A multidimensional analysis of student writing across levels and disciplines

Hilary Nesi

Coventry University

h.nesi@coventry.ac.uk

The British Academic Written English (BAWE) corpus, created as part of an investigation of student writing in British higher education[1], contains 2,761 student assignments, produced and assessed for university degree coursework and fairly evenly distributed across 35 disciplines and four levels of study. About half the assignments in the corpus were graded at a level equivalent to 'distinction' (70% or above), and half at a level equivalent to 'merit' (between 60% and 69%). The majority (1,953) were written by L1 speakers of English. As some assignments contained multiple pieces of coursework, the total number of separate texts in the corpus is 2,897. A fuller account of the corpus design is provided in Nesi (2008) and Alsop and Nesi (forthcoming).

Our ESRC-funded project included an examination of departmental documentation, interviews with tutors and students and genre analysis, resulting in the categorisation of texts into 13 broad genre families (see, for example, Gardner 2008). A further investigatory strand involved multidimensional analysis of the corpus, with the support of Douglas Biber and his team at Northern Arizona University. Biber considers linguistic co-occurrence to be central to the study of register, and assumes that if a group of linguistic features co-occur with high frequency they must share an underlying communicative function. His approach has evolved from early studies such as that of Chafe (1982) which posited sets of linguistic features likely to co-occur in specified groups of texts. Biber's method is more complex than this, however, in that it counts the frequency of specified linguistic features and their relative distribution in a corpus of texts, and draws on these findings to ascertain the extent to which groups of texts are similar or different from each other.

In preparation for the analysis, the BAWE corpus was tagged for 67 linguistic features, grouped into 16 grammatical/functional categories:

1. Tense and aspect markers
2. Place and time adverbials
3. Pronouns and pro-verbs
4. Questions
5. Nominal forms
6. Passives
7. Stative forms
8. Subordination features
9. Prepositional phrases, adjectives and adverbs
10. Lexical specificity
11. Lexical classses
12. Modals
13. Specialized verb classes
14. Reduced forms and dispreferred structures
15. Coordination
16. Negation

Patterns of co-occurrence of these features were then mapped across disciplinary groups, levels of study, and genres in the BAWE corpus, using factor analysis to identify co-occurrence patterns (called 'dimensions').

This part of the multidimensional analysis was purely quantitative, but the linking of dimensions to functions is open to interpretation. The BAWE corpus was analysed both in terms of the dimensions originally identified by Biber (1988), and in terms of new dimensions of variation specific to the corpus. The findings presented in this

---

paper, however, only relate to the following 1988 dimensions, already associated with communicative functions in a number of prior studies (for example Biber 1988, Conrad & Biber 2001):

Dimension 1: *Involved v. Informational*, with 1st and 2nd person pronouns and Wh questions as positive features.

Dimension 2: *Narrative v. Non-narrative*, with past tense verbs, 3rd person pronouns, public verbs such as "said", present participial clauses and perfect aspect verbs as positive features.

Dimension 3: *Explicit v. Situation-dependent*, with Wh relative clauses, phrasal coordination and nominalizations as positive features.

Dimension 4: *Persuasive*, with modals, suasive verbs and conditional subordination as positive features.

Dimension 5: *Abstract v. Non-abstract*, with passives, conjuncts, and passive adverbial and postnominal clauses as positive features.

Table One shows dimension scores across levels. Levels 1, 2 and 3 represent the first, second and final years of undergraduate study, and Level 4 represents taught Masters programmes.

Table One: Scores across dimensions and levels of study

	Involved[1]	Narrative[2]	Explicit[3]	Persuasive[4]	Abstract[5]
Level 1	-12.8	-2.7	5.1	-1.4	5.9
Level 2	-13.9	-2.8	5.6	-1.5	6.2
Level 3	-14.8	-3.0	5.7	-1.4	6.4
Level 4	-17.3	-3.2	6.4	-2.0	5.4

Scores for Dimensions 1 and 2 were uniformly negative. Dimension 1 differentiates between oral (verbal) and written (nominal) styles. In Biber's 1988 analysis the most "involved" component of his corpus, conversation, scored +35, while general academic (published) prose scored -15. BAWE writing had similarly high negative scores with progressively fewer involved features at each level, indicating less focus on the author and greater focus on the object of study. On Dimension 2, romance fiction scored 7 in Biber's 1988 analysis, and general academic writing scored between -2 and -3. BAWE writing had similar scores, at each level placing a greater emphasis on the building of arguments rather than narrating.

