Modals in the construction of research articles: A cross-disciplinary perspective

Matthew Peacock

City University of Hong Kong (China)

enmatt@cityu.edu.hk

Abstract

This paper describes a corpus-based analysis of variation in the distribution and function of modals and their role in the expression of "stance" in a corpus of 600 research articles (RAs) across twelve disciplines. Stance is an expression of attitudes, judgments, or assessments towards the truth of propositions (Biber et al., 1999), and part of the important function of claiming and confirming membership of discourse communities and therefore in constructing identity. Three functional categories of modals perform a valuable role in the construction of stance: Possibility/Ability, Obligation/Necessity, and Prediction (Biber et al., 1999). However, very little research seems to have investigated variation across disciplines or their use in the RA. The corpus was analysed using WordSmith Tools (Scott, 2004), followed by manual checking of the function of every occurrence. Inter- and intra-rater agreement was also checked. Many statistically significant disciplinary differences were found, along with numerous marked differences with individual modals. Further examination of the corpus revealed considerable disciplinary variation in the patterns and verbs associated with the target modals, and a number of sub-functions of the topics covered by the modals. Conclusions are that modals perform an important role in the construction of stance.

Keywords: modals, corpus analysis, English for Specific Purposes, genre analysis, research articles.

Resumen

El uso de expresiones modales en artículos de investigación: una perspectiva crosdisciplinar

En este trabajo se analiza la variación en la distribución y función de las expresiones modales y su papel en la expresión de un determinado punto de vista

en un corpus de 600 artículos de investigación pertenecientes a doce disciplinas académicas diferentes. Se entiende por punto de vista la expresión de actitudes, juicios de valor o valoraciones sobre la veracidad de una proposición (Biber et al., 1999), recurso comúnmente utilizado a la hora de reivindicarse como miembro de una comunidad discursiva y por lo tanto de construir una identidad. Tres categorías funcionales de expresiones modales desempeñan un valioso papel en la construcción de un punto de vista: Posibilidad/Habilidad, Obligación/Necesidad y Predicción (Biber et al., 1999). Sin embargo, hasta la fecha no se ha investigado mucho sobre cómo estas categorías de expresiones modales pueden variar dependiendo de la disciplina académica o sobre cómo se utilizan en artículos de investigación. El corpus fue analizado usando WordSmith Tools (Scott, 2004). A continuación se comprobó manualmente la función de cada expresión modal. También se comprobó la categorización de dichas funciones utilizando un observador independiente y posteriormente se realizó una comprobación adicional por parte del investigador. Se encontraron numerosas diferencias entre disciplinas que pueden considerarse significativas desde el punto de vista estadístico, así como numerosas y significativas diferencias en el uso de las diferentes expresiones modales. Un posterior análisis del corpus reveló la existencia de una considerable variación disciplinar en el uso de colocaciones y de verbos asociados con los modales empleados, así como un número de sub-funciones de los temas asociados con las expresiones modales. Se concluye que los modales desempeñan un importante papel en la expresión de determinados puntos de vista.

Palabras clave: expresiones modales, análisis de corpus, Inglés para Fines Específicos, análisis de género, artículos de investigación.

Introduction

This paper describes a corpus-based analysis of discipline variation in the distribution and function of modals, for example "can" and "would", and their role in the construction of stance. The corpus was 600 research articles (RAs) across twelve disciplines: Biology, Business, Chemistry, Computer Science, Economics, Environmental Science, Language and Linguistics, Law, Neuroscience, Physics and Materials Science, Psychology, and Public and Social Administration. The starting point of this research was the idea that modals may play an important role in the construction of stance in RAs. The RA was chosen for this research because of its significance for the spread of knowledge. Williams (2002) says discourse communities develop codes for communication through the use of patterns and that this code, rather than individual words, is one of their defining characteristics. RAs have been

called the key medium for authenticating findings and disciplines (Hyland, 1996), and the preferred method for communication among discourse communities (Williams, 1998). Their language defines these communities.

Stance is defined as the expression of attitudes, judgments, or assessments towards the truth of propositions (Biber et al., 1999: 966). It is represented by "linguistic mechanisms" (Biber, 2006: 97-98) and "lexical markers that convey the attitudes of the speaker/writer (…) toward our message (and) the degree of certainty" (Reilly, Zamora & McGivern, 2005: 186). It has also been described as the "expression of attitudes, feelings, judgments, or commitment concerning the propositional content of a message" (Biber & Finegan, 1989: 93). Suggestions, claims, and propositions are an important part of the RA; and in RAs the construction and expression of stance is part of the important function of claiming, confirming, and expressing membership of and position in the discourse community of peers, academics, and other researchers, and therefore in constructing identity.

Modals in general are said to be the most common markers of stance (Biber et al., 1999; Reilly, Zamora & McGivern, 2005). Biber et al. (1999) that three functional categories of modals play an important role in the construction of stance: Possibility/Ability for example "could"; Obligation/Necessity for example "must"; and Prediction/Volition for example "will". Additionally, modals are by far the most common markers of stance in university registers (for example, classroom teaching, texbooks – Biber, 2006). However, very little previous research seems to have investigated discipline variation in modal usage and their role in the construction of stance.

Next, four previous studies will be described, in order of importance. First, Biber et al. (1999) examined modal frequency in the academic prose section of the Longman Spoken and Written English (LSWE) corpus, book extracts plus RAs, 2.6 million words each: seventy-five book extracts (mostly technical trade books) from 13 disciplines, and RAs from 15 disciplines. Biber et al. do not give exact figures, but report the following results: Possibility/Ability 7200 per million words (pmw), Obligation/Necessity 3200 pmw, and Prediction/Volition 3600 pmw. Item results were "can" 3000, "could" 800, "may" 2800, "might" 600, "must" 1200, "should" 1200, "have to" 400, "need to" 200, "will" 2200, "would" 1400, and "shall" 200. "(Had) better", "have (got) to", "(be) supposed to" and "ought to" were each less than 100. However, occurrences of Volition (or intention: for example "This paper will/is going to discuss …") were not excluded. This

sub-function can be called future "will" and is not part of the expression of stance.

