

# Current Trends in Corpus-based Discourse Analysis in Britain

Yasunori Nishina

## Abstract

In this paper, I will first summarize several historical and recent works by past and present researchers at the University of Birmingham (hereafter, the UoB). Few of these researchers would call themselves corpus linguists. Rather, they would describe themselves as discourse analysts, lexicographers, or ESP practitioners, in view of the fact that they more commonly utilize corpus linguistic approaches in the identification and application of qualitative aspects of authentic language use. As such, I will explain how the purpose of their research is not always quantitative in nature.

The historical tradition in the Birmingham school follows an idea, first suggested by J. R. Firth, that “a word is always contextual.” In the second half of the paper, I will show how this essence still lives in much of the current research at the UoB, linking it with my own work on the interface between corpus linguistics and academic discourse analysis. In particular, I will demonstrate how corpus linguistics and discourse analysis can harmonize with one another. I will then argue that this combination can be an effective approach for identifying the link between language and culture and particularly revealing the detailed cultural values of a particular discipline.

## 1. The UoB’s Perspective of Corpus Linguistics

### 1.1 Corpus Linguistics for the UoB

Teubert and Cermakova (2007) assume that corpus linguistics sees both language and meaning as “a social phenomenon” concerned with the message and meaning (p.37). However, Chomskyan and cognitive linguistics consider language to be a psychological and mental phenomenon concerned with understanding “what happens in the mind in the process of encoding and decoding a message” (p. 39). Understanding is a personal action by both speakers and hearers, who “translat[e] a word, a sentence, a text into the language of thought, into mentalese” (p. 38), corpus linguistics prioritizes different interests than Chomskyan and cognitive linguistics.

Indeed, even among members of the UoB, determining whether corpus linguistics is a method (e.g., Thompson & Hunston, 2006) or a particular theory of language (Teubert, 2005) remains a divisive problem, although the majority of the UoB members see corpus linguistics as a method for textual analysis. Various approaches and interests in seeking meaning in the message have emerged within corpus linguistics, including sophisticated quantification techniques (e.g., Biber, 1988; Gries, 2008), details of individual word use (e.g., Teubert, 2008), and the writer’s stance by pattern (e.g., Charles 2004, 2006 a, 2006 b).

## 1.2 A Brief History in Corpus Linguistics at the UoB

John Sinclair was a leading proponent of the early corpus project entitled English Lexical Studies in Edinburgh in 1963, more recently referred to as the OSTI Report (Krishnamurthy, 2004). He was also the first person to conduct a lexical investigation based on a corpus, defining the important role of collocation by following fundamental ideas put forth by Palmer and Hornby (1933) and Firth (1957) (for details, see Teubert & Cermakova, 2007, pp. 53–54). In other words, when Sinclair moved to the UoB, corpus linguistics at the UoB gave its first cry.

Since the 1970s, the UoB has developed major corpora, including the Birmingham Collection of English Text (17 million words) developed in the 1980s and the Bank of English (BoE) in the 1990s. The UoB also launched the COBUILD project with HarperCollins and has developed corpus-based lexicography since the 1980s. Sinclair's innovative corpus-based studies contributed to both practical aspects in developing a better dictionary (Sinclair, 1987) and theoretical aspects via his innovative approach to word-meaning in English (Sinclair, 1991).

## 1.3 Traditional Concept in Birmingham School

The main idea in Sinclair's works (see details in Stubbs, 2009) follows the traditional idea for a word-meaning in Firth (1935): "the complete meaning of a word is always contextual, and no study of meaning apart from a complete context can be taken seriously" (p. 37). The significant studies in Sinclair (1991, 2004) elaborate upon the ways in which meaning cannot belong to individual words, and word meaning is always defined by its co-texts and expressed by 'unit of meaning' (Sinclair 1991, 2004). After Sinclair passed away, Susan Hunston and Wolfgang Teubert deepened his idea of unit of meaning in their research and teaching. Teubert (2003), for instance, points out that "no word has a meaning except when it is encountered in context" (p. 9), while Hunston (2011) assumed that "the meaning of any word cannot be identified reliably if the word is encountered in isolation" (p. 14).

The idea that "a word is always contextual" is comprehensively illustrated in Sinclair's corpus-based studies on lexical items (Sinclair, 1991; Stubbs, 2001). To identify the unit of meaning, many of Sinclair's studies prioritized the observing of sequences of words of varying degrees of fixedness with flexible boundaries as well as identifying the semantic similarity in the co-texts of a target word and phrase. Unit of meaning is, for instance, realized through the collocation between words (e.g., *food + assistance*) and colligation between words and grammatical categories (e.g., ADJ + *about* / ADJ + PREP). Corpus linguistics revealed that texts are constructed from such sequences of words (e.g., 55.38% in whole texts; Erman & Warren, 2000) and that lexis and grammar are closely linked as patterns (Hunston & Francis, 1999).

Corpus linguistics further revealed that such sequences of words are imbued by semantic and discursal functions and express writer's stance in texts (e.g., Charles, 2004, 2006a, 2006b; Hunston & Sinclair, 2000; Nattinger & DeCarrico, 1992; Pawley & Syder, 1983). Such discursal features of a word (or sequence of words) are termed as semantic preferences between words and lexical sets (e.g., *in + TIME*) and semantic prosody that expresses "the consistent aura of

meaning” of a word by its collocates (Louw, 1993, p. 157) as well as the speaker’s attitude (e.g., *naked eye*).

[Instances of Positive Semantic Prosody]

- *provide* expresses the speaker’s approval of what they regard as valuable things, such as *help, relief, food, shelter, information* (Stubbs 1995, 2001).
- *undergo* expresses the speaker’s sympathy with someone who is forced to suffer an unpleasant experience: *had to undergo an operation* (Stubbs 2001).

[Instances for Negative Semantic Prosody]

- *budge* expresses the speaker’s frustration at a failed attempt to move something: *it wouldn’t budge* (Sinclair 1998).
- *naked eye* expresses the speaker’s difficulty in seeing things because they are small or far away: *hardly visible to the naked eye* (Sinclair 1996).

However, corpus linguistics also indicates that these features of unit of meaning are probabilistic. In the case of semantic prosody, for instance, the verb “afford” is only “more-often-than-the-average negative” (Hunston 2011, pp. 81), although it occurs equally frequently in positive and in negative clauses. This is largely because the overall ratio of positive to negative clauses is about 9:1 (Halliday, 1993; Matthiessen, 2006); thus, a lexical item that occurs evenly in the negative and positive is highly skewed toward the negative.

## 2. Birmingham Today

### 2.1 Members and Major Publications

This section shortly introduces some of the current academic staff members (most of them are applied linguists) at the UoB who utilize corpus linguistic approaches in their language studies. These staff members include (research interests are indicated in parentheses) :

Susan Hunston (Pattern Grammar, Discourse Analysis), Wolfgang Teubert (Critical Discourse Analysis, Lexicology), Geoff Barnbrook (Local Grammar), Nick Groom (Academic Discourse, EAP), Suganthi John (Academic Discourse, EAP), Oliver Mason (Computer Linguistics), Alison Sealey (Social Linguistics, First Language Acquisition), Rosamund Moon (Lexicography), Paul Thompson (Academic Discourse, EAP), Caroline Tagg (Text Analysis), Crayton Walker (Collocation), Martin Hewing (Pedagogical Grammar), Philip King (Translation Studies), and David Willis (Lexical Syllabus)...

Based on their research interests, they apply corpus for a wide range of language issues. They have also published many texts—often referred to as “bibles” —in their fields, including applied linguistics, such as:

Corpus, Concordance, Collocation (John Sinclair), Trust the text (John Sinclair), Linear Unit Grammar: Integrating Speech and Writing (John Sinclair), English Collocation Studies: The OSTI Report (John Sinclair), Pattern Grammar (Susan Hunston, Gill Francis), Corpora in Applied Linguistics (Susan Hunston), Corpus Approach to Evaluation (Susan Hunston), Evaluation in Text (Susan Hunston), System and Corpus: Exploring Connections (Susan Hunston), Fixed Expressions and Idioms in English (Rosamund Moon), Introducing Metaphor (Murray Knowles, Rosamund Moon), Corpus Linguistics: A Short Introduction (Wolfgang Teubert), Lexicology and Corpus Linguistics (Wolfgang Teubert), Language and Computers (Geoff Barnbrook), Meaningful Texts (Geoff Barnbrook, Parnilla Danielsson), and Academic Writing (Maggie Charles, Susan Hunston)...

