Revealing disciplinary variation in student writing: a multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP)

Jack A. Hardy[1] and Ute Römer [1]

Abstract

The purpose of this study was to uncover sets of co-occurring, lexico-grammatical features to help to characterise successful student writing. The writing was captured by the Michigan Corpus of Upper-level Student Papers (MICUSP, 2009) and was taken from sixteen disciplines. MICUSP is a corpus of A-graded, upper-level student papers of different disciplines and paper types (O'Donnell and Römer, 2012; and Römer and O'Donnell, 2011). Following Biber (1988), we used a multi-dimensional analysis to identify dimensions of frequently co-occurring features that best account for cross-disciplinary variation in MICUSP. The four functional dimensions of MICUSP appear to distinguish between: (*1*) Involved, Academic Narrative *versus* Descriptive, Informational Discourse; (*2*) Expression of Opinions and Mental Processes; (*3*) Situation-Dependent, Non-Procedural Evaluation *versus* Procedural Discourse; and (*4*) Production of Possibility Statement and Argumentation. Along with a description of the methodology, this paper defines the features that constitute the factors, which have been labelled based on their communicative functions. Similarities and differences at the disciplinary and genre-specific levels are discussed as are the implications for discipline-specific and register-based pedagogies.

1. Introduction

The study of disciplinary specificity has become increasingly relevant as more writing instruction for college-level students moves from only

[1] Department of Applied Linguistics and ESL, Georgia State University, PO Box 4099, Atlanta, GA 30302, USA.
   Correspondence to: Jack A. Hardy,   *e-mail*: jhardy12@gsu.edu

literature-based composition courses in English departments to include content-based courses taught by scholars in different disciplines (Thaiss and Porter, 2010). Undergraduates are now expected to write 'in the discipline' or 'across the curriculum' (see Bazerman *et al*., 2005; and Russell, 1990, for a history of Writing Across the Curriculum). An effect of these changes is that students need to write in a way that conforms to the practices of a discipline they may not (yet) be familiar with. Another line of research has investigated many disciplines at once (Hyland, 2008) and involves the analysis of linguistic features as they are distributed across various registers of writing and/or categories of writers (e.g., Hyland, 2004, 2008). For example, areas of interest have included investigating professionals *versus* graduate *versus* undergraduate; and native *versus* non-native speakers of English. The overarching purpose of this line of research is not only descriptive since research in the sub-fields of English for Specific Purposes (ESP) and English for Academic Purposes (EAP) have a long pedagogical tradition of helping to bridge the gaps between the expectations of the powerful, central figures of a discourse community and those on the periphery who hope to work their way to the centre. Swales and Feak's (2004) textbook guide for writers in graduate school, for example, offers insights from professional academic writing to help readers to become more aware of, and competent in, the writing practices of their disciplines.

When developing a study to help undergraduate writers, however, the writing of professionals (e.g., published research articles) might not make for the most appropriate models. Instead, one should look for models that students will be more likely to be expected to produce. For example, a psychology professor may assign his or her students to write responses to a research article. Such responses would be graded as such, and the students would not be expected to write in the same register as their input. One solution to the problems of discipline and level discrepancies of models is to study the writing of students (Johns, 2009). Analyses of a corpus that has been developed to take into consideration what successful university writing models look like (e.g., in terms of their linguistic composition, topics/subject matter and format and style, among other features) may directly address writing instructors' needs for teaching materials that provide linguistic frequency or distributional data that could be shared with students. To this end, we explored the linguistic properties of a large corpus of A-graded student writing from sixteen disciplines from the Michigan Corpus of Upper-level Student Papers (MICUSP). MICUSP contains advanced student papers that are categorised primarily according to discipline and text type (O'Donnell and Römer, 2012; and Römer and O'Donnell, 2011).

We use this corpus as a way to study empirically authentic student writing across disciplines. It is not our intention to prescribe how students *should* write in these disciplines, because we recognise that there is a thin line between description and prescription in this context. One of the benefits of the methodology employed in this study, however, is that it deals with continua and related dichotomies. In a sense, we intend to show variation in

successful student writing in different disciplines without making claims for absolutes.

1.1 Multi-dimensional analysis of academic writing

Many corpus-based studies of academic writing have looked at predictive or correlational data showing the relationship between textual features and quality of writing scores given by instructors or raters. A substantial number of studies have identified linguistic features (e.g., subordination, prepositions, linking adverbials) that are predictive of quality scores given by instructors/raters and those that distinguish differences between disciplines and students' demographics (e.g., proficiency levels, graduate *versus* undergraduate). The identification of linguistic features that are, statistically, found to be indicators of writing quality has captivated researchers because of its potential pedagogical import (Grant and Ginther, 2000). However, correlational and predictive indicators on the relationship between linguistic distributions and the quality of writing have been somewhat inconclusive (Jarvis *et al*., 2003). It is clear that contextual and disciplinary differences, including the roles of topics and prompts, contribute to variations in academic texts (Beers and Nagy, 2009).

        In the vein of the other articles contained in this special issue, this study utilises multi-dimensional (MD) analysis – a procedure pioneered by Biber (1988). This application is based on the assumption that 'differences in registers include patterns of co-occurring lexico-grammatical features' (Halliday, 1988: 162). These systematic differences occur because individuals make lexical and grammatical choices that are appropriate to the register in which they are speaking and writing (Biber and Conrad, 2001: and Hymes, 1984). It is important to note that these co-occurrences are not one-dimensional. A description of a text as 'informal' or 'formal', for example, would not capture the full multi-dimensional variation that can be used to understand differences between registers and their texts. Once the factors, or dimensions, are extracted, one can then analyse qualitatively groups of texts or individual texts to understand more fully the variation that is present in the corpus. This approach to register variation was originally developed to define broader distinctions across general registers of spoken and written English. We apply this procedure to uncover disciplinary variation among the more specialised register of successful writing of upper-level college students as captured in MICUSP.

1.2 The focus of this paper

Since the publication of Biber's (1988) influential work on variation across speech and writing, similar analytical procedures have been used to uncover

variation in more specific contexts. Studies on professional and academic written discourse utilising the MD approach continue to increase in number (e.g., Biber, 2006; Gray, 2011; and Reppen, 2001) but have, up to this point, not been extended to unpublished advanced student writing. In order to understand disciplinary variation in successful, upper-level student writing, this study identifies the co-occurring lexico-grammatical features of MICUSP. Although our primary focus is on disciplinary differences, we also aimed to investigate variation that could be accounted for in terms of text type (e.g., response papers, reports and critiques) and student level (graduate *versus* undergraduate). In the following sections, we discuss our methodology and results, briefly describing MICUSP, the MD approach, and our primary multivariate data with results of tests of group/mean differences. We focus on the interpretation of four resulting dimensions. For each dimension, we then offer a description and discussion of the variation as seen through a comparative dimensional scale and text samples.

2. The Michigan Corpus of Upper-Level Student Papers

MICUSP is a corpus of samples of proficient student academic writing that was compiled by a team of researchers and students at the English Language Institute of the University of Michigan, Ann Arbor (O'Donnell and Römer, 2012; and Römer and O'Donnell, 2011). The corpus enables researchers, language teachers and students to investigate the written discourse of proficient, advanced-level native and non-native speaker student writers at a large American research university. It also provides students with a wide selection of A-graded papers that may serve as models for their own academic writing. MICUSP is currently available through a user-friendly online search interface, known as MICUSP Simple, which can be accessed free of charge.[2] Offline versions of the corpus (in annotated XML and plain text format) are scheduled to be published on a CD-ROM, accompanying a MICUSP resource book (Römer and O'Donnell, forthcoming). Figure 1 shows a screenshot of the MICUSP Simple search and browse page.

