Verbal and mental processes in academic disciplines

Nesi, H. and Holmes , J.

Author post-print (accepted) deposited in CURVE August 2012

CURVE is the Institutional Repository for Coventry University

http://curve.coventry.ac.uk/open

Chapter 3

Verbal and Mental Processes in Academic Disciplines

Jasper Holmes and Hilary Nesi

1   Introduction

Interview surveys have shown that university lecturers from different disciplines look for different attributes in the writing of their students, and describe and evaluate academic activities in discipline-specific ways (Lea & Street, 2000; Nesi & Gardner, 2006). This chapter aims to identify some important disciplinary differences at clause level in student assignments which have been awarded good grades, and have therefore met lecturers' expectations, at least to some extent. Using keyword analysis, which is a corpus linguistic technique, followed by closer analysis of clause meanings in context, it considers the ways in which student writers position themselves as members of their discourse community, as purveyors of hard or soft and pure or applied knowledge. The identification of discipline-specific clausal features helps to further our understanding of the way disciplinary knowledge is conceptualized and expressed, and may also inform the design of discipline-specific writing programmes for novice academic writers.

2   Process Types in Academic Writing

According to the Systemic Functional Linguistics (SFL) model (Halliday, 1994), the world of experience has three forms of representation, each realized by two Process types. 'Outer' experience is represented as actions or events, and is realized in Material and Behavioural Processes. 'Inner' experience is represented as reaction and reflection, and is realized in Mental and Verbal Processes. The third form of representation, 'generalization', is the relationship between these experiences, and is realized in Relational and Existential Processes. Mental and Verbal Processes usually involve at least one animate participant, a 'Senser' or a 'Sayer' and human agency in these types of Processes is usually recoverable from the text, even if it is disguised by the use of metaphor, as in 'This dissertation considers an alternative view' (John, this volume), where the Mental Process is attributable to the writer. In Material Processes, on the other hand, the 'Actor'

may be animate, inanimate or abstract and agency may be hidden by means of a passive construction, while in Relational Processes human activity can be entirely disguised through nominalization (Halliday, 1994: 352–353), which metaphorically transforms congruently worded clauses such as 'he argued' or 'he composed a piece of music' into nominal groups such as 'the argument' or 'the musical composition', which are then related to new entities in clauses such as 'the argument is valid' or 'the composition was in binary form'.

Transitivity mechanisms allow writers to adjust their presence or distance in the text according to their communicative goals, as noted by John (this volume) who discusses the effect of Process on writer visibility, and Tang (this volume) who examines writers' means of self-expression. Because different transitivity choices achieve different communicative effects, we expect the distribution of Processes to vary according to domain, genre and context. John (this volume), for example, found Material and Relational Processes to be the most frequent in the methodology sections of MA dissertations in the field of Applied Linguistics, while Martinez (2001) found that Material Processes dominated in the Method sections of scientific research articles whereas Relational Processes dominated in the Results and Discussion sections. Love (1993) noted the high frequency of Relational and Existential Processes in geology textbooks, while Babaii and Ansary (2005) reported more Relational and Existential Process types in physics book reviews than in sociology and literature, and more Material Process types in sociology and literature book reviews than in physics.

This study concentrates on students' lexical choices relating to Verbal and Mental Processes. As these are processes which involve human agency it was assumed that they could shed some light on the students' sense of scholarly identity. Martinez (2001: 241) found that the Verbal and Mental Processes in research article introductions helped writers achieve their goal of 'contextualizing . . . previous research, reviewing theories, ideas and previous findings'. In John's examples of MA dissertations (this volume) Mental Processes were often found to involve subjective interpretation and result in greater visibility for the writer, whereas the Sayer in Verbal Processes was usually a cited authority behind which the student writer could hide. It is reasonable to assume that Verbal and Mental Processes will play similar roles in the types of university assignment we will examine, given that students often have to introduce and review others' theories and findings, even if, prior to the dissertation, their ultimate objective is often to demonstrate their acquisition of academic knowledge, rather than to occupy a research niche in the manner described by Swales (1990).

