尙以書面語料為主，輔以口語語料，文體之均衡以及全面性客語口語語料庫之建構，是本研究目前無法涵蓋之部分，仍有待將來的努力。

客語是個表意豐富的語言，具有許多值得研究的語言學課題，但現今之客語研究一直缺乏有系統之分析，本研究以建構客語語料庫為出發，從語料的蒐集、判讀、輸入，到語料庫的建構及使用介面功能介紹，展示出一個有系統的研究步驟，為日後相關客語語料庫為本的相關研究議題提供一個很寶貴的參考資源及可參考的研究方法，有助於客語研究的蓬勃發展及研究議題的深入探討。就實用性而言，本研究就客語語料做全面性的收集、整理，並建構一個可供檢索語料的語料庫，實際語言使用的範例，對客語教學的教材研發及編撰均極有助益。就客語資源而言，本計畫目前已聯合國立新竹教育大學葉瑞娟教授、黃漢君教授與吳睿純教授等諸位客語研究學者共同建構此語料庫，不但避免語料重複收集、人力浪費等缺點，並共創客語研究合作環境之最大效益。然而，語料庫的建構費時費力，但聚沙成塔，本研究計畫盼拋磚引玉，能引起大家對客語語料庫更多的關注，也能藉此讓語料庫發揮更大的功能。

## 參考文獻

### ● 語料庫語言學相關研究

- Ahrens, Kathleen, Shirley Yuan-Hsun Chuang, and Chu-Ren Huang, 2003. Sense and meaning facets in verbal semantics: A MARVS perspective. *Languages and Linguistics* 4.3: 469-484.
- Boas, Hans C. 2003. *A Constructional Approach to Resultatives*. Stanford: CSLI Publications.
- Boas, Hans C. 2008a. Resolving form-meaning discrepancies in Construction Grammar. *Constructional Reorganization*, ed. by J. Leino, 11-36. Amsterdam/Philadelphia: John Benjamins.
- Boas, Hans C. 2008b. Determining the structure of lexical entries and grammatical constructions in Construction Grammar. *Annual Review of Cognitive Linguistics* 6: 113-144.
- Bybee, Joan, Revere Perkins and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: The University of Chicago Press.
- Chang, Li-Li, Keh-Jiann Chen, and Chu-Ren Huang. 2000. A lexical-semantic analysis of Mandarin Chinese verbs: Representation and methodology. *International Journal of Computational Linguistics and Chinese Language processing* 5.1: 1-18.
- Chen, Keh-Jiann, and Chao-Jan Chen. 1998. A corpus based study on computational morphology for Mandarin Chinese (語料庫爲本的中文複合詞構詞律模型研究). *Quantitative and Computational Studies on the Chinese Language*, ed. by Benjamin K. T'sou, Tom B.Y. Lai, Samuel W. K. Chan, and William S-Y. Wang, 283-306. HK: City University of Hong Kong.
- Chui, Kawai, and Huei-ling Lai. 2008. The NCCU Corpus of Spoken Chinese: Mandarin, Hakka, and Southern Min. *Taiwan Journal of Linguistics* 6.2: 119-142.
- Chui, Kawai, Huei-ling Lai, and Hui-chen Chan. 2008. The NCCU Corpus of Spoken Chinese: Mandarin, Hakka, and Southern Min. Proceedings of the 7th Annual Wenshan International Symposium (Theme: Spoken Corpus: Language Documentation and Applications), 68-100. May 19. National Chengchi University, Taipei.
- Chung, Siaw-Fong. 2007. A Corpus-driven Approach to Source Domain Determination. Ph. D. Dissertation of National Taiwan University, Taiwan.