As in the samples of members and publications, the interests of the UoB members are grouped into several areas. Susan Hunston, Wolfgang Teubert, and Caroline Tagg are interested in discourse analysis applying corpus linguistic methodologies. Rosamund Moon focuses on the application of the corpus for lexicographical works, scrutinizing the meaning and function of fixed expressions and idioms, including metaphors. Paul Thompson, Maggie Charles, Nicholas Groom, and Suganthi John investigate the academic discourse of a particular genre or a particular discipline from epistemological and educational viewpoints. Oliver Mason is deeply interested in devising a computer program for corpus analysis (Mason, 2000) and methodologies for the extraction of meaningful data from corpora (e.g., automatic extraction of patterns; Mason & Hunston, 2004). As these examples indicate, the UoB members utilize the corpus linguistics methodologies to analyze unit of meaning and discourse features of a word or sequence of words qualitatively.

## **2.2 Research Centers and Corpus Resources at the UoB**

The Center for Corpus Research (CCR) at the UoB was created to promote the use of corpus analysis in research, teaching, and learning.<sup>1</sup> The CCR offers language resources (e.g., the BoE), facilities (e.g., a computer suite with various computer software), and technical advice to people interested in corpus-based textual studies. The CCR also hosts training workshops, seminars, and conferences on corpus research and applications for teaching. The CCR has engaged in various projects (e.g., the COBUILD project with HarperCollins and the BoE corpus project), including Moon's lexicographical project with the BoE, Thompson's academic corpus projects for the British Academic Spoken English (BASE) and British Academic Written English (BAWE), and Hunston's projects focused on a corpus-driven approach to a lexical grammar of English and corpus approaches to the study of evaluation.

The Dictionary Research Center (DRC), originally located at the School of English in the University of Exeter, was transferred to the UoB in the fall of 2001.<sup>2</sup> The DRC aims to promote lexicographical activities and interests at the UoB and was involved with HarperCollins in the COBUILD project from 1980 to 2000. The DRC has been participating in the Johnson Dictionary project, including the sourcing of Johnson's citations, since 1988, in conjunction with Cambridge

UNIVERSITY OF BIRMINGHAM

Corpus Hub at Birmingham

Welcome » Overview »

» Welcome » Concordancer » Collocates

Hide sidebar

Signed in as: Guest user

Corpus Hub At Birmingham

Welcome to the Corpus Hub at Birmingham. This site is still under active development - please send in any comments (no matter how grave or how trivial) using the "send feedback" button in the top-right. More information about the technology used in CHAB is available on our development site.

» Welcome

» Tabs

» Sidebar

» Signing In

» Active Development

» News

**Quick Start**

Click here to be guided through the first steps in searching a corpus.

**Introduction to CHAB**

Corpus linguists study language by looking for patterns in *corpora* (large bodies of text collected for specific purposes). A typical corpus contains several million words of text on a particular subject, or gathered from a particular location, or some other grouping worthy of study. By examining such large bodies of text, it's possible to spot usage patterns, even when they're quite rare. The most common tool used in corpus linguistics is the *concordancer*, a specialised search program that displays results in *KWIC* (Key Word in Context) format. The KWIC format makes usage patterns easy to see by displaying search results as *concordance lines*, aligned such that the word you searched for (the "key word") is in the same place on every line.

One of the most common uses of corpora is to spot *collocates* - words which, when used together, mean something different than when the words mean on their own. For example, "cheese" is a dairy product and "chalk" is a powdery white mineral, but "chalk and cheese" isn't a powdery white dairy product. Modern concordancers such as CHAB help you to find collocates by sorting concordance lines and using statistics to search for probable matches, but there is still significant disagreement about the use of statistics in the corpus linguistics community.

This page describes how to navigate around CHAB. If you get lost, you can get back to this page at any time by clicking on "welcome", directly below the University of Birmingham logo. Pages in CHAB are made up of a *main column* (which this text is in), a *sidebar* on the left, and the *tab bar* directly above. Clicking on tabs and sidebar items will take you to the page associated with that link.

Figure 1. Corpus hub at the UoB

University Press. The recent research activities include corpus-based lexicography, bilingual and multilingual lexicography, the lexicographical description of collocation, the language of definitions, perceptions of dictionaries, and metaphor and dictionaries.

Various corpus interfaces are available for members of the UoB, including the BoE and the Corpus Hub at Birmingham (CHAB).<sup>3</sup> In particular, CHAB is a collection of software tools made available under open source licenses. CHAB consists of various components, including a web-based concordancer called the Birmingham Concordancer,<sup>4</sup> a search engine that powers the concordancer and a collection of corpora that supply the search engine,<sup>5</sup> and a set of tools that compile the corpora (i.e., the SCAN Toolkit).<sup>6</sup>

### 3. Corpus Studies by UoB Members

#### 3.1 Case 1: Teubert and Cermakova's globalization

Teubert and Cermakova (2007) examined the meaning of *globalization* in the BoE. The study revealed that the data in the BoE may be biased toward British English since *globalization* occurs 486 times but its British variant *globalisation* has 1,447 occurrences. The most frequent collocates for globalization are *anti-*, *world*, *against*, *means*, *economic*, *international*, and *business*; no interesting findings stem from these. However, its discourse is likely to show a negative tone (especially negative emotions), such as "a lot of debating, worrying and talking about globalization," "many protests, demonstrations and campaigns against globalization," "the challenge, the impact, the pressures, the forces and effects of globalization," and "anti-globalization and anti-globalization protesters" (p. 92). Indeed, the figure for the neutral tone is about one third of the total but only about one tenth of the positive tone alone. According to Teubert and Cermakova, the difference

in positive/negative tone lies in the register variation: Texts from newspapers often have a skeptical (i.e., negative) tone, whereas the tone of professional and academic writing is more matter of fact (i.e., neutral) (p.96). In addition, *globalisierung* (*globalization* in German) expresses a different meaning from the English term: *in der Tat bedeutet Globalisierung Amerikanisierung*—globalization means Americanization (p. 98).

### 3.2 Case 2: Moon's fixed expression and idioms

Moon (1998) investigated lexical, syntactic, semantic, discursal, and social variations of fixed expressions and idioms (FEIs) in the Oxford Hector Pilot Corpus (18 million words). She found that the density of metaphors and proverbs seems to be greater in journalism than in other text types, and pure idioms seem to be less common in spoken interaction than often thought. Her data also suggest that some individual items have individual genre preferences (Moon, 1998, p. 309).

Among the many findings in this study, Moon used a Birmingham-style analysis to investigate the ratio of FEIs used literally and idiomatically. The data from the BoE indicate that idiomatic uses are overwhelmingly more common than literal ones except for *break the ice* and *in hot water*. Moon also scrutinized the figures for polysemous senses of a particular FEI. For instance, the FEI *break the ice* has two idiomatic meanings: disperse awkwardness in social interaction (189 tokens) and, in the sports contexts, score for the first time in a game or season (12 tokens). Since Table 1 only presents the overall picture for the lemma *break* occurring within a five-word window of *ice*, with no further syntagmatic restrictions, Moon also highlighted the different ratios in the detailed structures of this FEI.

**Table 1** Literal and Idiomatic Meanings for FEIs  
(Moon, 1998, p. 182)

	Literal frequency	Idiomatic frequency
<i>beat about/around the bush</i>	0	109
<i>break the ice</i>	65	201
<i>in hot water</i>	181	178
<i>kick the bucket</i>	7	42
<i>let the cat out of the bag</i>	0	91
<i>(out) on a limb</i>	4	249
<i>(skate) on thin ice</i>	0	241
<i>spill the beans</i>	2	198

As Table 2 indicates, in the case of the string with *the*, the ratio of literal meanings to idiomatic meanings is around 1:6, whereas in the case of the string with an adverb particle, almost all tokens are literal. In other words, the definite article *the* and adverb particles are key elements for distinguishing between idiomatic meanings and literal meanings for this FEI. Thus, Moon succeeded in demonstrating that form and meaning are well linked.