The papers included in MICUSP were written by students at four different levels of study: final year undergraduate, and first-, second- and third-year graduate level. Papers of different types, ranging from essays to lab reports, have been collected from a wide range of disciplines within four academic divisions (namely, Humanities and Arts, Social Sciences, Biological and Health Sciences, and Physical Sciences). The corpus hence enables analyses of disciplinary, developmental and genre-related phenomena associated with student writing. Each paper in MICUSP also captures metadata on the student's gender and nativeness status
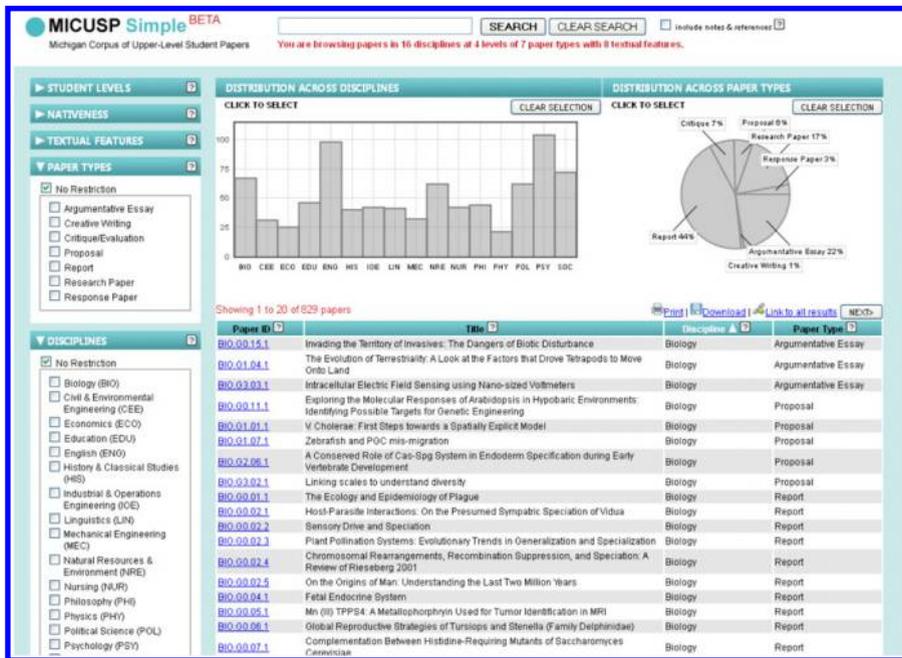
---

[2] See: http://search-micusp.elicorpora.info

Figure 1: Screenshot of MICUSP Simple[3]

(including information on first language background in the case of non-native speakers).

MICUSP consists of 829 student papers from sixteen disciplines, making up over 2.6 million words. Table 1 provides an overview of the distribution of MICUSP papers across disciplines and academic divisions.[4]

As shown in Table 1, numbers of papers from the individual disciplines range from 21 to 104. Some of the strongest disciplines in terms of papers and word counts are Psychology, English, Sociology and Biology. The overall word count is highest for the Social Sciences division (978,254), followed by the Humanities and Arts (734,437) and Biological and Health Sciences (511,550), and is lowest for Physical Sciences (392,288).

MICUSP papers have been categorised according to paper types (e.g., argumentative essay and report), following a systematic data-driven text-type analysis of all the papers in the corpus (see Römer and O'Donnell, 2011: 167–74). The papers fall into the following seven text type categories: argumentative essay, creative writing, critique/evaluation, proposal, report, research paper and response paper.

---

[3] See: http://search-micusp.elicorpora.info/
[4] For further details on the composition of MICUSP and divisions of papers by student level, gender, and nativeness status, the reader is referred to Römer and O'Donnell (2011) and Römer and O'Donnell (forthcoming).

Academic division	Discipline	Papers	Tokens
Humanities and Arts	English	98	268,733
	History and Classical Studies	40	182,629
	Linguistics	41	155,047
	Philosophy	44	128,028
	Σ 223		Σ 734,437
Social Sciences	Economics	25	78,070
	Education	46	150,282
	Political Science	62	210,783
	Psychology	104	323,326
	Sociology	72	215,793
	Σ 309		Σ 978,254
Biological and Health Sciences	Biology	67	176,124
	Natural Resources and Environment	62	176,653
	Nursing	42	158,773
	Σ 171		Σ 511,550
Physical Sciences	Civil and Environmental Engineering	31	98,918
	Industrial and Operations Engineering	42	124,973
	Mechanical Engineering	32	123,335
	Physics	21	45,062
	Σ 126		Σ 392,288
All divisions / disciplines		Σ 829	Σ 2,616,529

Table 1: Distribution of papers across academic divisions and disciplines (Römer and O'Donnell, 2011: 164)

3. Method: factor analysis

In order to identify the primary disciplinary features of texts from MICUSP, the normalised values of numerous linguistic features were computed for each text in the corpus. Texts were tagged for part of speech (POS) and additional semantic categories using the Biber tagger (Biber, 1988, 2006). The procedure for Exploratory Factor Analysis with Promax Rotation was conducted using PASW v. 18.[5] The resulting factors were then interpreted as underlying dimensions of functional linguistic variation. For this study, rates of occurrence were computed for 110 tagged features across 825 (out of 829) texts from MICUSP (four texts were dropped due to tagging errors). Only fifty-four linguistic features were retained in the final test after multiple runs

---

[5] More well known as SPSS: from 2009 to 2010 this statistics computer program was known as PASW. The current version is known as IBM SPSS.

combining three- to six-factor solutions and a range of cut-off cumulative loading scores and percentage of overlapping variances. The solution for four factors was selected as optimal in this study. These factors account for 35 percent (or 'very good') overlapping variance (Comrey and Lee, 1992). The next set of factors did not provide significant variance. We used a cut-off loading of + or − 0.30 for a feature to be included in a dimension.

4. Results and discussion

Table 2 lists the co-occurring features of the four dimensions of MICUSP.

After the extraction of factors, factor scores were computed for all texts, disciplines and demographic categories in the MICUSP corpus. An individual factor score (per text) in a dimension shows, primarily, where the text plots in a dimension scale when compared to other texts. The sections below present our functional analyses and our interpretation of the four dimensions of student academic papers.

Since we investigated many different disciplines and other variables were involved (e.g., text type and student level), we decided to analyse and present our data in a way that is new to MD analysis studies. In addition to showing the dimension scores for each discipline in MICUSP, we also analysed and compared the fifty texts with the highest and lowest factor scores across these categories − a total of 100 texts per dimension. This allowed us to give a more precise and comprehensive picture of the poles of each dimension than if we had examined only one or two texts. In addition, One-Way Analysis of Variance (ANOVA) tests of significance are provided under the comparative figures below.

4.1 Dimension 1: involved, academic narrative *versus* descriptive, informational discourse

We have labelled the two poles of Dimension 1 as (+) Involved, Academic Narrative and (−) Descriptive, Informational Discourse. By examining 100 texts with the most extreme +/− scores in this dimension, we are better able to understand the function of the dominant features of co-occurrence of Dimension 1. Many verb categories and complement clauses co-occur with personal pronouns and past tense verbs on the positive side; while nouns, nominalisations and adjectives appear on the negative side. This dimension mirrors many of the features in Biber's (1988) Dimension 1: 'Involved vs. Informational Production' (with the exception of narrative features such as past tense verbs).