Second, Vázquez Orta (2010) analyses the use of the modals "can", "could", "may", "might", "must", "will", "should", "would", and "shall" in 48 Business Management RAs, half by L1 English authors and half by L1 Spanish authors. The total frequency for all these modals was 7000 pmw for the English authors, and 5300 pmw for Spanish authors. The most common modals in order of frequency for the former were "may", "could", "can", "would", and "might", and "can", "will", "may", "would", "could", and "might" for the latter. Vázquez Orta (2010: 93) suggests that "lack of modalisation seems to be a natural tendency for Spanish writers", adding that this adds to their difficulties writing in English and publishing. He does not, however, discuss the function of these modals in the expression of stance.

Third, Neff et al. (2004), as part of a larger study of writer stance, examine the modals "might", "may", "could", and "can" in a corpus of 60 L2 and L1 argumentative essays written in English. Neff et al. report that non-native speakers (NNS) overused "can" in comparison to native speakers (NS), and underused "might", "may", and "could". They speculate (p. 82) that the overuse of "can" results from the fact that Spanish relies heavily on one verb, "poder", which corresponds to the English "can" and "could" and sometimes "might" and "may". Finally, these scholars assert that NNS have problems modalizing statements.

Rationale for the present research

It is suggested that modals may play an important role in constructing stance and therefore in expressing the very important RA functions of putting forward suggestions, claims, and propositions. Modals would thereby function for claiming, confirming, and expressing membership of and position in a discourse community: and would therefore be valuable persuasive devices and an important part of research writing including RAs, and worth investigating further. And if modals are important, they must also be acquired by aspiring NNS research writers. Bhatia (2000: 147) says a strong justification for genre research is that it informs the teaching of research writing, especially for writers who wish to join academic discourse communities.

Taylor and Chen's (1991) suggestions that scientific discourse is heavily restricted by disciplinary cultures, and that much more attention must be given to these differences, have gained considerable ground in recent years. There seems to have been very little research into discipline variation in modal usage and the construction of stance and it is suggested the area is increasingly important due to the fast-growing numbers of research writers around the world, particularly NNS. It is proposed that the area has not received the attention it warrants and that further research is needed, to assess variation in modal usage across a number of science and non-science disciplines. The present research can provide this information and the results should also tell us much more about the nature of RAs, and help teachers of research writing inform learners of appropriate patterns.

Research method

The distribution, frequency and function of modals in 600 research articles was investigated across twelve disciplines, six science and six non-science. The disciplines were classified as science or non-science by asking experts to discuss the classification. There was no controversy regarding Biology, Chemistry, Computer Science, Environmental Science, Neuroscience, and Physics and Materials Science (classified by the experts as sciences), or Business, Language and Linguistics, Law, and Public and Social Administration (classified as non-sciences). Regarding Economics, experts said they consider it a non-science, pointing out that it is located in university Faculties of Humanities. There was more debate with Psychology. Experts said that while this topic has initiated some discussion, they do not classify it as a science, adding that only Clinical Psychology is considered a science among the 15 or more areas of psychology, and that university Psychology departments are not normally located in Science Faculties. Therefore Psychology was classified as a non-science for the present research. The modals and their functional categories (following Biber et al., 1999: 485) are:

(1) Possibility/Ability: "can", "could", "may", "might".

(2) Obligation/Necessity: "must", "should", "(had) better", "have (got) to", "need to", "ought to", "be supposed to".

(3) Prediction: "will", "would", "shall", "be going to".

Biber et al's list was adopted as it is the standard work on the topic. "(Had) better", "have got to", "ought to", and "shall" were excluded from the analysis after preliminary searches showed that they are virtually absent from the present corpus. Also, as noted above, Biber et al. call category three 'Prediction/Volition': and as Volition or future "will" is not part of the expression of stance, it was excluded from the analysis. Examples from the present corpus of "would" and "will" used for "Prediction" follow: "A general or remedial English course *would* be extremely difficult to organise" (Language and Linguistics). "The fact that the software industry is unlikely to soon become the next tobacco or asbestos industry does not mean that the legal system *will* not intrude. Near term likely intrusions *will* be forced by: More contractual requirements …" (Computer Science). "One *would* expect to find deficits that are fairly limited to certain perceptual or conceptual domains" (Language and Linguistics). These modals function to express commitment to the truth of the propositions presented.

Research aims and research questions

The aims of this research were, within the corpus, to:

(1)   Investigate the frequency of all modals in the target categories.

(2)   Investigate disciplinary variation.

(3)   Investigate function.

The following questions are directly addressed:

(1)   How frequently do RA authors use modals across twelve disciplines? Are there any interdisciplinary differences?

(2)   How do modals function across twelve disciplines?

The RA Corpus

The corpus was 600 RAs published between 2000 and 2008, 50 from each discipline – see Table 1. The twelve disciplines were selected because they represent a range of subjects and also have large numbers of research writers, mostly NNS, around the world. This increases the usefulness of this

research regarding recommendations for teaching. Various leading journals were selected from each discipline (see Appendix). Visits were made to the relevant departments and two sources from each asked to name principal journals from their field.

Discipline	No. of RAs	Total word length
Biology	50	286,440
Business	50	329,599
Chemistry	50	182,472
Computer Science	50	359,003
Economics	50	364,710
Environmental Science	50	249,874
Language and Linguistics	50	320,847
Law	50	372,748
Neuroscience	50	303,098
Physics and Materials Science	50	226,253
Psychology	50	381,709
Public and Social Administration	50	306,624
All twelve disciplines	600	3,683,377

Table 1. Lengths of disciplinary corpora.

