

Corpora and teaching academic writing: Exploring the pedagogical potential of MICUSP

Ute Römer
Georgia State
University

EAP teaching, phraseology, MICUSP,
data-driven learning, student writing

This paper introduces a new corpus resource for teaching English for Academic Purposes (EAP) and for carrying out research into advanced student writing across a range of academic disciplines: MICUSP, the Michigan Corpus of Upper-level Student Papers. It exemplifies how MICUSP can be used as a source of materials in the context of data-driven learning in the EAP classroom. The paper also presents selected results from two MICUSP-based studies: an analysis of attended and unattended *this* and a study of the distribution of common phraseological items across student papers. It is argued that these results are of pedagogical relevance and can contribute to an improvement in EAP teaching.

1. Introduction

One core topic within the field of corpus linguistics and language teaching and one area in which TaLC participants from around the world have been active over the last ten years or so is the teaching of English for Academic Purposes (EAP) with a focus on academic writing. Up to now, applied corpus research in this area has mainly focused on learner writing as captured in learner corpora such as the International Corpus of Learner English (ICLE) and on published expert writing as covered in subsets of large reference corpora such as the British National Corpus (BNC) or the Corpus of Contemporary American English (COCA).¹ Pedagogically-oriented studies on these two types of data include Altenberg and Granger (2001), Coxhead (2000), Nesselhauf (2005), and Simpson-Vlach and Ellis (2010). Advanced student writing, that is writing produced by upper-level undergraduate and graduate students, has not yet been given much attention – mainly due to a lack in appropriate resources.²

This paper serves to introduce the Michigan Corpus of Upper-level Student Papers (MICUSP) as a new resource that provides EAP and English as a Second Language (ESL) teachers, students, and corpus researchers access to 829 A-graded writing samples produced by senior undergraduate and first- to third-year graduate students in 16 different disciplines at a large American research university. Consisting of advanced and successful student writing across disciplines, levels and text types, MICUSP presents data that can be used by instructors in the EAP classroom and serves as a model to less advanced student writers. Students seeking to improve their academic writing skills can access MICUSP as a tool to explore patterns of academic writing in a self-tutoring fashion. Linguistic researchers can use MICUSP to explore the vocabulary and phraseology, grammatical structures, text and discourse patterns in a previously unavailable type of data.

After a brief description of the corpus, its design and teacher-inspired annotation, this paper will discuss how MICUSP can be used to enhance the teaching and learning of academic writing. It begins with an illustration of how MICUSP is accessible through MICUSP Simple, a search and browse interface to the corpus that has been designed with users in mind who are *not* trained in corpus analysis but may wish to find out more about advanced student writing across academic disciplines. It also exemplifies how MICUSP Simple can be used directly by teachers and learners in the EAP/ESL classroom. The final part of the paper discusses examples of indirect pedagogical applications of the corpus and presents selected results from two MICUSP-based case studies, one on the use of attended and unattended *this* (Römer & Wulff 2010; Wulff, Römer & Swales forthcoming), the other on the distribution of frequent phraseological items across disciplines and text positions (O'Donnell & Römer in preparation).

¹ The BNC and COCA are both accessible on the site provided by Mark Davies at Brigham Young University: <http://corpus.byu.edu/>; for information about ICLE, see <http://www.uclouvain.be/en-ceed.html>.

² One notable exception here is the British Academic Written English (BAWE) corpus. For information about the corpus, see <http://wwwm.coventry.ac.uk/researchnet/BAWE/>, Alsop and Nesi (2009) and Nesi (2011).

2. The Michigan Corpus of Upper-level Student Papers (MICUSP): Design and annotation

MICUSP was designed to provide a global snapshot of high-quality student writing assignments from a large American research university, the University of Michigan, Ann Arbor. Papers of different types (argumentative essay, creative writing, critique/evaluation, report, research paper, research proposal, response paper) have been collected from a wide range of disciplines within four academic divisions (Humanities and Arts, Social Sciences, Biological and Health Sciences, Physical Sciences). The 16 disciplines included in MICUSP are, in alphabetical order: Biology (BIO), Civil and Environmental Engineering (CEE), Economics (ECO), Education (EDU), English (ENG), History and Classical Studies (HIS), Industrial and Operations Engineering (IOE), Linguistics (LIN), Mechanical Engineering (MEC), Natural Resources and Environment (NRE), Nursing (NUR), Philosophy (PHI), Physics (PHY), Political Science (POL), Psychology (PSY), and Sociology (SOC). Figures 1 and 2 show the distribution of MICUSP papers across disciplines and text types, indicating the number of texts in each category (e.g. 67 papers in Biology; 186 argumentative essays).