It is of note that the Involved, Academic Narrative (+) end of this dimension appears to be level specific. Of the fifty highest loading texts, undergraduate students produced forty-six. In terms of academic divisions,

Dimension	Features
1	**Positive: involved, academic narrative production** all verbs (not including auxiliary verbs), mental verbs, first-person pronouns/possessives, verbs of activity, private verbs, infinitive verbs, sum stance to comp. clauses controlled by verbs, *that*-deletion, *to*-complement clauses controlled by verbs of desire, intention and decision, nominal pronouns, wh-clauses, all wh-words, third-person pronouns, past tense verbs, factive adverbs **Negative: descriptive, informational production** singular nouns, nominalisations, nouns, all adjectives, attributive adjectives, word length
2	**Positive: expression of opinions, attitudes, emotions and mental processes** sum stance that comp. clauses, sum stance that comp. clauses controlled by verbs, *that*-complement clauses controlled by a factive verb, *that*-complement clauses controlled by a non-factive verb, *that*-complement clauses controlled by a verb of likelihood, private verbs, mental verbs, public verbs (e.g., *assert*, *complain*, *say*), predicative adjectives, *that*-deletion, sum stance to comp. clauses, sum stance that comp. clauses controlled by adjectives, sum stance to comp. clauses controlled by adjectives, all verbs (not including auxiliary verbs), verbs of communication, *that*-complement clauses controlled by an adjective of likelihood **Negative:** (no negative features)
3	**Positive: situation-dependent, non-procedural evaluation** sum stance adverbs, factive adverbs, adverbs (not including emphatics, hedges, amplifiers, time, place), third-person pronouns, verbs of communication, all conjunctions, pronoun *it*, nominal pronouns, verbs (uninflected present, imperative and third person), public verbs (e.g., *assert*, *complain*, *say*), *that*-deletion, *that*-complement clauses controlled by a verb of likelihood, all wh-words, wh-pronoun, relative clause - object position with prepositional fronting ('pied piping') **Negative: procedural** nouns, agentless passive verbs, all passives
4	**Positive: production of possibility** verb *be* (uninflected, present tense, verb and auxiliary), verbs (uninflected present, imperative and third person), modals of possibility, modals of prediction, predicative adjectives, all verbs (not including auxiliary verbs), sum stance to comp. clauses controlled by adjectives, sum stance that comp. clauses controlled by adjectives, infinitive verbs **Negative:** past tense verbs

Table 2: Composition of the positive and negative features of Dimensions 1, 2, 3 and 4

Physical Sciences is, notably, absent. For the Biological and Health Sciences, only three texts are in this set of fifty. On the other hand, thirty-one and sixteen texts come from the Humanities and the Social Sciences, respectively. The text types that are most represented in this set are reports (eighteen), response papers (ten), argumentative essays (eight) and creative writing (six). Unlike the disparity of student level in the positive pole for Dimension 1, the texts that loaded highly on the negative pole, Descriptive, Informational Discourse (–), were more balanced by undergraduate (twenty-two) and graduate (twenty-eight) students. However, in terms of academic divisions, this group of texts is relatively homogeneous in that it mostly excludes the humanities (four). Instead, Biological and Health Sciences (twenty-one), Physical Sciences (fourteen) and Social Sciences (seven) dominate as being descriptive and informational. In terms of text types, reports (twenty-seven) and research papers (sixteen) are represented the most.

Returning to the entire sample of 825 texts, we can now see the full picture of disciplinary variation in the corpus according to Dimension 1. Figure 2 shows the combined dimension scores of the sixteen disciplines. Results were analysed using one-way ANOVA, which showed a significant effect for discipline ($F(15, 809) = 18.65, p < 0.001$).

Amongst these results, it is worth highlighting the two disciplines on the extreme positive side of Dimension 1: Philosophy and Education. Using a post-hoc test (Tukey's HSD), we determined that these two differed significantly from all other disciplines save each other. Student writing in Philosophy and Education (as captured in MICUSP) appears to be significantly more involved and narrative in style than the writing in other disciplines, even compared to other positively loading disciplines. Text samples from papers in Philosophy (Example A) and Education (Example B) show the ways in which student writing in these disciplines can be highly involved and narrative. Philosophy writing, especially, appears to involve a lot of stance verbs (e.g., *believe*) in its evaluations of ideas or ways of thinking. Similarly, it appears that writing in Education also incorporates involvement using stance and mental verbs (e.g., *I liked...*, *I think...*). However, what places this discipline highly in this category is the use of narrative language. Many of the Education assignments in MICUSP involve a report of how a lesson went, including the writer's understanding of that lesson (Example B). Even for assignments in which students wrote about prospective lesson plans (i.e., for events in the future), there was often a sense of sequence or narration to them.

(A)   Socrates then concludes that group (D) does not exist, since those people, by desiring what they believe to be harmful (bad) things are desiring to be miserable and unhappy. No one wants to be miserable and unhappy, so no one desires what he believes to be bad. (A)–(C) actually desire what they believe to be good, and group (D) does not exist, so no one desires what he believes to be bad. I feel compelled to say here that although Socrates actually
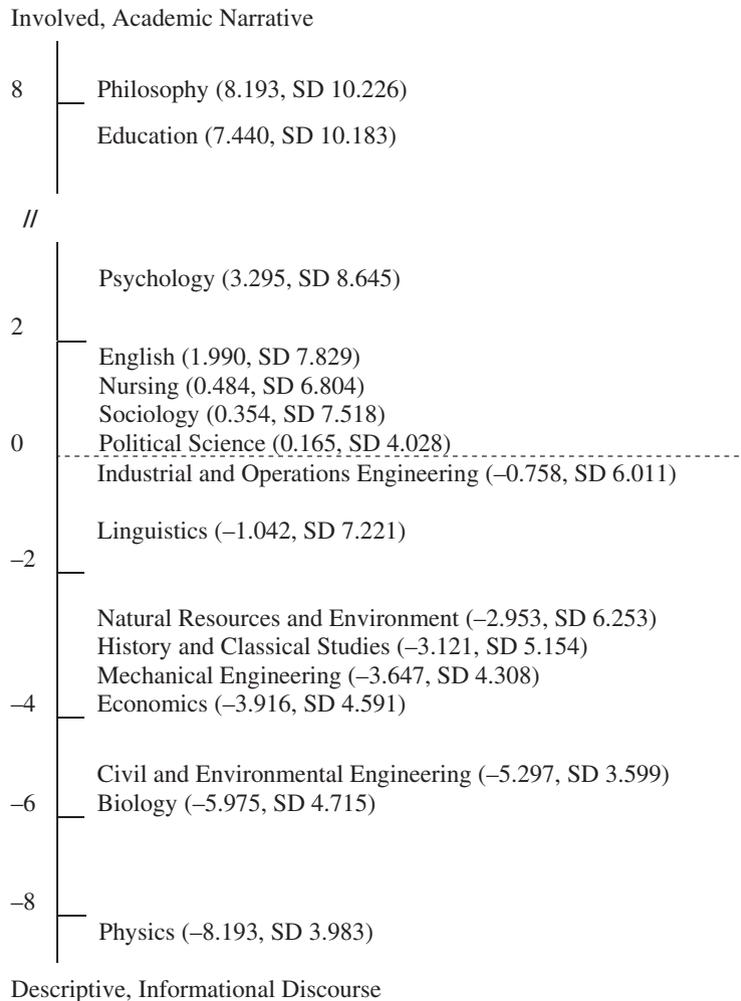
Involved, Academic Narrative

```
8  __   Philosophy (8.193, SD 10.226)

          Education (7.440, SD 10.183)


   //

          Psychology (3.295, SD 8.645)

2  __
          English (1.990, SD 7.829)
          Nursing (0.484, SD 6.804)
          Sociology (0.354, SD 7.518)
0  ......Political Science (0.165, SD 4.028)..........................
          Industrial and Operations Engineering (–0.758, SD 6.011)

          Linguistics (–1.042, SD 7.221)
–2 __

          Natural Resources and Environment (–2.953, SD 6.253)
          History and Classical Studies (–3.121, SD 5.154)
          Mechanical Engineering (–3.647, SD 4.308)
–4 __   Economics (–3.916, SD 4.591)

          Civil and Environmental Engineering (–5.297, SD 3.599)
–6 __   Biology (–5.975, SD 4.715)


–8 __
          Physics (–8.193, SD 3.983)
```

Descriptive, Informational Discourse

Figure 2: Comparison of dimension scores for disciplines in Dimension 1: (+) Involved, Academic Narrative *versus* (–) Descriptive, Informational Discourse (ANOVA: group, $F(15, 809) = 18.65$, $p < .001$)

claims that "no one wants what is bad" (78), what he means is that no one wants what he believes is bad.