3   Classification of Academic Knowledge

A standard typology for the classification of academic knowledge distinguishes between 'pure' or 'applied' and 'hard' or 'soft' disciplines (Becher & Trowler,

2001). In general, the natural sciences and mathematics are classed as hard-pure, the science-based professions such as engineering are classed as hard-applied, the humanities are classed as soft-pure and the social professions such as education and law are classed as soft-applied.

Knowledge in hard-pure disciplines is quantitative and tends to develop steadily and cumulatively; new findings derive linearly from an existing body of knowledge. Soft-pure knowledge, on the other hand, is qualitative and new developments in these disciplines tend to derive from the combination and recombination of existing work and results (Becher, 1989: 13; Becher & Trowler, 2001: 39). This accords with Hyland's (2000: 37–40) claim that writers in soft disciplines use more, and more varied, reporting verbs than writers in hard disciplines, because they need to support their arguments with references to other researchers whose works are known and respected. In the hard disciplines causal and logical relationships are relatively easily established from observations and quantitative data, and so there is less need to cite the opinions of others. Hyland (2000: 28) also finds that writers in the soft disciplines use more 'discourse act' reporting verbs such as ASCRIBE, DISCUSS and STATE, whereas writers in the hard disciplines prefer to use 'research act' reporting verbs such as OBSERVE, DISCOVER and CALCULATE.

The distinction between pure and applied disciplines depends on the extent to which the discipline is concerned with theory, or practice. Applied knowledge builds on theory, but is ultimately practical; it is concerned with 'knowing how' as opposed to 'knowing that' (Becher, 1989: 15). Scholars in the science-based professions aim to produce products and techniques, and those in the social professions aim to produce protocols and procedures. Applied methods operate in the real world rather than under experimental conditions where variables can be carefully controlled, and for this reason they always entail some qualitative judgement, even when the discipline is science based.

All this suggests that Verbal and Mental Processes across the discipline types are likely to vary, especially in the pure fields where the distinction between soft and hard knowledge is greatest. Students' use of the lexical items associated with these Processes should indicate the extent to which they are positioned within one of the four quadrants into which hard, soft, pure and applied disciplines fall.

4   Keywords and Keyness

The four disciplinary areas can be examined using WordSmith Tools software (Scott, 2004), which enables identification of keywords (KWs). Scott and Tribble (2006: 56) explain 'keyness' as 'what the text "boils down to" . . . once we have steamed off the verbiage, the adornment, the blah blah blah'. The technique for identification of KWs described by Scott (1997) and Scott and

Tribble (2006) requires both a reference corpus and one, or more than one, study corpus, often a subset of the larger reference corpus. Keyness is obtained by statistical comparison of word frequencies in these two types of corpora; the standard default setting for WordSmith Tools requires a minimum of three occurrences of each keyword in the study corpus, and a Log Likelihood statistic (Dunning, 1993) with a p value of 0.000001. A word is deemed to be positively key if its frequency in the study corpus is unusually high and negatively key if its frequency in the study corpus is unusually low.

Scott and Tribble (2006: 59–69) illustrate the process of keyword analysis using Romeo and Juliet as a study corpus and the entire collection of Shakespeare's plays as a reference corpus. Their analysis shows that KWs which occur significantly more frequently in the study corpus than in the reference corpus reflect important themes specific to Romeo and Juliet. A keyword analysis, then, provides an opportunity to examine the typical uses of lexical items associated with Verbal and Mental Processes.