The papers included in MICUSP were written by male and female native and non-native speaker students at four different levels of study: final-year undergraduate, and first-, second- and third-year graduate level. Altogether 829 papers, totaling almost 2.6 million words, have been collected through an online submission interface that captures the required metadata pertaining to a student's linguistic background and academic experience. Each of the papers in MICUSP has been marked up in TEI-compliant XML (see Burnard 2005) and maintains the structural divisions

Figure 1. Distribution of MICUSP papers across academic disciplines

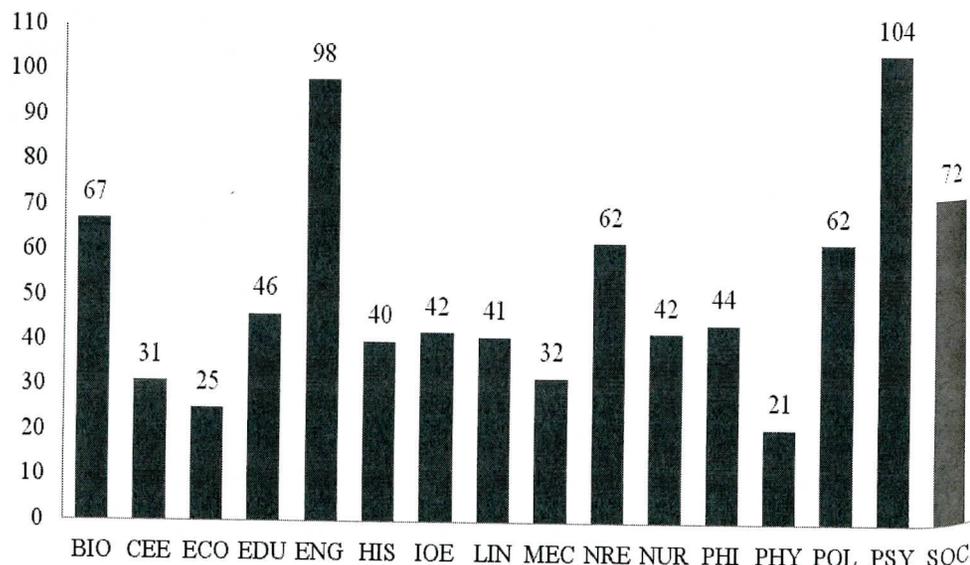
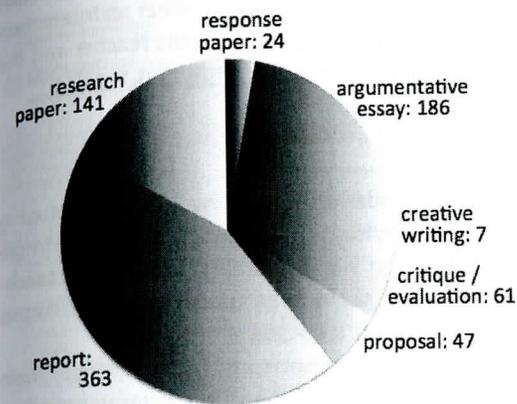


Figure 2. Distribution of MICUSP papers across text types



(sections, headings, paragraphs) of the original paper as well as emphatic features like quotation marks or italics.

Supplementing the encoding of text-structural phenomena, we carried out a two-step pedagogical annotation which was inspired by conversations with university EAP teachers. Step one consisted of a systematic analysis of the papers to classify their text type (e.g. report, research paper; see Figure 2). In step two we annotated each paper for whether or not it contained any of the following eight textual features: abstract, definitions, discussion of results, literature review, methodology section, problem-solution pattern, reference to sources, and tables, graphs or figures. All these features were considered important in teaching academic writing by the EAP teachers we talked to. A file header that has been added to each MICUSP file includes, among other things, information about the text types and features, the discipline and the student's level, native-speaker status, and gender, which makes it possible to filter the set of papers according to certain criteria and to carry out customized searches in subsections of the corpus, for example only in English essays written by native-speaker final-year undergraduate students. Further details on the compilation and design of MICUSP can be found in Ädel and Römer (2012), O'Donnell and Römer (2012), and Römer and O'Donnell (2011).

