科学研究費助成事業

平成 28 年

研究成果報告

平成 28 年 6月 21 日現在
と
开究種目: 基盤研究(C)(一般)
开究期間: 2013 ~ 2015
果題番号: 25370619
F究課題名(和文)Development of the Specialized Corpus of Civil Engineering Research Articles (SCCERA)
开究課題名(英文)Development of the Specialized Corpus of Civil Engineering Research Articles (SCCERA)
开究代表者
アレクサンダー ギルモア(Gilmore, Alexander)
東京大学・工学(系)研究科(研究院)・准教授
研究者番号:2 0 5 8 6 9 8 7

交付決定額(研究期間全体):(直接経費) 2,300,000円

研究成果の概要(和文): 土木工学研究に関する記事(SCCERA)のコーパスは専門語800万語対応が完了、土木学科の スタッフと学生にオンラインで利用できる東京大学のシステムです。コーパスはスタッフ会議を通じて部門内・大学院 生へ提供され、アカデミックライティングコースで進められています。専門的なコーパスの構築・その提供が可能とな り、その情報説明を行う「SCCERAの紹介」と題する入門ガイドを部門内で作成し配布しました。当部門のメンバーから のフィードバックは良好です: コーパスはまた、英国と米国での会議で国際的に推進されており、これに関する研究論 文の二編はさらに土木研究コミュニティに情報を発信するために生産されてきました。

研究成果の概要(英文): The 8 million word Specialised Corpus of Civil Engineering Research Articles (SCCERA) is now complete and available online to staff and students in the Department of Civil Engineering, University of Tokyo, for corpus queries. The corpus has been promoted within the department through staff meetings and in the academic writing courses provided for post-graduate students. An introductory guide, entitled 'An introduction to SCCERA', explaining how the specialised corpus was constructed and what information it can provide, was produced and distributed within the department. The corpus has also been promoted internationally at conferences in the UK and USA and two research papers have been produced to further disseminate information to the civil engineering research community.

研究分野: corpus linguistics

キーワード: specialized corpora civil engineering second language writing

1.研究開始当初の背景

Language corpora have been available to linguistics community since the the mid-1960s when the one million-word Brown Corpus of American English was originally constructed. The first 'corpus-informed' dictionary (the American Heritage Dictionary) quickly followed, and today, all of the dictionaries produced by the major publishers, as well as many grammar reference books (e.g. Sinclair 1990; Carter & McCarthy 2006) and textbooks (e.g. Willis & Willis 1998), are based on large general corpora (Kennedy 1998). The value of corpora lies in their ability to show us how language is really used in discourse communities; traditional materials relied heavily on native-speaker intuitions, which are notoriously unreliable, and therefore ran the risk of providing learners with a distorted view of the target language (Wolfson 1986; Biber, Conrad & Reppen 1998).

Whilst the 'mega-corpora' available today have been crucial in providing a solid foundation for our understanding of more general lexico-grammatical patterning in English, they are less helpful for analysis of the language used in specific academic or professional contexts. Large variability has been found to exist between different disciplines in terms of word frequencies, collocational patterns and rhetorical moves. For example, Hyland (2008), comparing 4-word lexical bundles from the fields of Biology, Electrical Engineering, Applied Linguistics and Business Studies, calculated that over half of the extended collocations in each discipline did not occur in the other subject areas examined: 4-word bundles like as shown in figure or it can be seen appeared

to be unique to the Electrical Engineering sub-corpus. Given these wide discrepancies in the linguistic characteristics of different academic disciplines, it would seem sensible to use specialized corpora as the starting point in the design of English for Specific Purposes (ESP) materials. This project aims to produce a Specialized Corpus of Civil Engineering Research Articles (SCCERA), which can help to identify key linguistic or rhetorical patterns in the field. These insights will then be used produce to corpus-informed materials for academic writing classes in Engineering Faculties throughout the world, in order to help non-native speaking post-graduate students and academic staff acquire the specific rhetorical practices of their target discourse community.

