Stance and engagement: A corpus-based analysis of academic spoken discourse across science domains

Wenhsien Yang

Department of Applied English,
National Kaohsiung University of Hospitality and Tourism,
Kaohsiung, Taiwan
yangwenhsien@mail.nkuht.edu.tw

Abstract

Disciplinary norms or conventions could affect how a text is structured and what lexico-grammatical choices are preferable. However, relatively, discourse studies on academic spoken texts are still much underrepresented in the literature, in particular, examinations across various disciplines. Hence, the present study attempts to go some way towards filling this gap by analysing academic speech to investigate if variations exist between the soft and hard sciences with reference to Hyland's (2005) 'stance and engagement' interaction model in academic discourse. The results indicate that, unlike the distinct diversities in written discourse, the employment of hedges, boosters, self-mention and pronouns used to refer to speakers and audience are less diverse across disciplines in spoken discourse. However, with regard to word frequency and ranking, subtle differences in the use of these devices are still identified. It is believed that the various ways in which the different disciplines shape their arguments and construct their knowledge through discourse contribute to these subtle variations (Hyland & Bondi, 2006). Implications and suggestions for researching the markers to represent a speaker's stance and the audience's engagement in academic spoken discourse across disciplines as well as for teaching academic speech in ESP courses are also discussed.

1 Introduction

Research on written academic discourse has been extensive in the past few decades. A great number of studies have investigated it from a genre-based perspective (Bhatia, 2004). However, surprisingly, research on academic spoken discourse is relatively far less than that on written discourse, not to mention the fact that analysing academic speech across disciplinary variations is still rather underrepresented. Besides, academic speech is believed to be a more heterogeneous, contradictory and varied discourse, which could potentially reveal rather different phenomena from academic writing. Hence, a number of extensive databases of academic speech have been compiled to represent how academic socialisation takes place through the speech acts of graduates, academics and professionals. These include the Michigan Corpus of Academic Spoken English (MICASE) and the British Academic Spoken English (BASE) corpus (Lindemann & Mauranen, 2001)

Academic spoken genres mainly encompass the speech in lectures, conference presentations, or seminars. Many studies on spoken discourse have adopted a genre-based approach, focusing on either the macro-level rhetorical structure (Aguilar, 2004; Allison & Tauroza, 1995; Dudley-Evans, 1994; Flowerdew, 1994; Young, 1994) or micro-level lexico-grammatical devices such as lexis or phrases (Carter-Thomas & Rowley-Jolivet, 2008; Dunkel & Davis, 1994; Hansen, 1994; Lindemann & Mauranen, 2001; Thompson, 1994; Wulff, Swales & Keller, 2009). Very few have further explored the metadiscourse functions in academic speech such as the attitude, interaction or judgement of the participants (Flowerdew, 1992; Mauranen, 2002). What is more, little or no attention has been paid to the examination of either discourse or metadiscourse functions in academic speech and their comparison across disciplinary variations.

Thus, the present study aims to bridge this gap. To be specific, the following questions are investigated:
  (1) What are the keywords employed in academic speech across disciplines?
  (2) What are the words deployed to fulfil the 'stance and engagement' model in academic speech? and
Are there any differences in the keyword selection between the soft and hard sciences in representing 'stance and engagement'?

2 Literature Review
2.1 Stance and engagement

As Hyland (2004b) argues, what academics principally do in their academic activities is evaluate, a process which can be reflected in both written and spoken discourse. Thus, researchers may try to deliberately manipulate language to construct a credible identity and social relationship with the audience by claiming solidarity with recipients, evaluating their production, and acknowledging alternative viewpoints. This would make presenting a convincing argument to control the level of personality in texts one of their major concerns (Hyland, 2004b). The notion of evaluation to represent one's judgements, feeling or views about something (Hunston & Thompson, 2000) has been constantly refined by several researchers as *attitude* (Halliday, 1994), *appraisal* (White, 2003), *stance* (Hyland, 1999), and *metadiscourse* (Hyland & Tse, 2004b). In Hyland's (2005) latter work, he proposed the *stance and engagement* model, referring to the use of rhetorical devices to maintain social interaction between researchers and their audience and for researchers to present persuading judgements. These language choices are greatly affected by different researchers' assumptions regarding

the reality of existence and knowledge. Thus "every instance of evaluation has to be seen as an act socially situated in a disciplinary or institutional context" (Hyland, 2005: 175).

By *stance*, Hyland (2005) means the way in which researchers present their voice or personality and convey their judgements, opinions and commitments, while *engagement* means how researchers acknowledge the presence of their audience and relate to them in the text. Presenting *stance* can rely on the employment of four elements: *hedges, boosters, attitude markers* and *self-mentions*. Stance is rather researcher-oriented and concerns how the researcher projects him/herself in the text, while engagement is more participant-oriented, focusing on how to bring the audience into the discourse to "anticipate their possible objections and engage them in appropriate ways" (pp. 182). It is composed of five main elements: *reader pronouns*[1]*, personal asides, appeals to shared knowledge, directives* and *questions*. Although *stance* and *engagement* apparently represent the writer/speaker and the audience respectively, Hyland (2005) also argues that they are in fact two sides of the same coin and can overlap, as they both facilitate the interpersonal dimension of discourse. In addition, their marking is also highly contextualised owing to diverse assumptions of knowledge shared by the members of specific communities. Thus, researchers in different disciplines may employ rhetorical choices to represent themselves, their work and their audience in different ways. It is worthy of note that Hyland's *stance and engagement* proposition was initially applied to examine academic written texts, but is also assumed to be applicable to studying academic spoken texts, as academic speech activities also extensively involve speakers negotiating a balance between authority and concession with their audience (Poos & Simpson, 2002), in particular where their interaction requires much more directness and promptness in speaking acts.