On Dimension 3 BAWE writing had high positive scores, at each level becoming less situation-dependent and more "elaborated", with fewer references to the real world outside the text. Scores for Dimension 5 also rise across Levels 1 to 3, but fall slightly at Level 4. This might be because some of the more professional genres produced at Masters level require greater emphasis on the agent. In Biber's 1988 analysis general academic writing scored between 4 and 5 for Dimension 3, and between 5 and 6 for Dimension 5.

The greatest difference between BAWE writing and general academic writing as analysed by Biber (1988) is on Dimension 4. Biber's general academic writing texts were unmarked for this dimension (i.e they had a score of 0); the negative BAWE scores indicate that students present information in a more factual manner, with fewer overt attempts at persuasion, perhaps because they have less need to convince their readers of the centrality of their topics or of the validity and significance of their academic activities.

Unsurprisingly, scores for Dimension 4 varied according to genre, as shown in Table Two. Case studies, empathy writing and narratives had the highest scores on this dimension, whilst reports and literature surveys were the least persuasive.

Table Two: Scores across dimensions and genre families

	Involved[1]	Narrative[2]	Explicit[3]	Persuasive[4]	Abstract[5]
Essay	-14.3	-2.5	6.2	-1.8	5.9
Methodology Recount	-15.8	-3.7	4.5	-2.5	7.3
Critique	-14.8	-3.1	6.0	-1.6	6.4
Explanation	-15.4	-3.6	5.0	-2.3	5.8
Case Study	-16.4	-2.9	5.8	-0.5	4.5
Exercise	-12.1	-3.9	4.6	-1.3	5.7

Design Specs	-13.1	-4.0	4.1	0.7	6.8
Proposal	-16.4	-3.8	6.3	1.3	4.8
Narrative Recount	-4.8	-1.1	3.8	-0.7	4.0
Research Report	-16.2	-3.1	5.5	-2.4	7.2
Problem Question	-12.0	-2.8	5.2	1.6	6.4
Literature Survey	-17.9	-2.6	6.3	-3.4	5.0
Empathy Writing	-11.5	-2.7	4.5	0.8	4.5

The Narrative Recount was also fund to be an outlier on the other dimensions, being the most involved, narrative-like and situation-dependent of the genres, and the least abstract. Narrative Recount was one of the least common genres in the corpus (75 examples, with roughly equal distribution across the four disciplinary groupings). Included in this genre family are biographies, short stories, reflective recounts and simple recounts of chronologically-ordered events. A contrasting genre family would be the critique, for example, which places less emphasis on the author, the agent and the outside world, and more emphasis on argument.

Table Three shows dimension scores across disciplinary groups. Different colour shades within each dimension indicate statistically significant differences (p = <0001). Arts and Humanities writing is the most narrative-like and the least persuasive, Social Sciences writing is the most elaborated, Life Sciences writing is the most nominal and "object" as opposed to "author" centred, and Physical Sciences writing the most abstract and least narrative-like.

Table Three: Scores across dimensions and disciplinary groups

	Involved[1]	Narrative[2]	Explicit [3]	Persuasive[4]	Abstract[5]
Arts and Humanities	-13-4	-2.1	5.7	-2.3	5.5
Social Sciences	-15.3	-3.0	6.5	-1.3	6.2
Life Sciences	-15.6	-3.0	5.7	-1.5	5.7
Physical Sciences	-13-4	-3.7	4.4	-1.2	6.5