2.研究の目的

The aim of this research project was to design and create the Specialized Corpus of Civil Engineering Research Articles (SCCERA), using data from high-impact articles in peer-reviewed, influential journals, cited in the Science Citation Index (SCI). The pedagogic motivation behind the project was to identify linguistic and rhetorical patterns common to this specialized genre, and to help non-native academic staff and post-graduate students in engineering departments where English is not the first language write effective research documents that meet the expectations of the international scientific community.

3.研究の方法

Phase 1: The initial phase involved consultation with academic staff and

post-graduate students in the Department of Civil Engineering at the University of Tokyo, as well as colleagues abroad, on the make-up of the specialized corpus (SCCERA).

Phase 2: The second phase involved construction of the specialized corpus (SCCERA). This time-consuming process involves a number of steps:

- a) Scanning and conversion of 'hard' (paper) copies of research articles (RAs) into text-only electronic files, using Optical Character Recognition (OCR) software.
- b) Conversion of downloaded RAs from
 Portable Document Format (.pdf) to
 text-only files, using Microsoft Word.
- c) Removal of unwanted aspects of the RAs, including dates of acceptance, author affiliation, contact information, or layout features.
- d) Labeling of text-only files to facilitate storage and retrieval functions in the completed corpus. Files can be identified by characteristics such as: specialty (Coastal Engineering, Remote Sensing, etc.); journal; English variety (British, American, Indian, etc.); sub-section (titles, abstract, keywords, introduction, methods, results, discussion, conclusion, section headings, figure/table captions, acknowledgments and references).
- e) Part-of-speech annotation (POS tagging) of the corpus using CLAWS software.

Phase 3: The third phase involved preliminary analysis of the SCCERA corpus using WordSmith Tools, Version 6.0 (Scott 2011). This sought to answer key research questions associated with the lexicogrammatical characteristics of civil engineering research articles. The analysis

included an investigation of the following features across different text files. RA sub-sections or sub-corpora within SCCERA: word frequencies; keywords and key keywords; common collocation patterns (including 3-, 4- and 5-word bundles); common grammatical structures; standardized type/token ratios; pedagogically significant concordance lines in order to, for example, disambiguate near-synonymous words (Lee & Swales 2006: 63).

Phase 4: The fourth phase of the investigation aimed to explore the potential pedagogic applications of the findings from the corpus/discourse analysis carried out SCCERA. above using An indirect. corpus-informed, approach facilitated the production of materials for an English for Specific Academic Purposes (ESAP) writing course for civil engineers, which were trialed students with post-graduate in the Department of Civil Engineering at Tokyo University.

Phase 5: The final phase of the project involved dissemination of the research results through workshops, conference presentations within Japan and abroad, and publication of research articles.

4.研究成果

The 8 million word Specialised Corpus of Civil Engineering Research Articles (SCCERA) is now complete and available online to staff and students in the Department of Civil Engineering, University of Tokyo, for corpus queries. The corpus has been promoted within the department through staff meetings and in the academic writing courses provided for post-graduate students. An introductory guide, entitled 'An introduction to SCCERA', explaining how the specialised corpus was constructed and what information it can provide, was produced and distributed within the department. Feedback from members of the department has been very positive: out of a total of 30 post-graduate students or professors who responded to a Survey Monkey online questionnaire, 67% believed that SCCERA was useful or very useful for their studies. The corpus has also been promoted internationally at conferences in the UK and USA and two research papers have been produced to further disseminate information to the civil engineering research community.

< 引用文献 >

Biber, D., S. Conrad & R. Reppen (1998). Corpus linguistics: Investigating language structure and use. Cambridge: Cambridge University Press.
Carter, R. & M. McCarthy (2006). Cambridge grammar of English. Cambridge: Cambridge University Press.