2.2 Academic discourse across disciplinary variations

Different disciplines have their specific cultures, and these cultures may be best understood in their arbitrary conventions. Patterns of the language practices in different disciplines are "closely matched with the relevant characteristics of their associated domains of enquiry and they can transcend the institutional boundaries within any given system" (Becher, 1994: 153) as they help mediate their contexts and construct situated arguments in the disciplines that create them (Hyland, 2004b). As Hyland (2004a: 93) reminds us, "[although] the hard-soft distinction is by no means clear cut, it does offer a useful way of examining general similarities and differences between fields." In general, knowledge in the natural/hard science domain tends to be relatively analytical, structured and cumulative to establish empirical uniformities, while the soft science domain emphasises interpretation, diversity, and mutual understanding, and also allows for more tolerance on the part of readers (Becher, 1989; Hyland, 2000; Hyland 2004a). The hard science disciplines rely more on clear criteria to establish or refute hypotheses, and thus the deployment of evidential markers is rather common in this domain. On the other hand, with less reliance on general understandings and proven quantitative methods to verify claims, explicit evaluation and engagement with synthetic strategies to generate persuading discourse and personal credibility becomes central in the soft sciences (Hyland, 2005).

In comparing or contrasting the similarities and variations of rhetorical deployment across disciplinary variation, analysts, in particular Hyland, have conducted extensive studies on academic writing texts. In a 240 academic article corpus, Hyland (2004a, 2004b) found that in

---

[1] The term 'reader pronouns' is also used to refer to 'listener pronouns' in the present study.

the more discursive soft disciplines, there were many more occurrences of reader-oriented markers, in particular *reader pronouns, questions* and *asides*, than were found in hard discipline texts. On the other hand, the hard sciences have the highest proportion of using *physical directives* (over 80%) to instruct the audience to conduct research processes or to take some action in the real world, while the soft disciplines tend to use more *textual directives* to guide readers through discussions. Hyland explained that this is probably because *directives* allow scientists in the hard disciplines "to engage and lead an audience through an argument to a particular conclusion without expressing a clear authorial identity" (Hyland, 2004a: 102). Besides, nearly 70% of self-mention appeared in soft science papers owing to their extensive use of first personal pronouns. Yet, relatively, researchers in the hard science domain tend to downplay their personal role and highlight their universalistic and conceptual knowledge to generate the replicability and generality of their findings. In short, the way writers present their arguments, control their stance, and relate to their audience reflects the different social and epistemological conventions of their science domain (Hyland, 2000).

Other studies on rhetorical choices in academic writing genres across the soft-hard disciplines can mostly be found, just to name a few, in Hyland (1998, 2000, 2001, 2002, 2007, 2008) for academic research articles, in Hyland and Tse (2004a) and Yang (2012) for dissertation acknowledgements, in Yang (2013a) for academic textbook blurbs, in Yang (2013b) for academic calls for papers, and in Hyland and Tse (2009), Diani (2009) and Groom (2009) for book reviews. All of these studies exemplify how the different science disciplines make meanings and elucidate various contexts for interpretation with individual conventions within their disciplinary cultures (Hyland, 2004a).

So far, a great amount of research comparing the differences in various academic writing genres across disciplines has been carried out, and has identified the influences of disciplinary variations on how texts are constructed. However, little or no attention has been paid to academic spoken texts using similar approaches. One of the few exceptions is Poos and Simpson's (2002) comparison of the use of hedging across disciplines using an academic spoken corpus (MICASE). They found that there is more to hedge about in the soft than in the hard sciences. A closely related explanation offered by them is that norms of interaction in the soft sciences "call for presenting alternate points of view, stating and eliciting opinions, carefully crafting arguments, and allowing for multiple possibilities—all of which can and do involve the use of various hedging strategies" (pp. 14). Thus, academic discipline can be a strong indicator of the frequencies of using hedges in speech events. Furthermore, another study conducted by Simpson-Vlach (2006) also confirmed that not only are hedging devices more commonly employed in the soft than in the hard sciences, but variations in lexical and phraseological items other than hedges also occur across disciplines.

2.3 Keyword analysis
Phraseology and epistemology of academic disciplines are closely interlinked, and word analysis is a general methodology for analysing corpus-driven corpora (Groom, 2009). Lexis is believed to play a central role in authorship attribution and statistical approaches to style, and this is especially true for keywords (Davis, 2009; Kirk, 2009). Keywords are crucial as they can reflect what the text is really about in a targeted situation, or in other words, avoid trivia and insignificant detail (Scott & Tribble, 2006). Keyword analysis helps researchers identify differences between genre texts, determine the content of texts, and identify textual and rhetorical styles in a given set of academic texts across science domains (Baker, 2009;

Groom, 2009). It not only indicates the significance (or *keyness*) and importance of lexis to the texts described, but "often provides a way of identifying which words best distinguish the texts of a particular author or group of authors from another" (Hyland, 2012: 68). Besides, keywords (or the text's *aboutness*) are also closely associated with the cultures, assumptions, and value systems in academic discourse (Bondi, 2010), and can serve as valid indicators of the epistemology of a discipline in constructing, formulating, negotiating and disseminating knowledge (Malavasi & Mazzi, 2010). Thus, keyword analysis has increasingly become of interest in the study of the rhetorical features of academic discourse (Marion & White, 2005). Numerous studies investigating keywords and keyness in specialised discourse can mostly be found in Bondi and Scott's (2010) edited book or in Yang's (2011, 2013a. b.) serial research.

3 Methodology
3.1 Corpora
Two corpora were used for analysis in the present study. One, serving as the study corpus, is the corpus of British Academic Spoken English (BASE)[2] and the other, functioning as the reference corpus, is The London-Lund Corpus of Spoken English (LLC) (Svartvik, 1990). BASE is a collection of academic spoken texts including 160 lectures and 40 seminars from four broad disciplinary groups, namely arts and humanity, social studies and sciences, life and medical sciences, and physical sciences. It contains 1,614,654 tokens in total (see Table 1). For the purpose of this study, the first two disciplinary groups were classed as the 'soft sciences' while the latter two were grouped together as the 'hard sciences'. In comparison with BASE, LLC includes 1,798,044 tokens, collected mainly from casual dialogues and monologues in conversations, discussions or (non-academic) speeches.