RAs were randomly chosen from each journal by giving each a number and drawing numbers from a box. No attempt was made to only choose NS writers for the corpus. Only empirical data-driven RAs with the Introduction-Method-Results-Discussion format were chosen, following Hyland (1998) who notes that this is an important genre. Essays, discussions and RAs by writers already chosen were not used. It is suggested that the disciplinary corpora are sufficiently representative because of their size and because of the use of discipline sources to choose journals.

Investigating the corpus

Analysis was done in the following steps, using the Concord function of WordSmith Tools 4.0 (Scott, 2004). Explanation of steps 1 and 2 follows:

Step 1: Investigate disciplinary variation in the frequency of the target modals.

Step 2: Manually and individually check the function of every occurrence of all the target modals by reading the relevant sentence and surrounding sentences. One reason for doing this was to check whether or not every occurrence functioned to construct stance.

> Another was that "will" does not always function in this way, as noted above: all such functions were excluded from the count. It is also noted that it is possible that authors use modals to tone down their findings either in order to be modest or polite, or to follow academic writing conventions within disciplines.

Statistical significance was set at $p<.05$ and was tested with the log-likelihood calculator (Rayson, Berridge & Francis, 2004).

For step 1, high-frequency is defined as 40 pmw or higher, following Biber et al. (2004: 376): "we take a conservative approach (…) [a] frequency cut-off of 40 times per million words to be included in the analysis".

Regarding step 2, "function" means "operates" or "acts". Individual manual checking of the function of every occurrence is vital. The importance of doing this is stressed by Tognini-Bonelli (2004), who asserts that while frequency can be obtained from statistical analysis, context is vital for understanding function. Williams (2002: 60) proposes that a "microscopic study" must be carried out before categorisation can be done. Two evaluators carried out step 2: this writer and a local university lecturer. To measure inter-rater agreement, the second coder independently evaluated the function of every occurrence. To measure intra-rater agreement, this writer reassessed the function of every occurrence after one month. Inter-rater and intra-rater agreement, after discussions, were both 100%.

Results

The results for all functional categories and individual modals in the whole corpus can be seen in Tables 2 and 3:

The twelve target modals were very common – their average frequency was 736 pmw. The tables reveal considerable disciplinary variation, and large numbers of statistically significant differences. One clear and broad difference is between the sciences and non-sciences: the former show significantly lower frequencies, apart from just one science, Computer Science.

Target modals	All 12 disc.	Bus.	Eco.	Lang. & Ling.	Law	Psy.	Pub. & Soc. Adm.
Possibility / ability							
can	2229	**2575***	2356	**2407***	2186	*1496***	2186
may	1624	**1838***	*1378***	1694	**2809***	1620	1762
could	790	840	*585***	871	**934***	774	**971***
might	514	*447***	*434***	**658***	**1100***	**836***	578
TOTAL	5156	**5701***	*4753***	**5631***	**7029***	*4727***	5497
Obligation / necessity							
should	790	**1018***	*653***	775	**1518***	813	831
must	338	**441***	*277***	361	**735***	*245***	325
need to	196	**413***	*68***	**369***	193	*98***	**429***
(have) to	184	**226***	154	**295***	190	193	**357***
(be) supposed to	8	8	14	**27***	11	20	5
TOTAL	1516	**2106***	*1166***	**1827***	**2648***	*1369***	1947
Prediction							
would	1197	1240	1186	**1353***	**2683***	**1379***	1319
will	952	**1170***	**1231***	1017	**2012***	*617***	**1355***
(be) going to	11	18	3	21	3	3	**50***
TOTAL	2161	**2427***	**2421***	**2391***	**4698***	*1999***	**2724***
TOTAL – all modals	8833	**10234***	*8339***	**9849***	**14374***	*8095***	**10168***

Table 2. Discipline differences: Non-sciences frequency per million words
* = statistically significant difference: **bold** is significantly higher, *italics* is significantly lower

Target modals	All 12 disc.	Bio.	Chem.	Com. Sc.	Env. Sc.	Neu.	Phy. & Mat. Sci.
Possibility / ability							
can	2229	*1419***	*1767***	**3779***	*2037***	*1411***	**2518***
may	1624	1589	*775***	1517	*1440***	*1374***	*1253***
could	790	851	692	*612***	*658***	**959***	*669***
might	514	440	*162***	*290***	*225***	477	*268***
TOTAL	5156	*4300***	*3396***	**6198***	*4361***	*4221***	*4708***
Obligation / necessity							
should	790	*407***	*353***	**1024***	*512***	*617***	*547***
must	338	*265***	*123***	395	*214***	*165***	320
need to	196	*85***	*43***	231	*141***	*49***	*20***
(have) to	184	*57***	*18***	**234***	*101***	197	*67***
(be) supposed to	8	*0***	4	*0***	*0***	*0***	*0***
TOTAL	1516	*814***	*541***	**1884***	*968***	*1029***	*954***
Prediction							
would	1197	*577***	*512***	*1080***	*1047***	*922***	*547***
will	952	*322***	*191***	1052	**1120***	*337***	*564***
(be) going to	11	*0***	*0***	**31***	*0***	*0***	*0***
TOTAL	2161	*899***	*703***	2164	2166	*1259***	*1111***
TOTAL – all modals	8833	*6012***	*4640***	**10246***	*7495***	*6509***	*6772***

Table 3. Discipline differences: Sciences frequency per million words.
* = statistically significant difference: **bold** is significantly higher, *italics* is significantly lower

Four of the most striking areas of individual discipline variation visible in Tables 2 and 3 are:

(1)　the high frequencies for Business in Obligation/Necessity ("should", "must", "need to", and "(have) to");

(2)　the very high frequencies for Law in Possibility/Ability ("may", "could", "might"), Obligation/Necessity ("should", "must"), and Prediction ("would", "will");

(3)　the high frequencies for Computer Science in Possibility/Ability ("can"), and Obligation/Necessity ("should");