5   Method

For this study our reference corpus was the entire British Academic Written English (BAWE) corpus,[1] a 6.5 million word collection of student assignments which have been awarded high grades when assessed as part of degree coursework at three British universities. The corpus holdings are distributed fairly evenly over four disciplinary groups (Arts and Humanities, Life Sciences, Physical Sciences, Social Sciences) and four levels of study (from first year undergraduate to taught masters level) (see Alsop & Nesi, 2009; Ebeling & Heuboeck, 2007; Gardner & Holmes, this volume; Nesi, 2008). The main study corpora were two subsets of the BAWE corpus: 96 history assignments (309,761 words) and 68 physics assignments (196,487 words). At a second stage of analysis, KWs in these two 'pure' disciplines were compared with further subsets of the BAWE corpus in 'applied' disciplines: engineering (238 assignments, 599,687 words), medicine (80 assignments, 214,226 words), and hospitality, leisure and tourism management (HLTM) (93 assignments, 296,709 words). Becher and Trowler (2001: 39) note that it is not always straightforward to determine a priori whether a particular discipline is pure or applied, since different researchers and different university departments give different emphasis to different aspects of their field. However, for the purposes of this study HLTM, Medicine and Engineering were selected as representative of the applied disciplines, based on the kinds of assignments submitted to the BAWE corpus by students in these disciplines, and also on interview data gathered in the early stages of the project 'An Investigation of Genres of Assessed Writing in British Higher Education' (see Nesi & Gardner, 2006).

While undertaking our analysis, we took into account that the term 'Process' has two senses in SFL: it can refer to what goes on in the whole clause, and it can refer to 'that part of the proposition encoded in the Verbal Group' (Bloor & Bloor, 1995: 110). In our study we considered both the Verbal Group and all clausal elements that suggested the Processes of 'saying' and of 'internal cognition'. Berber Sardinha's (2000) formula for predicting the number of KWs in a corpus was found to yield around 1,500 KWs for our history corpus alone, and over 500 KWs even if we used a very stringent level of significance (p<0.0000000001). This number was clearly too great for the detailed analysis we had in mind, and we therefore restricted our data to those 'saying' and ~~of~~ 'internal cognition' words identified in WordNet (Fellbaum, 1998).

First, a list of word forms was extracted from WordNet containing all the hyponyms and troponyms of relevant senses such as 'argument', 'belief', 'cognition', 'opinion', 'say', 'state' and 'tell'. This initial list was shortened by removing some words that are not used epistemologically (e.g. BARK, GROWL, HISS) and some which did not appear in our chosen corpus (e.g. AVER, DISAVOW, OPINE). The reduced list of 122 words was then expanded by adding all inflectional forms for each lemma, resulting in the word forms shown in Table 3.1.

Table 3.1   Verbal and mental process words, including inflected forms

abstract	confirmation	explanation/s
account/s	consider/s/ing/ed	fact/s
acknowledg/e/es/ing/ed	consideration/s	falsification
add/s/ing/ed	contradict/s/ing/ed	falsif/y/ies/ying/ied
advance/es/ing/ed	contradiction/s	find/s/ing/found
announc/e/es/ing/ed	criticis/e/es/ing/ed	findings
announcement/s	criticism/s	grounds
answer/s/ing/ed	criticiz/e/es/ing/ed	identif/y/ies/ying/ied
apparent/ly	declaration/s	indicat/e/es/ing/ed
argu/e/es/ing/ed	declar/e/es/ing/ed	indication/s
argument/s	defend/s/ing/ed	infer/s/ring/red
ascertain/s/ing/ed	defense	inference/s
assert/s/ing/ed	definitely	information
assertion/s	demonstrat/e/es/ing/ed	justif/y/ies/ying/ied
assum/e/es/ing/ed	demonstration/s	know/s/ing/n/knew
assumption/s	detect~~/ed~~/s/ing/ed	knowledge
bas/e/es/ing/ed	determin/e/es/ing/ed	not/e/es/ing/ed
belief/s	discover~~/ed~~/s/ing/ed	notic/e/es/ing/ed
believ/e/es/ing/ed	discovery/discoveries	observ/e/es/ing/ed
calculat/e/es/ing/ed	disproof	opinion/s
calculation/s	disprov/e/es/ing/ed	premise/s
claim/s/ing/ed	distinctly	present/s/ing/ed
clear/ly	evidence	proof/s
cognition	evident/ly	proposal/s
confirm/s/ing/ed	explain/s/ing/ed	propos/e/es/ing/ed

