

出國報告 (出國類別：國際學術會議)

赴波蘭華沙參加第五屆學習者語料
庫研究國際學術研討會 (The 5th
Learner Corpus Research
Conference)

服務機關：海軍軍官學校

姓名職稱：黃蘭棻副教授

派赴國家：波蘭

出國期間：108年9月6日-9月15日

報告日期：108年9月19日

摘要

本次至波蘭華沙與科技部國際合作研究人員捷克查理大學Tomáš Gráf博士，進行研究會議三天，包括討論107年度科技部專題研究計畫新收錄之語料、結案報告和後續論文發表，以及108-109年度計畫之研究工作分配。之後兩人共同參加第五屆學習者語料庫研究國際學術研討會(The 5th Learner Corpus Research Conference)並發表論文，報告如何使用錯誤分析，檢視不同英語程度之學習者口語語料之正確性，以及未來整合新收錄語料的計畫。該研討會每兩年舉辦一次，今年由學習者語料庫學會主辦，華沙大學應用語言學院承辦，為此領域重要的國際研討會。本次研討會有來自波蘭當地和世界各國的研究學者、研發人員和語言教師，共同參與，內容相當豐富，與本人近幾年的研究，有高度相關性，也對往後兩年研究計畫的執行，多有幫助。

關鍵詞：學習者語料庫、錯誤分析、口語語料

目次

一、目的	1
二、過程	2
三、心得及建議	5
四、附錄	7
附錄 1：活動照片	7
附錄 2：發表論文之英文摘要	9

一、目的

本次出國目的有二項，第一、與捷克查理大學 Tomáš Gráf 博士召開國際合作研究會議，交換分析結果、檢討科技部專題研究計畫執行過程和討論成果發表事項。第二、參與第五屆學習者語料庫研究國際學術研討會並共同發表研究成果。

二、過程

本次出國參加國際合作研究會議和國際學術研討會之過程概述如下：

9月6日下午自左營搭乘高鐵前往桃園機場，轉搭荷蘭航空經荷蘭阿姆斯特丹轉機至波蘭華沙蕭邦機場。9月7日抵達後先至會場附近旅館準備國際合作研究會議資料。之後與捷克查理大學 Graf 博士進行多次國際合作研究會議，分別討論 107 年度科技部專題研究計畫新收錄的語料、研討會論文發表、語料分析、待投稿之論文初稿，以及 108-109 年度兩年期科技部研究計畫之工作分配。

9月11日中午至華沙大學應用語言學院報到，參加第五屆學習者語料庫研究國際學術研討會(The 5th Learner Corpus Research Conference)所提供的第一場工作坊 Learner Corpus Research: Merging Industry and Academic Interests。此場次由劍橋大學出版社兩位英語教材研發經理講授如何使用約五千五百萬字劍橋學習者語料庫，來研究不同母語背景和不同英語程度之學習者書面語料。

12 至 14 日大會安排 4 場專題講座、47 個論文報告、12 個工作報告、13 個海報論文和 2 個軟體展示。與會者超過一百位，分別來自二十四個國家。第一天專題演講講者為比利時魯汶天主教大學 Paquot 教授，題目為 Expanding the scope of complexity research in SLA:

a phraseological perspective，詳細介紹近幾年來如何使用學習者語料庫研究方法，探究詞組多樣性和成熟度。第二天專題演講由美國北亞利桑那大學 Plonsky 教授講授 Methodological reform and learner corpus research，Plonsky 教授研究範圍相當廣泛，且有很多期刊編輯經驗，他分析將近四百個學習者語料庫研究案例，評論六十幾種研究方法，分享此領域的研究現況，也提出幾個建議事項，尤其在統計檢驗、語料分享和研究仿製的議題。第三天專題演講為 Corpora & Cambridge Assessment English: A widening perspectives，講者 Brenchley 博士來自英國劍橋英語檢測機構，主要介紹劍橋語料庫資源和其研究如何應用於英文寫作評分。除了三場專題講座，本人出席多場論文發表，主題為語料複雜性、語料庫與語言教學、英語口語流暢度、英文正確性、第二語言習得等相關研究。

13 日下午場本人與 Gráf 博士共同發表之論文，題目為 Accuracy in spoken learner English at B2 and C1 levels (and future inclusion of A2 and B1 levels)，會中報告英語程度為歐洲語言能力參考指標(CEFR) B2 和 C1 等級之英語口語語料錯誤分析結果和 A2、B1 新收錄語料的基本規模。

