

Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing

Gardner, S., Nesi, H. & Biber, D.

Published PDF deposited in Coventry University's Repository

Original citation:

Gardner, S, Nesi, H & Biber, D 2018, 'Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing' *Applied Linguistics*
<https://dx.doi.org/10.1093/applin/amy005>

DOI 10.1093/applin/amy005

ISSN 0142-6001

ESSN 1477-450X

Publisher: Oxford University Press

© The Author(s) (2018). Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

Discipline, Level, Genre: Integrating Situational Perspectives in a New MD Analysis of University Student Writing

^{1,*}SHEENA GARDNER, ¹HILARY NESI, and ²DOUGLAS BIBER

¹Coventry University, ²Northern Arizona University

*E-mail: sheena.gardner@coventry.ac.uk

While there have been many investigations of academic genres, and of the linguistic features of academic discourse, few studies have explored how these interact across a range of university student writing situations. To counter misconceptions that have arisen regarding student writing, this article aims to provide comprehensive linguistic descriptions of a wide range of university assignment genres in relation to multiple situational variables. Our new multidimensional (MD) analysis of the British Academic Written English (BAWE) corpus identifies clusters of linguistic features along four dimensions, onto which academic disciplines, disciplinary groups, levels of study, and genre families are mapped. The dimensions are interpreted through text extracts as: (i) Compressed Procedural Information versus Stance towards the Work of Others; (ii) Personal Stance; (iii) Possible Events versus Completed Events; and (iv) Informational Density. Clusters of linguistic features from the comprehensive set of situational perspectives found across this framework can be selected to inform the teaching of a ‘common academic core’, and to inform the design of programmes tailored to the needs of specific disciplines.

1. INTRODUCTION

A long-standing question for those teaching academic writing to university students centres on the extent to which instruction should be general or specific to particular disciplines (Ferris 2001; Hyland 2002); the debate about English for General versus English for Specific Academic Purposes (EGAP versus ESAP) continues to this day (de Chazal 2013; Flowerdew 2016). Research on the discipline- and genre-specific nature of published academic writing is substantial, as the pages of journals such as *JEAP* and *ESPJ* attest, but research findings regarding the nature of published academic writing may not be very useful to writing tutors advising student writers, particularly at undergraduate level. The literature suggests progression routes leading students from more general to more discipline-specific writing (Johns, 2008; Gardner 2016), and it is now widely recognized that pre-university or first-year composition

teaching for multiple disciplines may tend towards EGAP, while in-sessional upper-level teaching for specific disciplinary contexts may tend towards ESAP. Questions remain, however, regarding the nature of a 'common core' of features relevant to all types of academic writing, applicable in a wide range of EAP teaching contexts.

Earlier work on academic genres across disciplines in student writing (Nesi and Gardner 2012; Gardner and Nesi 2013) has identified the disciplines in which academic genres such as essays, lab reports, and case studies occur. Knowing that essays are frequent in History and in Sociology, for example, tells us that in both disciplines students are expected to demonstrate their powers of independent thinking and build an argument using evidence from discipline-appropriate sources. It does not, however, tell us whether abstract nominalization, complex noun groups, or stance adverbials are equally important resources in both these disciplines. Similarly, knowing that case study genres are frequent in Health and in Business does not tell us whether there is a common core of language that might be used when analysing either a company or a patient case, or when making recommendations for either business or medical interventions.

Large-scale studies of English corpora have found that specific features (such as imperatives, phrasal verbs, attributive adjectives, or stance adverbials) are more or less frequent in academic prose compared to conversational, fictional, and media registers (Biber, Johansson, Leech, Conrad and Finegan 1999). Studies of student writing have also investigated the incidence of specific linguistic features, such as shell nouns (Nesi and Moreton 2012), lexical bundles (Durrant 2017), and phrasal and clausal complexity (Staples, Egbert, Biber and Gray 2016). These studies add incrementally to our understanding of academic prose, but in their broad treatment of academic registers and their focus on specific linguistic features, they also run the risk of misleading practitioners. The fact that phrasal verbs or first-person pronouns are less frequent in academic than in other registers does not mean they should always be replaced in all academic texts because it is quite possible that they occur frequently in some academic situations, and not at all in others. Equally, the fact that long nominal groups with abstract head nouns are frequent in academic registers generally does not mean that all types of long nominal group are frequent across all types of student writing.

Unfortunately, although features associated or disassociated with academic registers have been treated as markers of writing development in EAP contexts, from Hong Kong (Crosthwaite 2016) to the UK (Issitt, 2017), such measures do not account for the way such features might cluster and disperse across the range of disciplines, genres, and levels of study in student writing situations.

The dominant approach to identifying clusters of features that occur in texts from contrasting situational contexts is multidimensional (MD) analysis. It was developed in the early 1980s as a research methodology for describing the patterns of linguistic variation that distinguish among registers (see Biber 1988). Early applications were used to describe the relations among spoken

and written registers in English, where written academic registers were found to be more explicit and abstract, and to have less interpersonal and affective content and fewer narrative concerns than spoken registers or fiction. The methodology was extended to investigate variation among university registers (Biber 2006), where marked differences emerged between oral and procedural discourse (in service encounters, office hours, study groups, and classroom management) and more literate and content-focused discourse (in textbooks and course packs). In this 2006 study, all the spoken university registers were found to be characterized by a focus on personal stance.

More recently, the third phase of MD research on academic Englishes has tended to concentrate on distinctions within domains. Some studies have focused on a particular aspect of the academic situation, such as academic level (Biber, Conrad, Reppen, Byrd and Helt 2002), academic genre (Nesi and Gardner 2012, Hardy and Friginal 2016), or academic discipline (Biber *et al.* 2002, Biber 2006, Hardy and Römer 2013). Analysis of academic genres has identified certain types of student writing, for example proposals and literature surveys, as being particularly informationally dense, and other types such as narratives and creative writing as being more 'involved', containing features more typical of the spoken language. Analysis of academic disciplines suggests that the hard sciences are more informational, while Humanities disciplines are more involved. More specifically, MD analyses of the Michigan Corpus of Upper-level Student Papers (MICUSP), a corpus of American student writing, found that at the extremes of each of the four MICUSP dimensions, the linguistic clusters in Physics texts contrast most with those in Philosophy (Hardy and Römer 2013), just as those in research report genres contrast consistently with those in creative writing (Hardy and Friginal 2016). Building on this work, we think it is possible to discriminate more finely between situational types of student writing by working with a larger data set and combining examination of a greater number of situational variables.

The aim of the current study is to relate linguistic features to situational perspectives on student academic writing, to enhance our understanding of the way linguistic features cluster in different writing situations, and inform academic writing teachers, curriculum planners, and materials developers involved in the teaching of English for General or Specific Academic Purposes.

Following an introduction to the British Academic Written English (BAWE) corpus and the situational variables related to genre, discipline, and level of study (Section 2), details of the MD methodology used in this study are explained (Section 3). Section 4 presents our findings, starting with an overview of the features that cluster at the poles of each dimension (4.1), followed by a mapping of the levels and disciplinary groups, as well as the more specific disciplines and genre families, along each of the four dimensions (4.2–4.5). Through discussion and annotated examples, the character of each dimension will be revealed. The final section (5) reviews the entire new framework and the insights gained from interpreting the factor analysis from these multiple situational perspectives.

2. SITUATIONAL VARIABLES: GENRE, DISCIPLINE, AND LEVEL

In line with the aim of informing a progression from general to specific situations, the theoretical constructs underpinning this analysis allow for degrees of specificity, as will now be described in terms of genre, discipline, and level, with particular application to the BAWE corpus of university student writing.¹

The BAWE corpus was developed as a resource for investigations of successful British university student writing at the beginning of the 21st century. Assignment texts were selected for the corpus if they had received a top grade when assessed by discipline tutors as part of regular degree-level coursework. Care was taken to ensure that no one discipline, level of study, or individual student was over-represented. The contents of the corpus are described in detail in [Nesi and Gardner \(2012\)](#). The rest of this section explains the situational variables of genre, discipline, and level in relation to the corpus, and statistical details are presented in the 'Methodology' section below (see [Table 1](#)).

The development of a classification of genres for the BAWE corpus was informed by explorations of writing contexts through document examination and interviews with students and professors. Genres with similar purposes and staging were grouped into 13 'families'. For example, an expository essay and a discussion essay are classified together in the Essay genre family; a book review and a product evaluation are classified in the Critique genre family; and an annotated bibliography and a literature review are classified in the Literature Survey genre family. The genre families are described in [Nesi and Gardner \(2012\)](#) according to five broad social purposes: Explanations and Exercises allow students to demonstrate their knowledge and understanding; Essays and Critiques provide opportunities for students to cultivate their independent thinking and powers of critical evaluation; Methodology Recounts, Literature Reviews, and Research Reports develop students' research capabilities; Case Studies, Design Specifications, Problem Questions, and Proposals help prepare students for future professional practice; Narrative Recounts and Empathy Writing enable students to reflect on their own practice and communicate with a readership beyond their course. This means that we can map clusters of linguistic features onto specific genres, genre families, or groups of genre families that share a broad social purpose.