	Soft sciences	Hard sciences
Tokens running for words	864,819	749,835
Total tokens for word list	864,766	749,817
Word types	24,130	18,146
TTR (Tokens/Word types ratio)	2.79	2.42
Standardised TTR (STTR)	35.68	33.14
STTR Standard deviation (Std. dev.)	63.07	65.66
Mean/ Word length in characters	4.46	4.34
Word length Std. dev.	2.48	2.43
Mean in words	7146.83	6942.75
Std. dev.	3307.14	3084.96

Table 1. Words used by the soft and hard sciences in the BASE corpus

3.2 Analysis
The concordancing tool WordSmith (Scott, 2008) was used to generate the wordlists, keywords and concordance of the two corpora. "Keywords analysis centres on the qualitative concordance analysis of a set of words which have been identified by a computational

---

[2] The transcriptions used in this study come from the British Academic Spoken English (BASE) corpus project. The corpus was developed at the Universities of Warwick and Reading under the directorship of Hilary Nesi and Paul Thompson. Corpus development was assisted by funding from BALEAP, EURALEX, the British Academy and the Arts and Humanities Research Council.

procedure as being statistically significant, or 'key', in a specialised corpus, when compared against a larger and more general reference corpus" (Groom, 2008: 128). Hence, to generate two sets of keywords, the soft and hard disciplines in BASE were respectively used as the individual study corpora, while LLC was treated as the referred corpus. In addition, to comply with Hyland's stance and engagement model, the lexis deployments of the two science domains were manually searched in the two wordlists with the help of the concordancing tool. The focus of this study is on the use of hedges, self-mention, personal pronouns and boosters in the texts. The lexis related to these four elements has been identified by previous studies such as Hyland (2005) and Poos and Simpson (2002).

4 Results and Discussion
4.1 Keywords in academic speech across disciplines
In total, there are 3,236 and 3,046 keywords generated respectively in the soft and hard sciences, of which 3,153 and 2,950 are overused keywords, while the remaining 83 and 96 are underused. It is the overused keywords that represent the uniqueness or *aboutness* of the present corpus compared to the LLC. The top 50 overused keywords in each science domain are discussed further (see Appendix 1). When cross-examined, 37 of these 50 overused keywords were found to be identical in both domains, i.e. 74%, which suggests that the preferred lexis for utterance is divergent to some extent across disciplines. Among the identical categories, firstly, the indicative personal pronoun *you* is highly ranked in both domains, which is largely different from academic writing where *you* is rarely used, due to the lack of involvement of the audience in academic written texts (Hyland, 2005). However, in spoken discourse the pronoun *you* is extensively employed to bring the listener into the discourse. In addition, the pronoun *we* is a frequent device used to bind the writer and reader in academic writing (Hyland, 2005), whereas it is mainly used to refer to the speaker(s) in academic speech. A possible explanation is that readers are always absent and tend to be specialists, whereas listeners are present and very likely have mixed backgrounds, that is, both specialists and non-specialists are often present on speech occasions. In other words, compared to writers speaking to an ideal, future, or putative audience, speakers face an audience whose level of knowledge is uncertain to them (Aguilar, 2004). Thus, *we* can help represent an inclusive voice, while *you* signals an exclusive distance between the speaker and listener.

Secondly, the determiners such as *the, this,* or *that* are also highly ranked in the speech of both science domains. Indeed, both definite articles and demonstratives are a very common category in English parts of speech. Their overuse here indicates the fact that relevance and clarity usually override brevity (Aguilar, 2004); hence, these determiners recurrently help the speaker to clarify, emphasise, remind and centre the ideas expressed verbally. This device is used more frequently in the soft than the hard disciplines.

Thirdly, similar to the previous research, *just* and *if* are also two highly-ranked keywords in both domains. As Lindemann and Mauranen (2001) argue, the word *just* in speech often serves as a minimiser, limiter or mitigator, and is ambiguous with blurred boundaries between categories. Thus, it is always associated with metadiscourse and hedging. *If*-conditionals are rather valuable devices in academic discourse as they not only build up a hypothesis for envisaging alternatives, but also allow space for researchers to set up an alternative argument to handle and situate the research claim (Carter-Thomans & Rowley-Jolivet, 2008). In speech, *if*-clauses are useful for the speaker to redirect the audience in a polite way when structuring the speech is required. Furthermore, this device also helps the speaker signal to the audience

when to focus on visual designs such as slides, handouts or screens, which commonly accompany academic presentations. The present corpus shows that the percentages of these two keywords used in the hard sciences are slightly higher than in the soft sciences. One possible reason is that making hypotheses could be a more conventional practice in research in the hard sciences.

Lastly, the difference in the top 50 keywords between the two science domains could more clearly explain the variations in word preference of the two areas. In soft disciplines, plurality, possibility, and acknowledging the existence of the past and humans are valued. Thus, words such as *as*, *know* (combined with *you* into '*as you know*'), *people*, *was*, *more*, *some*, *kind* or *like* (the preposition) would become key. However, in the hard sciences, singularity, absoluteness, causality, the present and the future are relatively emphasised; hence, words such as *one*, *two*, *then*, *now*, *will* or *going* are more frequently employed. To conclude this section, it is found that the top 50 keywords used in both sciences are largely similar (74%), but slightly different preferences were still identified, probably due to the differences in reality and knowledge assumed by the two science domains.