**Table 2 Structure and Meanings for Break the Ice (Moon, 1998, p. 183)**

Structure	Literal meaning	Interaction meaning	Sports meaning
<i>break the ice</i>	18	152	10
<i>break ice</i>	6	1	1
<i>break the ice</i> + adverb particle	10	0	0
<i>break ice</i> + adverb particle	6	0	0
<i>the ice is broken</i>	1	31	1
<i>the ice breaks</i>	6	4	0
<i>ice breaks</i>	5	0	0
<i>the ice breaks</i> + adverb particle	7	1	0
<i>ice breaks</i> + adverb particle	6	0	0

### 3.3 Case 3: Tagg's text messages

Tagg (2009) investigated a wide range of lexical and phrasal items peculiar to text messages (TMs) using approximately 11,000 messages. She identified practices of eye dialect and omission, such as *8* for ate; *ne* for any; *y* for why; *no* for know; *mess* for message; *bac* for back; *beta* for better; *d* for the; *2 nite* for tonight; *tho* for although; *cumin* for coming; and *4 ever* for forever (e.g., *Mam said dont make ne plans for nxt wknd; Been? Y D' you want 2 know?*). For example, *at the moment* (27 tokens) also has an alternative variant *at the mo* (33 tokens) in TMs, which occurs more frequently than the basic form. The definite article *the* is also often omitted as *at moment* (3 tokens) and *at mo* (23 tokens) (e.g., *I'm in town at mo; Can't return the camera at mo; Yes, on bus at mo*).<sup>7</sup> Tagg (2009) pointed out that the figure for the definite articles *the* is relatively low in text messages, which is in contrast with other registers; one reason for this is the various alternative (spellings) variants of the basic form (Tagg, 2009, p. 244). The phrase *by the way*, for instance, frequently occurs as *btw* in TMs.

ian network! All fine here. Btw, am back friday morn  
wot the details Yeh yeh. btw charlottes access have  
tty fab thanks 2 me! Hehe btw craig where were  
monday, probably all day, btw did you know this we  
ws on what ur doin 2 nite?! Btw do i nt get x x x's  
elly, how r u? Hows work, btw how r they 4 staff  
Tee hee. But ta for invite. BTW, running late and mi  
iff and its cold! What does btw stand for? I'm just  
Concordance lines for *btw*

Tagg (2009) also revealed the functions of discourse markers peculiar to TMs. For instance, *ha ha*, *haha*, *hahaha*, *tee hee*, and *lol* all refer to "laughter and presumably acknowledge amusing comments made by interlocutors" (p. 311), whereas *ha* refers to "triumph along the lines of 'I win!' rather than laughter" (p. 311):



- **Lol!** U drunkard! Just doing my hair at d moment. Yeah still up 4 tonight. Wats the plan?  
Xxx
- **Tee hee.** Off to lecture, cheery bye bye.
- **Ha!** I wouldn't say that I just didn't read anything into way u seemed. I don't like 2 be judgemental.... i save that for fridays in the pub!
- **Ha.** You don't know either. I did a clever yet simple thing with pears the other day, perfect for christmas.

A final example is the case for negative semantic prosody imbued by a sequence of words in TMs. For instance, the usage of *a bit of a* has a peculiar discourse function that “reverts to the more negative connotation also implied by *a bit*” (Tagg, 2009, p. 236).

Well done. Often a bit of a challenge to do it  
rd. i know this is a bit of a cheek especially  
eems to be. I had a bit of a cold but gettin bet  
y touch ! feel like a bit of a cow. But ta for th  
. I know I' ve been a bit of a drip but I will start  
dy wife are having a bit of a get together satur  
at might have been a bit of a mean message to  
thing has proved a bit of a mistake i think- lo  
il out tonight for a bit of a piss up, he's not  
cancel lunch. Have a bit of a problem. Call you  
Concordance lines for *a bit of a*

### 3.4 Case 4: Fang's international community

Fang (2009) analyzed the discursual meaning of the phrase *international community* in a British newspaper (*The Guardian*) as a representative of western discourse communities and a Chinese newspaper (*People's Daily*) as a representative of Asian discourse communities by investigating its collocational and grammatical behaviors. According to Hunston (2002), corpus use contributes to studying the link between “language and ideology” and seeing “existing power relations” (p. 109) ; Fang (2009) further pointed out that corpus linguistics can shed sufficient light on “the interpretative basis” of critical discourse analysis and provide reliable outcomes from analyses by finding meaningful features or patterns of language (p. 53). Fang's study resulted in the conclusion that the same lexical item is used to refer to different meanings in the two discourse communities studied, thus presenting “how underlying ideologies are inscribed in and mediated through the linguistic system” (Caldas-Coulthard, 1996, p. x).

To summarize Fang's (2009) findings, the phrase *international community* in the English newspaper most often referred to powerful and leading nations (i.e., developed western countries such as the United States and the United Kingdom) as the salient pattern of this phrase is (V/Prep.) + *the international community* + *to-inf.*, in which *to-inf.* is followed by *put pressure*,



*make efforts to stop the blood she, punish the North, stop the massacre, capture them, intervene to stop,* and so on. On the other hand, the meaning of this phrase in the Chinese newspaper varied widely depending on the context, although it focused on four types. In particular, when highlighting the importance of China, the *international community* refers to third world countries, but when reporting conflicts between China and other nations, it refers to “the right side” or “allies” of China that “support China and condemn the opposition” (p. 58). The pattern *international community + with regard to* highlights this peculiarity:

- The Japanese Foreign Minister’s speech ... neglects the neighboring countries’ and the **international community’s** serious criticism **with regard to** the Japanese leader’s visiting Yasukuni Shrine...
- Bill Clinton’s speech reflects the mutual consensus of the **international community with regard to the** Taiwan issue.

The *international community* in the first example refers to the countries that support China and condemn Japan. The second example also has the same connotation with reference to the countries that support China on the issue of Taiwan and condemn Taiwan’s separations. Fang concluded that a certain power relation is hidden in the usage of *international community* according to its context, by which the ideology of a world is clearly diverged into “a leading developed world” and “a developing world that receives aid from developed countries” (p. 58).

### 3.5 Case 5: Hunston’s semantic sequences

Semantic sequences are “recurring sequences of words and phrases that may be very diverse in form and which are therefore more usefully characterised as sequences of meaning elements rather than as formal sequences” (Hunston, 2008, pp. 271–272). In other words, concordance lines present the way in which multiple instances of a given word or a sequence of words occur in broadly similar contexts, where broadly similar contexts is described as semantic sequences (Hunston, 2011). Hunston (2011) prioritized the investigation of the New Scientist corpus, concluding that her study never demonstrated what language is like or even what the English used in the New Scientist is like, but rather how things are often said in the specialized texts.

For instance, the word *distinguishing* is frequently followed by a prepositional phrase beginning with *between* and containing a plural noun phrase: This is mere grammatical information. However, the sequence *distinguishing between* frequently follows an expression of difficulty or importance. In other words, this sequence is often realized in a semantic sequence: difficulty or importance + *distinguishing between* + two or more similar things (Hunston, 2011, p. 90). Thus, semantic sequence is a broader concept than that of a phrase or unit of meaning characterizing given discourses (p. 91). Hunston (2008, p. 285) further exemplified another instance of semantic sequence in which the adjective pattern followed by *that* clause – it + v-link + ADJ (e.g., *clear/apparent/obvious/evident*) + *that* – in the New Scientist corpus typically expresses a claim on the part of the writer. The sequence *it is clear that*, for instance, occurs in three distinct semantic

sequences:

- LOGICAL BASIS + *it is clear that* + CLAIM
- CONSENSUAL INFORMATION + *it is clear that* + CLAIM
- *It is clear that* + CLAIM + EXCEPTION/CAVEAT

## 4. A Current Trend in Corpus Linguistics at the UoB

### 4.1 Corpus-based (Academic) Discourse Analysis

Discourse analysis has traditionally been seen as a top-down approach as it focuses on entire texts and their cultural context, particularly identifying patterns that extend across sentences and paragraphs. As such, the top-down discourse analysis pays too little attention to the significance of individual phraseologies and makes generalizations from too few examples. On the other hand, corpus linguistics has generally been interpreted as working from the bottom up as it is likely to utilize techniques that decontextualize individual texts and highlight recurrent patterns of smaller linguistic items. Consequently, bottom-up corpus methods have been perceived as placing too much significance on individual textual elements and their frequency of occurrence.

A current trend in corpus linguistics at the UoB is to approach corpus linguistics and discourse analysis as not independent activities, but rather approaches that harmonize with one another in a complementary way. Applied linguists at the UoB assume that the “corpus-assisted discourse study” (e.g., Partington, 2004, p. 19) is an effective and enriched approach for identifying the link between language and cultural values of a particular discipline (e.g., Charles 2004, 2006a, 2006b; Groom, 2007; Hunston, 2000; Nishina 2010); such a combined approach would offer even alternative theoretical approaches as well rather than complementary methodologies.

