

LSP Learner Corpus: Exploring Students' Abstracts of Medical Mock Experiment

Shozo YOKOYAMA
Faculty of Medicine
University of Miyazaki
Miyazaki, Japan
yokoyama@med.miyazaki-u.ac.jp

Chizuko SUZUKI
Graduate School of Humane Studies & the Department of English
Nagasaki Junshin Catholic University
Nagasaki, Japan
suzuki@n-junshin.ac.jp

Tamao ARAKI
Faculty of Nursing
Miyazaki Prefectural Nursing University
Miyazaki, Japan
taraki@mpu.ac.jp

Akira SAWAGUCHI
Faculty of Medicine
University of Miyazaki
Miyazaki, Japan
sawa@fc.miyazaki-u.ac.jp

In our previously published study, we investigated the corpus of scientific articles (PubMed Central <http://www.ncbi.nlm.nih.gov/pmc/>) consisting of a total of more than 5 million words in the field of medicine and nursing. The data was accumulated and categorized according to the following specialized subfields: genome biomedicine, clinical surgery, and nursing and public health. All articles were POS-tagged, so that the frequency of verbs and modal verbs in each rhetorical section of the different subfields could be identified respectively using WordSmith Tools (Scott, 2004). The results of the analysis were then used to develop a web-based concordance search interface and Moodle-based courseware containing a basic familiarization exercise for these verbs and sample sentences with a Japanese translation.

This research originally sought 'correct' and 'practical' ways of helping students write scientific research papers. Since then, by utilizing a more learner-centered perspective, we have gradually come to notice that there would be great advantages in analyzing learner written data in order to understand their error patterns, which could help students acquire better writing competence. To do so, we first collected learner data and compiled a small corpus database entitled Students' Abstracts of Medical Mock Experiment (SAME) corpus. The data was obtained from first year PhD students in the Graduate School of Medicine of the

University of Miyazaki, in 2011. The number of participants was 20. And the token number of the learners' corpus was 3,692.

The students were required to produce an English abstract of at least 200 words on a given mock medical experiment. The abstract was required to be written with attention to the IMRD (Introduction, Materials and Methods, Results and Discussions) structure or the Background, Methods, Results and Discussions structure, but had to be finalized as a single meaningful, cohesive paragraph, not as separate sections. The topic of the experiment was a helicobacter pylori infection and the outline of the experiment was briefly explained in Japanese.

Technical terminology was also provided in both Japanese and English so that we could focus primarily on an analysis of the use of basic academic vocabulary and discourse markers. The abstracts were allotted to students as assignments, and the students were allowed to use dictionaries and/or electronic spell checkers, but the manuscripts could be submitted as either handwritten or electronic manuscripts. Each manuscript was collected and compiled electronically as a text file. The data was analyzed using WordSmith Tools.

A close comparison between the contents that should be expressed and the learners' actual lexical choices revealed that many participants often lacked the effective use of formulaic expressions, proper understanding of collocations, and an awareness of academic discourse as a distinct register/genre. We will also discuss how these findings may contribute to supporting students in medical academic writing.

References

- Gilquin, G., Granger, S. & Paquot, M. (2007). 'Learner corpora: The missing link in EAP pedagogy.' *Journal of English for Academic Purposes* 6: 319-335.
- Scott, M. (2004). *Wordsmith tools*. Oxford: Oxford University Press.
- Vihla, M. (1999). *Medical Writing: Modality in Focus*. Amsterdam: Rodopi.
- Yokoyama, S., Suzuki, C., Yasunami, S. & Kawakita, N. (forthcoming). 'A direct application of medical corpora to academic writing: A specialized concordance search interface and Moodle-based courseware.' In Boulton, A. & Thomas, J. (eds). *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press.