- Chung, Siaw-Fong, and Kathleen Ahrens. 2008. MARVS revisited: Incorporating sense distribution and mutual information into near-synonym analyses. *Languages and Linguistics* 9.2: 415-434.
- Chung, Siaw-Fong, Kathleen Ahrens, and Chu-Ren Huang. 2005. Source domains as concept domains in metaphorical expressions. *Computational Linguistics and Chinese Language Processing* 10.4: 553-570.
- Croft, William. 2003. Lexical rules vs. constructions: A false dichotomy. *Motivation in Language: Studies in Honour of Günter Radde*. ed. by, H. Cuyckens, T. Berg, R. Dirven and K.-U. Panther, 49-68. Amsterdam/Philadelphia: John Benjamins.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67.3: 547-619.
- Fillmore, Charles. 1970. The grammar of Hitting and Breaking. *Readings in English Transformational Grammar*, ed. by R. Jacobs & P. Rosenbaum, 120-133. Waltham and London: Ginn.
- Goldberg, A. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.
- Goldberg, A. E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele Eva and Ray Jackendoff. 2004. The English Resultative as a Family of Constructions. *Language* 80: 532-565.
- Gries, Stefan Th. 2010. Useful statics for corpus linguistics. *A Mosaic of Corpus Linguistics: Selected Approaches*, ed. by Aquilino Sánchez and Moisés Almela. Frankfurt am Main: Peter Lang.
- Hong, Jia-Fei. 2010. Corpus-driven Linguistic Approaches to Sense Prediction. Ph.D. Dissertation of National Taiwan University, Taiwan.
- Hong, Jia-Fei, and Chu-Ren Huang. 2008. A corpus-based approach to the discovery of Cross-Strait lexical contrasts. *Language and Linguistics* 9.2: 221-238.
- Huang, Chu-Ren, Kathleen Ahrens, Li-Li Chang, Keh-Jiann Chen, Mei-Chun Liu, and Mei-Chih Tsai. 2000. The Module-Attribute Representation of Verbal Semantics: From semantics to argument structure. *Computational Linguistics and Chinese Language Processing* 5.1: 19-46.
- Huang, Chu-Ren, Keh-Jiann Chen, and Zhao-Ming Gao. 1998. Noun class extraction from a corpus-based collocation dictionary: An integration of computation and qualitative approaches. *Quantitative and Computational Studies on the Chinese Language*, ed. by Benjamin K. T'sou, Tom B.Y.Lai, Samuel W.K. Chan, and William S-Y. Wang, 339-352. HK: City University of Hong Kong.
- Huang, Chu-Ren, and Jia-Fei Hong. 2005. Deriving conceptual structures from sense:

A study of near synonymous sensation verbs. *Journal of Chinese Language and Computing* 15.3: 125-135.

- Kay, Paul. 1995. Construction Grammar. *Handbook of Pragmatics*, ed. by J. Verschueren, Jan-Ola Östman, and J. Blommaert, 171-177. Amsterdam and Philadelphia: John Benjamins.
- Kay, Paul, and Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The What's X Doing Y? construction. *Language* 75.1: 1-33.
- Lai, Huei-ling. 2008. A Corpus-based exploration of research topics on Hakka grammar. 行政院客家委員會 96 年度補助大學暨獎助學術研究計畫成果發表會. June 6. National Central University, Taoyuan.
- Lai, Huei-ling. 2009. The NCCU Corpus of Spoken Chinese: Hakka Corpus. Multi-Development and Application of Languages and Linguistics. May 15-16. National Cheng Kung University, Taiwan. (Invited speech)
- Lai, Huei-ling, and Shu-mei Chiang. 2007. The family of Hakka idiomatic *Gin2 X Gin2 Y* constructions: Paratactic constructions with subordinate meanings. Paper presented at the 10th International Cognitive Linguistics Conference, Poland: Jagiellonian University, July 15-20.
- Lai, Huei-ling, and Shih-min Li. 2008. Hakka Corpus: Construction and application (客語語料庫的建構及應用). Paper presented at the Eighth International Conference of Hakka Dialects. November 22-24. National Central University, Taiwan.
- Lien, Chinfa. 2005. Phase and aspect markers in Li Jing Ji. *Essays in Chinese Historical Linguistics: Festschrift in Memory of Profess Fang-kuei Li on his Centennial Birthday*, ed. by Pang-Hsin Ting and Anne O. Yue, 393-420. Taipei: Institute of Linguistics, Academia Sinica.
- Liu, Mei-Chun, Chu-Ren Huang, Charles Lee, and Ching-Yi Lee. 2000. When endpoint meets endpoint: A corpus-based lexical semantic study of Mandarin verbs of throwing. *Computational Linguistics and Chinese Language Processing* 5.1: 81-96.
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Mayer, Mercer. 1969. *Frog, Where Are You?* New York: Dial Books for Young Readers.
- Mei, Tsu-lin (梅祖麟). 1979. The etymology of the aspect marker tsi in the Wu dialect. *Journal of Chinese Linguistics* 7.1: 1-14.
- Palmer, Frank R. 1979. *Modality and the English modals*. London: Longman.
- Stefanowitsch, Anatol, and Stefan Th. Gries. 2003. Collocations: investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 6.1: 1-20.