9 月 14 日下午研討會結束後，本人即前往波蘭華沙蕭邦機場搭乘荷蘭航空班機，經荷蘭阿姆斯特丹返回台灣桃園機場，再搭乘高鐵返校。

三、心得及建議

根據主辦單位報告，此次國際學術研討會最多的論文報告是與學習者語言複雜度相關，但這些論文皆是研究書面英文，至今仍然極少學者開始研究英語口語複雜度，但這讓我和國際合作研究學者 Gráf 博士更想盡速研究此議題，也更積極去思考如何進一步再探索我們所建置的學習者英語口語語料庫。

會中有兩篇論文報告是有關語料庫於英語教學的應用方法，學者指出教學場域的實際授課老師雖然對語料庫研究表示認同，但卻對使用語料庫自編教材的方法，不予肯定，主要是因為市面上教材多元又成熟，在教學負荷沉重之下，不願意再多花時間，使用語料庫編制教材，這些初步研究引發本人將語料庫帶入教學的研究方向。

學習者英語口語之研究與書面語相較，數量仍是不足，最主要原因是資源稀少，口語語料往往需要耗費好幾十倍的時間建置，也因如此，研究成果多是書面語。本計畫所參與的 Louvain International Database of Spoken English Interlanguage (LINDSEI) 國際專案是少數學習者口語英語語料庫，目前有二十二個國家參與建置，此次研討會除了本人與 Graf 博士報告台灣和捷克兩個子語料庫的研究成果，還有德國、比利時和挪威子語料庫的研究人員，分別發表相關研究成果。

感謝科技部經費支持本人於國際學術研討會中發表研究結果，亦可與各國相關領域的專家學者交流討論，同時能多方面吸取最新的研究方向和方法，能從更多元的角度來使用新建置的學習者語料庫，進行後續 108 和 109 年度科技部專題研究計畫。

四、附錄

附錄 1：活動照片



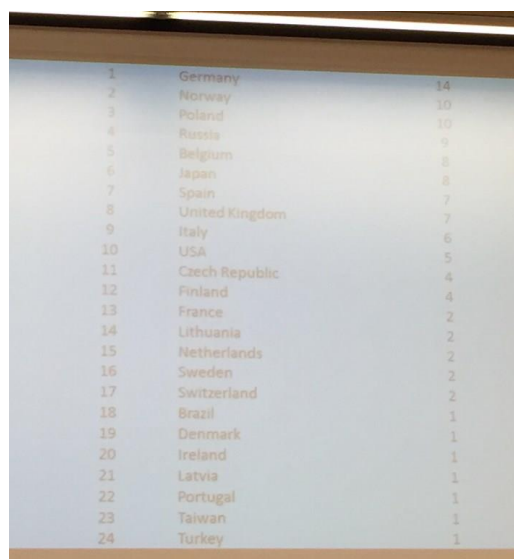
照片 1：與捷克查理大學 Tomáš Gráf 博士討論國際合作研究計畫



照片 2：論文發表



照片 3：大會主辦人華沙大學應用語言學院 Agnieszka Leńko-Szymańska 博士



1	Germany	14
2	Norway	10
3	Poland	10
4	Russia	9
5	Belgium	8
6	Japan	8
7	Spain	7
8	United Kingdom	7
9	Italy	6
10	USA	5
11	Czech Republic	4
12	Finland	4
13	France	2
14	Lithuania	2
15	Netherlands	2
16	Sweden	2
17	Switzerland	2
18	Brazil	1
19	Denmark	1
20	Ireland	1
21	Latvia	1
22	Portugal	1
23	Taiwan	1
24	Turkey	1

照片 4：主辦單位提供與會者國籍和人數



照片 5：本人與 LINDSEI 德國和捷克子語料庫建置人員 Goetz 博士(左) 和 Graf 博士(右)



照片 6：專題講者劍橋英語檢測機構 Brenchley 博士



照片 7：大會資料



照片 8：本人研習證書

附錄 2：發表論文之英文摘要

Accuracy in spoken learner English at B2 and C1 levels (and future inclusion of A2 and B1 levels)

Tomáš Gráf, Charles University, Prague, Czech Republic
Lan-fen Huang, Republic of China Naval Academy, Taiwan

One of the many aspects of spoken learner language mentioned in the CEFR (Council of Europe, 2001; 2018) is accuracy. It characterizes B2 speakers as having “a relatively high degree of grammatical control” and no longer producing errors which cause misunderstandings. C1 speakers are characterized by “a high degree of grammatical accuracy” with rare, difficult to spot, and generally corrected errors. As this cursory description deals with only grammatical accuracy, the present study explores the accuracy of B2 and C1 spoken learner English more broadly, including grammatical, lexical and lexico-grammatical errors, and aiming to identify which types of errors characterize each level, and which errors either disappear or persist at level C1. Also considered is the effect of task design, and of the two typologically different L1s of the learners. This research is unique in that it exploits corpora which have been assessed for proficiency and thus provides an empirical basis for the understanding of CEFR B2 and C1 accuracy, and identifies particular areas of difficulty for speakers at these levels.