4.2 Markers representing the speaker's *stance* and *engagement* in academic speech

In this section, the devices of self-mention, hedging and boosters in the speaker's stance are discussed. First, as shown in Table 2, the speakers in the soft sciences tend to use hedges more frequently than those in the hard sciences by nearly 3,000 occurrences. The major functions of hedges are to present information as personal opinion rather than as validated fact, and to allow the audience space to dispute or refute the speaker's interpretation (Hyland, 2005). Thus, identical to the previous study (Poos & Simpson, 2002), the present results also confirm that hedging devices appear more commonly in academic spoken discourse of the soft sciences than of the hard sciences. This imbalanced deployment indicates "less assurance about what colleagues could be safely assumed to accept" (Hyland, 2005: 188) in the soft science disciplines, while reporting systems in the hard science disciplines are relatively formalised so hedging use is minimised. However, interesting findings of the present research regarding differences between the two domains include nearly double the frequency of the use of *I think* (1,930 v. 1,118), five times the use of *kind of* (1,530 v. 387) and double the occurrence of *sort of* (2,437 v. 1,181), which reflects the epistemological emphasis on personal interpretation, uncertainty and the acceptance of diversity in the soft sciences.

	Soft sciences	Hard sciences		Soft sciences	Hard sciences
may	764	637	*assume*	81	151
might	990	784	*believe*	272	100
could	1,218	923	*argue*	286	41
perhaps	443	327	*I think*	1,930	1,118
probably	433	598	*kind of*	1,530	387
apparently	62	46	*sort of*	2,437	1,181
possibly	304	288	*..far as I know*	6	10
likely	137	157	*we feel that…*	64	29
seem	365	202	*if*	4,238	5,211
suggest	232	110	*if I…*	163	262
indicate	37	57	*if we…*	251	577

| *appear* | 96 | 99 | *if you* | 2,059 | 2,332 |
| *suppose* | 256 | 185 | **Total** | **18,654** | **15,812** |

Table 2. Frequency of the hedges used by the two science domains in the BASE wordlist

Table 3 shows the frequency of using self-mention in both domains. Speakers may use the first personal pronouns to present propositional, affective and interpersonal information, and they are more often employed in soft science papers as they help writers identify themselves with the particular argument they are making and thus obtain an individual credit (Hyland, 2005). However, the current results show a different perspective; in total, *self-mention* is slightly more commonly used in the academic speech of hard scientists. A further examination of the associated lexis of *self-mention* found that in the hard disciplines, *I mean, I'm going to, I'm sure, I'm saying, we have to* (followed by a verb)*,* and *we have* (leading to an explanation of research procedures) are very common phrases used in the hard sciences, which apparently signifies a stronger authoritative voice without permitting too much space for negotiating or refuting. Yet, in the soft disciplines, *we can* (leading to a suggestion), *we don't* (followed by an open discussion *e.g. We don't want you to assume that we have…*), *I think, I guess* or *I suppose* are more frequently deployed, which also indicates that interpretation, diversity and ambiguity are the basis of knowledge construction in this domain. Hence, in spoken discourse, disciplinary variations in *self-mention* are still identified, although this element, unusually, has a slightly higher occurrence in the hard disciplines compared to its deployment in academic writing (Hyland, 2005).

The last element of *stance* studied in this research is *boosters*. Both *hedge*s and *boosters* are communicative strategies to increase or reduce the force of statements, but *boosters* permit speakers or writers "to express their certainty in what they say and to mark involvement and solidarity with their audience" (Hyland, 2000: 87). As seen in Table 4, speakers in the soft sciences tend to use *boosters* more frequently than those in the hard sciences. Appropriate and interchangeable employment of hedges and boosters can balance objective information, subjective judgement and interpersonal negotiation in the discourse, thus making the claims more acceptable (Hyland, 2005). The present research shows that both elements more usually appear in the speech of the soft rather than the hard sciences, which may imply that speakers in the soft sciences would more liberally use these rhetorical devices to create a convincing argument for their participants because in their fields there is less control of variables but greater possibilities for diverse outcomes. Therefore, the speakers need to clearly express their judgements to establish an understanding with their audience (Hyland, 2005). Yet, speakers in the hard sciences generally follow their succinct styles of reporting facts, which mostly results from rigorously-controlled experiments, and thus may minimise the deployment of these two strategies, opening the way for negotiation.

	Soft sciences	**Hard sciences**
I (X, would, will, am, have)	12,424	11,847
we (X, would, will, am, have)	7,289	11,577
Total	**19,713**	**23,423**

Table 3. Frequency of the usage of self-mention in the two science domains in BASE

	Soft sciences	**Hard sciences**
clearly/ clear	221/ 224	146/ 150
certainly/ certain	287/ 389	213/ 192
definitely/ definite	50/ 20	27/ 11
obviously/ obvious	361/ 103	442/ 122
the fact that/ in fact	268/ 350	134/ 351
always	425	336
show that	41	20
substantially/ substantial	7/ 22	5/ 31
demonstrate	19	21
Total	**1,679/1,593**	**1,344/1,234**

Table 4. Frequency of the usage of boosters in the two science domains in BASE

The only element of *engagement* studied in the present research is the use of the second personal pronouns *you* and *your* (see Table 5), one major lexis of *reader pronouns*. This element is the most explicit way for speakers to bring their audience into a discourse, and *you* and *your* are, indeed, the clearest way in which a speaker can acknowledge the audience's presence (Hyland, 2005), which seems unavoidable in a face-to-face speech situation, compared to in a written text. Hence, it would be more understandable to see that the occurrences of these two words in both science domains are nearly identical; in other words, there is no obvious variation between the two domains. This phenomenon is rather different from the case reported in academic writing (Hyland, 2005), where the reader pronouns are relatively far less employed in the hard sciences, as the writers naturally downplay the presence of their readers. Another difference is the use of *we*, also considered as one of the reader pronouns when used to include both reader and writer, which is a very frequent engagement feature in writing texts; however, as previously mentioned, *we* is majorly deployed to refer to the speakers themselves as a *self-mention* rather than as a *reader pronoun* in spoken texts. Thus, it can be argued that the use of *you* and *your* does not boldly reflect disciplinary variations in academic spoken discourse so far. Probably, the real presence of the audience increases this usage in both science domains.