Academic disciplines can also be viewed along a continuum of specificity. There are four broad disciplinary groupings in the BAWE corpus classification, each of which is represented by texts from around seven specific disciplines. For example, Arts and Humanities includes English, History, Linguistics, Philosophy, and Classics; Life Sciences includes Agriculture, Biology, Food Science, Health and Psychology; Social Sciences includes Business, Economics, Law, Sociology, and Politics; while Physical Sciences includes

pure subjects such as Mathematics and Physics as well as applied subjects such as Computer Science and Engineering. The disciplines are represented by successful student assignments across the levels of study.

The aim in the BAWE corpus was to capture student assignment writing from taught courses rather than research courses. The texts are therefore from four levels of study at undergraduate and taught masters levels. The levels of study in British universities reflect a progression of expectation, perhaps more so than in the American system where greater cross-disciplinary optionality is possible. The break in continuity tends to come between Level 3 (representing the final year of undergraduate study) and Level 4 (representing taught courses at master's level). Level 4 students tend to come from a variety of backgrounds, often from other countries and often post-experience, or from a different discipline (e.g. moving into an MBA from a degree in Economics, or into Applied Linguistics from a degree in English). Finally, it is worth noting that while in some courses students will produce the same genres throughout, in others the upper-level writing is quite different, with a shift towards more research- or professionally-oriented genres.

3. METHODOLOGY

In this analysis we used the entire 6.5 million word BAWE corpus, comprising 2,760 assignments written by 812 students for around 1,000 different modules in over 30 disciplines, representing 300 degree courses from four universities in England. These assignments were grouped into 13 genre families. The corpus includes comparable numbers of texts at each level of study from first-, second-, and final-year undergraduate courses and from taught post-graduate courses and comparable numbers of assignments from each of the four disciplinary groups: Arts and Humanities, Life Sciences, Physical Sciences, and Social Sciences² (Table 1).

Table 1: Number of BAWE corpus assignments in levels of study and disciplinary groups

| | Level 1 | Level 2 | Level 3 | Level 4 | Total |
|---------------------|---------|---------|---------|---------|------------------|
| Arts and Humanities | 239 | 228 | 160 | 78 | 705 |
| Social Sciences | 207 | 197 | 162 | 201 | 776 ² |
| Life Sciences | 180 | 193 | 113 | 197 | 683 |
| Physical Sciences | 181 | 149 | 156 | 110 | 596 |
| Total | 807 | 767 | 591 | 586 | 2760 |

The numbers of assignments per genre family and discipline are more variable, and are indicated in Figures 2, 4, 6 and 8 below.³

The texts in the BAWE corpus were coded using the Biber tagger for c. 150 lexico-grammatical characteristics (see Biber *et al.* 1999). We then computed the rate of occurrence (per 1,000 words) for each linguistic feature in each text. This information provided the basis for the MD analysis of variation, the procedures for which have been documented in several previous publications (Biber 1988, Friginal 2013). In brief, the notion of linguistic co-occurrence is given formal status in the MD approach through a statistical factor analysis (or principal component analysis), which quantitatively identifies the sets of linguistic features that frequently co-occur in texts; these are referred to as the linguistic 'dimensions' of variation. Dimension scores are then computed for each text, by summing the standardized rates of occurrence for each of the linguistic features grouped on a dimension. Finally, mean dimension scores (and standard deviations) are computed for each text category (e.g. disciplinary group, level of study). Plots of these mean dimension scores allow linguistic characterization of any given category, comparison of the relations between categories, and a fuller functional interpretation of the underlying dimension.

Based on the theoretical claim that linguistic co-occurrence patterns reflect underlying functions (see Egbert and Biber 2017), the dimensions are interpreted to identify the communicative functions associated with each dimension. The interpretation process is based on consideration of the set of linguistic features co-occurring on each dimension, the similarities and differences among text categories with respect to the dimension (shown by their mean dimension scores), and detailed analysis of the ways in which co-occurring linguistic features function in individual texts. The functional interpretation is then summarized with a descriptive label for each dimension, such as 'Oral versus literate discourse' or 'Personal stance'.

For the present study, we began with the lexico-grammatical features identified by the Biber tagger. We eliminated variables with low communalities in the preliminary factor analysis runs because they had low shared variance with the overall factor structure and thus contributed little to the analysis. But there are additional considerations that influence the selection of features for the final factor analysis because there is considerable overlap among many of these features. That is, lexico-grammatical characteristics can be analysed at many different levels of specificity, and it is important to avoid hierarchical inclusion of features that represent the same domain of English grammar. For example, the tagger includes analysis of three specific classes of modal verbs (possibility modals, necessity modals, and prediction modals) as well as a count for total modal verbs. If all four of these variables had been included, the exact same domain of linguistic variation would have been represented twice. To the extent possible, specific lexico-grammatical features were retained in the factor analysis rather than more general superordinate grammatical features. In addition, redundancies were eliminated by

combining some variables, and dropping other variables that had low overall frequencies.

Thirty-nine linguistic variables were retained for the final analysis (see [Supplementary Material Appendix Table A1](#)). Readers are referred to [Biber *et al.* \(1999\)](#) and [Biber \(2006\)](#) for descriptions of these individual linguistic features. A four-factor solution was selected as optimal. This decision was based on scree plot inspection, and the interpretability of the factors extracted in different solutions. The factor solution accounts for 39.3 per cent of the cumulative shared variance.⁴ Factors were rotated using a Promax rotation, which resulted in generally small correlations among the dimensions. We now present the results of the factor analysis and explain how the dimensions have been interpreted through a consideration of higher and lower scoring texts along each dimension.

4. RESULTS AND DISCUSSION

4.1 Linguistic features in the BAWE dimensions

[Appendix Table A1 \(Supplementary Material\)](#) gives the factor loadings for the 39 linguistic features retained, on each of the four dimensions. From this we can extract [Table 2](#), which shows those features with the most salient (± 0.35) loadings at the positive and negative ends of the four dimensions. For example, we can see that there are six salient features that cluster at the positive end of the first dimension (premodifying nouns, common nouns, passives, action verbs, concrete nouns, and quantity nouns), but it is not immediately obvious which genres, disciplines, or levels of student writing will contain such clusters.

Before we look at how these clusters of features map onto texts, it is worth noting that while Dimensions 1 and 3 have clusters of salient features at their positive and negative poles, Dimensions 2 and 4 can best be characterized by the features located towards the positive poles alone. The negative ends of these dimensions are simply characterized by the absence of the features at the positive end.

To help interpret the factor analysis and label the resulting dimensions, we ranked the 2,760 assignment texts for each dimension, and examined their situational characteristics (discipline, genre family, level, etc.).⁵ We also manually examined high and low scoring assignments, and used corpus queries to search for texts with clusters of features via SketchEngine.⁶ We were guided in our interpretation by previous research, including [Biber \(2006\)](#), and our understanding of the contexts of student writing, acquired from earlier work ([Nesi and Gardner 2012](#)).

In what follows we will examine each dimension in turn and see how student writing from different disciplines, levels, and genre families is distributed over the dimensions.

Table 2: Most salient feature loadings on four dimensions

| Dimension 1 | Dimension 2 | Dimension 3 | Dimension 4 | | |
|----------------------------|-------------|----------------------------|-------------|--|-------|
| Premodifying nouns | 0.69 | Mental verbs | 0.75 | Word length | 0.87 |
| Common nouns | 0.60 | Stance verbs + that clause | 0.60 | Modal verbs | 0.56 |
| Passives | 0.56 | Stance verbs + to clause | 0.54 | Verb to be | 0.51 |
| Action verbs | 0.53 | <i>That</i> deletion | 0.52 | Subordinating conditional conjunctions | 0.40 |
| Concrete nouns | 0.52 | Communication verbs | 0.47 | | 0.35 |
| Quantity nouns | 0.43 | First-person pronouns | 0.40 | | |
| | | Past tense verbs | 0.39 | | |
| Communication verbs | -0.39 | | | | |
| Stance adverbials | -0.39 | | | | |
| Proper nouns | -0.40 | | | | |
| Stance nouns + that clause | -0.44 | | | Perfect aspect | -0.37 |
| Third-person pronouns | -0.55 | Prepositions | -0.44 | Past tense verbs | -0.83 |

4.2 Dimension 1: Compressed procedural information versus stance towards the work of others

The linguistic features that cluster at the positive end of Dimension 1 are nouns as premodifiers, common nouns, passives, action verbs, concrete nouns, and quantity nouns (see Table 2). These features highlight the importance of nouns in this cluster, and action verbs. At the opposite end of Dimension 1, we find third-person pronouns, stance nouns with *that* clauses, proper nouns, stance adverbials, and communication verbs. We can see specific contrasts (e.g. between common and proper nouns, or between action and communication verbs), but what is interesting about the dimensions is how these features cluster (so common nouns occur with action verbs, where proper nouns occur with communication verbs) and the stance prosody identified through nouns and adverbials.