- Linguistics* 8.2: 209-243.
- Traugott, Elizabeth Closs. 1995a. Subjectification in grammaticalization. *Subjectivity and Subjectivisation*, ed. by D. Stein and S. Wright, 31-54. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs. 1995b. The role of the development of discourse markers in a theory of grammaticalization. Paper presented at the 12th International Conference on Historical Linguistics. Manchester: University of Manchester.
- Traugott, Elizabeth Closs. 2005. Constructions in Grammaticalization. *The Handbook of Historical Linguistics*, ed. by Brian D. Joseph and Richard D. Janda, 624-647. Oxford: Blackwell.
- Tsai, Mei-chi, Chu-Ren Huang, Keh-Jiann Chen, and Kathleen Ahrens. 1998. Towards a Representation of Verbal Semantics -- An approach based on near-synonyms. *International Journal of Computational Linguistics and Chinese Language processing* 3.1: 61-74.
- Verspoor, Marjolijn, and Wander Lowie. 2003. Making sense of polysemous words. *Language Learning* 53.3: 547-586.
- Wang, Chueh-Chen, and Lilian M. Huang. 2006. Grammaticalization of Connectives in Mandarin Chinese: A corpus-based study. *Languages and Linguistics* 7.4: 991-1016.
- 陳前瑞，2009。〈“著”兼表持續與完成用法的發展〉，《語法化與語法研究（四）》，北京：商務印書館，頁 1-22。
- 鄭縈，2005。〈台灣客語動詞“有”的語法特點與語法化〉，《臺灣語言與語文教育》第 6 期，頁 31-47。
- 江敏華，2007。〈東勢客家話與時間或體貌有關的「緊」字〉，「第七屆客家方言國際學術研討會」論文，香港：香港中文大學，2007.01.20-21。
- 謝杰雄，2006。語料庫的建置與台灣客家語 VP 研究。新竹教育大學台灣語文與文學教育研究所碩士論文。
- 謝杰雄，2008。說明與演示一個線上客語語料庫。第八屆國際課方言學術研討會，11 月 22 日-23 日，中央大學，台灣。
- 黃居仁，陳克健，賴慶雄（編著），1996。《常用量詞詞典》。台北：國語日報社。
- 黃居仁，陳克健，賴慶雄（編著）。1997。《國語日報量詞典 Mandarin Daily Dictionary of Chinese Classifiers》。台北：國語日報社。
- 蔣紹愚，2005。〈關於漢語史研究的幾個問題〉，《漢語史學報》第 5 輯，頁 1-12。
- 蔣紹愚，2006。〈動態助詞“著”的形成過程〉，《周口師範學院學報》第 23 卷第 1 期，頁 113-117。
- 李詩敏，賴惠玲，2010。〈臺灣客語方言完成體與持續體之關係探索：語料庫為本的觀點〉，「第九屆國際客家方言研討會」論文，北京：中協賓館，

2010.10.20-21。

賴惠玲（計畫主持人），2007。「客語語法研究議題的開發：以語料庫為本」。

臺灣行政院客委會研究計畫。

賴惠玲，2008。面對瀕臨死亡的語言我們能做些什麼？以客語為例。國家科學委員會「週日閱讀科學大師」系列講座 12 之 6，12 月 21 日，高雄，國立科學工藝博物館。