3. Using MICUSP in teaching academic writing

In discussing how MICUSP can be beneficially used in teaching academic writing, I will distinguish two different types of corpus applications: direct applications and indirect applications (see Römer 2008). Indirectly, corpora, and especially frequency information derived from them, can help with decisions about *what* to teach and *when* to teach it. Indirect corpus applications thus have an effect on the teaching syllabus and the design of teaching materials including textbooks, dictionaries and grammars. Corpora can also be accessed directly by learners and teachers in the classroom, and so become interactive learning and teaching tools. Direct applications mainly affect

how something is taught and learned. They actively involve the learner and teacher in the process of working with corpora and concordances. This method, in which there is either an interaction between the learner and the corpus or, in a more controlled way, between the teacher and the corpus is commonly referred to as data-driven learning (DDL) (see e.g. Johns 1991, 1994).

3.1. Direct pedagogical applications of MICUSP

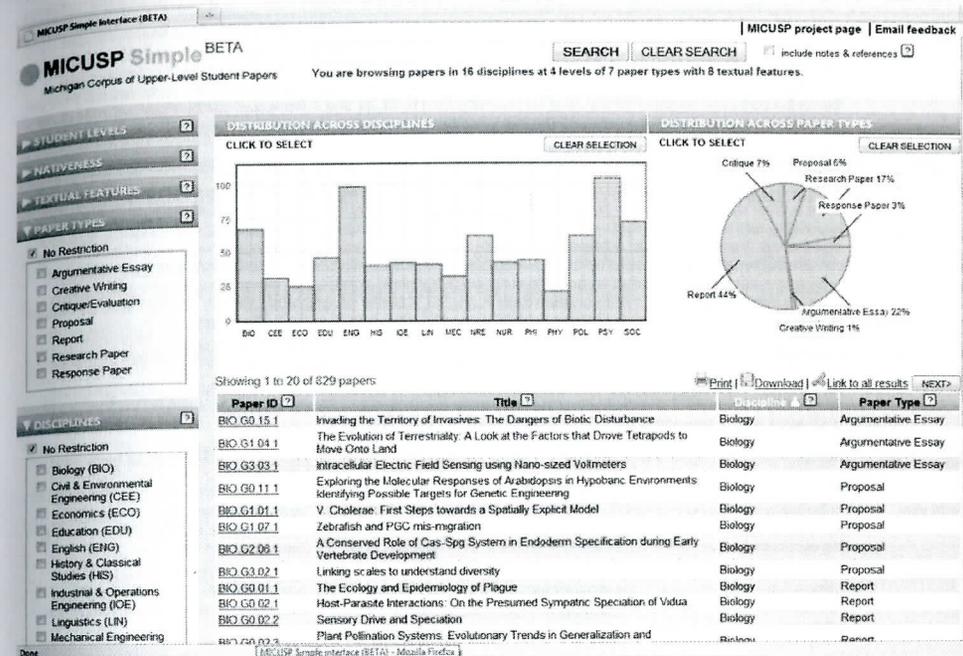
The MICUSP Simple online search and browse tool provides a straightforward way of bringing an academic writing corpus into the EAP classroom.³ With its user-friendly interface, simple functionality and interactive design, it puts MICUSP directly in the hands of learners and teachers. MICUSP Simple does not require any previous experience in corpus use or training in corpus analysis.⁴ Figure 3 shows a screenshot of the MICUSP Simple interface.

The two basic functions in MICUSP Simple are: (1) browsing papers; and (2) carrying out simple word and phrase searches. MICUSP Simple allows you to browse the papers in the corpus according to a series of categories, by interacting with the graphs and by using the checkbox selections on the left hand side (see Figure 3). Each of these feature sets can be hidden or expanded by clicking on the header bar for the feature. For example, in Figure 3, the checkboxes for student levels, nativeness and textual features are hidden while those for paper types and disciplines are visible. Selecting one or more of the departments from the histogram graph or the left hand select list will allow you to view papers belonging to those disciplines only. The histogram and the discipline checkboxes are linked together so making (de-)selections on one will be reflected immediately in the other. The paper type checkboxes and the clickable pie chart are also linked together.