Table 3 summarizes the statistical results for a general linear model (GLM) analysis (in SAS) of mean Dimension 1 scores across disciplinary groups, levels of study, and genre families. The results show that all three independent variables are statistically significant predictors of Dimension 1 scores (see *p* values) and that disciplinary group and genre family are important predictors of Dimension 1 scores (with R^2 values greater than 40 per cent).

When the means for the four disciplinary groups and four levels of study are examined (Figure 1), we can see that the differences in disciplinary group means are greater than those in levels of study, with Physical Sciences scoring plus 7 (7.235) compared to Arts and Humanities at minus 7 (−6.930), while all the levels of study means are close to 0. The letters (ABCD) show that whereas there is no significant difference between the means at Levels 1 and 2 (both have the same letter, ‘C’), there are significant differences between the means of each of the disciplinary groups.

Interestingly a visual examination of level in the ranking of individual texts along this dimension shows that the texts at the positive end are from across the levels of study, while those at the negative pole are predominantly from Levels 1 and 2 (53 of the last 60 texts are from Levels 1 and 2).

Table 3: GLM results for Dimension 1 (Compressed procedural information versus stance towards the work of others), comparing mean differences across disciplinary group, level of study, and genre family

| Independent variable | DF | <i>F</i> -value | Significance | R^2 (per cent) |
|----------------------|----|-----------------|--------------|------------------|
| Disciplinary group | 3 | 695.5 | $p < .0001$ | 43.1 |
| Level of study | 3 | 15.7 | $p < .0001$ | 1.7 |
| Genre family | 12 | 201.7 | $p < .0001$ | 46.8 |

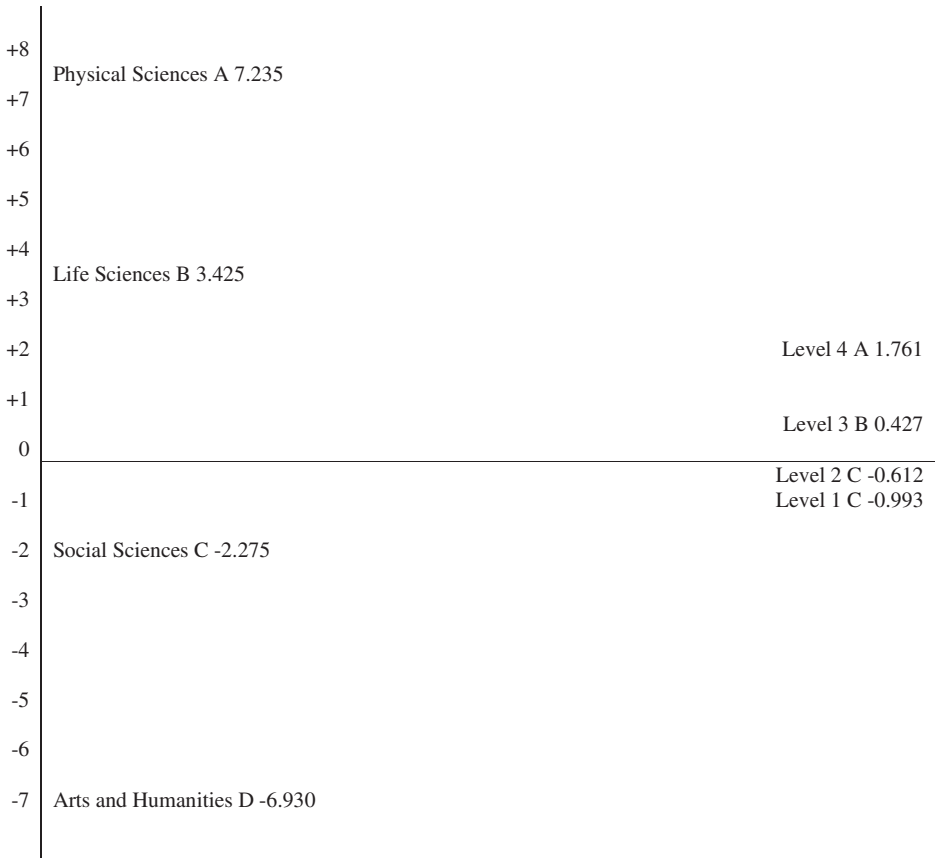
Compressed Procedural Information*Stance towards the Work of Others*

Figure 1: Dimension 1 mean scores for disciplinary groups and academic levels

To better understand how specific disciplines and genre families contribute to these results, Figure 2 plots the mean scores of the disciplines (which are adjacent to the y -axis) and genre families (which are in italics) along the first dimension. The number (n) of assignment texts for each group is indicated in brackets, beside the mean score.

Figure 2 shows how the disciplines of Food Science, Chemistry, Engineering, and Meteorology cluster at the positive end of Dimension 1, together with the Methodology Recount and Design Specification genre families, all of which have means greater than +8.

Extracts 1a–b illustrate how nouns as premodifiers, common nouns, passives, action verbs, concrete nouns, and quantity nouns cluster in Food

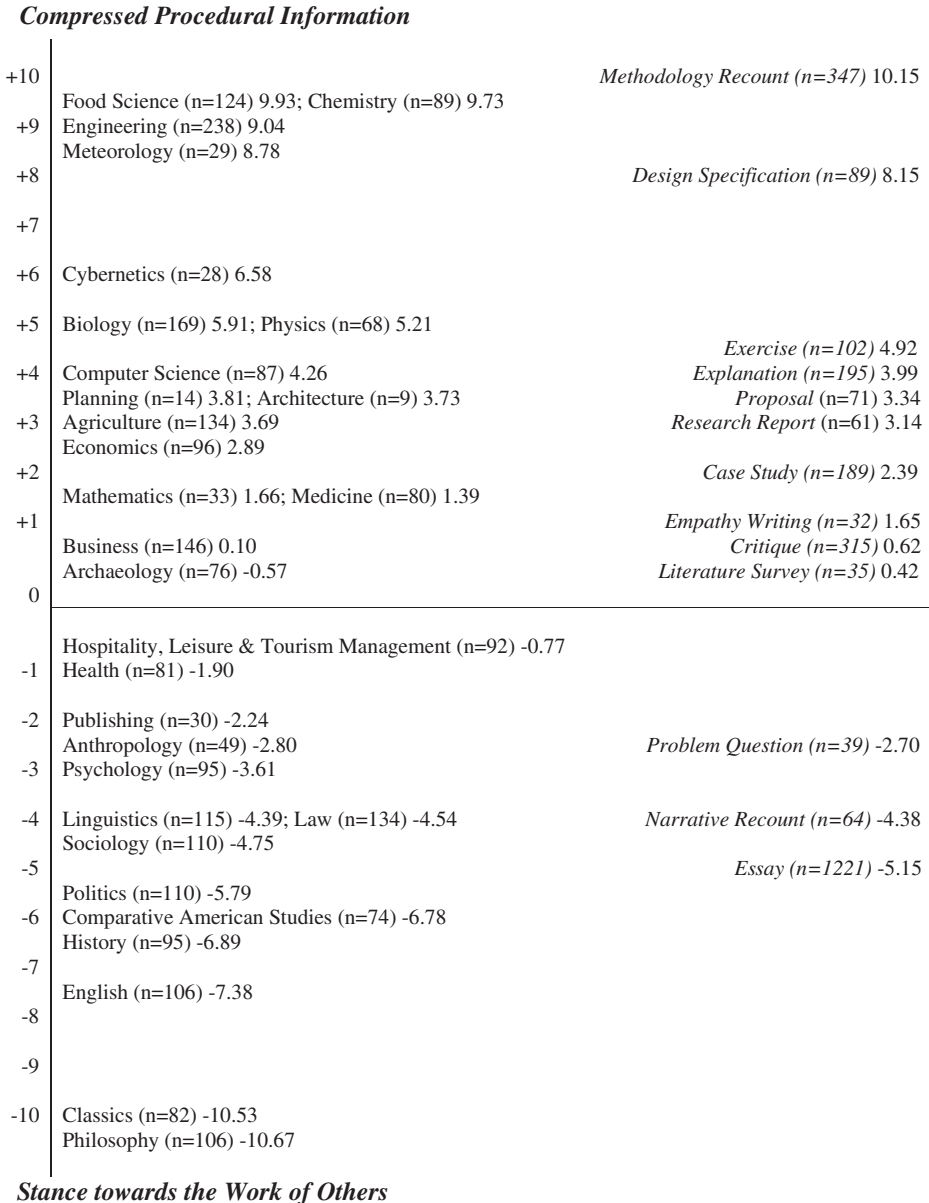


Figure 2: Dimension 1 mean scores for disciplines and genre families

Science, Chemistry, and Engineering, and in science reports (Methodology Recounts and Design Specifications).

Extract 1a.