### 4.2 Phraseologies and Evaluation

Phraseologies in the current Birmingham school stem from the theory of pattern grammar in Hunston (2006), Hunston and Francis (1999), and Hunston and Sinclair (2000). Pattern (or local) grammar begins with a pattern, then looks at the meaning of a word within it. Other approaches, such as FrameNet (e.g., Fillmore, 1977a, 1977b, 1985; Fontenelle, 2003) and verb classes (e.g., Levin 1993), first look at the meaning of a word, then its frame (i.e., word-meaning is independent of its frame). For instance, *recover* and *heal* are categorized in a same group in FrameNet, whereas they are treated quite differently in pattern grammar. Meanwhile, *recover from* and *suffer from* are treated as similar in pattern grammar whereas they are categorized into different groups in FrameNet (Hunston, 2011, p. 128). Some significant works based on pattern grammar have already been accomplished in practical fields (e.g., Francis et al., 1996, 1998).

Evaluation is another important issue for recent studies at the UoB. Discourse analysts have sought linguistic resources that express interpersonal meanings in discourse; such personal, private, and (inter-)subjective meanings in texts are conceptualized as evaluation (Hunston,

2011; Hunston & Thompson, 2000). Evaluative utterances express a personal opinion in a positive or negative way (Thompson & Hunston, 2000, p. 1), particularly in the writer's attitudes as ways of judgment or feeling (Martin & White, 2005, p. 42). Concepts similar to evaluation have been assumed in various relevant fields, such as appraisal (Martin & White, 2005), stance (Biber & Finegan, 1989; Conrad & Biber, 2000) and metadiscourse (Hyland, 2000; Hyland & Tse, 2004).<sup>8</sup>

Although lexical items express evaluative meanings, such as nouns (e.g., *success*), verbs (e.g., *fail*), adjectives (e.g., *excellent*), and adverbs (e.g., *unfortunately*), lexical bundles also demonstrate such writer's stance, such as *no doubt*, *in fact*, and *according to* (Biber et al., 1999). In the same token, some lexico-grammatical patterns also structure an evaluative function in discourse, as in 'it was adjective (e.g., *nice*, *kind*, *good*, *selfish*, *foolish*) of person to do something' (for details, see Hunston, 2011). Prepositions in particular serve to classify information; some interact with evaluative meaning (Hunston, 2008, 2011; Hunston & Sinclair, 2000; Nishina, 2010). The text-based approach stresses that, since evaluation is an action performed in discourse and expressed rather implicitly than explicitly by units of meaning (Hunston, 2011, pp. 7–14), it is always contextually determined (p. 24). In other words, the corpus approach only assists in the identification of evaluative meaning by pinpointing the recurrence of a particular lexical item; human interpretation is always necessary for its comprehensive analysis.

### 4.3 Disciplinary Discourse Analysis at the UoB

With an increased concern for English for Academic Purposes (EAP), some members at the UoB have attempted to compare texts from different genres and/or disciplines written by academics (e.g., research articles) as well as, in some studies, those written by students (e.g., essays or dissertations). This trend has to some extent stemmed from Hyland's studies. Hyland (2000) pointed out that writers in soft disciplines use more and varied discourse act reporting verbs (e.g., *ascribe*, *discuss*, *state*) than those in hard disciplines as they need to support their arguments with reference to important studies by other experts. Meanwhile, writers in the hard disciplines tend to express causal and logical relationships and use research act reporting verbs (e.g., *observe*, *discover*, *calculate*) as they need to interpret quantitative data gained from their ongoing studies. For this reason, they are unlikely to refer to the opinions of others, which contrasts with those in soft disciplines (p. 28).

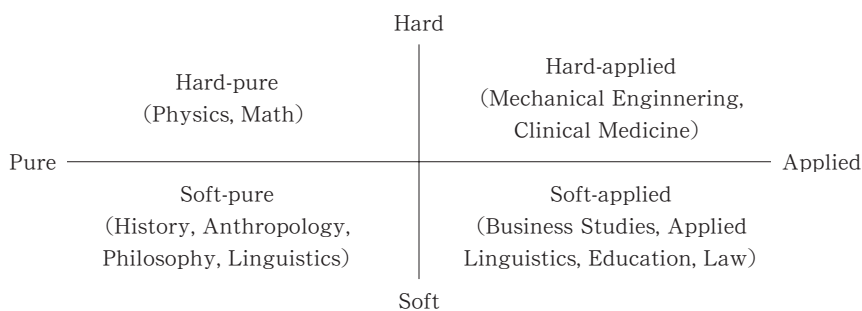


Figure 2 Representation of Knowledge Domains

Disciplinary group	Nature of knowledge
<b>Hard-pure: Pure science</b>	Cumulative; atomistic, concerned with universals; impersonal; value-free; clear criteria for knowledge verification and obsolescence; consensus over significant questions to address, now and then in the future; result in discovery/explanation
<b>Soft-pure: Humanities &amp; pure social sciences</b>	Reiterative; holistic; concerned with particulars; qualities, complication; personal; value-laden; dispute over criteria for knowledge verification and obsolescence; lack of consensus over significant questions to address; results in understanding/interpretation
<b>Hard-applied: Technologies</b>	Purposive; pragmatic; concerned with mastery of physical environment; applies heuristic approaches; uses both qualitative and quantitative approaches; criteria for judgement are purposive, functional; results in products/techniques
<b>Soft-applied: Applied social sciences</b>	Functional; utilitarian; concerned with enhancement of [semi-] professional practice; uses case studies and case law to a large extent; results in protocols/procedures

**Figure 3 The Nature of Knowledge and Disciplinary Grouping in Becher & Trowler (2001, p. 36)**

Recent studies on EAP at the UoB have sought to interpret the discourse functions of a sequence of words peculiar to a discipline to find the link between language and epistemological knowledge (Becher & Trowler, 2001), as indicated in Figures 2 and 3. This approach is due to the fact that patterns are linked to the ideology of each discipline (Charles, 2004). For example, Charles (2003, 2006a, 2006b, 2009) attempted to identify the link between language and epistemology by scrutinizing the discourse features of a word or pattern among theses from a physical science discipline (i.e., materials science) compared to those from a social science (i.e., political science). Charles (2003) focused on the role of anaphoric nouns. In politics, an anaphoric noun indicates the work against which one's own work reacts, meaning writers are more likely to evaluate that information negatively; meanwhile, in the materials science, writers are more likely to express a positive stance toward such information as anaphoric nouns indicate the work on which their own work is built. Such findings are well linked to the epistemological features of cumulative construction of knowledge in physical science disciplines (e.g., materials science) and recursive knowledge-building in social sciences (e.g., political science) (p. 322). Charles (2009) also exemplified how epistemological knowledge and corpus findings are well linked with one another. Becher and Trowler (2001) noted that knowledge in the social sciences is likely to be concerned with particulars whereas that in the natural sciences deals primarily with universals. They compared the figure for restrictive adverbs in two disciplinary corpora:

**Table 3 Restrictive Adverbs (per 0.1 mil.) (Charles, 2009, p. 154)**

Restrictive adverb	Politics	Materials
Only	143.7	118.0
Simply	23.7	12.3
Just	18.4	6.7
Merely	21.6	2.7
Total	207.4	139.7

The narrower focus is epistemologically required for the study of particular events and entities in politics but not in materials sciences, as indicated by the more frequent use of restrictive adverbs in Table 3.

From a different perspective, Hunston (2011) presented interesting figures for fact and hypothesis among disciplines in the academic section of the BNC:

**Table 4 Fact versus Hypothesis in Academic Disciplines (Hunston, 2011, p. 109)**

Register	Per mil.	<i>fact</i>		<i>hypothesis</i>
		'in fact' per mil.	'fact' less 'in fact'	per mil.
Politics, Law, Education	597.58	184.90	412.68	16.0
Humanities	643.61	294.91	348.7	36.1
Social Science	572.57	244.84	327.73	79.5
Natural Science	339.98	148.40	191.58	79.7
Engineering	341.11	189.50	151.61	91.4
Medicine	191.99	74.55	117.44	122.5

The disciplines with the most facts use the word *fact* least often in their discourse, whereas the figure for *hypothesis* presents the opposite tendency. Medicine, natural science, social science, and engineering fields use *hypothesis* most; humanities and politics, law, and education use it least (Hunston, 2011, p. 109).

#### 4.4 A Short Summary of Nishina (2010)

Finally, this section shortly summarizes the essence of Nishina (2010), who investigated the pattern ADJ PREP N in research articles of the epistemologically same fields—namely, the two soft-applied disciplines of applied linguistics and business studies (more than 2.5 million words in each field). The study included a quantitative analysis (i.e., Correspondence Analysis, hereafter CA) and a qualitative analysis (i.e., corpus-based discourse analysis using the framework of appraisal theory [Martin & Rose, 2007]). The two research questions for the study are:

- (1) What sort of differences are quantitatively and qualitatively identified in the pattern use between two disciplinary corpora through a corpus-based approach?;
- (2) What sort of insights into disciplinary culture can be gained and generalized from an investigation of the pattern **ADJ PREP N** in applied linguistics and business studies?