Table 3.1　Continued

proposition/s	representation/s	support/s/ing/ed
prove/e/en/es/ing/ed	representative	suppos/e/es/ing/ed
rationalis/e/es/ing/ed	respond/s/ing/ed	tell/s/ing/told
rationaliz/e/es/ing/ed	response/s	theor/y/ies
realisation	retort/s/ing/ed	think/s/ing/thought
realization	say/s/ing/said	thoughts
realis/e/es/ing/ed	see/s/ing/saw	unambiguous/ly
realiz/e/es/ing/ed	show/s/ing/n/ed	uncertain
reason/s/ing/ed	stat/e/es/ing/ed	unclear
recognis/e/es/ing/ed	statement/s	understand/s/ing/~stood
recognition	stipulat/e/es/ing/ed	undoubtedly
recogniz/e/es/ing/ed	stipulation/s	unlikely
repl/y/ies/ying/ied	suggest/s/ing/ed	verification
represent/s/ing/ed	suggestion/s	verif/y/ies/ying/ied

Using WordSmith Tools Version 4.0 (Scott, 2004), we then proceeded to identify from this list those words which were 'key' in our corpus of student writing.

The findings and discussion are presented below in two parts, first focusing on the pure disciplines of history and physics, and then contrasting these with the findings for the applied disciplines.

6　Findings and Discussion

6.1　Pure disciplines: History and Physics

Table 3.2 gives the KWs that occur with significantly different frequencies in the history and physics corpora as opposed to the reference corpus (p<0.00001).

Some words at lower keyness values not listed in Table 3.2 displayed similar frequency patterns across both study corpora. DISCOVERY, for example, was positively key in history (10.57) and physics (5.35), while TELL, CONSIDERATION and CONSIDER were negatively key in history (–9.61, –15.87, and –16.41) and physics (–5.91, –8.19 and –6.13). The most striking point about the lists in Table 3.2, however, is that so many of the positive KWs in history are negatively key in physics, and vice versa. ARGUE, BELIEF, SUPPORT and CLAIM were significantly more common in the history assignments but significantly less common in the physics assignments than in the BAWE corpus as a whole. DETERMINE, KNOW, CALCULATE, FIND and SHOW behaved in the opposite way.

It should be noted that our methods of lemmatizing and of retrieving KWs concealed the distinction between homonyms. For example the method did not distinguish between nouns and verbs such as CLAIM/n and CLAIM/v, or, more importantly, the derivationally unrelated STATE/n and STATE/v. Moreover,

Table 3.2 Verbal and mental process keywords in History and Physics

History		Physics	
Keyword	**Keyness**	**Keyword**	**Keyness**
Positive keywords			
		CALCULATE	350.5819
		THEORY	318.5182
ARGUE	257.5518		
		DETECT	193.5404
		FIND	185.4318
		KNOW	179.1548
		OBSERVE	178.9922
		SHOW	127.7502
		DETERMINE	89.02161
BELIEF	75.51513		
ASSERT	58.30611		
SUPPORT	56.45473		
		DISCOVER	55.50309
STATE	46.53033		
BELIEVE	41.65763		
		ABSTRACT	30.03328
CLAIM	27.87113		
		EXPLAIN	24.58973
PROPOSAL	23.85093		
CRITICISM	22.27682		
Negative keywords			
NOTICE	−19.8002		
		OPINION	−20.3714
		ARGUMENT	−20.8476
IDENTIFY	−29.2359		
		SUGGEST	−32.9894
DETERMINE	−34.0833		
ADD	−34.1799		
		RESPONSE	−37.5827
		BELIEF	−39.5892
BASE	−40.8858		
KNOWLEDGE	−43.1839		
CALCULATION	−50.2794		
KNOW	−52.8485		
		SUPPORT	−58.5883
CALCULATE	−99.0087		
FIND	−99.7343	CLAIM	−99.7004
		ARGUE	−135.263
SHOW	−177.149		
INFORMATION	−214.336		