The average fluoride concentration in local tap water was found to be $1500 \leq \text{g/l}$ [3], and in brewed tea worldwide varied from c. 600 to $3000 < \text{g/l}$ [4]. A series of standard fluoride solutions encompassing this range were made from a 0.1 M NaF stock solution of 0.4200 g reagent grade NaF (Aldrich) dissolved in 100 ml distilled water at 295 K.

(Chemistry Methodology Recount 0415c, 16.2 on Factor 1)

Extract 1b

The Satellite Scoreboards have been designed to rotate, thereby giving a wider field of view to the spectators. The manufacturing/mechanical team calculated the required torque and speed to move a Satellite Scoreboard. A motor was chosen that would satisfy these requirements. The motor was also required to operate from either 5 V or 12 V DC, since these were the two power supply voltages provided to each scoreboard.

(Engineering Design Specification 0146c, 9.6 on Factor 1)

Here we see nouns as premodifiers (*fluoride concentration, tap water, Satellite Scoreboards, power supply voltages*), common nouns (*tap, water, tea, field, team*), passives (*was found to be, was chosen*), action verbs (*made, rotate, move*), and concrete nouns (*water, tea, solution, motor*). These texts tend to be densely written, with long scientific nominal groups (noun premodifiers, common, concrete, quantity nouns) and a focus on concisely reporting experimental procedures through passive action verbs. We have therefore labelled the positive pole of Dimension 1 'Compressed Procedural Information'.

In stark contrast to the Compressed Procedural Information found in the science reports above, [Figure 1](#) shows that the negative features on Dimension 1 are concentrated in Essays in the Arts and Humanities disciplines of History, English, Classics, and Philosophy. For example:

Extract 2a

Lord Henry is a man whose theories are exotic and enticing but also often dangerous, yet he has little conception of their practical application. He proclaims hedonism as a way of life, yet lives a rather mundane life himself, seemingly fulfilled enough by the London social scene. It seems then that whilst his intelligence and wit are evident, his understanding of the human soul is distinctly lacking and thus he has no sense that his desire to 'dominate' Dorian is immoral. In fact he takes an almost perverse pleasure from observing the effect his words have upon the vulnerable Dorian in the scene just after the painting is finished.

(History Essay, 0252t, -14.8 on Factor 1)

Extract 2b

Despite Aeneas' seeming desire to stay with Dido, he still proves his dedication to his greater cause by suggesting to her that he had no intention of lingering in Carthage and that his love lies with the future of his Trojan people. He also backs his argument with the simple fact that leaving Carthage is beyond his control; the gods had demanded his devotion to the future of Rome. Despite his claims, he has the choice as to whether or not he follows his destiny, and it is by his own will that he pursues it.

(Classics Essay, 6192b, –13.6 on Factor 1)

Extract 2c

Plato claims that order in the state will be maintained through the 'nurture and education' (Rice, 1952, p.57) of the Guardians and the propaganda used by the Guardians. He is able to claim that they will only be concerned for the welfare of the state and that they will be perfect rulers because they have been taught so well. Any attempt to show this to be impossible, or example of a Guardian not behaving in this way would not be a problem for Plato, because he would be able to propose that the education had not been adequate. As a perfect education system would be impossible to realise in the real world, so therefore would be the possibility of these perfect Guardians.

(Philosophy Essay, 3019h, –12.2 on Factor 1)

Extracts 2a–c are typical of first- and second-year undergraduate Humanities Essays that seek to interpret the lives and works of significant individuals and places. We call this pole of Dimension 1 'Stance towards the Work of Others'. Here we see third-person pronouns (*he, her, it*), stance nouns (*theory, argument, fact, attempt, problem*), stance adverbials (*seemingly, only, so well*), proper nouns (*Lord Henry, Dorian, London, Aeneas, Dido, Carthage, Rome, Plato, Rice, Guardians*), and communication verbs (*proclaim, claim, state, propose*). Here we also see longer sentences, expanding through conjunctions (*but, yet, and*) in Extract 2a, and through different kinds of *that* clause in Extract 2b. Such features contribute to the more expansive style of this academic discourse, particularly when compared with the compressed language of the science reports in Extracts 1a–b.

4.3 Dimension 2: Personal stance

The linguistic composition of the positive end of Dimension 2 includes mental verbs, stance verbs with *that* and with *to* clauses, *that* deletion, communication verbs, first-person pronouns, and past tense verbs. It is thus similar to the negative end of Dimension 1 in its inclusion of stance features and communication verbs but differs in the nature of the stance features, and the inclusion of mental verbs and first-person pronouns. Thus although the two poles both include stance features, the clusters differ markedly, which suggests that stance

features should not be taught ‘en bloc’, but rather in relation to the clusters in which they occur and the situational variables of the texts in which they are frequent.

Table 4 summarizes the statistical results for the GLM analysis of disciplinary group, level of study, and genre family as predictors of Dimension 2 scores. The results show that Dimension 2 mean differences across all categories are statistically significant and important (especially for genre family, with an R^2 value over 20 per cent).

Contrary to Dimension 1, Dimension 2 scores for assignments at each level decrease (rather than increase) steadily with each year of study, indicating that students steadily express stance to lesser extents as they progress in their university educations. However, the larger differences are still found across the disciplinary groups, as indicated by the ABCD letters in Figure 3.

Unlike in Dimension 1, however, the disciplines are not ranked primarily according to disciplinary group. For instance, Philosophy, from Arts and Humanities, is at the positive pole of the dimension next to Health from Life Sciences, while English is next to Mathematics, and Medicine is next to Law (Figure 4).

Dimension 2 is similar to the negative pole of Dimension 1 in that both have a functional association with the expression of stance. However, the two differ in their particular functions: the negative pole of Dimension 1 is associated with evaluation of the work of others, while Dimension 2 is associated with evaluative language used to describe personal experiences and opinions. We have thus named Dimension 2 ‘Personal Stance’.

The co-occurring linguistic features associated with Dimension 2 include mental, stance, and communication verbs, as well as first-person pronouns, past tense verbs, and *that* deletions (see Table 2). These features are perhaps more often associated with informal spoken language than written academic texts. In the BAWE corpus they are found where students report and reflect on their personal experiences, or propose professional solutions to simulated ‘real world’ scenarios. Narrative Recounts have an exceptionally high score on this dimension, distinguishing them from all other types of academic writing in the corpus. This genre family includes reflective writing, as in Extract 3a; this genre can be surprisingly difficult for students to master because ‘many of the

Table 4: GLM results for Dimension 2 (Personal Stance) mean differences across disciplinary group, level of study, and genre family

| Independent variable | DF | F-value | Significance | R^2 (per cent) |
|----------------------|----|---------|--------------|------------------|
| Disciplinary group | 3 | 78.6 | $p < .0001$ | 7.9 |
| Level of study | 3 | 24.1 | $p < .0001$ | 2.6 |
| Genre family | 12 | 61.3 | $p < .0001$ | 21.1 |

Personal Stance

| | | |
|----|--|-------------------------------------|
| +3 | Arts and Humanities A 2.959 | |
| +2 | | |
| +1 | | Level 1 A 1.066 Level 2 AB 0.386 |
| 0 | | |
| -1 | Social Sciences B -0.217 Life Sciences C -1.098 | Level 3 B -0.124 |
| -2 | Physical Sciences D -1.960 | Level 4 C -1.844 |

Figure 3: Dimension 2 mean scores for disciplinary groups and academic levels

features which contribute to the success of reflective writing flout academic conventions within the Western higher education “essayist” tradition’ (Nesi and Gardner 2012: 229).

Extract 3a

When we got to the hospital we realised ^ we were not needed and the injured were being taken to another hospital. Just before midnight I thanked the doctors for the kindness o they had shown me over the past eight weeks and said goodbye. I would love to recommend my elective because I did thoroughly enjoy it but I will have to state truthfully that Egypt is currently not safe to visit.

(Medicine Narrative Recount 0065g, 20.3 on Factor 2)

Extract 3b

Due to the lack of force used to actually attempt to acquire the ‘phone (the force used was entirely independent of this act), I think it unlikely that attempted robbery would be the charge. Amy’s attempt was a complete one (meaning that she carried out the whole act, but simply did not reach the outcome ^ she had desired).

(Law Problem Question 0143e, 7.9 on Factor 2)

In Extracts 3a–b we see examples of mental verbs (*realised, think*), stance verbs (*enjoy, love, desire*), communication verbs (*said, state*), first-person pronouns (*I, we*), past tense verbs (*realised, thanked, said, carried out*), and *that* deletions (indicated by ^).