Some of the recent academic discourse studies have prioritized the pattern *it* v-link ADJ *that*/*to*-inf. (e.g., Charles, 2004; Groom, 2007) since the *that*-clause and *to*-inf. are stance markers. However, as Chafe and Danielewicz (1986) and Chafe (1982, 1985) suggested, both adjectival subordination constructions and adjectival prepositional phrases (and patterns) are used to express the idea's unit integration and expansion. Such adjectival prepositional patterns are comprehensively listed in Francis et al. (1998).

Data for the corpora include (1) the corpora source; (2) data for ADJ PREP N; (3) high-frequency ADJ in ADJ PREP N; and (4) CA plot for ADJ PREP N.

Current Trends in Corpus-based Discourse Analysis in Britain

Journal Name (Applied Linguistics)	Published Year	Texts	Tokens	Types
Language and Culture (Cognitive)	2005-6	36	338,256	11,287
International Journal of Corpus Linguistics (Corpus)	2005-6	34	314,375	17,061
Journal of English for Specific Purposes (ESP)	2005-6	39	312,483	14,252
Language Learning (LL)	2005-6	35	364,731	14,655
Studies in Second Language Acquisition (SLA)	2004-5	34	348,910	12,515
Language and Speech (Speech)	2003-4	29	322,596	11,732
TESOL Quarterly (TESOL)	2005-6	43	341,187	16,194
Language Testing (Testing)	2003-5	39	324,462	11,722
Journal Name (Business)	Published Year	Texts	Tokens	Types
Journal of Accounting Research (Accounting)	2005	27	332,721	9,470
Quarterly Journal of Business and Economics (Economics)	2000-5	63	304,420	9,914
Journal of Business Finance and Accounting (Finance)	2005	66	407,018	11,874
American Business Law Journal (Business Law)	2000-5	34	362,292	15,157
Journal of Small Business and Management (Management)	2000-5	65	334,237	14,256
Journal of Business and Industrial Marketing (Marketing)	2004-5	54	296,501	11,611
Business and Society (Society)	2003-5	44	336,610	15,314
Journal of Business Strategy (Strategy)	2001-5	83	294,878	15,527

Figure 4 Source for corpora

8 High-Frequent Patterns						8 Low-Frequent Patterns					
ADJ about N			ADJ of N			ADJ against N			ADJ between N		
	ALC	BC		ALC	BC		ALC	BC		ALC	BC
Tokens	84	116	Tokens	827	726	Tokens	4	3	Tokens	4	5
Types	36	50	Types	53	67	Types	3	3	Types	1	2
TTR	0.429	0.431	TTR	0.064	0.092	TTR	0.750	1.000	TTR	0.250	0.400
ADJ for N			ADJ on N			ADJ as N			ADJ by N		
	ALC	BC		ALC	BC		ALC	BC		ALC	BC
Tokens	727	714	Tokens	1635	2107	Tokens	52	76	Tokens	50	93
Types	71	74	Types	23	27	Types	19	19	Types	21	26
TTR	0.098	0.104	TTR	0.014	0.013	TTR	0.365	0.250	TTR	0.420	0.280
ADJ from N			ADJ to N			ADJ as to N			ADJ over N		
	ALC	BC		ALC	BC		ALC	BC		ALC	BC
Tokens	476	498	Tokens	5001	5886	Tokens	8	9	Tokens	2	3
Types	27	28	Types	175	179	Types	5	9	Types	2	3
TTR	0.057	0.056	TTR	0.035	0.030	TTR	0.625	1.000	TTR	1.000	1.000
ADJ in N			ADJ with N			ADJ at N			ADJ toward(s) N		
	ALC	BC		ALC	BC		ALC	BC		ALC	BC
Tokens	1540	1617	Tokens	1561	3076	Tokens	55	39	Tokens	25	19
Types	69	81	Types	76	81	Types	14	12	Types	3	5
TTR	0.045	0.050	TTR	0.049	0.026	TTR	0.255	0.308	TTR	0.120	0.263

Figure 5 Data for ADJ PREP N

Current Trends in Corpus-based Discourse Analysis in Britain

ADJ about N						ADJ of N							
Order	ALC	Freq. BC	Freq. BNC	Per. 2.5M		Order	ALC	Freq. BC	Freq. BNC	Per. 2.5M			
1	confident	11	knowledgeable	11	concerned	37	1	aware	206	aware	110	aware	157
2	clear	8	optimistic	8	worried	33	2	independent	91	independent	77	full	139
3	enthusiastic	6	cautious	7	sorry	10	3	typical	78	capable	68	capable	107
4	explicit	5	serious	6	sure	9	4	capable	70	representative	59	short	54
5	positive	4	uncertain	6	happy	7	5	representative	60	indicative	49	free	35
ADJ for N						ADJ on N							
Order	ALC	Freq. BC	Freq. BNC	Per. 2.5M		Order	ALC	Freq. BC	Freq. BNC	Per. 2.5M			
1	available	97	responsible	172	responsible	169	1	based	1448	based	1835	dependent	48
2	responsible	97	available	85	available	72	2	dependent	82	dependent	128	available	26
3	necessary	86	necessary	77	ready	56	3	high	25	conditional	32	keen	16
4	appropriate	72	appropriate	41	suitable	48	4	correct	15	contingent	18	high	13
5	useful	47	liable	34	good	47	5	low	14	high	11	intent	6
ADJ from N						ADJ to N							
Order	ALC	Freq. BC	Freq. BNC	Per. 2.5M		Order	ALC	Freq. BC	Freq. BNC	Per. 2.5M			
1	different	242	different	266	different	82	1	due	737	related	1286	similar	84
2	far	43	available	42	available	54	2	related	608	due	668	available	75
3	distinct	30	distinct	33	free	24	3	similar	452	similar	444	close	66
4	absent	19	far	28	far	21	4	sensitive	218	subject	241	used	62
5	separate	18	separate	15	distinct	16	5	relevant	175	equal	202	open	56
ADJ in N						ADJ with N							
Order	ALC	Freq. BC	Freq. BNC	Per. 2.5M		Order	ALC	Freq. BC	Freq. BNC	Per. 2.5M			
1	involved	280	involved	264	involved	199	1	associated	575	consistent	1307	concerned	115
2	present	153	significant	156	interested	139	2	consistent	361	associated	1211	wrong	42
3	interested	131	important	129	available	61	3	concerned	144	concerned	81	consistent	33
4	engaged	101	engaged	127	important	34	4	compatible	92	inconsistent	80	involved	31
5	similar	71	interested	125	present	32	5	familiar	91	firm	63	familiar	30