Once you have made selections, a status message at the top of the screen shows you what these are. The list of papers that match your selection are shown in the table in the bottom half of the screen. The paper identification code, title, discipline and paper type are given for each paper. If you hover the mouse pointer over the paper identification code, a popup box will appear that provides further information about the paper and its author. You can export your browse results in different formats by clicking one of the three links on the right immediately above the results table ('print', 'download', 'link to all results'). Clicking on the paper identification code will open up another tab or window in your browser and display the complete paper and related information. The browse view will also give you a word/phrase cloud that consists of key words and phrases in the text and thus provides a quick view on what the paper is about. You can also look at the (anonymized) original version of the paper in PDF format using the 'view original paper' link on this screen.

In addition to browsing papers, MICUSP Simple allows you to search for words and phrases in the whole corpus or just in papers that match your selections of discipline, paper type, student level and textual features. A user enters a word or phrase in the text search box at the top of the screen and clicks 'search'. The number of times the search term occurs and the number of papers

Figure 3. Screenshot of the MICUSP Simple search and browse interface



in which it is found are shown in the text right under the search box (see Figure 3). The histogram displays the results either in terms of actual occurrences of the search term or instances per 10,000 words; hovering over a bar reveals this number. The search results are presented in the table in the bottom half of the screen. As with the paper browsing function, it is possible to export your results in different formats. It is also possible to limit your search results by interacting with the graphs and by selecting checkboxes. For MICUSP Simple users we decided to provide results with an entire paragraph of context as opposed to a KWIC display which may be difficult to read for first-time or less experienced corpus users.⁵ Where there is more than one instance of the search term in the paragraph, the paragraph is only displayed once with all instances highlighted. Clicking on the paper ID for a specific result will open another browser tab and show the whole paper with all paragraphs and instances of the search term highlighted.

As this brief overview of the basic MICUSP Simple functions indicates, the interface can be used by teachers to quickly locate model papers of different types from different disciplines. It can also function as an interactive learning tool for learners who wish to explore patterns of academic writing in a self-tutoring fashion. In order to illustrate the direct pedagogical applications of the tool, let us look at two examples of DDL with MICUSP Simple. The context of the first example is teaching academic clusters, the context of the second one is teaching the writing of abstracts.

³ The MICUSP Simple URL is <http://search-micusp.elicorpora.info>. MICUSP Simple provides free and immediate access to all MICUSP papers, without needing to register.

⁴ For information about the MICUSP Simple usability testing which was carried out prior to the launch of MICUSP Simple, see O'Donnell and Römer (2011).

⁵ On the differences between reading a text and reading a concordance, see Tognini Bonelli (2001).

Figure 4. Part of the results for a MICUSP Simple search for *in figure*

Paper ID	Title	Discipline	Paper Type
BIO.G0.07.2 1 of 3 hits Show all	Plasmid Transfer, Genetic Stability and Nisin Resistance in Lactococci	Biology	Research Paper
BIO.G2.03.1	Cholera Seasonality, Rainfall, and Fadeouts: a Geostatistical Approach	Biology	Research Paper
CEE.G0.01.3	Gradation Analysis and Permeability Tests	Civil & Environmental Engineering	Research Paper
CEE.G0.05.1 All 2 hits Show less	Relative density testing for foundation soil of punch presses	Civil & Environmental Engineering	Research Paper
CEE.G1.02.1 1 of 2 hits Show all	Necking and Localization in Plane Stress / Strain	Civil & Environmental Engineering	Research Paper

In a session on teaching academic clusters, MICUSP Simple can be used to look at the distribution and use of frequent items, such as *on the other hand*, *for example*, *at the same time*, or *in figure*. The histogram and pie chart make it easy to see whether or not these items are roughly equally frequent across academic disciplines and whether or not they are typical of a particular paper type. It may also be interesting and worthwhile to examine where in a sentence or paragraph a common cluster most frequently occurs (which may tell us something about its function in the discourse). This can be easily gleaned from the paragraphs (with the search phrase highlighted) in the search results section of the screen. In addition, learners may profit from looking at the wider context of a frequent item and learn more about its appropriate use. For example, if they would like to know how successful student writers refer to figures in their papers, they can do a search for *in figure* and see that phrases used by MICUSP writers include *in Figure 1*, *the CSS plotted in Figure 4 shows that*, *shown below in Figure 2.1*, *provided in Figure 1*, and *as seen in Figure 2* (see screenshot in Figure 4). They will also note that *in figure* is most common in Engineering (especially Mechanical Engineering) and does not occur at all in disciplines like English or History.