The negative end of the second dimension is characterized by the absence of Personal Stance features. Here we find texts that aim to provide information as statements of objective truth, whether in explanations of theories and

Personal Stance

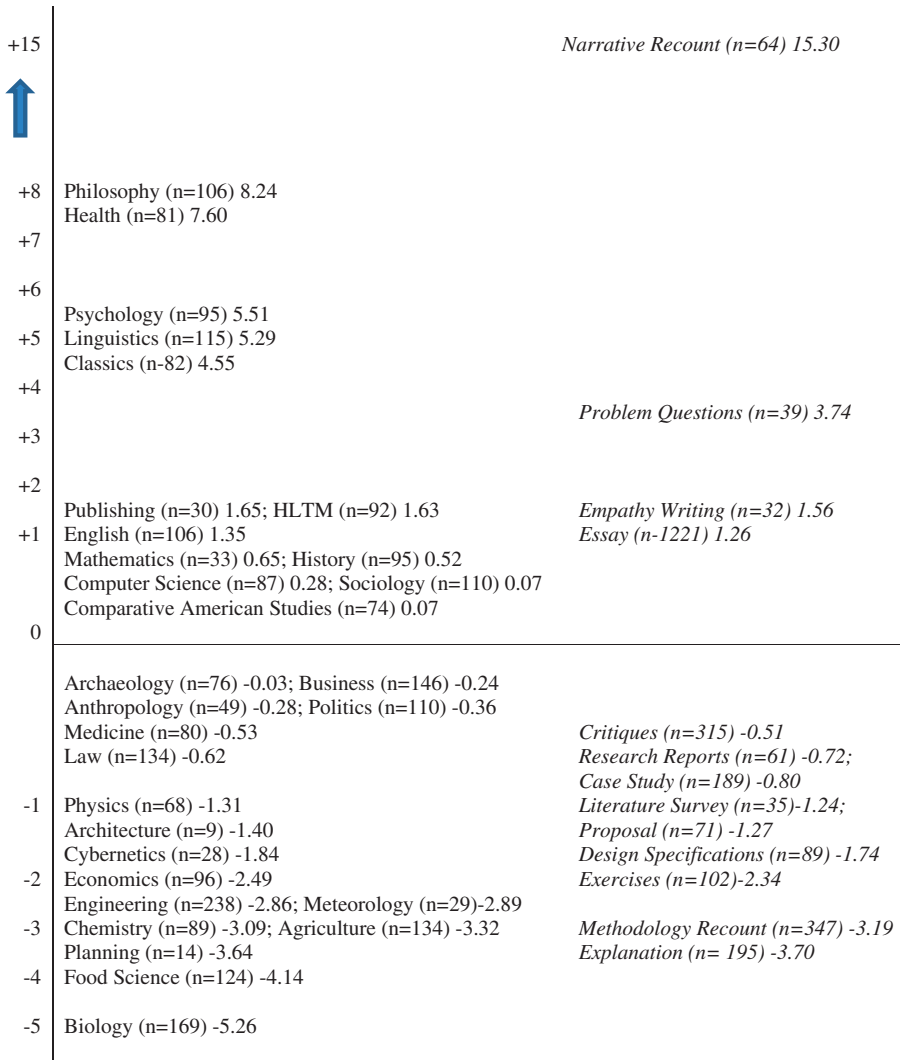


Figure 4: Dimension 2 mean scores for disciplines and genre families

classifications (Extract 4a) or descriptions of physical and temporal locations (Extract 4b).

Extract 4a

Bacteria are prokaryotes which possess simple chromosomes and no nuclear membrane. They are single-celled organisms and have simple structure. Fungi are eukaryotes which possess a true nucleus enclosed in a nuclear

membrane that contains their genetic material within complex chromosomes. They are either unicellular such as yeasts or multicellular such as moulds.

(Food Sciences Methodology Recount 6008p, -6.8 on Factor 2)

Extract 4b

A tree stands 4 m high and 2 m in front (south of) the proposed canopy roof. At different times of the day throughout the year the sun will cast a shadow of the tree onto the PV system installed on the proposed canopy roof. On most days this particular tree location forms shadows across the roof starting around midday and then on throughout the afternoon.

(Engineering Design Specification 6161d, -8.2 on Factor 2)

Neither of these texts suggests that there are any doubts or that alternative interpretations of the ‘facts’ would be possible. There is no mention of the writer as *I* or *we*. They are also quite different from Extracts 1a–b in their absence of past tense action verbs and passives.

4.4 Dimension 3: Possible events versus completed events

The linguistic features of Dimension 3 are predominantly verbs and dependent clauses. At the positive end we find present tense verbs, modal verbs, the verb *to be*, and subordinating conditional clauses. These are contrasted at the negative end with past tense verbs and the rather rare perfect aspect.

Table 5 summarizes the statistical results for the GLM analysis of disciplinary group, level of study, and genre family as predictors of Dimension 3 scores. The results show that Dimension 3 mean differences across disciplinary groups and genre families are statistically significant and moderately important (with R^2 values over 5 per cent).

There are no significant differences across levels of study, but as with Dimensions 1 and 2, there are significant differences across the four disciplinary groups (see Figure 5). Unlike Dimensions 1 and 2 where Physical Sciences were closer to Level 4 means, here it is Arts and Humanities and Level 4 texts that are the outliers, as both have negative means where all others are positive, though the values are relatively small.

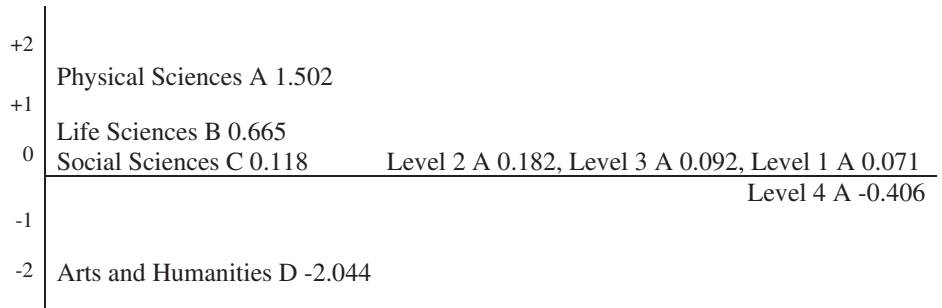
Again, a broader spread is seen when we look at the specific disciplines: Philosophy is markedly positive at 7.2, compared to History and Comparative American Studies at -8, while most of the disciplines and genre families are bunched between +5 and -4 (Figure 6).

In contrast with the mental verbs and stance features of Dimension 2, Table 2 shows that the lexico-grammatical clusters in Dimension 3 are very much focused on verb tenses, modality, and subordinate conditional clauses (usually *if...then*). We have interpreted this dimension as representing ‘Possible Events’. This constellation of features is common in disciplines such

Table 5: GLM results for Dimension 3 (Possible Events versus Completed Events) mean differences across disciplinary group, level of study, and genre family

| Independent variable | DF | F-value | Significance | R ² (per cent) |
|----------------------|----|---------|--------------|---------------------------|
| Disciplinary group | 3 | 58.0 | $p < .0001$ | 5.9 |
| Level of study | 3 | 1.6 | n.s. | – |
| Genre family | 12 | 20.7 | $p < .0001$ | 8.3 |

Possible Events



Completed Events

Figure 5: Dimension 3 mean scores for disciplinary groups and academic levels

as Computer Science, Philosophy, and Mathematics, as illustrated in Extracts 5a–c, as these disciplines tend to be associated with ‘timeless’ truths and hypotheses.

Extract 5a

There does not need to be an indication of whether the boiler is on or not, because if either heating or hot water is on then the boiler will have to be on. There is barely the need to have the holiday button and the features associated with it. If a person had lost the manual it would be quite difficult to change any settings.

(Computer Science Critique 0228 g, 22.5 on Factor 3)

Extract 5b

In this essay I will briefly outline the distinction between a belief in objective moral truths and a belief in moral relativity. I will then suggest that even if we accept one or other of these views we are not consequently tied to a