    PHI.G0.06.1, critique/evaluation, final year undergrad (UG), native speaker (NS)

    Dimension 1 score $= 53.534$

(B)   By asking a broad question all students could answer, I was able to elicit initial thoughts from many of my students and attend to the learning of all students. Seven students were able

to contribute their ideas about this design and many others had their hands raised. I also used "thumbs up or thumbs down" and "nod your head if…" questions to elicit answers from the entire class at once (07:48).

> EDU.G0.06.1, report, final year UG, NS
> Dimension 1 score $= 31.805$

At the other end of the continuum, Descriptive, Informational Discourse, we see more texts from Physical, Biological and Health Sciences. Disciplines such as Physics, Biology, and Civil and Environmental Engineering loaded highly negatively in this dimension. The student writing in these areas involves information being densely (e.g., high levels of nouns and nominalisation) and descriptively (e.g., many adjectives) expressed (Biber, *et al.*, 1999). Interestingly, in contrast to other similar analyses of published academic writing (e.g., Gray, 2011), the sample of student writing includes two adjectival features (all adjectives and attributive adjectives) in this informational pole of the first dimension. Example C is a text sample that loaded on the negative side of Dimension 1. This Biology report shows how, especially in the natural sciences, student writing can be both informational and descriptive. We can also see in this sample the lack of positive (involved, narrative) features (e.g., first-person pronouns, mental verbs and past tense verbs).

(C) Normally malaria is a curable disease, but only if treated properly. After an infectious bite there is an incubation period in the host that varies depending on the species of Plasmodium, before there is an onset of symptoms. The symptoms of malaria that a human host will go through can be categorized as either uncomplicated or severe. With uncomplicated malaria, the symptoms last between 6–10 hours and include a cold stage, a hot stage and then finally a sweating stage. Symptoms occur in a mixture of fever, chills, sweats, headaches, nausea, vomiting, body aches, and general malaise.

> BIO.G0.25.1, report, final year UG, NS
> Dimension 1 score $= -15.047$

4.2 Dimension 2: expression of opinions and mental processes

In contrast with Dimension 1, Dimension 2 has only positive features. There are not two opposing and mutually exclusive sets of features: instead, a negative factor score in this dimension means that a text has significantly fewer positive features. We have labelled Dimension 2 as (+) Expression of Opinions and Mental Processes. This dimension is characterised by the co-occurrence of stance features, complementation with verbs of likelihood, and

various verb categories (private/mental and communication). The expression of opinions in these texts also includes references to the writer's attitudes and emotions, and evaluation of academic topics. Mental and public verbs are often used with hedging features across many of the texts that loaded highly in Dimension 2.

Following the same process of examining the fifty highest loading and fifty lowest loading texts for Dimension 1, we found that the highest loading texts included many papers by Philosophy students (twenty). Of these highest loading texts, some disciplines were either mostly or completely absent. For example, other than two reports and one response paper written for Biology, no other texts from the Biological and Health Sciences and the Physical Sciences loaded highly. Also, in common with the results in Dimension 1, the most positively loading Dimension 2 papers came from undergraduate students (thirty-nine). The lowest loading texts were dominated by Physical Sciences and Biological and Health Sciences (thirty-eight). Fewer texts were from Social Sciences (eight) and Humanities (four). In terms of text types, the most common for this set of low loading texts were reports (thirty) and research papers (nine). Also, of the fifty texts with low Dimension 2 scores, most come from graduate student writing (forty). Figure 3 shows the combined dimension scores of the sixteen disciplines of MICUSP on a dimension scale.

Figure 4 shows how student writing in Philosophy is very different from the other disciplines in Dimension 2. Incorporating opinions, attitudes, emotions and mental processes may be a good description of student writing in Philosophy. Examples D and E are from Philosophy and show critiques/evaluations that loaded highly in this dimension. Notice the use of *that*-complement clauses showing stance (e.g., *it seems that. . . , it is easy to see that. . . , it might be that . . .* , and *it seems plausible that. . .*).

(D)   It seems that the list above includes all combinations of beliefs and desires, and that additional, more specific combinations of circumstances will fall under one of the categories. The task, then, is to prove that these people desire only good things, or that those who do not cannot exist.

It is easy to see that (A) and (B) are not problematic cases. People who clearly desire what they believe to be good things really do desire good things, regardless of whether they happen to be mistaken about whether or not a thing is good, as in case (B). People in (C) and (D) require a bit more explanation. A person in category (C) is using the word "bad" incorrectly to describe the object of his desires. If a person believes a thing to benefit him, then that thing cannot be believed (by him) to be bad. To think a thing is beneficial is to think it is good. A person cannot correctly believe a thing to be both bad and beneficial.

PHI.G0.06.1, critique/evaluation, final year UG, NS
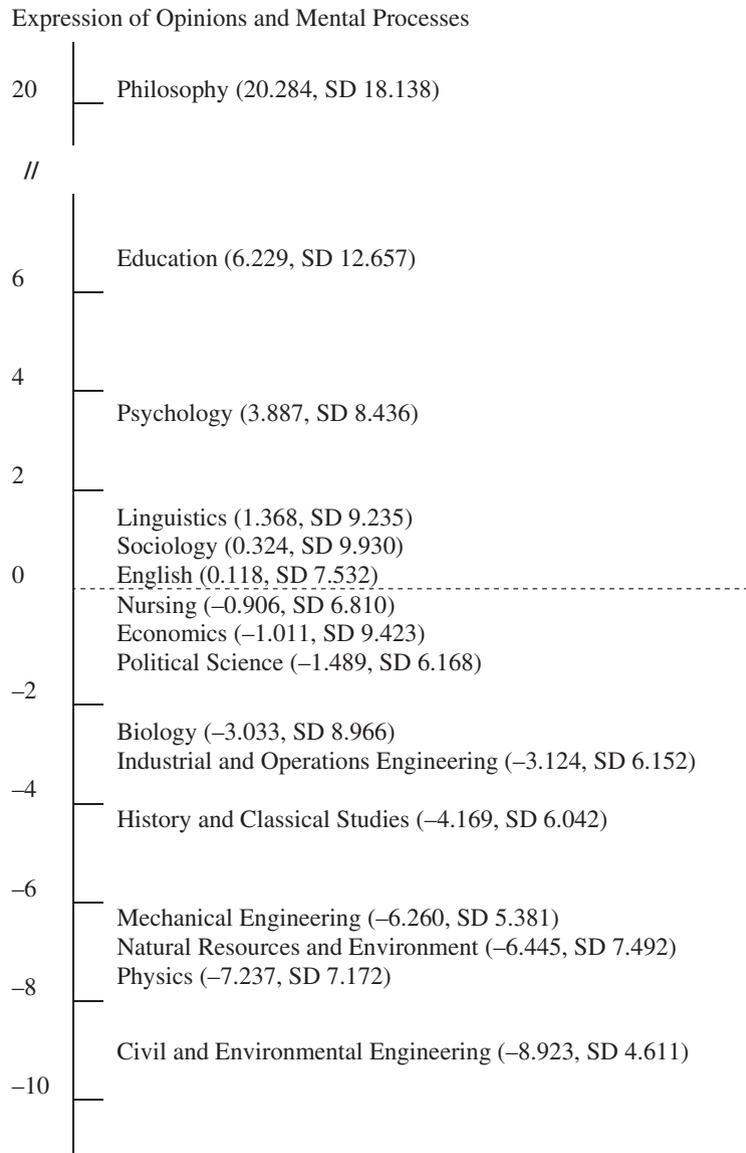Dimension 2 score $= 79.387$