賴惠玲（計畫主持人），2009。「客語語料庫之系統化建構與量化分析 99-C07」。臺灣行政院客委會研究計畫。

賴惠玲，李詩敏，2010。〈客語語料庫的建構及應用〉，《客語千秋：第八屆國際客方言學術研討會論文集》，頁 300-312。台北：文鶴出版社。

劉丹青，1995。〈無錫方言的體助詞“則”（仔）和“著”——兼評吳語“仔”源於“著”的觀點〉，《中國語言學報》第 6 期，頁 223-233。

劉美君，2002。國科會專題研究計畫「統整動詞語意 (I)：中文「動詞詞網」的建構」。

梅祖麟，1989。〈漢語方言裡虛詞“著”字三種用法的來源〉，《中國語言學報》第 3 期，頁 193-216。

聶家祺，劉容瑜，強舒嫩，羅婉君，2010。〈臺灣客語方言語料庫系統之建構〉，「第九屆國際客家方言研討會」論文，北京：中協賓館，2010.10.20-21。

錢乃榮，2002。〈進行體、持續體和存續體的比較〉，《中國語文研究》第 1 期（總 13 期），頁 45-52。

曹逢甫，黃雅榆，2002。客語危機以及客家人對客語和政府語言政策的態度，《客家公共政策. 研討會論文集》，行政院客家委員會。

曾慧馨，劉昭麟，高照明，陳克健，2002。以構詞律與相似法為本的中文動詞自動分類研究，《中文計算語言學期刊》第七卷第一期，頁一至二十八。

孫朝奮，1997。〈再論助詞「著」的用法及其來源〉，《中國語文》第 2 期，頁 139-146。

吳福祥，2004。〈也談持續體標記“著”的來源〉，《漢語史學報》第 4 輯，頁 17-26。

邢公畹，1979。〈現代漢語和台語裡的助詞“了”和“著”（下）〉，《民族語文》第 3 期，頁 206-212。

楊秀芳，1992。〈從歷史語法的觀點論閩南語「著」及持續貌〉，《漢學研究》第 10 卷第 1 期，頁 349-394。

葉秋杏，強舒嫩，2010。〈臺灣客語「怕」構式之語法化與詞彙化——以語料庫為本的研究〉，「第九屆國際客家方言研討會」論文，北京：中協賓館，2010.10.20-21。

張愛民，1992。《“有”字的意義與“有”字句式》，北京：語文出版社。

## ● 客語文本

邱一帆。2004。《油桐花下个思念》。桃園：華夏書坊。

- 邱一帆。2000。《田螺：客語詩集》。客家台灣文史工作室。
- 胡萬川編。1994。《東勢鎮客語故事集》。台中：台中縣立文化中心。
- 胡萬川編。1995。《東勢鎮客語故事集（二）》。台中：台中縣立文化中心。
- 胡萬川，王正雄編。1999。《東勢鎮客語故事集（五）》。台中：台中縣立文化中心。
- 胡萬川，洪慶峰編。2001。《東勢鎮客語故事集（六）》。台中：台中縣立文化中心。
- 胡萬川，陳嘉瑞編。2003。《東勢鎮客語故事集（七）》。台中：台中縣立文化中心。
- 胡萬川，黃晴文編。1996。《東勢鎮客語故事集（三）》。台中：台中縣立文化中心。
- 胡萬川，黃晴文編。1997。《東勢鎮客語故事集（四）》。台中：台中縣立文化中心。
- 范文芳。1998。《頭前溪个故事》。新竹：新竹縣文化局。
- 徐兆泉。2001。《客家笑科》。台北：南天。
- 徐兆泉編。2001。《台灣客家話辭典》。台北：南天。
- 徐兆泉譯。2000。《客家話小王子》。台北：南天。
- 徐登志，1997。《徐老師講古》。台中：大甲河文化。
- 楊時逢，1957。《臺灣桃園客家方言》。台北市：中央研究院歷史語言研究所。
- 劉惠萍，范姜烜欽編。2009。《花蓮客家民間文學集》。花蓮：花蓮縣文化局出版。
- 羅肇錦編，1999，《苗栗縣客語故事集（二）》，苗栗：苗栗縣立文化中心。
- 羅肇錦編，2002，《苗栗縣客語故事集（三）》，苗栗：苗栗縣立文化中心。
- 羅肇錦、胡萬川編，1998，《苗栗縣客語故事集》，苗栗：苗栗縣立文化中心。