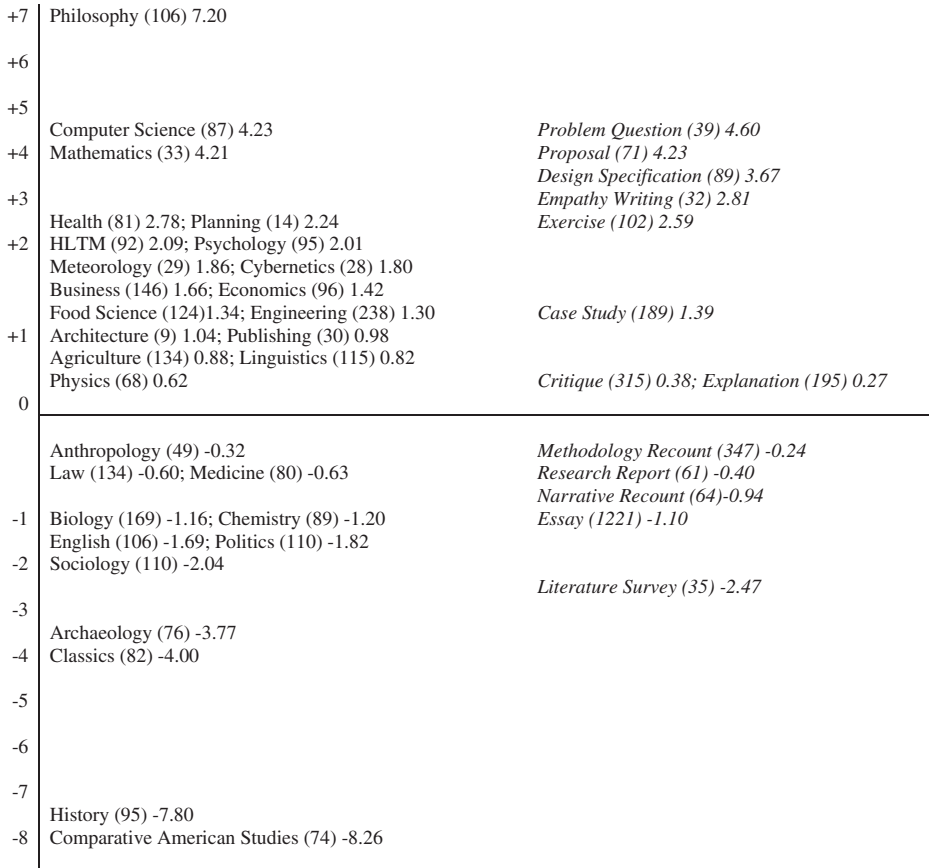
Possible Events**Completed Events**

Figure 6: Dimension 3 mean scores for disciplines and genre families

certain answer to the question of whether morality should be private. If we reject objective moral truths we may still be reluctant to adopt...

(Philosophy Essay 0294h, 13.9 on Factor 3)

Extract 5c

The algebraic mapping is Γ -invariant if and only if for each there exists some nonzero complex number such that In other words, f is Γ -invariant if and only if P and Q both transform by some common factor C under ...

(Mathematics Essay 0049a, 13.5 on Factor 3)

Modal verbs occur throughout these extracts (e.g. *may, should, will, would*), likewise subordinate conditional clauses introduced with *whether, if, and if and only if*. Many of the finite verbs are in the present tense (*accept, does, reject*), and the verb *to be* is also used (*be, is, are*).

Extracts 5a–c show ‘Possible Event’ clusters in Critiques and Essays. They are more likely to occur in Problem Questions, Proposals, and Design Specifications, however, as in Extracts 5d–f:

Extract 5d

This refusal by the school to view evidence submitted by X could give rise to one of the grounds of judicial review⁵, namely the right to a hearing. Even though X had a hearing, if he was unable to represent himself satisfactorily this may be a ground for review.

(Law Problem Question 0143f, 5.9 on Factor 3)

Extract 5e

Patients with diabetes and who require long-term (at least 1 month) total nutritional support as hospital in-patients will be invited to take part in the study. Explicit inclusion and exclusion criteria will clearly define who is eligible to enter the study, see Table 1. (Bury and Mead, 1998) The aim is to recruit 100 participants who will be randomly assigned into, either the intervention group or the control group. The participants will be randomised using a random numbers computer package. This randomisation will reduce bias and decrease the differences between the groups which may otherwise influence the results. (Bowling, 2002)

(Health Proposal 3119c, 9.2 on Factor 3)

Extract 5f

System Constraints

There are a few constraints to the system discussed so far: The system can only send one barcode at a time - It would be good if many barcodes could be scanned and then all sent together at the same time. This would speed up counter transactions however it would add to the complexity of the hardware and software.

(Computer Science Design Specification 0228a, 15.8 on Factor 3)

A more detailed investigation of modals across academic writing could explore the disciplinary patterns suggested by Extracts 5d–f, extending the investigation to include the use of *should* in Case Study recommendations in Business compared to Health (Gardner 2012).

Of the 81 texts with means of less than –10 on Dimension 3, 69 are Humanities Essays, of which 59 are from History, Classics, and Comparative American Studies. We call the negative end of this dimension ‘Completed Events’. It is characterized by simple past tense verbs with the support of the rarely used perfect aspect, features associated with recounts of historical events.

The repeated use of third-person past tense verbs, as in Extract 6a, contrasts with the use of first-person past tense verbs in personal narrative recounts, found in Dimension 2 (Extracts 3a–b) and the reporting of completed empirical research in the passive voice, found in Dimension 1 (Extracts 1a–b).

Extract 6a

Later the war between the Americans and the British became a world war as in 1779 the Spanish and the Dutch entered on the American's side. This caused dismay among the British at home and the large majority of the fleet returned to back home to protect from an invasion by combined French, Spanish and Dutch troops. The British roundly defeated this fleet, mainly comprised of French ships, on the 12th April 1782. Although Britain once again regained control of the seas, the attacks of the American privateers and the intervention of the French fleet came at a crucial time.

(American Studies Essay 0280b, –9.4 on Factor 3)

Completed Events features can also be found in specific sections of texts. Extract 6b makes repeated use of perfect aspect verbs in the conclusion of an Explanation—a pattern that would not be appropriate in the main body of the assignment.

Extract 6b

In conclusion, this essay has looked at the sectors and sub-sectors of the tourism industry and how British Airways fits into them as a company. It has discussed the problems that BA has faced over the last twelve months and the effects that these have had on the airline. Finally, it has looked at what BA is currently doing and is planning to do to rectify these problems to continue to grow and develop as a successful international airline.

(Conclusion section, Explanation, Hospitality, Leisure and Tourism Management 3041b, 1.36 on Factor 3)

As the most heavily weighted features on this dimension are modals, present tense verbs, and past tense verbs (see Table 2), all finite verbs are accounted for, so it is perfectly possible for texts to contain a balance of positive and negative features. This is what happens in Extract 6c, for example, which contains 9 present tenses and a modal (in CAPS) and 10 past tenses (underlined), and comes from an assignment with a ‘neutral’ Factor 3 score close to 0. The extract shows how writers can move between present and past tenses, and thus achieve an overall score close to 0.

Extract 6c

It is now widely accepted that the brain HAS the ability to create false memories. Craik and Tulving showed that items ARE more likely to be remembered if they ARE elaborated on and connecting to similar concepts already held in the brain (1975). Is it possible, then, that the brain CAN also falsely remember an item that is closely related to other items presented

to it? Roediger and McDermott presented participants with a recognition test, where they were read study lists in which all the words ARE related to a semantically associated critical lure word. They were then presented with a test list which comprised of words from the old list, the critical which words and new unrelated words. They were asked to identify from the test list which words they believed were old and which were new. Roediger and McDermott found that critical lures words were incorrectly recognised as old more frequently than the new, unrelated words (1995). In this experiment, we AIM to investigate the effect that the presence of the new, unrelated words HAS on the proportion of times that a critical lure is incorrectly identified as old.

(Psychology Methodology Recount 0037a, 0.28 on Factor 3)

Although the linguistic features in Dimension 3 are very familiar, they are also pervasive, and for this reason, this Dimension is perhaps more difficult to interpret than Dimensions 1 and 2. While Philosophy has the highest mean score for a discipline (at 7.2) at the positive end of Dimension 3, the features that cluster at the positive end express a range of functions across many different disciplines and types of texts. They are used to express logical and future possibilities, as well as to make suggestions and recommendations. In the middle of this dimension are found texts with a balance of present/modal and past tense verbs between sections, as in Extract 6b, or within sections, as in Extract 6c. Texts at the negative end of this dimension include the specific function of recounting past historical events.

4.5 Dimension 4: Informational density

The fourth and final dimension is characterized at its positive pole by long words, nominalizations, attributive adjectives, and abstract nouns (see Table 2). These are all features that are commonly associated with academic writing.

Table 6 summarizes the statistical results for the GLM analysis of disciplinary group, level of study, and genre family as predictors of Dimension 4 scores. The results show that all three independent variables are significant predictors of Dimension 4 scores associated with moderately important mean differences (R^2 values over 10 per cent for disciplinary group and level of study, and R^2 over 5 per cent for genre family).

The fourth dimension is the only one that identifies significant differences between all four levels of study and all disciplinary groups (Figure 7). It is interesting that the sequencing of the disciplinary groups, which was constant across the first three dimensions, has changed, so that Arts and Humanities texts are no longer adjacent to the Social Sciences but are at the opposite extreme, and now next to the Physical Sciences.

Here we see an opposition between the Social Science disciplines of Politics (2.75) and Economics (2.33) and the Arts and Humanities discipline of Classics (-5.23) (Figure 8).