	Soft sciences	**Hard sciences**
You	15,108	14,634
You're	1,323	1,304
You've	837	1,276
You'll	385	360
You'd	158	214
Your	1,710	1,804
Total	**19,521**	**19,595**

Table 5. Frequency of using the second personal pronoun *you* in the two science domains

4.3 *Keyness* in stance and engagement

Although we have compared some lexical or phraseological items of *stance* and *engagement* features across the two science domains, not necessarily all of them become *key* or have the same strength of *keyness* in its individual corpus. Tables 6, 7, and 8 exhibit the *keyness* of

hedges, personal pronouns and boosters ranked differently in the two domains. For instance, the hedges, *seem* and *feel*, are key in the soft disciplines but not in the hard sciences, while *assume, suppose* and *argu*e have higher divergence of keyness in the different domains. This preference still points out the influence of what the different sciences believe in terms of value system and knowledge construction. Apparently, speakers in the soft disciplines rely more frequently on using hedges to express their acceptance of negotiation and interpretation.

Besides, the use of personal pronouns also evidences that speakers in the soft disciplines are relatively more willing to project self in their speech to present identity and gain personal credibility. Yet, *we* has higher keyness in the hard disciplines because research in this domain often needs much more financial investment and collaborative work (Hyland, 2005). Similarly, general speech in the soft sciences tends to deploy *boosters* more commonly as they help highlight the sharedness, membership and engagement in interaction with the audience (Hyland, 1999). Together with the preceding discussion of the wordlists, the keywords in the elements of *stance* and *engagement* under examination both reflect some variation in linguistic choices between the two science domains.

	Soft sciences	Hard sciences		Soft sciences	Hard sciences
may	242	285	*assume*	1,114	433
might	117	141	*believe*	384	835
could	170	247	*argue*	709	1,450
perhaps	153	199	*I think*	74	192
probably	492	218	*kind of*	46	510
apparently	2,244	1,865	*sort of*	54	234
possibly	374	411	*feel*	1,787	X
likely	807	557	*if*	26	15
seem	771	X	*appear*	1,320	1,232
suggest	723	1,079	*suppose*	526	1,434
indicate	X	1,920			

Table 6. Keyness ranking of the hedges in BASE with reference to LLC

Note: x denotes not a keyword

	Soft sciences	Hard sciences
I	766	1,040
you	10	9
we	44	11

Table 7. Keyness ranking of the personal pronouns in BASE with reference to LLC

	Soft sciences	**Hard sciences**
clearly/ clear	297/ 442	446/ 706
certainly/ certain	574/ 302	850/ 975
definitely/ definite	2,465/ x	x/ x
obviously/ obvious	480/ 1,153	258/ 773
the fact that/ in fact	208	323
always	863	1,182
show that	307	335
substantially/ substantial	x/ 1,923	x/ 1,377
demonstrate	2,854	2,541

Table 8. Keyness ranking of the boosters in BASE with reference to LLC

Note: x denotes not a keyword

5 Implication and conclusion

This corpus-driven study investigated a less-attended academic discourse, namely academic speech, from the perspective of comparing the different use of *stance* and *engagement* across soft and hard science disciplines. Two corpora, BASE and LLC, were used to generate the wordlists and keywords with the help of a concordancing tool. Four elements, *hedges, self-mention, reader pronouns,* and *boosters* in the academic corpus were examined, and the results indicate that subtle differences in the use of these devices were identified between the two domains. It is believed that the various ways in which the different disciplines shape their arguments and construct their knowledge through discourse contribute to these subtle variations (Hyland & Bondi, 2006). So far, this study has shown how corpus analysis can be helpful in understanding the ways academics across disciplines talk, and it also offers some implications for ESP (English for Specific Purposes) practitioners and researchers.

First, the present analysis confirms that the discourse of the academy contains many discipline-specific regularities, preferences or assumptions, affecting the ways in which identities and arguments are presented conventionally and effectively. By making students aware of these rhetorical features, teachers can not only enhance their understanding of disciplinary cultures but can prepare them for producing their own convincing arguments in their field (Hyland, 2000). Academic speech is largely different from writing as it is instantaneously interactive. Usually, there is little time or opportunity to correct or ponder. Speakers are under pressure of presenting their arguments, negotiating with the audience and finishing in time, which makes a credible speech a challenge, in particular for novice research students. Thus, a tailor-made ESP course can be designed to accommodate learners' needs in developing communicative skills and strategies in spoken English for academic purposes.

Second, as Hyland (2004a) and Johns (1997) suggest, ESP teachers can raise postgraduate students' awareness of rhetorical consciousness by having them do small-scale corpus analysis of the texts they need to master. Teachers can draw students' attention to the features of academic speech they have to engage in, and direct them to discover what communicative strategies from their specific discipline they can apply to present convincing oral arguments. These tasks can enable learners "to recognise both the choices available to them and their impact" (Hyland, 2004a: 110).

Third, and finally, since this research could be one of the very few studies investigating the interaction model of stance and engagement in academic speech discourse, several further studies can be carried out to complement and build on it. For instance, the present study examined the use of words and keywords in four elements at a descriptive level; thus, a closer look at the pragmatic use of all the elements in the corpus would more clearly reflect how *stance* and *engagement* function in the texts of both science domains. Next, as two large corpora of academic spoken English have already been established, namely the American MICASE and the British BASE, it would be interesting to compare and contrast how stance and engagement elements are realised in these two *Englishes*. Lastly, another direction of examination can centre on the use of *stance* and *engagement* across the contexts in which English is used as a native and non-native language. This would shed light on how contextual differences and disciplinary variations may interplay to shape academic spoken discourse in English.

6 References

Alison, D., & Tairoza, S. (1995). The effect of discourse organisation on lecture comprehension. *English for Specific Purposes,* 14(2): 157-173.