Cheng, W. (2010). 'Hong Kong Engineering Corpus: Empowering professionals-in-training to learn the language of their profession'. In M.C. Campoy, M.L. Gea-Valor & B. Belles-Fortuno (Eds.), *Corpus-Based Approaches to English Language Teaching*. London: Continuum: 67-78. Connor, U. & T.A. Upton (Eds.) (2004). *Discourse in the professions: Perspectives from corpus linguistics*. Amsterdam: John Benjamins.

Feak, C.B. & J. Swales (2010). 'Writing for publication: Corpus-informed materials for postdoctoral fellows in perinatology'. In N. Harwood (Ed.), *English Language Teaching Materials*. Cambridge: Cambridge University Press, 279-300.

Hyland, K. (2000). *Disciplinary Discourses*. Harlow: Longman.

Hyland, K. (2008). 'As can be seen: Lexical bundles and disciplinary variation'. *English for Specific Purposes* 27: 4-21.

Kennedy, G. (1998). An introduction to corpus linguistics. New York: Longman.

Lee, D. & J. Swales (2006). 'A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora'. *English for Specific Purposes* 25: 56-75. Mudraya, O. (2006). 'Engineering English: A lexical frequency instructional model'. *English for Specific Purposes* 25: 235-256.

Orr, T. & A. Takahashi (2002). 'Constructing a corpus of fundamental engineering English for nonnative speakers'. In J. Williams (Ed.), *Proceedings of the IEEE International Professional Communication Conference*. Oregon, USA: 403-409.

Peacock, M. (2002). 'Communicative moves in the discussion section of research articles'. *System* 30(4): 479-497.

Pearson, J. (1998). *Terms in context*. Amsterdam: John Benjamins.

Rea Rizzo, C. (2010). 'Getting on with corpus compilation: From theory to practice'. *ESP World* 1(9): 1-22.

Römer, U. (2010). 'Using general and specialized corpora in English language teaching: Past, present and future'. In M. Campoy-Cubillo, B.
Belles-Fortuño & M. Gea-Valor (Eds.). *Corpus-based Approaches to English Language Teaching*. London: Continuum, 18-35.

Scott, M. (2011). WordSmith Tools Version 6, Liverpool: Lexical Analysis Software. Sinclair, J. M. (1990). Collins COBUILD English
grammar. London: Collins.
Swales, J.M. (1990). Genre Analysis: English in
Academic and Research Settings. Cambridge:
Cambridge University Press. Swales, J.M. & C.B.
Feak (1996). Academic writing for graduate
students. Ann Arbor: University of Michigan
Press.

Thurstun, J. & C. Candlin (1997). Exploring academic English: A workbook for student essay writing. Sydney: NCELTR Publications.
Willis, J. & D. Willis (1988). Collins COBUILD English course. London: Collins.

Wolfson, N. (1989). *Perspectives*. Boston MA: Heinle & Heinle Publishers.

5.主な発表論文等

[雑誌論文](計 2件)

Alexander Gilmore, "Developing a Specialized Corpus of Civil Engineering Research Articles (SCCERA)". *European Journal of Engineering Education* (submitted for review). 査 読有

Alexander Gilmore "The Language of Civil Engineering Research Articles: A Corpus-Based Approach". *European Journal of Engineering Education* (submitted for review). 查読有

[学会発表](計 2件)

"The language of civil engineering research articles – a corpus-based approach". September 3rd – 5th, 2015. 48th Annual Meeting of the British Association for Applied Linguistics, University of Aston, Birmingham, UK. "Building a Specialized Corpus of Civil Engineering Research Articles (SCCERA): No walk in the park". September $4^{th} - 6^{th}$, 2014. 47^{th} Annual Meeting of the British Association for Applied Linguistics, University of Warwick, Coventry, UK.

6.研究組織

(1)研究代表者
 Alexander Gilmore
 (アレクサンダー ギルモア)
 東京大学大学院工学系研究科・准教授
 研究者番号:20586987

(2)研究分担者
 Michael Handford
 (ハンドフオード マイク)
 東京大学大学院工学系研究科・教授
 研究者番号: 00466781