Table 6: GLM results for Dimension 4 (Informational Density) mean differences across disciplinary group, level of study, and genre family

| Independent variable | DF | F-value | Significance | R ² (per cent) |
|----------------------|----|---------|--------------|---------------------------|
| Disciplinary group | 3 | 128.9 | $p < .0001$ | 12.3 |
| Level of study | 3 | 126.5 | $p < .0001$ | 12.1 |
| Genre family | 12 | 12.5 | $p < .0001$ | 5.2 |

Informational Density

| | | |
|----|------------------------------|------------------|
| +2 | Social Sciences A 1.702 | Level 4 A 2.199 |
| +1 | Life Sciences B 0.408 | Level 3 B 0.338 |
| 0 | Physical Sciences C -0.561 | Level 2 C -0.413 |
| -1 | | Level 1 D -1.456 |
| -2 | Arts and Humanities D -1.794 | |

Figure 7: Dimension 4 mean scores for disciplinary groups and academic levels

We have labelled Dimension 4 ‘Informational Density’. It can be associated with the abstract theoretical concepts of the postgraduate (Level 4) Social Sciences, as in this example:

Extract 7a

Each of these can be applied to explaining the EU as a richly diverse and disparate polity. In the context of the EU, the prevailing interpretations are rational choice institutionalism, which regards institutions as a tool of state actors, helping them pursue their predetermined interests in overcoming ‘transaction costs’ and so forth, and historical institutionalism, which is ‘associated with a more generous interpretation of the influence of institutions’ whereby they act as the mediators through which actors interact. Moreover, for historical institutionalists, institutions also have some autonomy of their own, with the ability to shape and influence the behaviour of actors and thus the policy process.

(Level 4 Politics Essay 0255c, 9.2 on Factor 4)

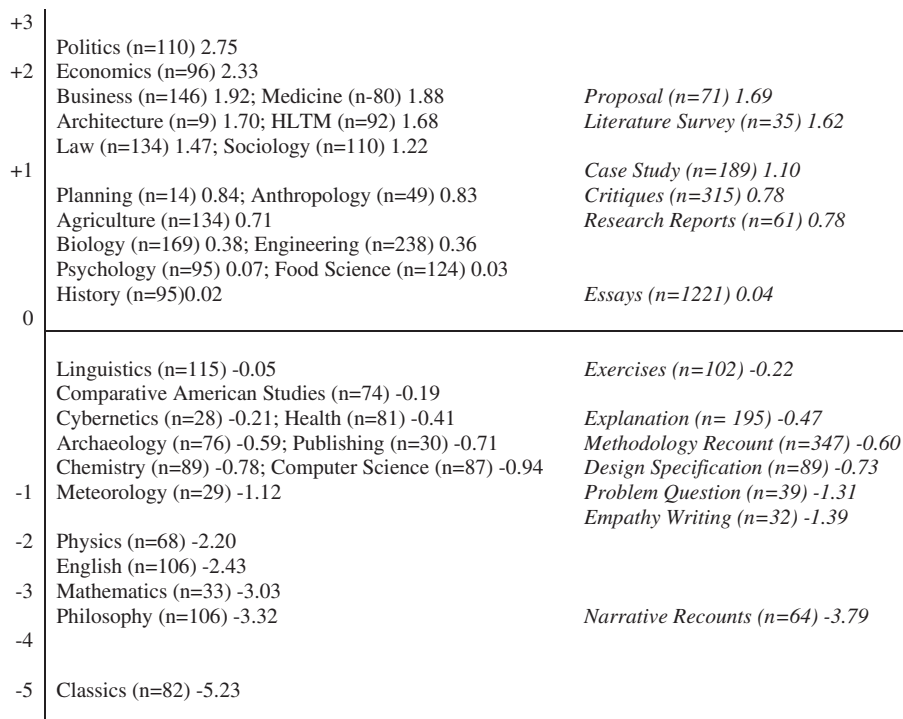
Informational Density

Figure 8: Dimension 4 mean scores for disciplines and genre families

Long words include *explaining* and *predetermined*; nominalizations include *choice*, *transaction*, and *interpretation*; attributive adjectives include *diverse* and *disparate*, *prevailing*, *rational*, and *historical*; and abstract nouns include *the EU*, *polity*, and *context*.

At the negative end of Dimension 4, we see language with a relative absence of long words, nominalizations, and abstract nouns. This can be found in *Empathy Writing*, where students have to write in non-academic genres (such as letters), using everyday language to address an imagined audience while also demonstrating their subject knowledge and expertise.

Extract 8a

Dear Ms Bongey,

I am glad that you found our meeting useful. I feel it is an important meeting for first time authors. However, I'm sorry we did not have time to address all your queries, but I hope this letter will clear up any points.

We have asked you to provide your manuscript on Microsoft Word, as many authors have this program, and the text can be easily imported into InDesign, a program that enables the designer to combine pictures and text, and arrange them on the page in the required format.

(Publishing Empathy Writing 3089a, -6.8 on Factor 4)

As with the other dimensions, there are sections of other more neutral scoring texts which have similarly low informational density. For instance, although an abstract, introduction, or conclusion to a student paper may have densely packed information, this may be ‘unpacked’ in the body of the assignment.

5. CONCLUSIONS

The present analysis has enabled us to identify and characterize with confidence clusters of lexico-grammatical features and their realizations in different writing situations (see Figure 9). When we bring the four dimensions together, a surprising realization is that a different aspect of the writing situation—disciplinary group, genre family, discipline, and level of study—is key to interpreting each dimension. The four disciplinary groups differ most significantly along Dimension 1 (Figure 1), while genre family differences are essential to understanding Dimension 2 (Figure 4). Disciplinary differences come to the fore in Dimension 3 (Figure 6), and the four levels of study only differ significantly along Dimension 4 (Figure 7). This confirms our theory that each of these situational features contributes to a rounded characterization of writing situations.

| Dimension 1 | Dimension 2 | Dimension 3 | Dimension 4 |
|---|--|---|--|
| COMPRESSED PROCEDURAL INFORMATION | PERSONAL STANCE | POSSIBLE EVENTS | INFORMATIONAL DENSITY |
| Physical & Life Science Methodology Recounts & Design Specifications ↑↓ | Narrative Recounts, (Reflective writing) ↑ Biology Explanations | Philosophy & Mathematics ↑↓ History, Classics & Comparative American Studies | Political Science Essays at Level 4 ↑ Empathy Writing (Non-academic language) |
| STANCE TOWARD THE WORK OF OTHERS | | COMPLETED EVENTS | |

Figure 9: Four dimensions of university student writing exemplified

The first dimension differentiates science reports from humanities essays. Most assignments in the Physical Sciences are reports (Methodology Recounts, Design Specifications, or Research Reports), and most Humanities assignments, particularly at the lower levels of undergraduate study, are Essays (Nesi and Gardner 2012: 51–2). The second dimension is perhaps surprising, in that the distinctive features of personal evaluative writing have not traditionally been considered central features of academic English, and yet this emerges from the MD analysis as the second strongest dimension, suggesting that the clustering of features represented here warrants more consideration. Narrative Recounts that include reflective writing are clearly an outlier on this dimension, but here and in previous studies (Nesi and Gardner 2012, Hardy and Römer 2013), personal stance features are also identified as typical of writing in Philosophy. On the other hand the absence of such features is typical of Biology Explanations, which represent a fourth distinctive cluster. The third dimension is locked into verb tense, where the present tenses are found in the timeless truths and hypotheses of Philosophy and Mathematics, while past tenses are prevalent in the narrative evidence of History and Classics. This third dimension is similar in terms of its linguistic characteristics to the fourth dimension in the MICUSP MD analysis (Hardy and Römer 2013; Hardy and Friginal 2016), although Physics and Research Reports in BAWE score close to 0 and are not associated with completed events as they are in MICUSP, suggesting differences in the balance of general statements, hedging, and reporting past events. The extent to which this reflects differences in regional varieties (British versus American English) and/or differences in the composition of the two corpora (e.g. differences in distribution across genres, disciplines, and levels of study) is worthy of investigation. The fourth dimension is important not only in capturing the dense abstraction typically found in upper-level social science theoretical discussions but also in its reminder that students may be required to write assignments quite lacking in such features.

In addition to examining how different assignments are situated along the four dimensions, we can look across the four dimensions to find evidence of two quite different types of stance, and two quite different types of compression or density. Two features of our methodology are relevant here. The differentiation of stance clusters can be partially attributed to the larger tagset which includes more stance and evaluation features, but also to the mapping of the disciplines, levels of study, and genre families onto the dimensions.

Stance in Dimension 1 is used to evaluate the work of others, typically in Essays, while stance and evaluation features in Dimension 2 cluster with first-person pronouns and are typically found in Narrative Recounts and Empathy Writing. Nesi and Gardner (2017) uncover the lexical features that are typical of the two varieties of stance in BAWE, and suggest that the Dimension 2 stance features may be specific to student writing, while those in Dimension 1 are also likely to be found in published professional academic writing. The

distinction between these two stance clusters should prove helpful in teaching, as contrastive lists of stance features can be extracted from the BAWE corpus (Nesi and Gardner 2017) to understand how the language of reflection and self-evaluation differs from the language of texts evaluating the work of others.

The same is true of the density features. The density at the positive end of Dimension 1, compressed procedural information, is most typical of scientific reports, while the density of Dimension 4 is most typical of the Social Sciences. The work of Staples *et al.* (2016) is of interest here, as it examines complexity and student progression in a sub-corpus of BAWE, finding that phrasal complexity, characterized by nominal modification and elaboration, increases with advances in academic level. It is of course possible for both types of density to occur in the same texts, but the MD analysis suggests this is not usually the case. Thus the evidence indicates that density is best taught in relation to these two distinct clusters, as appropriate to students' learning needs.