In teaching how to write an abstract instructors can make use of the 'textual features' filter option in MICUSP Simple and easily identify all MICUSP papers that contain an abstract. Selecting the textual feature 'abstract' in browse mode pulls up a set of 100 papers from twelve

Figure 5. MICUSP sample abstracts from Psychology (PSY.G1.03.1, top) and Physics (PHY.G1.01.1, bottom) papers

The Effects of Musical Training on Cognitive Ability

Abstract

Musical ability has long been shown to correlate with other measures of intelligence that extend to traditional academic performance. However, in much of the literature, it is unclear whether this correlation is due to a self-selection process derived from pre-existing, innate abilities, or whether musical education is to be credited. This paper will reveal that musical education is responsible for increases in cognitive ability, though there are also some self-selection factors as well. Among researchers who have demonstrated that exposure to music improves cognitive ability, there is disagreement as to which abilities are improved upon and how. Some researchers claim that verbal abilities are related to music cognition, and that this is mediated in the left planum temporale. Others state that spatial ability is related to music training, and that this is mediated in the right parietal lobe and in whole-brain networks that fire similarly to music and spatial tasks. This paper will review these possibilities and suggest that perhaps both accounts are right - music cognition is complex and draws upon several different cognitive capabilities that relate separately to pitch, rhythm, listening, and production. The paper then calls for future research to be performed using converging methods and experimental designs that elucidate the verbal/spatial debate and apply this information to cognitive development in children.

"Bose Einstein Condensation"

Abstract

In this paper we discuss the Bose Einstein Condensation, from 1924 to 1995. We address questions like how the idea came about, the connection between Bose and Einstein, the theory behind it and some of the interesting applications that come out of this science.

disciplines that can then be looked at individually (the set can of course be further narrowed down by applying additional filters). Presenting learners with a selection of sample abstracts from advanced student papers that have been submitted for a grade may help them learn how abstracts are normally structured and what their central functions or communicative stages are. In an exploratory fashion (with minimal teacher guidance), learners can identify phrases that commonly occur in an abstract, look at which tense forms are used and find out how sentences are connected. Learners can also observe that the length of abstracts may differ considerably across disciplines. Figure 5 displays two abstracts from first-year graduate student papers. The first one comes from a Psychology paper where it is common to have longer abstracts (of around 150 to 200 words), the second one from a Physics paper where abstracts are typically short and consist of only two or three sentences (around 50 words).

3.2. Indirect pedagogical applications of MICUSP

As mentioned above, corpus-derived evidence can have an effect on pedagogical practice in that it informs the teaching syllabus and reference works such as grammars or textbooks. Existing pedagogical descriptions are evaluated in the light of “new evidence” (Sinclair 2004: 271), and new decisions are made about the selection of language phenomena, the progression in the course, and the presentation of the selected items and structures (Mindt 1981; Römer 2005). In the context of teaching academic writing, MICUSP can be used as a source of such evidence. Since the data captured in the corpus are likely to be of relevance to learners and novice academic writers, findings from MICUSP-based research have a clear pedagogical value. I will now summarize a few central findings from two MICUSP-based case studies and discuss their pedagogical implications. The first study examines the use of attended and unattended *this* in MICUSP; the second focuses on common phraseological items and their preferred textual positions in MICUSP papers. Both case studies used offline versions of the corpus because the analytic steps involved went beyond the functionality offered by MICUSP Simple.⁶

Wulff et al. (forthcoming) used MICUSP to address the question: What governs the optional attendance of the determiner *this* by a noun phrase in academic student writing? Although *this* is one of the most frequent words in academic writing, occupying rank 11 in a MICUSP frequency list (see Römer & Wulff 2010), the factors that determine whether it is attended by a noun as in (1) or free-standing as in (2) have not received much attention in corpus research.