the method did not distinguish between Process types such as SUGGEST/ Verbal and SUGGEST/Relational; or SHOW/Verbal and SHOW/Relational. As Halliday points out (1994: 142) words like SUGGEST and SHOW may be judged to realize Verbal Processes if the subject (in the active voice) is a conscious being, and/or if the clause it is in is projecting (e.g. 'I suggested that there might be a risk'), or Relational Processes when the nominal elements are abstract (e.g. 'the data suggested a potential risk').

Therefore in order to explore the contrasts between the KWs more fully we also examined them in their wider context, using discourse analysis to make up for the fact that our corpus linguistics techniques did not allow for delicate analysis. For this purpose 20 examples of each of the positive KWs in each discipline were extracted and analysed.

The positive KWs from the history study corpus were found either to describe the interplay of claim and counter claim that constitute the practice of history, or to relate to the field of the history texts, describing past events where individuals and groups made claims and proposals, acted on beliefs and offered (or denied) support to each other. Agents were almost always human, and rarely implicit. Subjects, when present, were proper nouns or personal pronouns. ARGUE, ASSERT, BELIEVE, CLAIM/v and CRITICISM referred exclusively or chiefly to Processes whose agents were historians, including the student writers, while BELIEF, CLAIM/n, PROPOSAL and SUPPORT/n referred to Processes whose agents were historical figures or groups (SUPPORT was used chiefly to refer to political support offered by one group or state to another). In almost every case the forms *state* and *states* functioned as nouns, with the sense of 'nation', and therefore did not express the Process of 'saying'. The verb forms *stated* and *stating*, however, had interpersonal and epistemological or ideational roles in roughly equal numbers. Some examples of students' use of KWs expressing 'saying' and 'internal cognition' in history assignments are listed below. In these and subsequent examples the codes in square brackets are the unique identifiers of the BAWE corpus files.

1. . . . in his expressive Cold War jeremiad 'Gentleman: You are Mad,' social commentator Lewis Mumford **asserted** that 'madmen govern our affairs in the name of security'. [0005c]
2. Gareth Steedman Jones provides the seminal work. He **argued** that social movements, such as Chartism, could be constituted on ideological and political platforms . . . [0005a]
3. In this sense, it can be **argued** that for Marx and Engels, a primitive idea of democratic, or majority, rights served to justify a complex social theory of inevitable revolutionary struggle. [0003i]
4. I would **argue** that the 1917 revolution would not have occurred without it; [0010a]

5. Recent **criticism** has argued that the Cold War split has caused intellectu- als to make an over-simplified distinction between 'individualistic liberal- ism and state collectivism' . . . [0003j]

ns and beliefs expressed in these examples were only rarely supported by evidence (three times in the 20 occurrences of ARGUE: one time in the 20 occurrences of ASSERT: and in none of the 20 occurrences of BELIEVE). Nominalization, however, enabled arguments and beliefs to be explicitly eval- uated. In eight of the 20 occurrences of ARGUMENT, for instance, there was some form of evaluation, as in Example 6:

6. . . . in order to show the validity of the basic realist **argument** that there are two distinct realms of reality . . . [0004d]

hysics, all but three of the positive KWs referred to the establishment of facts from direct observation, measurement or calculation. The exceptions turned out not to express Verbal or Mental Processes. These were ABSTRACT (all examples referred to the abstract section of a research report), THEORY (in its usual role as the title of a section heading) and DETERMINE (when it was occasionally used to refer to causal relationships between states or events, as in Example 7).