These notes on stance and density illustrate how this new MD analysis can inform further investigation of clusters of linguistic features in student writing. By bringing together multiple situational perspectives to interpret the dimensions, we have been able to present an integrated picture (Figure 9) that makes sense of the dimensions in relation to the academic situations of the texts and thus lends itself more easily than previous single-perspective interpretations to further research and teaching applications. Writing programmes that focus, sometimes exclusively, on Essays will now be able to differentiate with confidence those features of upper-level informationally dense essays in the Social Sciences from those that are prevalent in lower-level humanities essays that express opinions on the work of others. The extracts in this article can be used as an exemplification of this. A general EAP programme may also wish to introduce other situational perspectives, such as procedural report writing (Dimension 1), reflective writing (Dimension 2), explanations (Dimension 2), and more, as we suggest these too would be part of a common core for multi-disciplinary general academic English.

NOTES

1 See www.coventry.ac.uk/bawe. The BAWE corpus was developed at the Universities of Warwick, Reading, and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (formerly of the Department of Applied Linguistics, Reading), and

Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800).

2 The level of nine Social Sciences assignments was not specified in the corpus metadata when the MD analysis was conducted.

3 The plan was to collect 32 assignments from each level of study in each

- discipline from four different modules. There are more in some multidisciplines such as Engineering, and fewer, for instance, where students wrote exams or produced creative artefacts.
- 4 This percentage is similar to the rates for other factor analyses of register variation; for example, the 7-factor solution in [Biber 1988](#) accounted for 51.9 per cent of the shared variance; the 4-factor solution in [Biber, Gray, and Staples 2016](#) accounted for 44 per cent of the shared variance; and the 10-factor solution in [Biber and Egbert 2016](#) accounted for 42.7 per cent of the shared variance.
 - 5 A spreadsheet with this metadata and other information about each assignment (module, grade, length, number of tables, etc.) is available with the corpus from the Oxford Text Archive, resource number 2539 (<http://ota.ahds.ac.uk/headers/2539.xml>), and via the BAWE website (www.coventry.ac.uk/BAWE).
 - 6 The BAWE corpus can be freely searched through the SketchEngine UK open-access site <https://the.sketchengine.co.uk/open>.

SUPPLEMENTARY DATA

Supplementary material is available at *Applied Linguistics* online.

Conflict of interest statement. None declared.

REFERENCES

- Biber, D.** 1988. *Variation across Speech and Writing*. Cambridge University Press.
- Biber, D.** 2006. *University Language: A Corpus-Based Study of Spoken and Written Registers*. John Benjamins.
- Biber, D.** and **J. Egbert.** 2016. 'Register variation on the searchable web: A multi-dimensional analysis,' *Journal of English Linguistics* 44: 95–137. DOI: 10.1177/0075424216628955.
- Biber, D., S. Conrad, R. Reppen, P. Byrd, and M. Helt.** 2002. 'Speaking and writing in the university: A multidimensional comparison,' *TESOL Quarterly* 36: 9–48. DOI:10.2307/3588359.
- Biber, D., B. Gray, and S. Staples.** 2016. 'Predicting patterns of grammatical complexity across language exam task types and proficiency levels,' *Applied Linguistics* 37: 639–68. <https://doi.org/10.1093/applin/amu059>.
- Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan.** 1999. *Longman Grammar of Spoken and Written English*. Pearson Education.
- Crosthwaite, P.** 2016. 'A longitudinal multidimensional analysis of EAP writing: Determining EAP course effectiveness,' *Journal of English for Academic Purposes* 22: 166–78. DOI: 10.1016/j.jeap.2016.04.005.
- de Chazal, E.** 2013. 'The general-specific debate in EAP: Which case is the most convincing for most contexts?,' *Journal of Second Language Teaching and Research* 2: 135–48.
- Durrant, P.** 2017. 'Lexical bundles and disciplinary variation in university students' writing: Mapping the territories,' *Applied Linguistics* 38: 165–93. DOI: 10.1093/applin/amv011.
- Egbert, J. and D. Biber.** 2017. 'Do all roads lead to Rome? Modeling register variation with factor analysis and discriminant analysis,' *Corpus Linguistics and Linguistic Theory* Available at <https://doi.org/10.1515/clt-2016-0016>.
- Ferris, D.** 2001. 'Teaching writing for academic purposes' in J. Flowerdew and M. Peacock (eds): *Research Perspectives on English for Academic Purposes*. Cambridge University Press, pp. 298–314.
- Flowerdew, J.** 2016. 'English for Specific Academic Purposes (ESAP) writing: Making the case,' *Writing and Pedagogy* 8: 1–32. DOI: 10.1558/wap.v8i1.30051.

- Friginal, E.** 2013. 'Twenty-five years of Biber's multi-dimensional analysis: Introduction to the special issue and an interview with Douglas Biber,' *Corpora* 8: 137–52. DOI: 10.3366/cor.2013.0038.
- Gardner, S.** 2012. 'A pedagogic and professional case study genre and register continuum in business and in medicine,' *Journal of Applied Linguistics and Professional Practice* 9: 13–35. DOI: 10.1558/japl.v9i1.13.
- Gardner, S.** 2016. 'A genre-instantiation approach to teaching English for specific academic purposes: Student writing in business, economics and engineering,' *Writing and Pedagogy* 8: 149–4. DOI: 10.1558/wap.v8i1.27934.
- Gardner, S.** and **H. Nesi.** 2013. 'A classification of genre families in university student writing,' *Applied Linguistics* 34: 25–52. DOI:10.1093/applin/ams024.
- Hardy, J.** and **E. Friginal.** 2016. 'Genre variation in student writing: A multi-dimensional analysis,' *Journal of English for Academic Purposes* 22: 119–31. DOI: 10.1016/j.jeap.2016.03.0.
- Hardy, J.** and **U. Römer.** 2013. 'Revealing disciplinary variation in student writing: A Multi-Dimensional Analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP),' *Corpora* 8: 183–207. DOI: 10.3366/cor.2013.0040.
- Hyland, K.** 2002. 'Specificity revisited: How far should we go?,' *English for Specific Purposes* 21: 385–95. DOI: 10.1016/S0889-4906(01)00028-X.
- Issitt, S.** 2017. 'Evaluating the impact of a pre-professional English for academic purposes programme: A corpus based study,' Ph.D. Thesis, University of Birmingham.
- Johns, A. M.** 2008. 'Genre awareness for the novice student: An ongoing quest,' *Language Teaching* 41: 237–52. DOI.org/10.1017/S0261444807004892.
- Nesi, H.** and **S. Gardner.** 2012. *Genres Across the Disciplines: Student Writing in Higher Education.* Cambridge University Press.
- Nesi, H.** and **S. Gardner.** 2017. Stance in the BAWE Corpus: New Revelations from Multidimensional Analysis. Corpus Linguistics 2017, University of Birmingham, 25–28 July 2017. Available at: <http://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2017/general/paper257.pdf>.
- Nesi, H.** and **E. Moreton.** 2012. 'EFL/ESL writers and the use of shell nouns' in R. Tang (ed.): *Academic Writing in a Second or Foreign Language: Issues and Challenges Facing ESL/EFL Academic Writers in Higher Education Contexts.* Continuum, pp. 126–45.
- Staples, S., J. Egbert, D. Biber, and B. Gray.** 2016. 'Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre,' *Written Communication* 33: 149–83. DOI: 10.1177/0741088316631527.

NOTES ON CONTRIBUTORS

Sheena Gardner is Professor of Applied Linguistics at Coventry University. Her research uses functional, corpus, and genre-based approaches to investigate the nature and use of academic English in educational contexts. Her publications include 'Genres across the Disciplines' with Hilary Nesi (Cambridge 2012), 'Multilingualism, Discourse and Ethnography' with Marilyn Martin-Jones (Routledge 2012), and 'Systemic Functional Linguistics in the Digital Age' with Siân Alsop (Equinox 2016). *Address for correspondence:* Sheena Gardner, School of Humanities, Coventry University, Coventry, CV1 5FB, UK. <sheena.gardner@coventry.ac.uk>.

Hilary Nesi is Professor in English Language at Coventry University. Her research activities concern the discourse of English for academic purposes and the design and use of dictionaries and reference tools in academic contexts. She was principal investigator for the projects to create the BASE corpus of British Academic Spoken English and the BAWE corpus of British Academic Written English. She is the co-author of 'Genres across the Disciplines: Student writing in higher education' (Cambridge University Press 2012).

Douglas Biber is Regents' Professor of Applied Linguistics at Northern Arizona University. His research on corpus linguistics, English grammar, and register variation (in English and cross-linguistic, synchronic, and diachronic) has resulted in over 220 research articles, 8 edited books, and 15 authored books and monographs.