Expression of Opinions and Mental Processes

```
20  ──  Philosophy (20.284, SD 18.138)


    //


 6  ──    Education (6.229, SD 12.657)


 4  ──
           Psychology (3.887, SD 8.436)


 2  ──
           Linguistics (1.368, SD 9.235)
           Sociology (0.324, SD 9.930)
 0  ──     English (0.118, SD 7.532)
 - - - - - - - - - - - - - - - - - - - - - - -
           Nursing (–0.906, SD 6.810)
           Economics (–1.011, SD 9.423)
           Political Science (–1.489, SD 6.168)
–2  ──
           Biology (–3.033, SD 8.966)
           Industrial and Operations Engineering (–3.124, SD 6.152)
–4  ──
           History and Classical Studies (–4.169, SD 6.042)


–6  ──
           Mechanical Engineering (–6.260, SD 5.381)
           Natural Resources and Environment (–6.445, SD 7.492)
–8  ──     Physics (–7.237, SD 7.172)


           Civil and Environmental Engineering (–8.923, SD 4.611)
–10 ──
```

Figure 3: Comparison of dimension scores for disciplines in Dimension 2: (+) Expression of Opinions and Mental Processes (ANOVA: group, $F(15, 809) = 25.56$, $p < 0.001$)

(E)   Now, it might be that the reasons I've suggested for Gilbert's inability to either make or fulfill predictions are problematic (they prohibit us from saying all that we could or would want to say of Gilbert). However, it seems plausible that his

inabilities could be explained in many ways, not all of which are
problematic. Or, perhaps it enough for us to be able to imagine
Gilbert's inabilities, without requiring an explanation of them.

>PHI.G0.06.7, critique/evaluation, final year UG, NS
>Dimension 2 score = 61.433

On the lower end of the spectrum, the disciplines that lack these features
(opinion and emotion) are mostly in the Physical Sciences. This is shown in
the text samples below, which are written by graduate students in Natural
Resources and Environment (Example F) and Mechanical Engineering
(Example G). In this style of writing, there are few instances of the writer
incorporating his or her own opinions, attitudes or mental processes. The
Mechanical Engineering text is entitled a 'literature survey report' in the
discipline, but, by describing a physical process, it reads as if it is similar
to a methods section in a research paper.

(F)   Initiated by Japan International Cooperation Agency (JICA)
       in 1990, the canal water improvement project has been
       implemented by the Bangkok Metropolitan Administration
       (BMA) to cope with the problem. Its underlying principle is to
       regulate the circulation of clean water to flush out and to dilute
       polluted water.

>NRE.G2.08.1, report, second year graduate student (G),
>Non-native speaker (NNS) – Thai
>Dimension 2 score = −20.629

(G)   The fabrication technique for stretchable electronic devices
       in the Rogers' lab entails a dry transfer printing technique
       wherein free standing micro and nano-sized patterned circuit
       components are picked up from their substrate using a
       conformal elastomer stamp. These electronic circuit components
       (transistors, photodiodes etc.) are subsequently, transferred to a
       specific locations on thin, flexible substrate.

>MEC.G3.03.1, report, third year G, NNS – Hindi
>Dimension 2 score = −17.890

4.3 Dimension 3: situation-dependent, non-procedural evaluation *versus* procedural discourse

We labelled the two poles in the third dimension as Situation-
dependent, Non-procedural Evaluation (+) and Procedural Discourse (−).
Procedural academic texts commonly feature nouns and passives (including
agentless passives and *by–* passives), whilst situation-dependent writing is
characterised by adverbials (Biber, 1988; and Biber *et al*., 1999), personal
pronouns and communication verbs. The set that loaded highly positively

included mostly Philosophy (twenty) and English (nineteen) papers. As with Dimension 1, these texts were, also, mostly written by final year undergraduates (forty-two), and of the eight written by graduate students, six came from philosophy. Of this set of fifty texts, the text type also proves to be informative. The two largest text types represented are argumentative essays (twenty-four) and reports (fifteen).

The lowest loading fifty texts are comprised mostly of writing outside the Humanities and Arts: Biological and Health (eighteen), Social (sixteen) and Physical Sciences (fifteen). There does not appear to be a pattern for student level within this set (twenty UG and thirty G). We do, however, observe a pattern for text type. As expected, this set mostly includes papers associated with empirical research, in which the explicit explanation of procedure or methodology is viewed as important: reports (twenty-seven), research papers (fourteen) and proposals (seven). Figure 4 shows the combined factor scores of the sixteen disciplines in Dimension 3.

The average dimension score comparison across disciplines again shows a distinction between the humanities and the sciences. Most humanities texts have positive scores while the sciences, often focussing on the discussion of methodologies and procedures, have negative average scores. Below, we provide samples from a report from Philosophy (Example H) and an argumentative essay from English (Example I). Notice that both are responses, in a way, to a particular course pack and novel, respectively. It is, thus, understandable that such papers would be situation-dependent: the writers can assume that their readers (the instructors) are familiar with the limited amount of source material, and their language adjusts accordingly. It should also be noted that this end of the dimension is not procedural. This style of writing, common in Philosophy and English, is not a recount of what the students read. Instead, the papers are used to evaluate (or argue) positions.

(H)   Due to these contradictions that were noted in the coursepack articles it appears that God can't be omnipotent according to the original definition that Descartes supported. If God is not omnipotent according to the original definition, then two possibilities could arise. Firstly, God may not exist since it appears he can't be omnipotent (and therefore wouldn't have one of the three necessary qualities), or omnipotence should be defined differently to better explain what is possible. Aquinas seemed to choose the second option and, so, we have a new definition of omnipotence to examine.

PHI.G0.17.2, report, final year UG, NS
Dimension 3 score = 33.411

(I)   Olivia says to her father that Sir William "took every opportunity of privately admonishing me against the artifices of Mr. Thornhill, who I now find was even worse than he represented
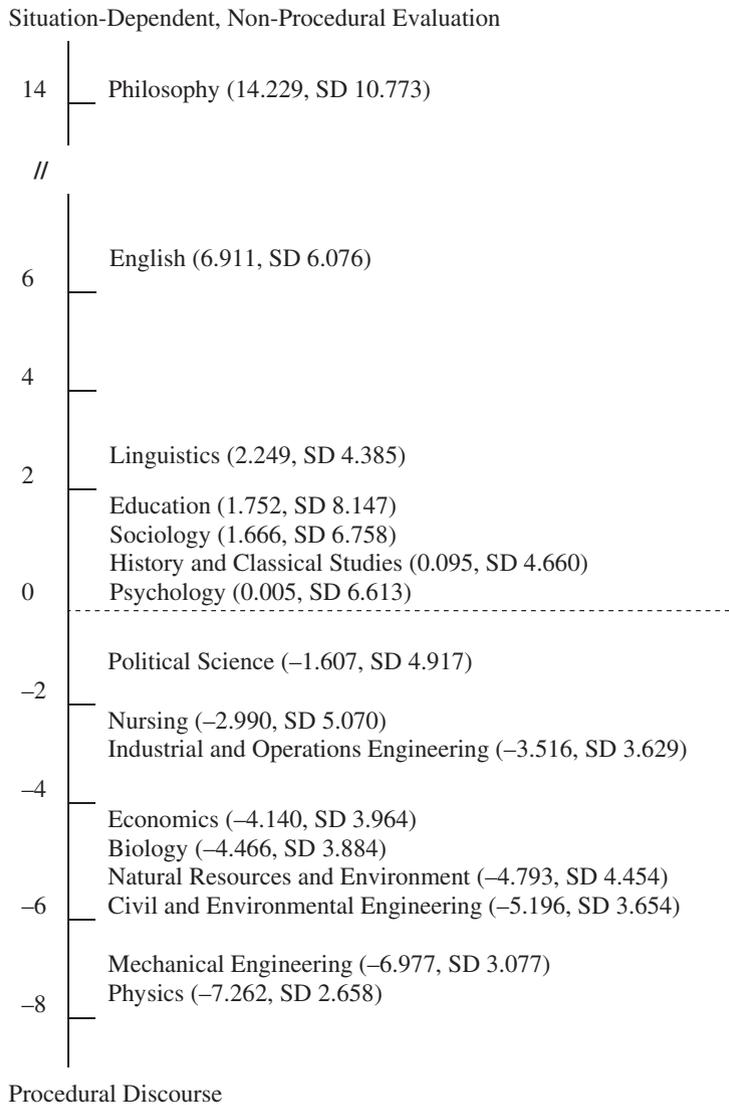
Situation-Dependent, Non-Procedural Evaluation