7. The colour of the particular area of phosphor that the electron is fired at **determines** the colour of the resultant light on the screen. [0051a]

Agents were for the most part human in both disciplines, but in physics human agency was much more likely to be implicit, as in Examples 8 and 9:

8. Gamma ray photons are uncharged and create no ionisation or excitation of any material they pass through and hence the methods of **determining** their energies are somewhat limited. [0051c]
9. . . . and this data was used to **calculate** a value for Planck's constant. [0074a]

Non-human agents representing physical phenomena, theories, models or textual elements such as tables or calculations were also present in the phys- ics assignments. The positive keyword SHOW, for example, was used with text-internal agents in 14 out of the 20 cases, although there was still always a human observer and a projecting clause, as in Example 10, indicating Verbal rather than Relational Process.

10. Table 1 **shows** that as the intensity was decreased, the stopping voltage measured increased. [0074a]

DETERMINE and EXPLAIN sometimes referred to causal relationships holding between real world phenomena, as in Examples 11 and 12 and were therefore not always part of Verbal or Mental Process clauses.

11. The energy carried by a wave is **determined** by its intensity. [0074a]
12. The rotational motion easily **explains** the Earth's diurnal motion. [6097b]

History and physics thus have distinct sets of KWs, used to signal the epistemological value of the propositions they introduce. This reflects a fundamental difference in disciplinary cultures. KWs in the physics assignments referred to causal, logical and evidential relationships between physical phenomena and between phenomena and propositions (in the form of models, theories and properties of physical objects or systems). Thus in physics the identities of agents are commonly suppressed, to emphasize the fact that knowledge is derived from replicable laboratory activities, observations and measurements rather than from interpretation or discussion. On the other hand the KWs in history were more likely to have explicit agents; the identities of the authorities and sources referred to were important in establishing their validity and relevance.

6.2 Applied disciplines: Medicine, Engineering and Hospitality, Leisure and Tourism Management

Table 3.3 presents the keywords for the three applied disciplines that were significant at the level $p<0.00001$.

The three lists are strikingly different from those in the pure disciplines (Table 3.2). Unlike the KWs in history and physics, the majority of positive KWs in the applied disciplines that also featured in our WordNet list (Table 3.1) indicated a degree of uncertainty regarding the proposition being expressed, as in Examples 13–15. Many also turned out not to express Verbal or Mental Processes.

13. There were no abnormalities in other systems, which **indicate** that this diagnosis is less likely. (0194h, Medicine)
14. In most cases it is **uncertain** whether the individual factors act as initiators or promoters, due to the complex interactions between them (5). (0047a, Medicine)
15. Although the shaft and thrust bearings are being designed to take 50 per cent body weight, it is **unlikely** that this will be thrust onto the drill in its lifetime. (0023e, Engineering).

Table 3.3 Keywords in HLTM, Medicine and Engineering assignments

HLTM		Medicine		Engineering	
Keyword	Keyness	Keyword	Keyness	Keyword	Keyness
Positive keywords					
		UNLIKELY	457.670		
				DEFINITELY	192.077
INFORMATION	179.730				
				APPARENT	171.312
		CONFIRM	164.213	CALCULATE	130.768
				UNLIKELY	119.399
		EVIDENCE	116.112		
		PRESENT	113.433		
				CALCULATION	100.183
		UNCLEAR	107.263		
		PROPOSED	107.263		
APPARENT	98.174				
		INDICATE	94.824		
		APPARENT	85.810		
DEFINITELY	79.766				
		UNCERTAIN	71.509		
		SUPPORT	61.418		
UNLIKELY	55.223				
		FINDINGS	48.883		
UNCERTAIN	42.951	APPARENTLY	42.905		
				UNDOUBTEDLY	41.530
UNDOUBTEDLY	36.815				
				UNCLEAR	31.147
APPARENTLY	30.679				
CRITICIZED	30.679				
UNCLEAR	30.679				
IDENTIFY	28.561	ARGUES	28.603	DETERMINE	28.737
Negative keywords					
				SUGGEST	−30.157
				CLEAR	−31.859
		KNOWLEDGE	−32.247	SAY	−32.387
		CLAIM	−34.498		
				ANSWER	−35.177
CLAIM	−37.772				
		FIND	−39.780	FINDINGS	−39.718
		INFORMATION	−43.675	CLAIM	−43.993
				ABSTRACT	−46.375
		ARGUMENT	−55.292	PRESENT	−55.266
				BELIEVE	−56.917
				CRITICISM	−60.525
				BELIEF	−65.731
				EXPLAIN	−67.214
		FACT	−70.340		