(4)　the many low category and individual item frequencies in five out of the six sciences – Possibility/Ability in Biology, Chemistry, and Environmental Science; Obligation/Necessity in Biology, Chemistry, Environmental Science, Neuroscience, and Physics and Materials Science; and Prediction in Biology, Chemistry, Neuroscience, and Physics and Materials Science. Chemistry, and to a lesser extent Biology, are particularly low in Obligation/Necessity and Prediction. However, in just six out of the 36 functional category totals, analysis may be confused by the possible use of synonyms or near-synonyms within individual categories, where one modal shows significantly higher usage and another significantly lower. For example, Psychology authors may be using "might" instead of "can", and "would" instead of "will"; Physics and Materials Science authors may be using "can" instead of "may", "could", and "might"; and Environmental Science authors may be using "will" instead of "would". However, in the other 30 category totals, there is no such potential confusion. Individual manual checking of the function of every occurrence of all the target modals, taking into account the possibility that authors use modals to tone down their findings through modesty or politeness or to follow discipline conventions, confirmed that apart from the exceptions with "will" noted above, they all functioned to construct stance. They were also in line with the Biber et al. (1999) functional categories. Examples from the corpus will be given in the next section.

Discussion and conclusions

Analysis of the corpus leads us to suggest that modals do play an important role in the construction of stance in RAs. The top ten modals in this corpus

together at 8813 pmw were less common than in Biber et al. (1999), though this is perhaps not surprising as they had a different corpus. Additionally, Biber et al. did not exclude occurrences of Volition: Volition comprised 22% of the occurrences of "will" and "going to". However, higher frequencies were found than in Vázquez Orta (2010), who reports 7000/5300 pmw for eight modals in Business Management RAs. In the present corpus the frequency was 9570 pmw for the same modals in Business RAs.

Closer examination of the corpus was then undertaken to investigate the striking discipline differences seen in Tables 2 and 3, which are not easy to explain. Other authors have followed Taylor and Chen's (1991) suggestion that discipline discourses are affected by disciplinary cultures. Hyland says (1999) that discipline differences reflect rhetorical constraints within a discipline, and Bondi (2006: 49; also see Malavasi & Mazzi, 2008) asks the very good and "daunting question" – "Where does the specificity of a discipline lie?" She says she cannot answer but speculates that part of the answer lies in the language used. Hyland (2008) also refers to language conventions among RA authors, and proposes that different disciplines value different kinds of arguments and also vary in what their readers already know and how they might be persuaded. Hyland says the result of this is that physicists do not write like philosophers or applied linguists, and theorizes that disciplines range along a cline with hard knowledge sciences and softer Humanities at opposite ends. His hypothetical cline describes sciences as empirical, objective, quantitative, showing linear and cumulative growth, utilizing experimental methods, not relying on rhetoric, and putting greater weight on methods, procedures and equipment: and Humanities as explicitly interpretive, qualitative, utilizing discursive argument and more fluid discourses, and putting greater weight on strength of argument to present claims. Hyland (2008), like Bondi (2006), hypothesizes that the specificity of disciplines lies in language, and also asserts that hard science writers use more modals such as "could" and "may" in order to objectify their research.

The discipline differences found were explored in the light of Hyland's (2008) hypothetical cline and other hypotheses. It is suggested that this can inform us about the research focus and the argument and persuasive structure of individual disciplines. The first step in this further analysis was using the Patterns function of Concord in WordSmith Tools (the Collocates and Cluster functions were much less useful for this) to isolate common patterns, and the verbs associated with those patterns, in the striking areas for Business and Law mentioned above – compared with the other eleven

disciplines. Table 4 shows the results, and Table 5 shows patterns and verbs for a selection of other modals and disciplines. The numbers in brackets in the "Sub-Function" columns refer to representative examples extracted from the corpus, which follow Table 5.

Discipline & notable discipline feature	Modal	Common Patterns	Sub-function	Associated verbs	Common patterns in the 11 other disciplines	Sub-functions in the 11 other disciplines	Associated verbs in the 11 other disciplines
Business - High frequencies, Obligation/Necessity	should	research ~, studies ~, projects ~, variables ~	Talking about research (1)	~examine, ~develop, ~consider, ~explore	values ~, users ~, Information ~ the system ~	Talking about RA topics (2)	~exist, ~avoid, ~serve, ~belong
	must	customers ~, entrepreneurs ~, companies ~, firms ~	Talking about RA topics (3)	~exist, ~develop, ~engage, ~use	systems ~, factors ~, information ~, theory ~	Talking about research (4)	~satisfy, ~equal, ~note, ~consider
	need to	entrepreneurs ~, companies ~, owners ~, firms ~	Talking about RA topics (5)	~manage, ~understand, ~consider, ~develop	data ~, studies ~, theories ~, findings ~	Talking about research (6)	~be, ~develop, ~carry, ~improve
Law - Very high frequencies, Possibility/Ability	may	victim/s ~, decision/s ~, offenders ~, court/s ~	Discussing courts and cases (7)	~cause, ~decide, ~choose, ~include	research ~, study ~, information ~, factors ~	Talking about research (8)	~differ, ~lead, ~contribute, ~choose
	could	judges ~, cases ~, laws ~, courts ~	Discussing courts and cases (9)	~include, ~make, ~increase, ~result	research ~, study/ies ~, results ~, model/s ~	Talking about research (10)	~be, ~afford, ~consider, ~replace
	might	decision/s ~, court/s ~, law ~, police ~	Discussing courts and cases (11)	~make, ~attempt, ~influence, ~become	participants ~, results ~, factors ~, studies ~	Talking about research (12)	~serve, ~develop, ~include, ~affect
Law - Very high frequencies, Obligation/Necessity	should	court/s ~, law ~, process ~, offender/s ~	Discussing courts and cases (13)	~continue, ~become, ~consider, ~allow	values ~, companies ~, firms ~, users ~	Talking about RA topics (14)	~treat, ~help, ~give, ~reflect
	must	court/s ~, judge/s ~, policy ~, evidence ~	Discussing courts and cases (15)	~remain, ~consider, ~show, ~determine	system/s ~, information ~, factor/s ~, theory ~	Talking about research (16)	~satisfy, ~equal, ~note, ~consider
Law - Very high frequencies, Prediction	would	case/s ~, abuse ~, parties ~, victim/s ~	Discussing courts and cases (17)	~involve, ~open, ~need, ~require	model ~, participants ~, study ~, results ~	Talking about research (18)	~begin, ~continue, ~depend, ~require
	will	victim/s ~, defendant/s ~, injurer/s ~, courts ~	Discussing courts and cases (19)	~be, ~act, ~engage, respond	study/ies ~, system ~, information ~, data ~	Talking about research (20)	~investigate, ~involve, ~stay, ~tend
Computer Science - High frequencies, Possibility/Ability	can	users ~, algorithms ~, model ~, system ~	Talking about RA topics (21)	~be, ~see, ~find, ~use	we ~, model ~, results ~, research ~	Talking about RA topics (22)	~be, ~lead, ~have, ~provide
Computer Science – High frequencies, Obligation/Necessity	should	users ~, system ~, nodes ~, research ~	Talking about RA topics (23)	~be, ~have, ~use, ~play	values ~, companies ~, projects ~, users ~	Talking about RA topics (24)	~treat, ~belong, ~give, ~become