14 — Philosophy (14.229, SD 10.773)

//

English (6.911, SD 6.076)
6

4

Linguistics (2.249, SD 4.385)
2
Education (1.752, SD 8.147)
Sociology (1.666, SD 6.758)
History and Classical Studies (0.095, SD 4.660)
0   Psychology (0.005, SD 6.613)

Political Science (–1.607, SD 4.917)
–2
Nursing (–2.990, SD 5.070)
Industrial and Operations Engineering (–3.516, SD 3.629)

–4
Economics (–4.140, SD 3.964)
Biology (–4.466, SD 3.884)
Natural Resources and Environment (–4.793, SD 4.454)
–6   Civil and Environmental Engineering (–5.196, SD 3.654)

Mechanical Engineering (–6.977, SD 3.077)
Physics (–7.262, SD 2.658)
–8

Procedural Discourse

Figure 4: Comparison of dimension scores for disciplines in Dimension 3: (+) Situation-Dependent, Non-Procedural Evaluation *versus* (–) Procedural Discourse (ANOVA: group, $F(15, 809) = 43.02$, $p < 0.001$)

him" (109). Knowing that Mr. Thornhill had "artifices" at all makes it hard to accept that Sir William did not know the full length of his nephew's "artifices" (109). Even if he did not know the full lengths to which Mr. Thornhill would go, there is really no acceptable length for letting someone, let alone a friend, be

> fooled. In other words, apprehensions are apprehensions and no matter how small, if he had reason to believe something bad would happen then that should have been enough to persuade him to take a stronger course of action.
>
> ENG.G0.45.1, argumentative essay, final year UG, NS
> Dimension 3 score = 26.111

As described above, the texts that loaded highly negatively on Dimension 3 are procedural. Many either have extensive methodology sections or read as agentless accounts of series of events. For example, Example J is from a Mechanical Engineering report by a third-year graduate student. Notice the extensive use of nouns (underlined) and passive constructions. This text sample, although not referencing an experiment conducted by the author, appears to be something like a formal literature review. With information-dense language, the writer is able to exhibit his understanding of procedures that others have followed.

(J)   An <u>extension</u> of this <u>device</u> was obtained by incorporating organic <u>semiconductor</u> based thermal <u>sensors</u> to it in order to achieve greater sensory <u>properties</u> for the proposed artificial <u>skin</u>. Another <u>strategy</u> involves <u>distribution</u> of rigid <u>subcircuit</u> <u>islands</u> over the flexible <u>polymer</u> <u>surface</u>, <u>fabrication</u> of active <u>circuit</u> <u>elements</u> on the <u>islands</u> and connecting them with stretchable <u>metal</u> <u>interconnects</u> (<u>gold</u> (Au)). In addition to the <u>planar</u> <u>gold</u> <u>interconnect</u> <u>film</u>; another <u>variant</u> was tried in which <u>gold</u> <u>film</u> was deposited on a prestretched <u>polymer</u> <u>membrane</u> leading to a wavy <u>geometry</u> upon the <u>relaxation</u> of the <u>substrate</u>. This <u>device</u> was found to be functional up to a <u>maximum</u> of 12 percent <u>strain</u> associated with repeated <u>stretching</u> and <u>relaxation</u>. The <u>need</u> to image non-planar (curved) <u>surfaces</u> using flexible, <u>large-area</u> <u>sensorarrays</u> led to the <u>development</u> of a low <u>temperature</u> thin <u>filmtransistor</u> <u>fabrication</u> <u>process</u> involving <u>deposition</u> of amorphous <u>silicon</u> and amorphous <u>silicon</u> <u>nitride</u> at a lower <u>temperature</u> on <u>plastic</u> <u>substrate</u> (120° C as against the conventional 250–300° C <u>temperature</u> <u>requirement</u>).

  MEC.G3.03.1, report, third year G, NNS – Hindi
  Dimension 3 score = −13.390

Many papers that loaded highly negatively (highly procedural) were constructed more like research proposals or research papers. Example K, for instance, is a report by a Nursing graduate student. In her report, she reviews literature and proposes a study. Again, because the purpose of her paper is to give explicit and detailed information about how she would conduct her study, it is understandable that the paper would incorporate many elements that are traditionally found in research paper methodology sections, such as the use of the passive voice.

(K)   For the purpose of the proposed study, a subset of the original dataset will be utilized. The subset is comprised of the family caregivers of women with recurrent breast cancer. The study will focus on data obtained at the three-month period. This time point was chosen because the investigators [sic] clinical experience and the literature suggests that the patient and family have had an opportunity to adjust to the initial shock of recurrent cancer and are now dealing with the cancer treatment, side effects and overall uncertainty of the disease and their future (citation). This time point is reinforced with the results of a study of 69 women with recurrent breast cancer by Bull and colleagues (citation).

        NUR.G3.03.1, report, third year G, NS
        Dimension 3 score $=-11.272$

4.4 Dimension 4: production of possibility statement and argumentation

On the positive side, Dimension 4 is characterised by the co-occurrence of all modal verbs (including possibility and prediction modals), present tense verbs and stance complement clauses. Past tense verbs loaded negatively in this dimension. The contrast between present and past tense verbs in academic writing is captured in the context of argumentation and the production of statements of possibility as compared with procedural texts which often focus on the description of completed research methods or approaches. Philosophy and Education papers have high positive average scores as opposed to negative average factor scores for texts from Mechanical Engineering and Physics. Texts from History and Classical Studies which are primarily focussed on discussions of past events have the lowest average negative score in this dimension.