Continued

Table 3.3   Continued

HLTM		Medicine		Engineering	
Keyword	Keyness	Keyword	Keyness	Keyword	Keyness
EVIDENCE	−74.028	SEE	−74.304		
				FACT	−78.302
		STATE	−97.101		
THEORY	−100.355			ARGUE	−119.530
STATE	−125.666				
		THEORY	−145.713	ARGUMENT	−161.501
				THEORY	−198.460
				EVIDENCE	−295.722
				STATE	−333.788

In HLTM assignments, UNCERTAIN also concerned the operation of businesses in an insecure environment:

16. Comprising mainly small businesses that rely on fluctuating demand, profitability is **uncertain** and production, wages and skills are low. (3013b, HLTM)

Most uses referred to uncertain inferences drawn from observations. Even those KWs that appear to have a high degree of certainty (CONFIRM, SUPPORT, DEFINITELY, UNDOUBTEDLY), however, often implied that the truth of the proposition had been established in the face of some doubt, as in Examples 17 and 18.

17. However, results using a higher piston velocity still would be needed to **confirm** this assumption. (0329f, Engineering)
18. Clinical signs such as pyrexia, dyspnoea and lung crackles were revealed during the physical examination which **supports** the diagnosis of a LRTI. (0047c, Medicine)

DEFINITELY and UNDOUBTEDLY were used to introduce assertions by the writer and had something of an interpersonal appeal, implicitly addressing objections that might be raised by the reader.

19. . . . SYSTECH Intl. can **definitely** do well in the marketplace . . . (0090a, Engineering)
20. **Undoubtedly**, since more new entrants are going into the industry, the industry competition will be stiffer. (3085c, HLTM)

As before, our methods of lemmatizing and of retrieving KWs did not distinguish between nouns and verbs with the same form, and it was necessary to examine them in context to see whether they formed part of Verbal or Mental Processes. SUPPORT/n, for example, turned out to function as a Participant in Material or Relational Process clauses, as in Example 21.

21. Die-cast aluminium would be used for the housing and lubricated ball
    bearings would be used for shaft **supports**. (0018d)

Becher and Trowler (2001: 36) describe applied approaches to knowledge as 'functional' and 'pragmatic', applying 'heuristic approaches' to develop products or procedures. These descriptions are consistent with our observations concerning the KWs; most of those from the applied disciplines reflected professional practices that use partial evidence to support the formation of opinions or decision making. The student writers used these words in order to determine causal relations (Examples 14, 20), diagnose properties of objects or systems (Examples 13, 17, 19) and assess likely outcomes (Example 15), always on the basis of imperfect evidence.

The data from Table 3.3 reveal both similarities and differences between the three applied disciplines. Becher and Trowler (2001: 36) identify medicine as a hard applied discipline, while HLTM is a social science and exemplifies a soft applied discipline. Nevertheless the two share many of the same KWs in Table 3.3. In contrast, engineering, the other hard applied discipline, yielded fewer positive KWs and more negative KWs. Engineering negative KWs included such words as ARGUE, BELIEF and SUGGEST, which were also negatively key in the hard pure physics assignments. These words are more likely to be positively key in soft disciplines because of the emphasis they place on the social development of knowledge and on the identity (and thus trustworthiness) of the researcher. Engineering and physics are both disciplines where argument and interpretation are less important than measurement and observation.