Table 4. The most striking disciplinary differences in modal usage, with common patterns and associated verbs.

Modal	Discipline	Common patterns	Sub-function	Associated verbs
can	Eco.	firm/s ~, revenue ~, model/s ~, markets ~	Talking about research (25)	~ punish, ~ expect, ~ lead, ~ borrow
	Lang. & Ling.	students ~, teachers ~, learners ~, writers ~	Talking about RA topics (26)	~ practice, ~ facilitate, ~ offer, ~ give
	Pub. & Soc. Adm.	housing ~, environment ~, managers ~, findings ~	Talking about RA topics (27)	~ help, ~ provide, ~ note, ~ afford
	Bio.	protein/s ~, plants ~, species ~, system/s ~	Talking about research (28)	~ serve, ~ be, ~ affect, ~ bind
	Chem.	sample/s ~, materials ~, effect/s ~, ions ~	Talking about research (29)	~ influence, ~ provide, ~ occur, ~ lead
	Phy. & Mat. Sci.	equation/s ~, values ~, stress ~, alloys ~	Talking about research (30)	~ influence, ~ have, ~ be, ~ give
may	Eco.	firms ~, market/s ~, policies ~, investors ~	Talking about research (31)	~ reduce, ~ help, ~ occur, ~ indicate
	Lang. & Ling.	student/s ~, English ~, writing ~, teacher/s ~	Talking about RA topics (32)	~ influence, ~ occur, ~ play, ~ serve
	Pub. & Soc. Adm.	maltreatment ~, children ~, public ~, health ~	Talking about RA topics (33)	~ reflect, ~ help, ~ affect, ~ lead
would	Bus.	firm/s ~, information ~, research ~, owners ~	Talking about research (34)	~ result, ~ allow, ~ support, ~ require
	Eco.	firm/s ~, bidder/s ~, revenue ~, labour ~	Talking about research (35)	~ lead, ~ make, ~ require, ~ change
	Lang. & Ling.	language ~, student/s ~, learner/s ~, course ~	Talking about research (36)	~ seem, ~ wish, ~ need, ~ take
will	Bus.	conflict ~, firm/s ~, market/s ~, product ~	Talking about RA topics (37)	~ take, ~ focus, ~ make, ~ affect
	Eco.	team/s ~, firm/s ~, regions ~, income ~	Talking about research (38)	~ see, ~ discuss, ~ consider, ~ refer
	Lang. & Ling.	language ~, student/s ~, teacher/s ~, English ~	Talking about research (39)	~ help, ~ begin, ~ enable, ~ examine

Table 5. A selection of other modal patterns and associated verbs.

Future studies *should* examine how factors such as environmental uncertainty and the nature of dependencies between different functional areas affect levels of involvement.

Representative examples from the corpus (modals have been highlighted with inverted commas):

(1) The Future studies "should" examine how factors such as environmental uncertainty and the nature of dependencies between different functional areas affect levels of involvement (Business)

(2) The implications of such a system "should" be more fully explored (Psychology)

(3) Companies "must" therefore develop the capability of creating new services that customers will value (Business)

(4) Obviously, additional factors "must" be considered to adequately account for child physical abuse risk (Public and Social Administration)

(5) Owners "need to" seriously consider the range and intensity with which they access various potential networks (Business)

(6) Personal and public theories "need to" be viewed as living, intertwining tendrils of knowledge which grow from and feed into practice (Language and Linguistics)

(7) Both the criminal and the potential victim "may" decide to have a gun available during the crime (Law)

(8) The lack of research "may" be due to the highly conserved structure of the TS active site (Biology)

(9) This law "could" increase the probability of being stopped (Law)

(10) We cannot rule out the possibility that these results "could" be due to the markedness of the consonants (Neuroscience)

(11) Similarly, police "might" be influenced by a suspect's prior criminal record (…) they "might" be more likely to pursue the arrest of a suspect who has a criminal history (Law)

(12) It is not clear how participants "might" react to such tokens (Psychology)

(13) Instead, courts "should" consider whether a custom of use is ingrained and what this norm reflects in society (Law)

(14) In such cases, users "should" change the minimum code clone size to '50' (Computer Science)

(15) The court "must" determine prospectively whether the copyright holder "should" control the market (Law)

(16) For the SV2 model, paths for both volatility factors "must" be simulated (Economics)

(17) Acts of wife abuse "would" be perceived as wrong if they were illegal (Law)

(18) The fuzzy logic results "would" be equivalent to classic automation results (Environmental Science)