## 附錄一 口語語料受錄者簽署同意書格式

### 同意書

賴惠玲教授所主持的研究計畫「政治大學漢語口語語料庫 — 客語口語語料庫子計畫」，目的是蒐集與整理口語語料，以提供各領域作為研究教學用途。

非常感謝您參與這個計畫。「政治大學漢語口語語料庫 — 客語口語語料庫子計畫」是網路上的語料庫，為了保護個人的隱私，我們會將您的談話內容和錄音的聲音檔作出適當的處理，例如人名、地名、電話號碼會變更，這些資訊在聲音檔中也會消音。

本人  同意  不同意 將本人的談話內容放在語料庫中。

本人  同意  不同意 將本人談話內容的錄音放在語料庫中。

本人  同意  不同意 將本人的談話內容的錄影放在語料庫中。

簽名： \_\_\_\_\_

日期： \_\_\_\_\_

## 附錄二 口語語料受錄者填寫個人背景資料表

### Speech Event Sheet

Genre: \_\_\_\_\_

Recorded Date: \_\_\_\_\_ Time: \_\_\_\_\_ Location: \_\_\_\_\_

Recorded by: \_\_\_\_\_ Duration of Recording: \_\_\_\_\_

Relationship between/among participants: \_\_\_\_\_

Setting: \_\_\_\_\_

#### SPEAKER INFORMATION

1. 姓名: \_\_\_\_\_ 性別:  女  男  
年齡: \_\_\_\_\_ 出生地: \_\_\_\_\_ 職業: \_\_\_\_\_  
常用的語言:  國語  台語  客語  其他 \_\_\_\_\_  
與其他參與者的關係: \_\_\_\_\_
2. 姓名: \_\_\_\_\_ 性別:  女  男  
年齡: \_\_\_\_\_ 出生地: \_\_\_\_\_ 職業: \_\_\_\_\_  
常用的語言:  國語  台語  客語  其他 \_\_\_\_\_  
與其他參與者的關係: \_\_\_\_\_
3. 姓名: \_\_\_\_\_ 性別:  女  男  
年齡: \_\_\_\_\_ 出生地: \_\_\_\_\_ 職業: \_\_\_\_\_  
常用的語言:  國語  台語  客語  其他 \_\_\_\_\_  
與其他參與者的關係: \_\_\_\_\_
4. 姓名: \_\_\_\_\_ 性別:  女  男  
年齡: \_\_\_\_\_ 出生地: \_\_\_\_\_ 職業: \_\_\_\_\_  
常用的語言:  國語  台語  客語  其他 \_\_\_\_\_  
與其他參與者的關係: \_\_\_\_\_

#### SIMPLE DESCRIPTION OF THE EXCERPT (a list of topics)

\_\_\_\_\_

附錄三 行政院客家委員會同意使用客家電視台節目之

公文

外國語文學院

行政院客家委員會 函

機關地址：110 台北市松仁路 3 號 8 樓  
聯絡人：盧燕萍  
聯絡電話：(02)87894567  
傳真：(02)87884710

116

臺北市文山區指南路 2 段 64 號

受文者：國立政治大學

發文日期：中華民國 96 年 1 月 15 日

發文字號：客會傳字第 0960000537 號

類別：普通件

密等及解密條件或保密期限：普通

附件：如說明

主旨：有關請本會授權客家電視影音資料供 貴校口語語料庫使用乙案，復請 查照。

說明：

- 一、復 貴校 96 年 1 月 10 日政院外字第 0960000419 號函。
- 二、有關 貴校擬收集客家電視影音資料以建置客語口語語料庫乙案，本會樂觀其成，並請先行擇抉合意節目，再行洽談後續授權事宜。
- 三、隨函檢送客家電視節目資料乙份供參。

正本：國立政治大學

副本：財團法人公共電視文化事業基金會、客家電視台、本會傳播媒體中心

主任委員 盧燕萍

趙山 1/18

1. 本行擬定中身先「客語口語語料庫」計畫主持人賴嘉玲於後。

2. 本行擬定是「閩語」有者。  
趙山 1/18

共 1 頁 第 1 頁

0741

總收文號第 0960000741 號

96 年 01 月 17 日



96500003

附錄四 圖片掃描自 Frog Story (Mayer, 1969)