Aguilar, M. (2004). The peer seminar, a spoken research process genre. *Journal of English for Academic Purposes,* 3: 55-72.

Baker, P. (2009). The question is, how cruel is it? Keywords, fox hunting and the House of Commons. In D. Archer (Ed.), *What's in a word-list? Investigating word frequency and keyword extraction* (pp. 125-136). Surrey: Ashgate.

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

Becger, T. (1994). The significance of disciplinary differences. *Studies in Higher Education,* 19(2): 151-161.

Bondi, M. (2010). Perspectives on keywords and keyness: An introduction. In M. Bondi & M. Scott (Eds.), *Keyness in texts* (pp. 1-20). Amsterdam: John Benjamins.

Bondi, M., & Scott, M. (2010). *Keyness in texts.* Amsterdam: John Benjamins.

Carter-Thomas, S., & Rowley-Jolivet, E. (2008). *If*-conditionals in medical discourse: From theory to disciplinary practice. *Journal of English for Academic Purposes,* 7: 191-205.

Davis, M. (2009). Word frequency in context: Alternative architectures for examining related words, register variation and historical change. In D. Archer (Ed.), *What's in a word-list? Investigating word frequency and keyword extraction* (pp. 53-68). Surrey: Ashgate.

Diani, G. (2009). Reporting and evaluation in English book review articles: A cross-disciplinary study. In K. Hyland & G. Diani (Eds.), *Academic evaluation: Review genres in university settings* (pp. 105-121). Basingstoke: Palgrave MacMillan.

Dillon, G. (1991). Contending rhetorics: Writing in academic disciplines. Bloomington, IN: Indiana University Press.

Dudley-Evans, A. (1994). Variations in discourse patterns favoured by different disciplines and their pedagogic implications. In J. Flowerdew (Ed.), *Academic listening, Research perspectives* (pp. 146-158). Cambridge: Cambridge University Press.

Dunkel, P. A., & Davis, J. N. (1994). The effects of rhetorical signalling cues on the recall of English lecture information by speakers of English as a native or second language. In J. Flowerdew (Ed.), *Academic listening, Research perspectives* (pp. 55-74). Cambridge: Cambridge University Press.

Flowerdew, J. (1992). Definition in science lectures. *Applied Linguistics,* 13: 202-221.

Flowerdew, J. (1994). *Academic listening, Research perspectives*. Cambridge: Cambridge University Press.

Groom, N. (2009). Phraseology and epistemology in academic book reviews: A corpus-driven analysis of two humanities disciplines. In K. Hyland & G. Diani (Eds.), *Academic evaluation: Review genres in university settings* (pp. 122-142). Basingstoke: Palgrave MacMillan.

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

Hansen, C. (1994). Topic identification in lecture discourse. In J. Flowerdew (Ed.), *Academic listening, Research perspectives* (pp. 131-145). Cambridge: Cambridge University Press.

Hunston, S., & Thompson, G. (2000). *Evaluation in text*. Oxford: Oxford University Press.

Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. *Text & Talk,* 18(3): 349-382.

Hyland, K. (1999). Disciplinary discourse: Writer stance in research article. In C. Candlin & K. Hyland (Eds.), *Writing: Texts, processes and practices* (pp. 99-121). London: Longman.

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

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20(3): 207–26.

Hyland, K. (2002). Options of identity in academic writing. *ELT Journal,* 56(4): 351-358.

Hyland, K. (2004a). A convincing argument: Corpus analysis and academic persuasion. In U. Connor & Upton, T. A. (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 87-112). Amsterdam: John Benjamins.

Hyland, K. (2004b). *Disciplinary discourse: Social interactions in academic writing*. Ann Arbour, MI: The University of Michigan Press.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies,* 7(2): 173-192.

Hyland, K. (2007). Different strokes for different folks: Disciplinary variation in academic writing. In K. Flottum (Ed.), *Language and Discipline Perspectives on Academic Discourse* (pp. 89–108). Newcastle: Cambridge Scholars Publishing.

Hyland, K. (2008). "As can be seen": Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1): 4–21.

Hyland, K. (2012). *Disciplinary identities: Individuality and community in academic discourse.* Cambridge: Cambridge University Press.

Hyland, K., & Bondi, M. (2006). *Academic discourse across disciplines*. Bern: Peter Lang.

Hyland, K., & Tse, P. (2004a). "I would like to thank my supervisor": Acknowledgements in graduate dissertations. *International Journal of Applied Linguistics,* 14(2): 259-275.

Hyland, K., & Tse, P. (2004b). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics,* 25(2): 156-177.

Hyland, K., & Tse, P. (2009). Discipline and gender: Constructing rhetorical identity in book reviews. In K. Hyland & G. Diani (Eds.), *Academic evaluation: Review genres in university settings* (pp. 87-104). Basingstoke: Palgrave MacMillan.

Johns, A. M. (1997). *Text, role and context: Developing academic literacies.* Cambridge: Cambridge University Press.

Kirk, J. M. (2009). Word frequency: Use or misuse? In D. Archer (Ed.), *What's in a word-list? Investigating word frequency and keyword extraction* (pp. 17-34). Surrey: Ashgate.

Lindemann, S., & Mauranen, A. (2001). "It's just real messy": The occurrence and function of *just* in a corpus of academic speech. *English for Specific Purposes,* 20: 459-475.

Malavasi, D., & Mazzi, D. (2010). History v. marketing: Keywords as a clue to disciplinary epistemology. In M. Bondi & M. Scott (Eds.), *Keyness in texts* (pp. 169-184). Amsterdam: John Benjamins.

Martin, J. P., & White, P. P. R. (2005). *The language of evaluation: Appraisal in English.* Basingstoke: Palgrave MacMillan.

Mauranen, A. (2002). "A good question": Expressing evaluation in academic speech. In G. Cortese & P. Riley (Eds.), *Domain-specific English: Textual communities and classrooms* (pp. 115-140). Frankfurt: Peter Lang.