(19) The answers since Georgetown in 1964 have varied and it is unclear what weight courts "will" give to this interest (Law)

(20) Any message disseminated in a loosely coupled system "will" be subject to varying interpretations (Public and Social Administration)

(21) By successively splitting partitions, the system "can" use the incoming evidence to gradually focus in on the underlying states of interest (Computer Science)

(22) At a minimum, our results "can" be used as a benchmark for negotiations (Business)

(23) It is desirable that a data mining system "should" be able to present comprehensible results, in an accessible manner (Computer Science)

(24) Another implication of children's language-specific categorization is that they "should" treat two words as different if those words contain the sounds of a discriminated, native contrast (Psychology)

(25) However, attrition of firms "can" lead to sample selection bias if the disturbance of the selection equation is correlated with the disturbance of the growth equation (Economics)

(26) Paraphrasing exercises are sometimes provided so that students "can" practice restating the ideas of a given excerpt (Language and Linguistics)

(27) The physical environment "can" provide clues to the social and political situation that influences behavior (Public and Social Administration)

(28) Recombinant proteins "can" be expressed to three distinct locations (Biology)

(29) The low-temperature sol–gel syntheses that are generally used to prepare nanocrystalline materials "can" lead to the formation of high concentrations of defects (Chemistry)

(30) The STEX equation "can" also be solved for positive eigenvalues (Physics and Materials Science)

(31) The presence of stickiness at least on the demand side of the labour market "may" generate instability, as shown in Section 3 (Economics)

(32) Teachers "may" not be fully aware of how much feedback they give (Language and Linguistics)

(33) Maltreated children "may" be prone to impulsive, aggressive behavior when faced with frustration (Public and Social Administration)

(34) The positive and significant coefficients indicate that firms "would" enhance the degree of knowledge management when the organizational members trust each other (Business)

(35) In terms of our misperception hypothesis, high value bidders "would" thus be more likely to formulate pessimistic beliefs about the available supply (Economics)

(36) Similarly, the very essence of persuasive language "would" seem to depend on writers operating free of spatial (or in the case of sermons and speeches, temporal) limitations (Language and Linguistics)

(37) A change in a belief in a product "will" not only affect the perception about the producers of that specific product, but "will" also impose a set of values on those producers who are considered to belong to other countries (Business)

(38) The GDP share of the high-income regions "will" still fall, according to the projections, but by less (Economics)

(39) Analysis of their language "will" indeed produce striking results (Language and Linguistics)

Tables 4 and 5 also reveal considerable disciplinary variation in patterns and in associated verbs. The next step was to look more closely at function. While the primary function of all target modals falls into one of Biber et al's (1999) three functional categories of modals, it was possible to identify sub-functions of the topic types covered by the modals. Discipline variation is also apparent here. They fell into three categories - talking about research, talking about RA topic areas, and, specific to Law, discussing courts and cases. The next step was to examine the most striking differences seen in Tables 2 and 3. These are the following:

Business authors used 34% more Obligation/Necessity modals than the other four non-science disciplines (excluding Law), employing "should" to discuss RA topic areas, and "must" and "need to" to discuss their research. Presumably it is correspondingly more important and necessary in Business to express commitment to the truth of propositions in the Obligation/Necessity area, and to put greater weight on strength of argument to present claims.

Law authors' use of all modals over all three Biber et al. (1999) functional categories was a remarkable 54% higher than the other five non-science disciplines. Authors utilised them to discuss the broad area of courts and cases and more specifically victims, decisions, offenders, judges, laws, police, process, policy, evidence, abuse, and defendants. Seemingly, in all three functional categories, it is correspondingly much more important and necessary in Law than in any other discipline to utilise modals to express commitment to the truth of propositions concerning these discipline-specific topics, and to present argument in support of claims. Law authors used far more modals to achieve these functions than did writers in other disciplines.

Next, Computer Science authors' use of modals was a striking 63% higher than the other five science disciplines. Apparently, in two out of three

functional categories, Possibility/Ability and Obligation/Necessity, it is correspondingly very much more important in Computer Science than in the other five sciences to employ modals to express attitudes and assessments towards the truth of propositions, to express degrees of certainty, and to utilise argument in support of claims. Computer Science authors used significantly higher numbers of modals to achieve these functions than did the other science writers.

Finally, Chemistry and Biology show particularly low frequencies for Obligation/Necessity and Prediction. A closer examination of Chemistry and Biology RAs was then carried out to try to identify the reasons for this. After careful searches revealed that authors do not appear to cover Obligation/Necessity and Prediction in ways aside from the use of modals, it was concluded that these authors present and develop arguments in a different way, with far less reference to these functions. Authors discuss their methods, results and conclusions in a much more narrative and descriptive style, normally describing their research steps and their findings one by one, letting readers work out their claims in this way. It seems that this is sufficient for readers, who perhaps do not need to be openly told the connections between arguments and claims. Computer Science RAs, however, were much more similar to the non-science disciplines.

Randomly selected and representative examples from Chemistry, Biology, and Computer Science discussion and conclusion sections follow:

(40) In conclusion, LC–ESI-MSn was proved to be a rapid facultative analytical method for the characterization of steroidal saponins in the crude extract from D. nipponica, while multi-stage tandem mass spectrometry combined with electrospray ionization (ESI-MSn), particularly adducted metal ion, was helpful for elucidating the structural information. (Chemistry)

(41) Isatin was found to be an effective inhibitor for commercial copper corrosion in aerated 0.5 M H2SO4 at higher concentrations in the temperature range 25–55 °C. Some evidence was found for a different corrosion mechanism for copper in solutions containing isatin in contrast to aerated uninhibited solutions of 0.5 M H2SO4. The difference between gravimetric and potentiodynamic corrosion rates emphasizes the limitation of the Tafel line extrapolation method in the determination of corrosion rates. (Chemistry)

(42) The c-Fos peptide immunogen sequence used to generate c-Fos antibody displayed no homology to any protein species other than c-