- (1) This sentence is an example.
- (2) This is an example.

Also, with respect to this language point, style guides conflict with actual usage. While Markel (2004: 229) for example suggests that “in almost all cases, demonstrative pronouns should be followed by nouns”, demonstrative *this* actually occurs frequently without a noun phrase – in 36% of the cases in a published research articles corpus (Swales 2005), and in 43% of the cases in a pre-release version of MICUSP (Römer & Wulff 2010).

In order to provide clarification on this issue and help better understand when, i.e. in which contexts, *this* is attended or not, Wulff et al. (forthcoming) carried out a systematic quantitative and qualitative analysis of all sentence-initial instances of *this* in MICUSP (5,827 altogether). Their study involved: (i) a logistic regression analysis to determine the probability of attended versus unattended *this* on the basis of a set of predictor variables; (ii) a distinctive collexeme analysis to measure the level of distinctive association between verbs and (un)attended instances of *this*; and (iii) a pattern analysis of the most prominent *this*+verb clusters and their distribution across MICUSP disciplines, student levels and texts. The results point to a strong influence of the verb accompanying (un)attended *this* on whether or not the determiner is followed by a noun. A number of verbs were found to be distinctively associated with attended *this* (e.g. USE, EXAMINE, FIND, BASE, EXPLORE), while other verbs typically appear with unattended *this* and form clusters

⁶ Offline versions of MICUSP consisting of files in XML and txt format will soon be made available together with a MICUSP resource book (Römer & O'Donnell in preparation).

with the determiner (e.g. BE, MEAN, LEAD, IMPLY, SEEM). A closer analysis of a set of frequent *this*+verb combinations (including *this* + *is*, *this* + *means* and *this* + *seems*) showed that they are highly fixed and rarely occur with intervening nouns. It also highlighted a number of interesting distributional trends with respect to the clusters' use in papers from different disciplines and levels, and their preferences (or avoidances) to occur in a particular section of a paragraph or text. Sentence-initial *this seems*, for instance, is most frequent in Philosophy papers and strongly prefers paragraph- and text-final and -medial positions. *This seems* very rarely occurs at the beginning of a paragraph or text.

On the whole, the results of this case study point towards an ongoing delexicalization of *this*+verb clusters like *this is* and *this means* into textual organization markers. These findings are inconsistent with traditional pedagogical descriptions and their claim that unattended *this* is a mere 'vague reference' which should be avoided if at all possible. For the practice of teaching academic writing this means that advice on (un)attended *this* found in textbooks and writing manuals is clearly overgeneralized. To a large extent, it depends on the verb whether attended or unattended *this* is the right choice. The study suggests that instructional materials for both native and non-native speakers of English need to recognize that certain verb forms following *this* do not favor attending nouns. The study also points to a need for cross-disciplinary differences and text-positional preferences to be reflected in pedagogical language descriptions.

In another MICUSP-based study, O'Donnell and Römer (in preparation) focus on the phraseology of student writing. They identify common phraseological items in MICUSP (n-grams such as *in addition to*, and phrase-frames such as *it would be * to*) and relate them to text structure. The main question the authors ask is: Which items typically occur at the beginning, in the middle or at the end of a text, paragraph or sentence? The analysis is based on Hoey's (2005) observations on textual colligation, namely that words and phrases may carry with them associations for occurrences at a specific location in a text or textual unit; they may show a preference for or avoidance of certain textual positions.

In order to retrieve information about an n-gram's or a phrase-frame's preferred location in MICUSP, O'Donnell and Römer (in preparation) tokenized each paragraph in MICUSP texts into words and sentences, and compiled a database which captures for each of the roughly 2.6 million words in the corpus: (i) its paper ID, discipline and student level; (ii) up to eight words of the word's right-hand context; (iii) its word number within the sentence and the sentence length; (iv) its word number within the paragraph and the paragraph length; (v) its word number within the text and the text length; (vi) the number of the sentence it occurs in and the overall number of sentences within the paragraph; and (vii) the number of the paragraph it occurs in and the overall number of paragraphs within the text. The database then made it possible to extract, for each word in MICUSP, detailed positional data according to textual units (sentences, paragraphs) and texts.