Poos, D., & Simpson, R. (2002). Cross-disciplinary comparisons of hedging: Some findings from the Michigan Corpus of Academic Spoken English. In R. Reppen, S. Fitzmaurice & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 3-24). Amsterdam: John Benjamins.

Scott, M. (2008). *WordSmith Tools* (Version 5) [Computer software]. Liverpool: Lexical Analysis Software.

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

Simpson-Vlach. R. (2006). Academic speech across disciplines: Lexical and phraseological distinctions. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp.295-316). Amsterdam: John Benjamins.

Svartvik, J. (ed). (1990). The London Corpus of Spoken English: Description and Research. *Lund Studies in English 82.* Lund, Sweden: Lund University Press.

Thompson, S. (1994). Frameworks and contexts: A genre-based approach to analysing lecture introductions. *English for Specific Purposes,* 13(2): 171-186.

White, P. (2003). Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text,* 23(2): 2594-2598.

Wulff, S., Swales, J. M., & Keller, K. (2009). "We have about seven minutes for questions": The discussion session from a specialised conference. *English for Specific Purposes,* 28(2): 79-92.

Yang, W-H. (2011). Enhancing authenticity in composing informative and promotional texts by analysing keywords in a genre-based writing course. *Journal of English Studies,* 6: 141-161.

Yang, W-H. (2012). A genre analysis of PhD dissertation acknowledgements across disciplinary variations. *LSP Journal,* 3(2): 51-69.

Yang, W-H. (2013a). Keyness in academic textbook blurbs: Lexical variations across discipline. In *Proceedings of 2012 Language and Language Teaching Conference* (pp. 63-72). Pingtung, Taiwan: National Pingtung University of Science and Technology.

Yang, W-H. (2013b). Two-folded messages behind CFP: A cross disciplinary study. *International Journal of Language Studies,* 7(2): 83-108.

Young, L. (1994). University lectures-macro-structure and micro-features. In J. Flowerdew (Ed.), *Academic listening, Research perspectives* (pp. 159-176). Cambridge: Cambridge University Press.

7 Appendix
Top 50 keywords in the two science domains of BASE with reference to LLCS

Soft Science								
N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P
1.00	THE	48674.00	5.63	100.00	21445.00	1.19	41161.87	0.00
2.00	OF	29843.00	3.45	100.00	12609.00	0.70	25829.86	0.00
3.00	AND	24120.00	2.79	100.00	13182.00	0.73	16400.03	0.00
4.00	TO	23065.00	2.67	100.00	14099.00	0.78	13827.69	0.00
5.00	IN	17647.00	2.04	100.00	9485.00	0.53	12158.64	0.00
6.00	IT'S	5284.00	0.61	100.00	0.00		11907.02	0.00
7.00	THAT	17210.00	1.99	100.00	9387.00	0.52	11674.02	0.00
8.00	IS	12989.00	1.50	100.00	6510.00	0.36	9568.26	0.00
9.00	SO	7491.00	0.87	100.00	2248.00	0.13	8123.73	0.00
10.00	YOU	15108.00	1.75	100.00	10789.00	0.60	7356.25	0.00
11.00	ABOUT	4213.00	0.49	100.00	778.00	0.04	5778.81	0.00
12.00	THIS	7686.00	0.89	100.00	3789.00	0.21	5733.21	0.00
13.00	THAT'S	2501.00	0.29	100.00	0.00		5630.32	0.00
14.00	WHAT	5621.00	0.65	100.00	2649.00	0.15	4366.18	0.00
15.00	OKAY	1912.00	0.22	88.00	9.00		4196.03	0.00
16.00	DON'T	1770.00	0.20	98.00	0.00		3983.66	0.00
17.00	AS	5786.00	0.67	100.00	3139.00	0.17	3918.33	0.00
18.00	WHICH	4467.00	0.52	100.00	1847.00	0.10	3875.86	0.00
19.00	ARE	5047.00	0.58	100.00	2412.00	0.13	3869.20	0.00
20.00	BECAUSE	2612.00	0.30	100.00	419.00	0.02	3773.31	0.00
21.00	BUT	5822.00	0.67	100.00	3316.00	0.18	3741.85	0.00
22.00	OR	4404.00	0.51	100.00	2113.00	0.12	3362.72	0.00
23.00	KNOW	3928.00	0.45	100.00	1696.00	0.09	3288.93	0.00
24.00	THERE'S	1454.00	0.17	97.00	0.00		3272.10	0.00
25.00	FOR	5362.00	0.62	100.00	3224.00	0.18	3239.44	0.00
26.00	IF	4238.00	0.49	100.00	2174.00	0.12	3035.32	0.00
27.00	YOU'RE	1323.00	0.15	91.00	0.00		2977.16	0.00
28.00	NOT	4090.00	0.47	100.00	2131.00	0.12	2883.96	0.00
29.00	HAVE	5128.00	0.59	100.00	3300.00	0.18	2851.90	0.00
30.00	THEY	5630.00	0.65	100.00	4048.00	0.23	2696.57	0.00
31.00	I'M	1183.00	0.14	97.00	1.00		2646.62	0.00
32.00	LIKE	2948.00	0.34	100.00	1166.00	0.06	2645.92	0.00
33.00	ON	5628.00	0.65	100.00	4283.00	0.24	2476.31	0.00
34.00	FROM	3179.00	0.37	100.00	1493.00	0.08	2473.40	0.00
35.00	THEY'RE	1097.00	0.13	93.00	0.00		2468.39	0.00
36.00	DO	3291.00	0.38	100.00	1612.00	0.09	2463.15	0.00
37.00	WITH	4162.00	0.48	100.00	2599.00	0.14	2401.27	0.00
38.00	WAS	5785.00	0.67	100.00	4647.00	0.26	2334.27	0.00
39.00	IT	10897.00	1.26	100.00	11936.00	0.66	2299.92	0.00
40.00	YEAH	1710.00	0.20	75.00	342.00	0.02	2267.63	0.00
41.00	PEOPLE	2460.00	0.28	100.00	943.00	0.05	2260.09	0.00
42.00	BY	2841.00	0.33	100.00	1332.00	0.07	2213.36	0.00