Fos itself when an advanced BLASTp homology search was performed with the peptide sequence employing the Swissprot data base. It is therefore likely that p35 is a novel protein species. The differential expression pattern of p35 in P+ and P- cells rules in a role for p35 as a novel determinant of the promotion susceptible phenotype. (Biology)

(43) The carbon emissions coefficient, the carbon emitted in producing 1 MJ of energy in the form of wood from short rotation coppice, is estimated to be 0.0013 kgCMJ-1 based on the standard set of assumptions. Because wood fuel itself is 'carbon neutral', this represents the total emissions of carbon to the atmosphere as a result of using wood fuel. This estimate does not include any allowance for carbon sequestered in unutilized components of coppice which may rise to between 5 and 12 tCha-1 over 25 years. (Biology)

(44) This study and its findings may have important implications as the Greek society becomes increasingly permeated by ICT. For instance, the GCAS could be useful to Greek researchers as they pursue further research in the area of general attitudes and general behaviors toward computers. Greek researchers are encouraged to use the GCAS to further examine its reliability and validity. Future research should continue to test the possible relationships between computer attitudes and computer experience and confidence. (Computer Science)

The above analysis leads to the proposal that many or most of the patterns and associated verbs presented are prevailing terminology within disciplines. These differing patterns and associated verbs appear to arise from the topics discussed: or more explicitly, it is apparent that they cover differing terminology, topics, research methods, and discussions across the twelve disciplines. They are thus an important part of the meanings, and the functions, of these modals. The sharp discipline differences found lead to the conclusion that these meanings and functions differ by discipline, and that useful conclusions regarding modal usage cannot be constructed by looking at only one or two disciplines. This highlights the importance of discipline-specific research into modals. These modals and their patterns seem to represent one method through which disciplines differentiate themselves, as well as being an important part of these RAs and part of the defining code (Williams, 2002) of RAs. They represent disciplinary norms, and the patterns presented are apparently accepted within disciplines as standard ways for authors to present their research. This study has revealed some conventional forms in disciplinary corpora.

It is concluded that modals play a valuable role in the construction of stance, an important part of research writing. Authors do employ them to express attitudes, value judgments, and assessments towards their suggestions, claims, and propositions, and thereby accomplish the important functions of claiming and confirming membership of their discourse community, and constructing identity. It is also concluded, regarding modals, that scientific discourse is constrained by disciplinary cultures (Taylor & Chen, 1991; Hyland, 1999), that an important part of the specificity of disciplines lies in language (Bondi, 2006), and that disciplines do differ in argument structure and in what readers already know and how they are persuaded. In addition, Hyland's (2008) hypotheses were a helpful starting point for analysis of the present corpus where, concerning modals, it was found that Business and Law authors (for example) do not write like Biology or Chemistry authors. And broadly speaking the twelve disciplines did range along a cline with five of the six sciences, and the six non-sciences plus Computer Science, at different ends. However, Hyland's (2008) further hypotheses regarding sciences versus non-sciences were somewhat less useful in the present instance, as they did not predict the discipline differences found. While Biology, Chemistry, Environmental Science, Neuroscience, and Physics and Materials Science authors did show tendencies to rely less on rhetoric and to put greater weight on methods, procedures and equipment, it was also found that Economics (to a large extent), and Language and Linguistics and Psychology (to some extent), were empirical, objective, and quantitative, and put a lot of emphasis on methods and procedures. However, Business, Language and Linguistics, Law, Psychology, and Public and Social Administration were found to tend more towards interpretive and discursive argument, and to place greater weight on argument to present claims, than did the five sciences. However, Hyland's (2008) assertion that hard science writers use more modals such as "could" and "may" to objectify their research was not supported by the present research.

Implications for teaching

The present research can inform the teaching of research writing. Tables 4 and 5 provide discipline-specific lists of the highest-frequency collocations – in this case, the nouns and verbs – associated with our target modals. Considerable disciplinary variation is apparent. Collocations are an important part of language knowledge and they need be acquired, stored,

and processed as complete units (Schmitt, Grandage & Adolphs, 2004) by students. It is suggested that awareness of these discipline variations is important for teaching, especially to students of research writing, and that discipline-specific teaching of these collocations is desirable. This may be especially important for NNS, who may not know the importance of these particular collocations in academic English, and who may also be unaware of discipline conventions.

Further research can look at other areas of contrast between sciences and non-sciences, and among other disciplines: patterns and associated verbs may well be fruitful areas for further investigation. Lack of space precludes the presentation of further examples, and further analysis, from the present corpus. However, this study has provided indications that modals are an important method of achieving the goal of persuading readers of the authenticity of their claims, one of the primary aims of RA authors, and that there are interdisciplinary differences in how they do this. Analysis of the large corpus shows modals to be an important part of research writing. It is suggested that this research has added to an understanding of variation in modal usage across a number of disciplines, and hoped that the present findings improve knowledge of RAs, have relevance for the teaching of research writing, and help teachers prepare discipline-specific research writing materials.

References

Bhatia, V.K. (2000). "Genres in conflict" in A. Trosborg (ed.), *Analysing Professional Genres*, 147-161. Amsterdam: John Benjamins.

Biber, D. (2006). "Stance in spoken and written university registers". *Journal of English for Academic Purposes* 5: 97-116.

Biber, D., S. Conrad & V. Cortes. (2004). "If you look at…: lexical bundles in university teaching and textbooks". *Applied Linguistics* 25: 371-405.

Biber, D. & E. Finegan (1989). "Styles of stance in English: lexical and grammatical marking of evidentiality and affect". *Text* 9: 93-124.

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.

Bondi, M. (2006). "A case in point: Signals of narrative development in Business and Economics" in K. Hyland & M. Bondi (eds.), *Academic Discourse across Disciplines*, 49-74. Bern: Peter Lang.