Figure 6 High-frequency ADJ in the pattern

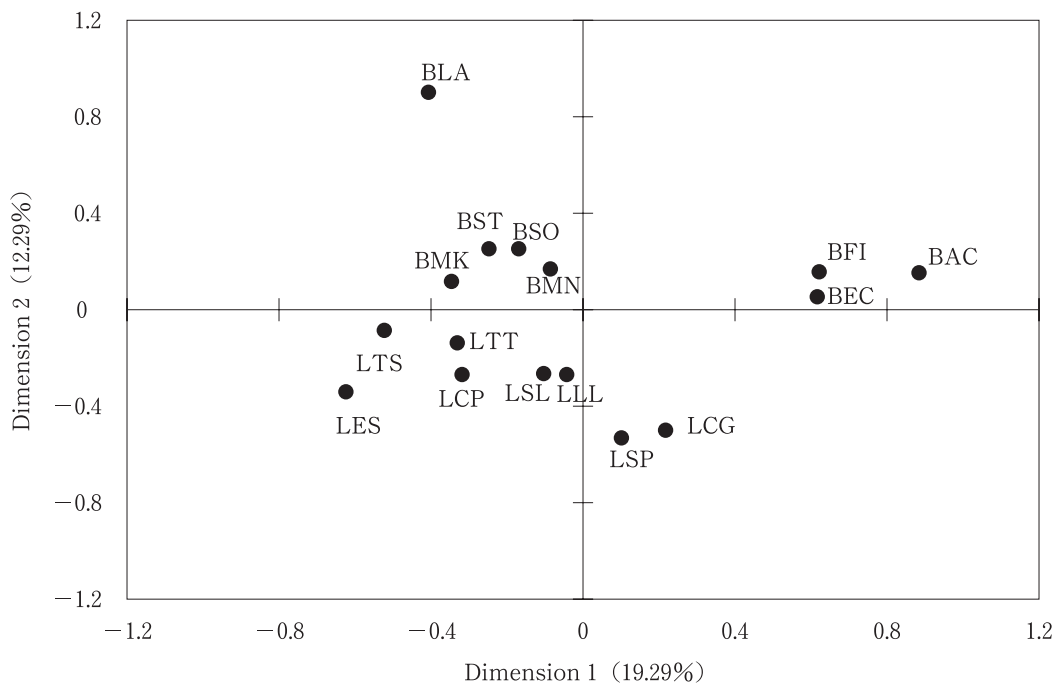


Figure 7 CA plot for sub-disciplines

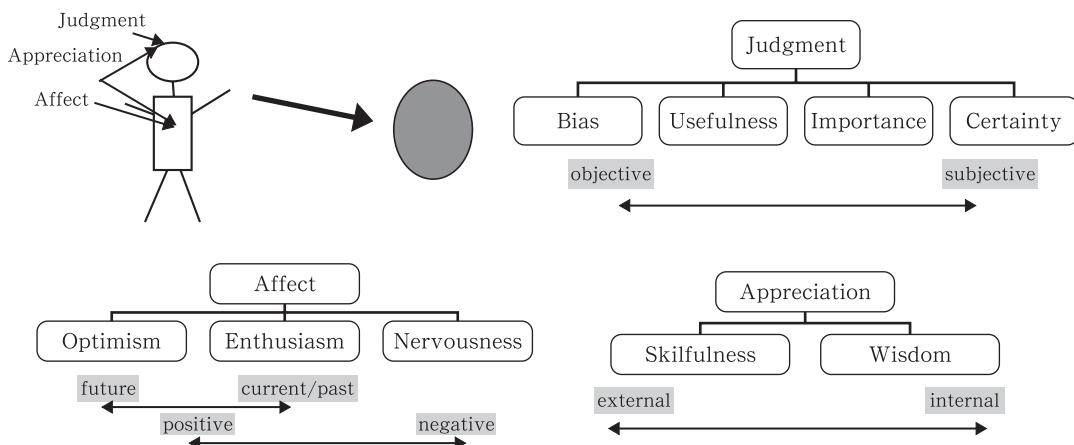


As Figure 7 indicates, the use of the patterns ADJ PREP N quantitatively somewhat differs in the two corpora, but it is unclear how they differ. To address this issue, semantic groups of each adjective pattern were first identified; then their discourse functions in the two corpora were examined. Table 5 indicates the semantic groups of ADJ *about* N.

**Table 5 Semantic groups of ADJ *about* N**

Order	ALC	Freq.	BC	Freq.
1	OPTIMISM	20	NERVOUSNESS	17
2	CERTAINTY	14	OPTIMISM	15
3	NERVOUSNESS	14	CERTAINTY	14
4	ENTHUSIASM	6	WISDOM	11
5	WISDOM	0	ENTHUSIASM	6
-----				
6	OTHER MEANINGS	30	OTHER MEANINGS	53
Total		84	116	

For instance, the OPTIMISM group most frequently occurs in ALC, whereas the NERVOUSNESS group occurs more in business studies. After extracting such data for each pattern, these groups were categorized into two large discourse functions expressed by the patterns *attitude* and *relation*. Following Martin and Rose (2007), *attitude* was further sub-categorized into the type of *judgment*, *affect*, and *appreciation*. The author also divided *relation* into *connection* and *attribution*. Finally, semantic groups of the pattern were grouped into each of the three sub-functions in *attitude* and the two sub-functions in *relation*. Semantic groups in *attitude* include *BIAS*, *USEFULNESS*, *IMPORTANCE*, and *CERTAINTY* in the *judgment* type; *OPTIMISM*, *ENTHUSIASM*, and *NERVOUSNESS* in the *affect* type; and *SKILLFULNESS* and *WISDOM* in the *appreciation* type. Semantic groups in *relation* include: *SIMILARITY*, *ASSOCIATION*, *DEPENDENCE*, *DISTANCE*, *DETECTABILITY*, *AFFECTEDNESS*, and *INVOLVEMENT* in the *connection* type and *ABILITY*, *RESPONSIBILITY*, *INHERENCE*, *UNIQUENESS*, *REPRESENTATIVENESS*, and *BA-SIS* in the *attribute* type.



**Figure 8 Attitude (how people see things) and its semantic categories**

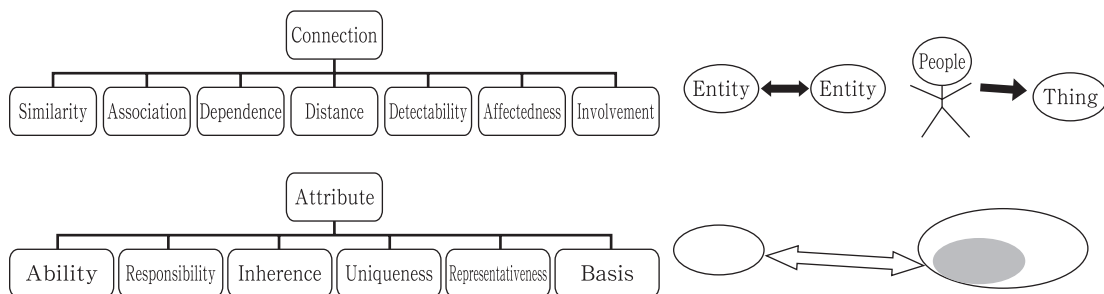


Figure 9 Relation and its semantic categories

As the space in the current paper is relatively limited, the following sub-sections highlight the study’s findings that link language and disciplinary values.

#### 4.4.1 Monogloss and heterogloss: ADJ about N

The voice of evaluative meaning is divided into two types: *monogloss* (or single voice), when “the source is simply the author,”<sup>9</sup> and *heterogloss*, when “the source of an attitude is other than the writer” (Martin & Rose, 2007, p. 49). Table 6 summarizes the proportions of these two types of *affect* evaluation in the pattern ADJ about N.

Table 6 Proportion of Monogloss and Heterogloss in Semantic Groups in Affect

	OPTIMISM		NERVOUSNESS		ENTHUSIASM		<i>affect</i> (Total)	
	<i>mono</i>	<i>hetero</i>	<i>mono</i>	<i>hetero</i>	<i>mono</i>	<i>hetero</i>	<i>mono</i>	<i>hetero</i>
ALC	35% (7)	65% (13)	29% (4)	71% (10)	0% (0)	100% (6)	12% (10)	88% (74)
BC	0% (0)	100% (15)	12% (2)	88% (15)	0% (0)	100% (6)	4% (5)	96% (111)

In most cases, the discourse function of the pattern ADJ about N is to express *affect* (i.e., people’s emotion) as a subtype of *attitude*. Table 6 indicates that *heterogloss* occurs much more frequently than *monogloss* in both corpora. However, it is also worth noting that *monogloss* occurs three times as frequently in ALC as it does in BC. In other words, applied linguists are more likely to express their emotion—namely, how they feel about their study. Identifying whether *affect* is writer-oriented or other-oriented is important for clarifying the ways in which each discipline treats such an ambiguous phenomenon in people because *monogloss* is the realization of subjective emotion whereas *heterogloss* is the realization of objective observation for an other’s emotion.

For instance, the semantic groups in *affect* type (e.g., the *ENTHUSIASM* group and the *NERVOUSNESS* group) reveal evaluations in ALC that focus on individual psychological consequences on the one hand and evaluations in BC that focus on material consequences on the other. Writers in both disciplines, for instance, often employ *NERVOUSNESS* phraseologies to describe their own or other people’s awareness of (and, thus, desire to avoid) the possible negative consequences of following a particular course of action. In ALC, this wariness typically relates to the

writer's (possible) own interpretation of research findings (i.e., *monogloss*), but in BC it invariably relates to corporate decision-making processes as an objective observation (i.e., *heterogloss*):

[Applied Linguistics]

- Because this lack of a difference is in fact a null effect, we should be **cautious about** interpreting these data. (ALC: Cognitive 2005\_21)
- As this study was not an experimental study manipulating specific task characteristics, we still must be **cautious about** the exact interpretation of the task effects. (ALC: Testing 2005\_1)
- Reflection on the other issues raised in the introduction would make us **cautious about** generalizing beyond the context of.... (ALC: ESP 2006\_3)

[Business Studies]

- An NIRI survey (2001) conducted after the effective date of Regulation FD, however, suggests that companies are not as **apprehensive about** information releases following implementation. (BC: Economics 2004\_7)
- ... companies face reputational penalties for reducing dividends and are therefore **cautious about** adjusting dividends. (BC: Management 2003\_10)
- Rather than excluding federally preemptive legislation, Calabresi's elevation of the common law as a vehicle to right statutory wrongs simply suggests that Congress might be **cautious about** using federal preemption as a means of imposing regulatory uniformity. (BC: Law 2000\_4)
- This decline made investors **anxious about** the new strategy, and they pressured management to back off of their seemingly risky direction. (BC: Strategy 2002\_6)

#### 4.4.2 Disciplinary prosody: Patterns expressing attitude

Next, disciplinary prosody—describing the semantic prosody peculiar to a particular discipline<sup>10</sup>—was examined. With the patterns expressing affect (e.g., *optimistic about* N), writers in BC promote stances toward business matters that are conversely cautious and even pessimistic at times; they also express concession and skeptical attitudes, views, or expectations.