The study uncovered some very interesting distributional preferences of phraseological items that are frequently used by MICUSP writers. To give just a few examples, the phrase-frame *it would be * to* (most commonly realized as *it would be interesting to*, *it would be hard to*, and *it would be difficult to*) clearly prefers to occur in sentence-initial and text-final positions but is roughly evenly distributed across the paragraph. Proficient student writers use this item towards the end of their papers to express evaluation and point in the direction of future research. The n-gram *in addition to* occurs most frequently at the beginning of sentences and paragraphs but

does not exhibit a preference for (or avoidance of) any particular position in a text. For learners this means that it is safe to use *in addition to* in any paragraph of a paper, but preferably in paragraph-initial position. Finally, the n-gram *in the future* reverses the sentence- and paragraph-initial trend observed for *in addition to* and mostly occurs in sentence-final position and towards the end of a paragraph. The item also clusters in the final 10% of the text where it is used to give an outlook on things that still remain to be done.

This second MICUSP-based case study is of pedagogical relevance because it highlights which phraseological items are most frequent in upper-level student writing across disciplines and where they most typically occur in a paper. In order to develop into successful members of an academic discourse community, novice academic writers need to identify commonly used phrases in the discourse of their discipline. It is also important for them (and their writing instructors) to know where in a text these common phrases are typically used by successful student writers and how they can help them structure a paper. Many of the items discussed by O'Donnell and Römer (in preparation) serve to express functions that are of central importance in academic writing, such as evaluating or organizing the discourse. Hence students should benefit from learning how to use some of these items appropriately in their papers.

4. Concluding remarks

Although much progress has already been made in providing better descriptions of language for learners and in bringing corpora and concordances into the EAP and ESL classroom, thus providing a more motivating and data-rich learning environment, the list of unresolved issues and tasks for the future of applied corpus linguistics is still rather long. Among other things (see Römer 2011 for a more exhaustive list), there is still a need for a wider range of pedagogically-relevant corpora, for easily (and freely) accessible, user-friendly corpus tools and search interfaces, and for textbooks and reference materials that incorporate corpus research findings and include genuine rather than invented language examples that are relevant for the learner.⁷

In this paper I have introduced a new corpus of proficient upper-level student papers from different academic disciplines: MICUSP. This corpus contains data that is highly relevant to novice academic writers (native and non-native speakers alike) who wish to improve their academic writing skills and learn more about the discourse strategies and language patterns used (and expected) in a particular discipline. Arguably, the type of writing captured in MICUSP serves as a better (and more realistic) model for undergraduate writers and international students than published expert writing. I have also demonstrated what basic functions are offered by MICUSP Simple, the free online search and browse interface to MICUSP, and exemplified how it facilitates direct pedagogical applications of the corpus in the EAP or ESL classroom. Moving on from the direct to the indirect use of MICUSP in teaching academic writing, I then summarized findings⁸ from two MICUSP-based studies and discussed their pedagogical relevance.

I believe that MICUSP could (and should in fact) be used by researchers and materials writers to inform academic writing guides and EAP textbooks. I also see the corpus as a valuable tool for learners and students who are looking for model papers of a certain type or from a certain discipline, and for teachers who would like to spice up their lessons with examples taken from proficient student writing. I hope that MICUSP will find multiple exciting applications in the EAP world and hence become a worthy sister corpus to the successful Michigan Corpus of Academic Spoken English (MICASE).⁸