Figure One shows the distribution of genre families in the corpus, most of which are represented across all the disciplinary groupings. Figures Two and Three, however, show that the overwhelming majority of Arts and Humanities texts and slightly more than half the Social Sciences texts belong in the 'essay' genre family, while Figures Four and Five show that texts from the Life Sciences and particularly the Physical Sciences are more evenly spread across a wider range of genres.
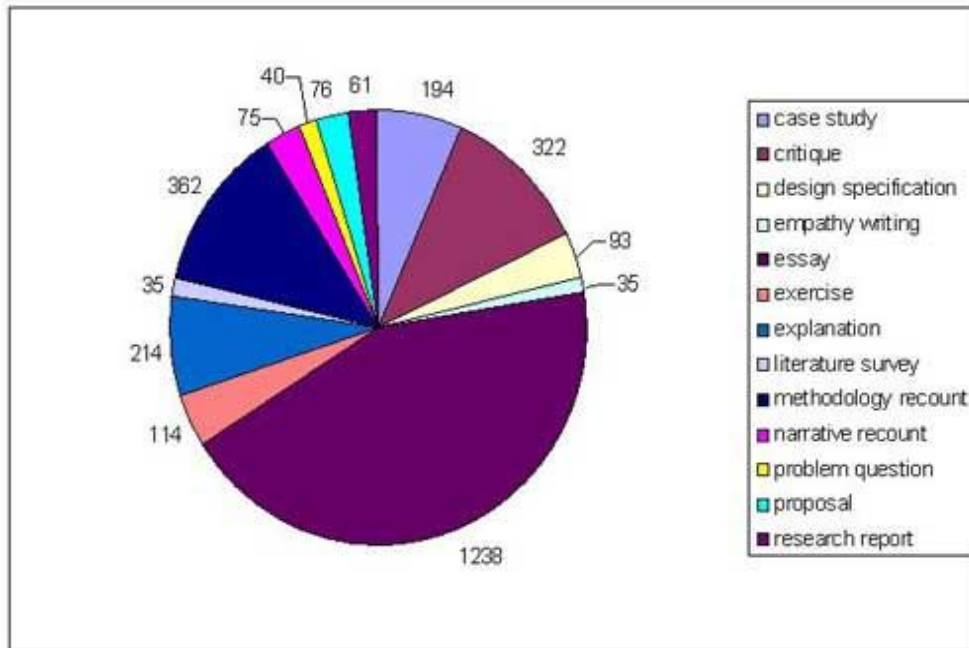
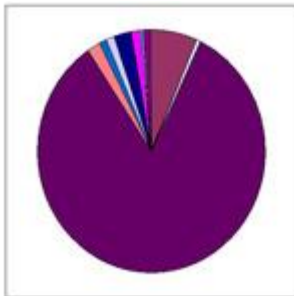**Figure One: Distribution of genre families**

Figure Two: Arts and Humanities
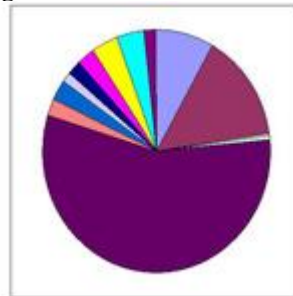

Figure Three: Social Sciences
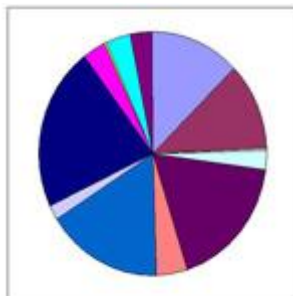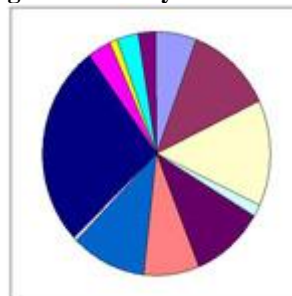

Figure Four: Life Sciences

Figure Five: Physical Sciences

Science students may produce less writing in the course of their studies, but it would appear that they need to write in a greater variety of registers, and therefore draw on a wider range of linguistic devices to achieve appropriate communicative effects.

References

Sian Alsop & Hilary Nesi. Forthcoming. Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora* vol 4 (1)

Douglas Biber. 1988. *Variation across Speech and Writing*. Cambridge University Press: Cambridge, UK.

Wallace Chafe. 1982. Integration and involvement in speaking, writing and oral literature. In Deborah Tannen (ed.) *Spoken and Written Language: Exploring Orality and Literacy*. Ablex: Norwood, NJ, USA. pp35-54.

Susan Conrad. & Douglas Biber. 2001. Multi-dimensional methodology and the dimensions of register variation in English. In Susan Conrad. & Douglas Biber (eds.) *Variation in English: multi-dimensional studies*. Pearson: Harlow, UK. pp13-42.

Sheena Gardner. 2008. Integrating ethnographic, multidimensional, corpus linguistic and systemic functional approaches to genre description : an illustration through university history and engineering assignments. In Erich Steiner &Stella Neumann (eds.) *Proceedings of the 19th European Systemic Functional Linguistics Conference and Workshop 23rd - 25th July 2007*. Saarbrücken, Germany. Available online at http://scidok.sulb.uni-saarland.de/frontdoor.php?source_opus=1688

Hilary Nesi. 2008. BAWE: An introduction ot a new resource. In A. Frankenberg-Garcia, T. Rkibi, M. Braga da Cruz, R. Carvalho, C. Direito & D. Santos-Rosa (eds) *Proceedings of the 8th Teaching and Language Corpora Conference*. ISLA: Lisbon, Portugal: pp239-246.