01 s the optimistic bias. Analysts may be optimistic about earnings, but if investors use the  
 02 t owners who own larger firms are more optimistic about future increased sales, it would n  
 03, and check cashing services were less optimistic about future sales than communications,  
 04 h sales over \$ 250,000 a year were more optimistic about future sales than owners of firms  
 05 ations. Sole proprietorships were less optimistic about growth than partnerships and corpo  
 06 ver events in their lives and are less optimistic about the future (Dembers & Brooks, 1989  
 07 olution mechanisms. (n 62) While Peck is optimistic about the role of the United Nations, sh  
 08 ncing, where firms that are relatively optimistic about their stock price prospects but co

Meanwhile, using these patterns of affect function (especially the *ENTHUSIAM* and

*OPTIMISM* groups), writers in applied linguistics express (or make efforts to gain) strong feelings of confidence or passion about ideas, abilities, solutions, activities, skills, and research consequences, which is also manifested in the fact that these patterns often co-occur with amplifiers (e.g., *quite, increasingly, extremely*), a comparative *more* (more than 50% of the total), or other language markers highlighting feelings of optimism or passion (e.g., *remain, hope, nevertheless, despite*).

- We remain **optimistic about** the contribution of neuropsychology to the understanding of semantic structure... (ALC: Cognitive 2005\_9)
- But I also hope they will nevertheless be **optimistic about** the ideas set out for planning at the end of this article. (ALC: TESOL 2005\_1)
- Despite the extra demands that the collaborative program makes on them, instructors are **enthusiastic about** participating in the program because they find that students enrolled in the program are generally more motivated, have a better attitude, and work harder. (ALC: ESP 2006\_20)
- Although no proficiency test was administered to either Natalie or Bernd, they had both been studying English for 2 years at a German university prior to taking part in the study, during which time they had to pass a number of exams to be eligible for the study abroad program, and were both extremely **enthusiastic about** English. (ALC: LL 2006\_11)

Based on these examples, it is assumed that the comparative value is linked to the specific business cultural norm—namely, a cautious attitude toward positive emotions.<sup>11</sup> Thus, the default assumption among research articles authors in business studies is not to have a positive emotion (e.g., optimistic feeling) about business activities, transactions or performances.

In the same token, several patterns of judgment function also express such disciplinary prosody. Turning to neutral meanings using *neutral about/on/between* N in the BIAS group, BC is likely to regard neutral business situations positively, as providing contexts for free and fair business competition unhindered by arbitrary biases, rules, or restrictions:

- The government will be **neutral about** the technology choices made by local companies..... After the WTO accession, the government will be **neutral on** technology choices, and companies that purchase technologies will make the decision based on their competitive strategies. The result is that technology development becomes less risky, because the company no longer needs to bet on one particular technology. (BC: Management 2003\_12)
- Mega-exchange, a model that is **neutral between** supplier and buyer and that: acts as a central trading hub to facilitate transactions between buyers and suppliers; is usually run by third-party market makers, where it gathers buyers and suppliers to enable efficient trading between them. (BC: Marketing 2005\_13)

Neutral conditions or relations are often regarded as favorable and profitable situations in

business discourse as no external pressures or constraints on business activity exist. The emphasis is on positive evaluations of situations in which no biases or preferences apply. Such instances express reversal prosody in a sense: The inherent meaning of a pattern is cancelled, and the opposite connotation is expressed by the disciplinary culture.

#### 4.4.3 Differences in message style: Hedges or clarity

The differences in two disciplines are further evident in message style. The patterns in the *USEFULNESS* group (e.g., ADJ for N) are, for instance, more likely to express a judgmental attitude toward things in hedged way with (semi-)modals and adverbs in ALC whereas they more likely to express their views confidently and non-ambivalently with no such modal meanings in BC. This difference lies in the distinction in the ways in which they view the world in the two disciplines: ALC has a more human-like, subjective view of the world while BC has a more objective, machine-like view of the world.

In the 20 examples with modals in ALC, most cases express the writer's attitude in a hedged or toned-down way, using *might* and *would*. On the other hand, in the 9 examples with modals in BC, most cases feature *will*, *may*, and *can*. Applied linguists are thus more likely to avoid making overt judgments than business researchers are. This difference is also apparent in instances featuring collocations with epistemic semi-modals.

[With modals in ALC]

g one. Although other metaphors may be appropriate for other learners, the learner per  
ctic priming what priming might be good for in everyday language 782 ALARIO ET AL. use  
controlled elicitation tests might be appropriate for pedagogically motivated inquiry, th  
terviewer, suggesting that it might be appropriate for use with learners of lower levels o  
f 5. She noted that the tasks might be appropriate for undergraduate students, but she was  
t of written and spoken texts might be appropriate for each level. Similarly, Urquhart and  
r fields (e.g. computation) may not be appropriate for describing real language data. (c)  
ified as a TCU because it would not be appropriate for the receiver to interrupt the calle  
h this level of attention might not be sufficient for awareness, for seeing the stimulus.  
post-hoc interpretations that woul be unsuitable for developing more specific hypotheses.  
It for everybody. PS: Yes, it would be fine for me if we could do it in Dutch. (Laughs) PB  
with barrier, hearing barrier would be sufficient for identification of that word. Thus, d  
esentations that in principle would be sufficient for differentiating known words from oth  
ic and articulatory relations would be sufficient for all distinctive features of the worl  
ening test was developed that would be suitable for the learners proficiency level. Thr  
prompts included wording that would be suitable for both male and female speakers. This wa  
erb of swinging, which would have been appropriate for the swinging event. It seems unlike  
corpus. This assumption may have been wrong for a few tokens only; thus, a possible lemma  
ertain a grammar that would have been appropriate for a Southern English listener instead

[With modals in BC]

a holistic approach to markets can be good for corporate profitability. One approach is to cooperate with government policy can be good for corporate profitability. To answer this question which specific HR functions may be eligible for outsourcing. It is important to distinguish would also imply that the time may be right for increased merger activity between stock markets and logistics considerations. It must be appropriate for the organization concerned. Organizations also argued that the program may not be suitable for other developing countries. 8 The number contributing to societal betterment will be good for business. Scholars offer two complementary probabilities that an employee will be eligible for health insurance at work. In the UK motorists advise a patient that exercise would be good for him, he has clearly both stated a fact about

[With semi-modals in ALC]

tests and their learning was found to be insufficient for explaining the complexities of teaching English. Because the test is meant to be appropriate for children in Grades 2 through 5, it is not good for Europe or the USA had to be good for KwaZulu (p. 22). From Pakistan, Shamim's message before the passage is said to be suitable for his or her use. When his or her score is 5% and 90%, the material is said to be suitable for use in supervised instruction. A Table 1072, 1857) in the study seemed to be adequate for the 3 PL MML estimation that usually requires a weight, though a pass is needed to be eligible for tertiary education. In contrast, the rates were, however, considered acceptable for the purposes of this study, since they are higher than Hong Kong, and as such is considered suitable for a graduating cohort irrespective of their letters, while e-mail is not considered appropriate for the conveyance of such letters. The results are shown below: (1) The industry was deemed suitable for testing our hypothesis [as it includes a range of knowledge, the words were deemed appropriate for this study. The resultant list of vocabulary items in the single model can be considered entirely appropriate for learners in different academic disciplines. The list of text-units in the ICE corpus seemed sufficient for this purpose. In the majority of cases, the material was suitable for Spanish. Although the stimuli seemed suitable for older infants it was uncertain whether they were suitable for younger children. 7). The positive characterization seems appropriate for our isiZulu / Ƙ / stimuli, which are

By the same token, collocating adverbials with this group in the pattern demonstrates that ALC is more likely to add modalizing comments of all kinds than business studies are. Table 7 summarizes these collocating adverbials, indicating that adverbials occur more than twice as frequently in ALC as they do in BC. By and large, the table shows that amplifiers, diminishers, commentative, and stance adverbs highlight how a writer feels about a judgment that he or she has made or is making. The fact that such features are much more strongly associated with ALC allows us to speculate that the subjectivity of the analyst is more foregrounded in applied linguistics than it is in business studies, where there seems to be more of an attempt to present arguments as having a logical inevitability to them.