References

- Ädel, A. 2010. Using corpora to teach academic writing: Challenges for the direct approach. In M.C. Campoy-Cubillo, B. Belles-Fortuño & L. Gea-Valor (eds), *Corpus-Based Approaches to English Language Teaching*. London: Continuum, p. 39-55.
- Ädel, & U. Römer. 2012. Research on advanced student writing across disciplines and levels: Introducing the Michigan Corpus of Upper-level Student Papers. *International Journal of Corpus Linguistics* 17(1): 1-32.
- Alsop, S. & H. Nesi. 2009. Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora* 4(1): 71-83.
- Altenberg, B. & S. Granger. 2001. The grammatical and lexical patterning of 'make' in native and non-native student writing. *Applied Linguistics* 22(2): 173-195.
- Burnard, L. 2005. Metadata for corpus work. In M. Wynne (ed.), *Developing Linguistic Corpora: A guide to good practice*. Oxford: Oxbow Books, p. 30-46.
<http://ahds.ac.uk/linguistic-corpora/>, accessed 07/11/11.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34(2): 213-238.
- Hoey, M.P. 2005. *Lexical Priming: A new theory of words and language*. London: Routledge.
- Johns, T.F. 1991. Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (eds), *Classroom Concordancing. English Language Research Journal* 4: 1-16.
- Johns, T.F. 1994. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (ed.), *Perspectives on Pedagogical Grammar*. Cambridge: Cambridge University Press, p. 27-45.
- Markel, M. 2004. *Technical Communication*. 7th edition. Boston: Bedford/St. Martins.
- Mindt, D. 1981. Angewandte Linguistik und Grammatik für den Englischunterricht. In P. Kunsmann & O. Kuhn (eds), *Weltsprache Englisch in Forschung und Lehre: Festschrift für Kurt Wächtler*. Berlin: Schmidt, p. 175-186.
- Nesi, H. 2011. BAWE: An introduction to a new resource. In A. Frankenberg-Garcia, L. Flowerdew & G. Aston (eds), *New Trends in Corpora and Language Learning*. London: Continuum, p. 213-228.
- Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.

⁷ See also Ädel (2010) for a discussion of challenges for DDL in teaching academic writing.

⁸ The MICASE project website with links to a corpus search interface, MICASE sound files and MICASE-based teaching materials is at <http://micase.elicorpora.info/>.

- O'Donnell, M. B. & U. Römer. 2012. From student hard drive to web corpus (part 2): The annotation and online distribution of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora* 7(1): 1-18.
- O'Donnell, M.B. & U. Römer. In preparation. Investigating the interaction between phraseological items and textual position.
- Römer, U. 2005. *Progressives, Patterns, Pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam: John Benjamins.
- Römer, U. 2008. Corpora and language teaching. In A. Lüdeling & M. Kytö (eds), *Corpus Linguistics: An international handbook (volume 1)*. Berlin: Mouton de Gruyter, p. 112-130.
- Römer, U. 2011. Corpus research applications in second language teaching. *Annual Review of Applied Linguistics* 31: 205-225.
- Römer, U. & M.B. O'Donnell. 2011. From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora* 6(2): 159-177.
- Römer, U. & M.B. O'Donnell. In preparation. *MICUSP: A corpus resource for exploring proficient student writing across disciplines*. Amsterdam: John Benjamins.
- Römer, U. & S. Wulff. 2010. Applying corpus methods to written academic texts: Explorations of MICUSP. *Journal of Writing Research* 2(2): 99-127. <http://www.jowr.org>, accessed 07/11/11.
- Simpson-Vlach, R. & N.C. Ellis. 2010. An academic formulas list: New methods in phraseology research. *Applied Linguistics* 31(4): 487-512.
- Sinclair, J.M. 2004. New evidence, new priorities, new attitudes. In J.M. Sinclair (ed.), *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, p. 271-299.
- Swales, J.M. 2005. Attended and unattended 'this' in academic writing: A long and unfinished story. *ESP Malaysia* 11: 1-15.
- Tognini Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Wulff, S., U. Römer & J.M. Swales. Forthcoming. Attended/unattended 'this' in academic student writing: Quantitative and qualitative perspectives. *Corpus Linguistics and Linguistic Theory*.

A direct application of medical corpora to academic writing: A specialized concordance search interface and Moodle-based courseware

English for medical purposes, writing, reading, concordancing, IMRAD

Shozo Yokoyama
University of
Miyazaki

Despite a relatively large number of research papers discussing corpus-based discourse analysis for medicine, few attempts have been made to uncover the rhetorical distinctiveness of subdisciplines such as Genome Bio-Science, Nursing, Public Health, and Clinical Surgery. The aim of this study is to apply the frequencies of verbs (including modal verbs) found in medical research articles (RAs) among these four subdisciplines to materials development by incorporating a specialized concordance search interface and Moodle-based courseware for medical students.

Chizuko Suzuki
Nagasaki Junshin
Catholic University

Seisuke Yasunami
Kumamoto
University

Naoko Kawakita
Miyazaki Prefectural
Nursing University

Collaborating
researcher:
Ryo Ohba,
Waseda University