    For Dimension 4, the set of fifty highly positively scoring papers includes more undergraduate (thirty-three) than graduate writers (seventeen). In terms of disciplines, eleven out of the sixteen disciplines are represented in this set. However, as per Dimension 2, the two disciplines most represented are Philosophy (seventeen) and Education (nine). The representation according to text types is as follows: argumentative essays (fourteen), reports (fourteen), critiques/evaluations (nine) and response papers (eight). The lowest scoring fifty papers are similarly heterogeneous in terms of disciplines. The most common discipline in this set is History and Classical Studies (eleven). Papers written by graduate students (thirty-seven) outnumber those written by undergraduates (thirteen). Also, in terms of text types, reports (twenty-six), research papers (eleven) and argumentative essays (ten) represent most of this set. Figure 5 shows the disciplinary variation within Dimension 4.
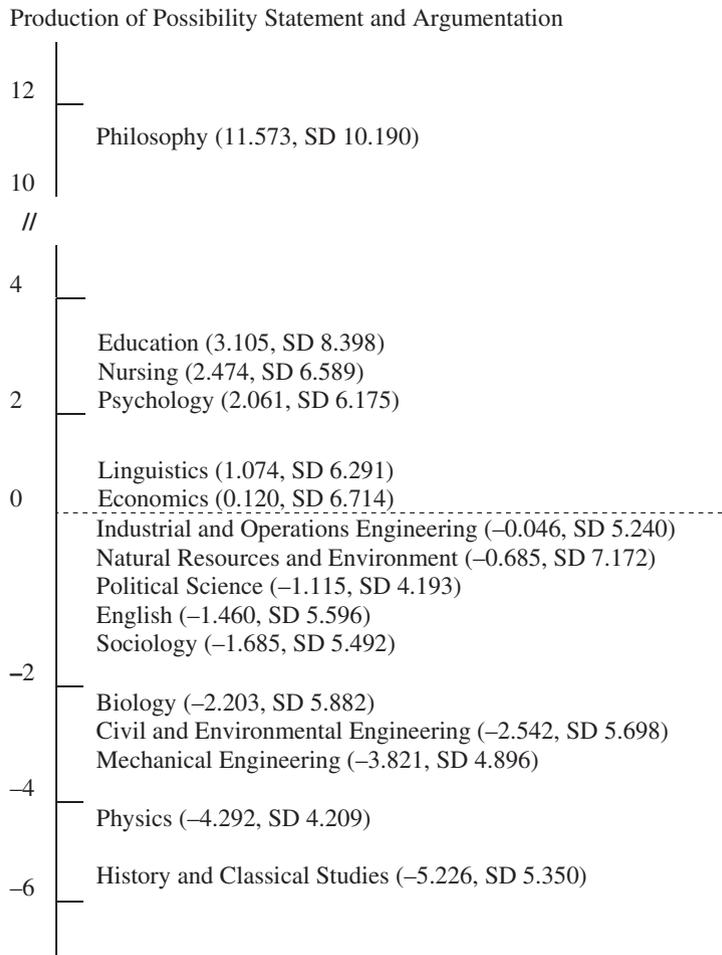
Production of Possibility Statement and Argumentation

```
12 ┤──
          Philosophy (11.573, SD 10.190)

10 ┤
 // ┤
  4 ┤──

          Education (3.105, SD 8.398)
          Nursing (2.474, SD 6.589)
  2 ┤──    Psychology (2.061, SD 6.175)

          Linguistics (1.074, SD 6.291)
  0 ┤──    Economics (0.120, SD 6.714)
 - - - - - - - - - - - - - - - - - - - - - - - - - -
          Industrial and Operations Engineering (–0.046, SD 5.240)
          Natural Resources and Environment (–0.685, SD 7.172)
          Political Science (–1.115, SD 4.193)
          English (–1.460, SD 5.596)
          Sociology (–1.685, SD 5.492)
 -2 ┤──
          Biology (–2.203, SD 5.882)
          Civil and Environmental Engineering (–2.542, SD 5.698)
          Mechanical Engineering (–3.821, SD 4.896)
 -4 ┤──
          Physics (–4.292, SD 4.209)

          History and Classical Studies (–5.226, SD 5.350)
 -6 ┤──
```

Figure 5: Comparison of dimension scores for disciplines in Dimension 4: (+) Production of Possibility (ANOVA: group, $F(15, 809) = 16.97$, $p < 0.001$)

In this display of variation, we can see that the disciplines did not load according to academic divisions. The posing of possibilities (philosophically or physically) would not be appropriate in all disciplines. History, for example, is concerned with reporting real events in the past, not things that could have happened. As per the other dimensions, Philosophy loaded at an extreme end of the spectrum. Examples A, D, E and H from Philosophy also have high Dimension 4 scores. Below, we give examples from texts that loaded highly positively. These texts from Education (Example L) and Nursing (Example M), are both instances of the student writer which offer some kind of interpretation. Interestingly, the Education example describes past events, but there is an emphasis on the writer describing what could have been altered or changed to improve her

teaching. It can, thus, be assumed that she will attempt to make those changes in the future. In Example M from Nursing, however, we see more of an advisory style of writing for the audience, reflecting the real life concerns associated with this discipline.

> (L)   Another possibility could have been for Jim to speak with someone in the faculty of his school or community who was a fluent speaker of Spanish. He could have used this person as a resource to help facilitate communication between Carlos' parents and himself in a conference. Alternatively, Jim could have also tried to get in touch with Carlos's sister. Regardless of which family members Jim was able to come in contact with, a conference would have allowed Jim the opportunity to voice his concerns about Carlos's progress in spelling and phonics. In retrospect, I believe I overlooked how large a problem it was that Jim did not use the aforementioned resources to his advantage.
> EDU.G0.08.1, argumentative essay, final year UG, NS
> Dimension 4 score $= 27.739$

> (M)   Good family communication can help to smooth the transition from child to adult. It is important to spend time with the adolescent, as a family, and one-on-one, parent to child. Often, the teenager will not want to spend time alone with the parents, for any of a variety of reasons. In these instances, it is helpful to take a moment to just remind her that you are available to talk at any time, if she is interested. When communicating with her, pay attention, try not to interrupt, act interested in what she is saying, and be respectful of the opinions expressed. It is okay to disagree, but should be done with tact.
> NUR.G3.02.1, report, third year G, NS
> Dimension 4 score $= 17.308$

At the other end of the spectrum, we can look at texts that have large numbers of past tense verbs and include fewer of the features of possibility, such as modals of possibility or prediction. This can be seen in Example N, which is taken from an argumentative essay from History and Classical Studies.

> (N)   Accompanying these different disciplinary conceptualizations of gender and sexuality were different normalization strategies and rhetorical devices. The biological focus of the West German sex education books allowed the authors to implicitly normalize certain behaviours, passing culturally determined beliefs off as natural and ergo inviolable. Specifically, it led to a focus on models of the norm. Thus, they cast certain behaviours and patterns of development as the biologically determined route, and suggested, furthermore, that any deviations from the norm

are biologically problematic, indeed *ipso facto* pathological.
    HIS.G1.03.1, argumentative essay, first year G, NS
    Dimension 4 score $=-12.568$

An examination of texts from other disciplines that loaded highly in this dimension is also revealing. An excerpt from the methodology section from a Biology research paper (Example O) shows multiple use of the past tense. Not surprisingly, the other sections of the paper are written mostly in the present. However, this text scored highly negatively on the positively loading features, such as infinitive verbs, non-auxiliary verbs and all modals.

(O)   DNA was extracted from tissue and feather samples using the Qiagen DNEasy Extraction kit. An addition of dithiothreitol was used for samples from feathers. Polymerase chain reaction (PCR) was carried out using two primers pairs for cytochrome B (CytB; Sorenson et al. 1999). PCR and sequencing was done following protocols in Mindell et al. (1997).
        BIO.G2.04.1, research paper, second year G, NS
        Dimension 4 score $=-14.348$

There is also a style of writing that is similar to the writing in History but that comes from other disciplines. The traditional 'library research paper', for example, might involve a student compiling information about past events. Depending on the degree of interpretation of the writer, there may be fewer modals of possibility, which could be used to hedge interpretations of past events. Example P, which is taken from an Education report, not only describes historical events, but also uses the past tense for reporting verbs, such as *asserted*.

(P)   In a famous 1840 address that served as a rallying point for immigrant Catholic dissatisfaction with the Horace Mann-inspired, Protestant-run education of the time, Hughes asserted that attendance at common schools constituted "a proximate danger to the religious faith of Catholic children," and that Catholic parents "could not in good conscience permit their children to attend these schools" – and upon hearing, the crowd "greeted his spirited suggestions with heavy applause."
        EDU.G1.10.1, report, first year G, NS
        Dimension 4 score $=-12.568$

5. Conclusion

This paper has explored the linguistic properties of successful, upper-level student writing found in MICUSP. First, we established the linguistic

composition of four statistically co-occurring factors of POS-tagged features across 825 texts from the corpus. Then, we interpreted the meanings and functions of these co-occurring features. The four functional dimensions that arose from the analysis distinguish between: (*1*) Involved, Academic Narrative *versus* Descriptive, Informational Discourse; (*2*) Expression of Opinions and Mental Processes; (*3*) Situation-Dependent, Non-Procedural Evaluation *versus* Procedural Discourse; and (*4*) Production of Possibility Statement and Argumentation. These textual meanings and functions have been identified by using interpretative and qualitative techniques from an analysis of text samples to comparisons of average factor scores within a dimension scale. Running a new multi-dimensional analysis following Biber's (1988) approach appears to have been effective in describing the underlying linguistic characteristics of 'quality' university-level writing in the United States. Logically overlapping features and dimensions identified in this study nicely complement other related studies, such as Biber (1988, 2006) and Gray (2011). Specific patterns and linguistic distributions unique to the sub-register of student writing were interpretable and could be utilised for future research.