Hyland, K. (1996). "Talking to the academy: forms of hedging in science research articles". *Written Communication* 13: 251-281.

Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins.

Hyland, K. (1999). "Disciplinary discourses: Writer stance in research articles" in C. Candlin & K. Hyland (eds.), *Writing: Texts, Processes and Practices*, 99-121. London: Longman.

Hyland, K. (2008). "Genre and academic writing in

the disciplines". *Language Teaching* 41: 543-562.

Malavasi, D. & D. Mazzi (2008). "'It can be assumed that, if our conceptual model is valid, then…': the construction of multiple identities in Economics vs. Marketing". *Linguistica e Filologia* 27: 157-179.

Neff, J., F. Ballesteros, E. Dafouz, F. Martínez & J.P. Rica (2004). "Formulating writer stance: A contrastive study of EFL learner corpora" in U. Connor & T. A. Upton (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, 73-89. Amsterdam: Rodopi.

Neff, J., E. Dafouz, M. Díez, F. Martínez, R. Prieto & J.P. Rica (2003). "Evidentiality and the construction of writer stance in native and non-native texts" in J. Hladky (ed.), *Language and Function* 223-235. Amsterdam: John Benjamins.

Rayson, P., D. Berridge & B. Francis (2004). "Extending the Cochran rule for the comparison of word frequencies between corpora" in G. Purnelle, C. Fairon & A. Dister (eds.), *Le Poids des Mots: Proceedings of the 7th International Conference on Statistical Analysis of Textual Data (JADT 2004)*, Volume II, 926-936. Louvain: Presses Universitaires de Louvain.

Reilly, J., A. Zamora & R.F. McGivern (2005). "Acquiring perspective in English: the development of stance". *Journal of Pragmatics* 37: 185-208.

Schmitt, N., S. Grandage & S. Adolphs (2004). "Are corpus-derived recurrent clusters psycholinguistically valid?" in N. Schmitt (ed.), *Formulaic Sequences*, 127-151. Amsterdam: John Benjamins.

Scott, M. (2004). *WordSmith Tools*, Version 4. Oxford: Oxford University Press.

Swales, J.M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Taylor, G. & T. Chen (1991). "Linguistic, cultural, and subcultural issues in contrastive discourse analysis: Anglo-American and Chinese scientific texts". *Applied Linguistics* 12: 319-336.

Tognini-Bonelli, E. (2004). "Working with corpora: Issues and insights" in C. Coffin, A. Hewings & K. O'Halloran (eds.), *Applying English Grammar: Functional and Corpus Approaches*, 11-24. London: Arnold.

Vázquez Orta, I.O. (2010). "A contrastive analysis of the use of modal verbs in the expression of epistemic stance in Business Management research articles in English and Spanish". *Ibérica* 19: 77-96.

Williams, G.C. (1998). "Collocational networks: interlocking patterns of lexis in a corpus of plant biology research articles". *International Journal of Corpus Linguistics* 3: 151-171.

Williams, G. (2002). "In search of representativity in specialised corpora: categorisation through collocation". *International Journal of Corpus Linguistics* 7: 43-64.

Matthew Peacock teaches in the Department of English at the City University of Hong Kong. His research interests include English for Specific Purposes, corpus analysis, research writing, genre analysis, and TEFL methodology. In 2001 he co-edited (with John Flowerdew) a collection from Cambridge University Press, *Research Perspectives on English for Academic Purposes*.

Appendix: Journals in the corpus.

Biology
- *Applied Soil Ecology*
- *Biochimica et Biophysica Acta*
- *Biomass and Bioenergy*
- *Chemistry and Biology*
- *Current Biology*
- *Journal of Biotechnology*

Language and Linguistics
- *English for Specific Purposes*
- *Journal of English for Academic Purposes*
- *Journal of Neurolinguistics*
- *Journal of Second Language Writing*
- *Language and Communication*
- *Language Sciences*
- *Speech Communication*
- *System*

Business
- *Industrial Marketing Management*
- *International Business Review*
- *International Journal of Information Management*
- *International Journal of Project Management*
- *International Journal of Research in Marketing*
- *Journal of Business Venturing*
- *Journal of International Management*
- *Journal of Operations Management*

Law
- *California Law Review*
- *Canadian Journal of Criminology*
- *International Review of Law and Economics*
- *Journal of Criminal Justice*

Chemistry
- *Analytica Chimica Acta*
- *Analytical Biochemistry*
- *Corrosion Science*
- *Inorganica Chimica Acta*
- *International Journal of Inorganic Materials*
- *Journal of Chemical Thermodynamics*
- *Journal of Organometallic Chemistry*
- *Journal of Solid State Chemistry*

Neuroscience
- *Cognition*
- *Brain and Cognition*
- *Neuropsychologia*
- *Neuroscience*

Computer Science
- *Computers in Human Behavior*
- *Computer Speech and Language*
- *Information and Software Technology*
- *International Journal of Human-Computer Studies*

Physics and Material Science
- *Acta Materialia*
- *Biomaterials*
- *Chemical Physics*
- *Corrosion Science*
- *International Journal of Fatigue*
- *Journal of Luminescence*
- *Journal of the Mechanics and Physics of Solids*
- *Physica C: Superconductivity*
- *Polymer*

Economics
- *Economic Modelling*
- *Journal of Economic Behavior and Organization*
- *Journal of Economics and Business*
- *Journal of Financial Economics*

Phsycology
- *Acta Psychologica*
- *Cognitive Psychology*
- *Journal of Anxiety Disorders*
- *Journal of Research in Personality*

Environmental Science
- *Applied Energy*
- *Atmospheric Environment*
- *Biomass and Bioenergy*
- *Ecological Modelling*
- *Environmental Pollution*
- *Global Environmental Change*

Public and Social Administration
- *Child Abuse & Neglect*
- *Evaluation and Program Planning*
- *Habitat International*
- *International Journal of Public Sector Management*
- *Social Science & Medicine*
- *World Development*