The data derives from the Czech (Gráf, 2015a) and Taiwanese (Huang, 2014) components of LINDSEI¹. These two subcorpora have been rated for accuracy by professional IELTS examiners who also received a special CEFR rater-standardisation training (Huang et al., 2018). The size of the dataset and the distribution of levels is shown in Table 1. B1 and C2 speakers were excluded from the analysis. The transcriptions of the remaining 89 speakers’ 15-minute interviews were error-tagged using the Louvain error-tagging manual (Dagneaux et al., 2008) extended by Gráf (2015) to include 59 error types at grammatical, lexical, lexico-grammatical, and syntactical levels and taking into account characteristic features of speech grammar and of the genres determined by the tasks.

¹ Louvain International Database of Spoken English Interlanguage (Gilquin et al., 2010)

	B1	tokens	B2	Tokens	C1	tokens	C2	tokens	Total
	(n)		(n)		(n)		(n)		tokens
Czech	0	0	13	24,165	35	66,305	2	5,499	95,969
Taiwanese	9	10,028	39	55,707	2	3,785	0	0	69,520
Total	9	10,028	52	79,872	37	70,090	2	5,499	165,489

Table 1. Size of the dataset: numbers of speakers and of tokens the speakers at different proficiency levels produced.

A total of 5,108 errors was identified. The comparison of error rates (errors per 100 words, henceforth phw) between the two levels of proficiency showed that B2 speakers produce errors at a higher frequency (6.7 errors phw at B2 and 1.9 errors phw at C1), and this was similar for the performance in the three different tasks (monologue, dialogue and picture description). Grammatical errors are the most frequent, followed by errors of lexical nature. Other types are much less frequent (see Table 2).

	Morphological errors	Grammar errors	Lexical errors	Lexico-grammatical errors	Word order errors	Infelicities
B2	7 (0.2%)	3,023 (72.5%)	701 (16.8%)	152 (3.7%)	196 (4.7%)	76 (1.8%)
C1	3 (0.3%)	558 (59.4%)	277 (29.5%)	55 (5.9%)	34 (3.6%)	13 (1.4%)

Table 2. Representation of error types at B2 and C1 levels.

As for grammar, article and verb tense errors are the most frequent for both groups. At C1, these are less frequent and also some types of errors common at B2 do not appear here. Persistent errors typically involve the use of articles, of the present perfect, and in tense agreements. Differences also occur between Czech and Taiwanese speakers. The latter produce errors at a higher frequency and are more prone to commit errors in areas which the English Vocabulary Profile classifies as B1. Lexical errors involve mostly those affecting single lexemes (esp. prepositions).

Qualitative analysis of a selection of these errors reveals that none of these errors impact intelligibility. While the CEFR claims that at C1 errors are generally corrected, this is not the case here: most of the errors still seem to be errors of competence rather than performance and the speakers do not appear to be aware of them and are thus suitable targets for pedagogical intervention.

References

- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment Companion Volume with New Descriptors*. Strasbourg Cedex: Council of Europe.
- Dagneaux, E., Denness, S., Granger, S., Meumier, F., Neff, J., & Thewissen, J. (2008). *The Louvain Error Tagging Manual Version 1.3*. Centre for English Corpus Linguistics, Université catholique de Louvain. Louvain-la-Neuve.
- Gilquin, G., De Cock, S., & Granger, S. (Eds.). (2010). *LINDSEI Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Gráf, T. (2015a). *Accuracy and fluency in the speech of the advanced learner of English*. (PhD thesis), Charles University, Prague. Retrieved from <https://is.cuni.cz/webapps/zzp/detail/151663>
- Gráf, T. (2015b). *Korpus LINDSEI_CZ*. Prague: Charles University.
- Huang, L.-f. (2014). Constructing the Taiwanese component of the Louvain International Database of Spoken English Interlanguage (LINDSEI). *Taiwan Journal of TESOL*, 11(1), 31-74.
- Huang, L.-f., Kubelec, S., Keng, N., & Hsu, L.-h. (2018). Evaluating CEFR rater performance through the analysis of spoken learner corpora. *Language Testing in Asia*, 8(14), 1-17.