43.00	WE'RE	983.00	0.11	92.00	0.00		2211.79	0.00
44.00	WE	5179.00	0.60	100.00	4082.00	0.23	2154.89	0.00
45.00	SOME	2331.00	0.27	100.00	887.00	0.05	2153.13	0.00
46.00	KIND	1577.00	0.18	87.00	297.00	0.02	2143.55	0.00
47.00	MORE	2181.00	0.25	100.00	774.00	0.04	2117.49	0.00
48.00	HE'S	909.00	0.11	78.00	0.00		2045.24	0.00
49.00	BE	5590.00	0.65	100.00	4836.00	0.27	1981.08	0.00
50.00	JUST	3053.00	0.35	100.00	1765.00	0.10	1926.45	0.00

Hard science								
N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P
1.00	THE	39187.00	5.23	99.00	21445.00	1.19	32915.53	0.00
2.00	AND	21840.00	2.91	99.00	13182.00	0.73	16484.86	0.00
3.00	OF	19810.00	2.64	99.00	12609.00	0.70	14124.82	0.00
4.00	SO	9684.00	1.29	99.00	2248.00	0.13	13782.12	0.00
5.00	TO	20172.00	2.69	99.00	14099.00	0.78	12943.62	0.00
6.00	IS	13297.00	1.77	99.00	6510.00	0.36	12088.58	0.00
7.00	IT'S	4880.00	0.65	98.00	0.00		11960.55	0.00
8.00	THAT	14915.00	1.99	99.00	9387.00	0.52	10724.89	0.00
9.00	YOU	14634.00	1.95	99.00	10789.00	0.60	8758.63	0.00
10.00	IN	13030.00	1.74	99.00	9485.00	0.53	7914.25	0.00
11.00	WE	8065.00	1.08	99.00	4082.00	0.23	7104.79	0.00
12.00	THAT'S	2893.00	0.39	99.00	0.00		7085.11	0.00
13.00	OKAY	2879.00	0.38	98.00	9.00		6935.19	0.00
14.00	THIS	7362.00	0.98	99.00	3789.00	0.21	6389.63	0.00
15.00	IF	5211.00	0.69	99.00	2174.00	0.12	5329.83	0.00
16.00	ABOUT	3453.00	0.46	99.00	778.00	0.04	4960.46	0.00
17.00	ARE	5138.00	0.69	99.00	2412.00	0.13	4807.13	0.00
18.00	DON'T	1767.00	0.24	99.00	0.00		4325.60	0.00
19.00	CAN	4249.00	0.57	99.00	1910.00	0.11	4109.98	0.00
20.00	WHAT	4779.00	0.64	99.00	2649.00	0.15	3872.48	0.00
21.00	WE'RE	1566.00	0.21	97.00	0.00		3833.26	0.00
22.00	I'M	1571.00	0.21	98.00	1.00		3829.48	0.00
23.00	HAVE	5279.00	0.70	99.00	3300.00	0.18	3796.69	0.00
24.00	DO	3750.00	0.50	99.00	1612.00	0.09	3749.77	0.00
25.00	HERE	2277.00	0.30	97.00	378.00	0.02	3664.93	0.00
26.00	THERE'S	1407.00	0.19	97.00	0.00		3443.85	0.00
27.00	TWO	2810.00	0.37	99.00	913.00	0.05	3368.52	0.00
28.00	BECAUSE	2172.00	0.29	99.00	419.00	0.02	3316.24	0.00
29.00	YOU'RE	1304.00	0.17	96.00	0.00		3191.61	0.00
30.00	GOING	3055.00	0.41	99.00	1255.00	0.07	3154.81	0.00
31.00	YOU'VE	1276.00	0.17	96.00	0.00		3123.05	0.00
32.00	WE'VE	1229.00	0.16	96.00	0.00		3007.96	0.00
33.00	THEN	2853.00	0.38	99.00	1214.00	0.07	2872.48	0.00
34.00	ONE	4432.00	0.59	99.00	3058.00	0.17	2853.48	0.00
35.00	YEAH	1805.00	0.24	81.00	342.00	0.02	2773.75	0.00
36.00	WHICH	3376.00	0.45	99.00	1847.00	0.10	2766.41	0.00
37.00	THESE	2190.00	0.29	98.00	693.00	0.04	2663.95	0.00
38.00	THEY'RE	1081.00	0.14	93.00	0.00		2645.58	0.00

39.00	FOR	4396.00	0.59	99.00	3224.00	0.18	2627.79	0.00
40.00	BUT	4426.00	0.59	99.00	3316.00	0.18	2574.85	0.00
41.00	WITH	3863.00	0.52	99.00	2599.00	0.14	2559.89	0.00
42.00	I'VE	1016.00	0.14	96.00	0.00		2486.44	0.00
43.00	JUST	3015.00	0.40	99.00	1765.00	0.10	2315.22	0.00
44.00	GET	2535.00	0.34	98.00	1235.00	0.07	2297.51	0.00
45.00	IT	9682.00	1.29	99.00	11936.00	0.66	2293.76	0.00
46.00	OR	3223.00	0.43	99.00	2113.00	0.12	2198.04	0.00
47.00	NOW	2233.00	0.30	99.00	1037.00	0.06	2103.19	0.00
48.00	WILL	2010.00	0.27	99.00	822.00	0.05	2080.90	0.00
49.00	NOT	3129.00	0.42	99.00	2131.00	0.12	2043.46	0.00
50.00	ON	4687.00	0.63	99.00	4283.00	0.24	2042.81	0.00

***