Once again our study of the KWs indicated a fundamental difference in disciplinary cultures, particularly between the pure and applied fields, but also between HLTM and medicine on the one hand, and engineering on the other.

7   Conclusion

In this study we have been able to quantify a distinction between student writing in hard, soft, pure and applied disciplines by looking at the keyness of selected lexical items. This distinction reflects a difference in knowledge construction, which, students who have achieved acceptable grades seem to have learnt to apply in their university coursework. One pedagogical implication of the findings is that there is a need to recognize discipline-specific ways of

thinking and saying at nominal group and clause level, where they can indicate the extent and explicitness of expressions of human agency, for example, and the degree of certainty with which propositions are put forward. For teaching and learning purposes it is important to present appropriate discipline-specific exemplars not only of entire texts, but also of lexical items and the grammar of the clause.

Notes

[1] 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 (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800).

References

Alsop, S., & Nesi, H. (in press). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora, 4*(1).

Babaii, E., & Ansary, H. (2005). On the effect of disciplinary variation on transitivity: The case of academic book reviews. *Asian EFL Journal, 7,* 113–126.

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines.* Milton Keynes: Society for Research into Higher Education and Open University Press.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd edn). Buckingham: Society for Research into Higher Education and Open University Press.

Berber Sardinha, T. (2000). Comparing corpora with WordSmith Tools: How large must the reference corpus be? *Proceedings of the Workshop on Comparing Corpora.* Annual Meeting of the ACL, Volume 9 (pp. 7–13). Morristown, NJ: Association for Computational Linguistics.

Bloor, T., & Bloor, M. (1995). *The functional analysis of English: A Hallidayan approach.* London: Arnold.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19,* 61–74.

Ebeling, S., & Heuboeck, A. (2007). Encoding document information in a corpus of student writing: The experience of the British Academic Written English (BAWE) corpus. *Corpora, 2,* 241–256.

Fellbaum, C. (1998). *WordNet: An electronic lexical database.* Cambridge, MA: MIT.

Halliday, M.A.K. (1994). *An introduction to functional grammar* (2nd edn). London: Arnold.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing.* Harlow: Longman.

Lea, M., & Street, B. (2000). Student writing and staff feedback in higher education: An academic literacies approach. In M. Lea & B. Stierer (Eds), *Student writing in higher education: New contexts* (pp. 1–14). Buckingham: Society for Research into Higher Education and Open University Press.

Love, A. (1993). Lexico-grammatical features of geology textbooks: Process and product revisited. *English for Specific Purposes, 12,* 197–218.

Martinez, I.A. (2001). Impersonality in the research article as revealed by analysis of the transitivity structure. *English for Specific Purposes, 20,* 227–247.

Nesi, H. (2008). BAWE: An introduction to 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* (pp. 239–246). Lisbon, Portugal: ISLA.

Nesi, H., & Gardner, S. (2006). Variation in disciplinary culture: University tutors' views on assessed writing tasks. In R. Kiely, P. Rea-Dickins, H. Woodfield & G. Clibbon (Eds), *Language, culture and identity in applied linguistics* (pp. 99–117). London: Equinox.

Scott, M. (1997). PC Analysis of key words – and key key words. *System, 25,* 233–245.

—(2004). *WordSmith Tools 4.0* [Computer software]. Oxford: Oxford University Press.

Scott, M., & Tribble, C. (2006). *Textual patterns: Keyword and corpus analysis in language education.* Amsterdam: Benjamins.

Swales, J. (1990). *Genre analysis: English in academic research settings.* Cambridge: Cambridge University Press.