Our decision to analyse the fifty texts that highly characterise the positive and negative poles of each dimension provided us with a unique insight into the disciplinary differences of academic writing from MICUSP. These fifty 'extreme' texts per scale show the influence of topics, contexts and expected writing conventions across disciplines in producing salient variation in university writing. For example, Dimension 1 (Involved, Academic Narrative *versus* Descriptive, Informational Discourse) distinguished Philosophy and Education texts from papers in Biology and Physics based on how writers encoded themselves, their ideas and opinions in their writing against production foci such as detailed nominal descriptions or delivery of technical information. Dimension 3 (Situation-Dependent, Non-Procedural Evaluation *versus* Procedural Discourse) captured the primary linguistic differences between process-orientated writing that is common in the sciences (e.g., from methods sections of research papers or laboratory reports) and situation-dependent discourses in essays and reports from Philosophy and English, which often involved reflection on and synthesis of readings. We also reported initial results of text type and student level (graduate *versus* undergraduate) comparisons across the four dimensions from the fifty high/low texts. Our exploration of these sub-categories of MICUSP provided interesting results that require more in-depth analysis in future studies. For example, many of the polar extremes were written by undergraduate writers. It would be interesting to longitudinally study the progress of student writers across these dimensions.

The results of this study have pedagogical applications, especially in the teaching of university-level academic writing. As noted earlier, writing instruction has increasingly shifted from English departments to writing intensive coursework housed in various 'content-focussed' departments (e.g., Biology, Engineering and Computer Science). There is clearly an objective

in these departments to develop a better understanding of student writing as it exists in the relevant content areas. Such an understanding of the linguistic properties of successful (or unsuccessful) writing could help instructors to address the needs of novices who are at the peripheries of their academic discourse communities. Related to this, the sub-fields of ESP and EAP, which make use of corpora in content-area teaching, have become very popular over the last few years (Belcher, 2009; Biber *et al.*, 2010; and Johns, 2009). Studies such as Flowerdew (2005), Gavioli (2005), Hinkel (2002), Hyland (2008), Jarvis *et al.* (2003) and Yoon and Hirvela (2004) to name only a few, recognise the valuable contribution of corpus-based data in the teaching of academic writing across disciplines, especially in increasing learners' awareness of the textual features of their own writing relative to target (i.e., successful) models. We hope to have shown in this paper that important insights of pedagogical relevance can be gained not just from analysing published academic writing but also from data-rich analyses of unpublished advanced student writing across text types, student levels and academic disciplines.

References

Bazerman, C., J. Little, L. Bethel, T. Chavkin, D. Fouquette and J. Garufis. 2005. Reference Guide to Writing Across the Curriculum. Available online, at: http://wac.colostate.edu/books/bazerman_wac/wac.pdf

Beers, S.F. and W. Nagy. 2009. 'Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre?', Reading and Writing: An Interdisciplinary Journal 22, pp. 185–200.

Belcher, D. (ed.) 2009. English for Specific Purposes in Theory and Practice. Ann Arbor, Michigan: University of Michigan Press.

Biber, D. 1986. 'Spoken and written textual dimensions in English: resolving the contradictory findings', Language 62 (2), pp. 384–414.

Biber, D. 1988. Variation across Speech and Writing. Cambridge: Cambridge University Press.

Biber, D. 2006. University Language: A Corpus-based Study of Spoken and Written Registers. Amsterdam: John Benjamins.

Biber, D. and S. Conrad. 2001. 'Introduction: multi-dimensional analysis and the study of register variation' in S. Conrad and D. Biber (eds) Variation in English: Multi-dimensional Studies, pp. 3–12. New York: Longman.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. The Longman Grammar of Spoken and Written English. London: Longman.

Biber, D., R. Reppen and E. Friginal. 2010. 'Research in Corpus Linguistics' in R. Kaplan (ed.) The Oxford Handbook of Applied Linguistics, pp. 548–67. (Second edition.) Oxford: Oxford University Press.

Comrey, A.L. and H. Lee. 1992. A First Course in Factor Analysis. (Second edition.) Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Ferris, D. 1994. 'Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency', TESOL Quarterly 28 (2), pp. 414–20.

Flowerdew, L. 2005. 'An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies', English for Specific Purposes 24, pp. 321–32.

Gavioli, L. 2005. Exploring Corpora for ESP Learning. Amsterdam: John Benjamins.

Grant, L. and A. Ginther. 2000. 'Using computer-tagged linguistic features to describe L2 writing differences', Journal of Second Language Writing 9 (2), pp. 123–45.

Gray, B.E. 2011. Exploring Academic Writing through Corpus Linguistics: When Discipline Tells Only Part of the Story. Unpublished Dissertation. NorthernArizona University.

Halliday, M.A.K. 1988. 'On the language of physical science' in M. Ghadessy (ed.) Registers of Written English, pp. 162–72. London: Pinter.

Hinkel, E. 2002. Second Language Writers' Text: Linguistic and Rhetorical Features. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hyland, K. 2004. Disciplinary Discourses: Social Interactions in Academic Writing. Ann Arbor, Michigan: University of Michigan Press.

Hyland, K. 2008. 'Genre and academic writing in the disciplines', Language Teaching 41 (4), pp. 543–62.

Hymes, D. 1984. 'Sociolinguistics: stability and consolidation', International Journal of the Sociology of Language 45, pp. 39–45.

Jarvis, S., L. Grant, D. Bikowski and D. Ferris. 2003. 'Exploring multiple profiles of highly rated learner compositions', Journal of Second Language Writing 12 (4), pp. 377–403.

Johns, A. 2009. 'Tertiary undergraduate EAP: problems and possibilities' in D. Belcher (ed.) English for Specific Purposes in Theory and Practice, pp. 41–59. Ann Arbor, Michigan: University of Michigan Press.

McNamara, D.S., S.A. Crossley and P.M. McCarthy. 2010. 'The linguistic features of quality writing', Written Communication 27 (1), pp. 57–86.

Michigan Corpus of Upper-level Student Papers. 2009. Ann Arbor, Michigan: The Regents of the University of Michigan.

O'Donnell, M.B. and 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), pp. 1–18.

Reppen, R. 2001. 'Register variation in student and adult speech and writing' in S. Conrad and D. Biber (eds) Variation in English: Multi-Dimensional Studies, pp. 187–99. New York: Longman.

Römer, U. and 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), pp. 159–77.

Römer, U. and M.B. O'Donnell. Forthcoming. MICUSP: A Corpus Resource for Exploring Proficient Student Writing across Disciplines. Amsterdam: John Benjamins.

Russell, D.R. 1990. 'Writing across the curriculum in historical perspective: toward a social interpretation', College English 52 (1), pp. 52–73.

Shaw, P. and E. Liu. 1998. 'What develops in the development of second-language writing?', Applied Linguistics 19 (2), pp. 225–54.

Swales, J.M. and C.B. Feak. 2004. Academic Writing for Graduate Students: Essential Tasks and Skills. (Second edition.) Ann Arbor, Michigan: University of Michigan Press.

Thaiss, C. and T. Porter. 2010. 'The state of WAC/WID in 2010: methods and results of the U.S. Survey of the International WAC/WID Mapping Project', College Composition and Communication 61 (3), pp. 534–70.

Yoon, H. and A. Hirvela. 2004. 'ESL student attitudes toward corpus use in L2 writing', Journal of Second-Language Writing 13, pp. 257–83.