**Table 7 Collocating Adverbials**

Order	ALC	Freq.	BC	Freq.
1	more	14	not	5
2	not	12	more	3
3	particularly	5	most	3
4	quite	3	no longer	2
5	most	3	particularly	2
6	entirely	2	especially	2
7	less	2	better	2
8	rhetorically	1	too	1
9	well	1	theoretically	1
10	very	1	less	1
11	so	1	extremely	1
12	developmentally	1	generally	1
13	easily	1	intrinsically	1
14	also	1	best	1
15	certainly	1		
16	eminently	1		
17	just	1		
18	largely	1		
19	necessarily	1		
Total		53		26

## [Applied Linguistics]

- Despite the acknowledgment that no single model can be considered entirely **appropriate for** learners in different academic disciplines, reasons are given to explain why constituent steps should be investigated in sufficient detail if ESP teachers are to provide a pedagogically meaningful model for second language learners in a particular discipline. (ALC: ESP 2006\_14)
- The proposed rating scale may be particularly **appropriate for** use by EAP learners... (ALC: ESP 2006\_21)
- This situation was obviously not **ideal for** addressing the research question more fully. (ALC: Testing 2004\_2)
- Bachman's definition of textual competence may be more **suitable for** us in this respect: Textual competence includes the knowledge of the conventions for joining utterances together to form a text... (ALC: Corpus 2005\_7)

## [Business Studies]

- Abraham, Seyyed, and Al-Elg (2001), in an overview of the stock markets in Bahrain, Kuwait, and Saudi Arabia, conclude that the three markets are **suitable for** international diversification purposes and also can be used to hedge against oil price fluctuations. (BC: Economics 2004\_8)
- The Conference champions its promulgated acts as worthy of state adoption based on its



conclusion that means-based uniformity is **appropriate for** this area of private law. (BC: Law 2000\_4)

- This period is **ideal for** an investigation of the steel industry since it traverses monumental changes from the 1970's decade of crises, the 1980's decade of catastrophe, and the 1990's renaissance (Stubbles, 1995). (BC: Strategy 2004\_4)
- The Kaiser-Meyer-Olkin (KMO) measure of sample adequacy indicated that the 28-item sample was not **adequate for** factor analysis (KMO measure=0.51). (BC: Management 2000\_22)

The instances in BC in particular describe the research target, approach, or consequence in a straightforward manner, since such usefulness has already been testified to by the previous studies, has been indicated by the research data (or specific logical reason), can be extrapolated from a given fact, or is common knowledge shared in the business discourse community.

#### 4.4.4 Disciplinary parameters in *attitude*

The discourse analysis for the patterns in the two disciplinary corpora can be used to identify the disciplinary parameters. Table 8 summarizes the disciplinary parameters speculated from the patterns for *attitude*.

**Table 8 Disciplinary Parameters in Attitude**

ALC	BC	Semantic Groups	Attitude Type
+neutral (-biased view)	+critical (+alternative view)	BIAS	Judgment
+human-like (+subjective)	+machine-like (+objective)	USEFULNESS	
+commentative	+concise	USEFULNESS	
+hedged (+unconfident)	+assertive (+confident)	USEFULNESS/IMPORTANCE	
+generalizing	+particularizing	IMPORTANCE	
+abstract target	+concrete target	IMPORTANCE	
+current target	+future target	CERTAINTY	
+optimistic (+subjective)	+cautious (+objective)	OPTIMISM/ENTHUSIASM	Affect
+emotion generated (+focus)	+emotion reported (-focus)	NERVOUSNESS	Appreciation
+generalizing	+particularizing	SKILFULNESS/WISDOM	
+internal	+external	SKILFULNESS	

Referring to the disciplinary discourse features of the patterns that express relational function as well, Nishina's (2010) results identified disciplinary cultures and values of applied linguistics and business studies from the pattern ADJ PREP N as follows:

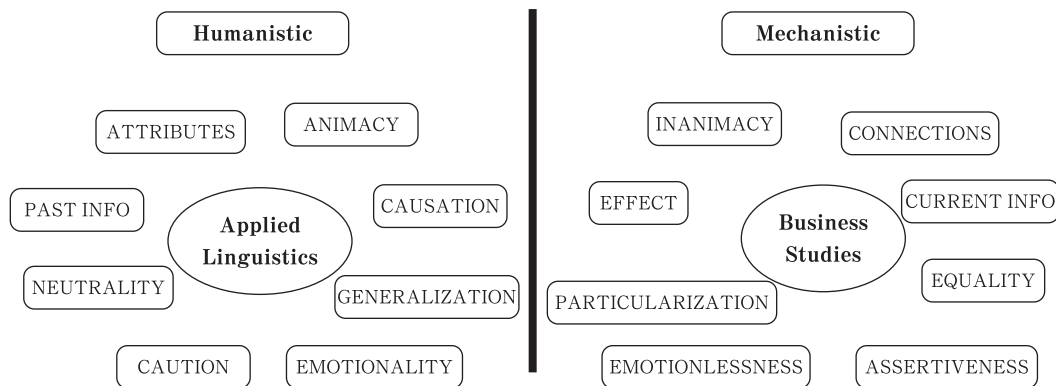


Figure 10 Disciplinary cultures and values of AL and BS

As shown in this study, one of the current trends in applied corpus linguistics is to endeavor to link the language use with the epistemological knowledge and disciplinary values.

## 5. Conclusion

### 5.1 Concluding Remarks

This paper first summarized historical and recent corpus-based studies by past and present researchers at the UoB, prioritizing how they commonly utilize corpus linguistic approaches to the identification of meaningful language phenomena (e.g., unit of meaning) from authentic texts. In particular, the discussion focused on how the purpose of such research is not fundamentally quantitative in nature. This paper also demonstrated how corpus linguistics and discourse analysis harmonize with one another and how this combination is an effective approach for identifying the link between language and culture. For instance, Nishina (2010) attempted to reveal the detailed cultural values of a particular discipline based on corpus-assisted discourse analysis. The primary essence in Birmingham school—namely, a word and a sequence of words are always contextual—still lives in much of the current studies at the UoB. This tradition will likely continue long into the future.

### 5.2 Future Corpus Research at the UoB

Hunston's recent studies tackle the interface among corpus linguistics, phraseology, and evaluative meanings. The direction of corpus research at the UoB will follow her studies and continue to extract qualitative discourse features from corpora. In particular, some ongoing and future topics for corpus studies at the UoB would be evaluation; semantic categories of lexicogrammatical patterns; links between language and (academic) culture; discourse features in sub-genres and sub-disciplines; and new word class in English. For instance, Hunston (2011) highlighted the importance of modal-like expressions in English largely because an English native speaker prefers modal-like expressions to modal verbs. As a result, learners of English consistently use modal auxiliaries more frequently than native speakers of English do (Aijmer,

2002). Since modal-like expressions express evaluative meanings, by listing such expressions, one set of texts can be measured against another to compare the amount and type of such evaluative language in each (Biber, 2006; Charles, 2006a, 2006b; Hyland, 2009). Therefore, it would be a qualitative and meaningful study to identify such evaluative features from corpora and scrutinize their discourse features.

### Notes

- 1 The current director of the CCR is Professor Paul Thompson; the technical director is Dr. Oliver Mason.
- 2 The current director of the DRC is Dr. Rosamund Moon.
- 3 For more information on CHAB, visit <https://arts-ccr-002.bham.ac.uk/chab/> (Note: the interface is currently under development).
- 4 The Birmingham Concordancer is a web application that offers various tools from modern concordancers.
- 5 The SQL-Based Corpus Analyser (SCAN) is a corpus search engine that stores corpora in a MySQL database.
- 6 A large collection of tools for compiling corpora have been developed as part of CHAB, particularly for use with SCAN.
- 7 The phrase *at mo* does not occur in the BoE and is peculiar to TMs (Tagg, 2009, p. 241).
- 8 For instance, Hyland and Tse (2004, p. 157) defined *metadiscourse* as “the *linguistic resources* used to organize a discourse or the writer’s stance towards either its content or the reader”; Conrad and Biber (2000, p. 57) defined *stance* as “a cover term for the expression of personal feelings and assessments.”
- 9 *Monogloss* can be detected through personal pronouns, adverbs, modals, and so on.
- 10 The hint of the concept for the disciplinary prosody in this study emerged from O’Halloran’s (2007, 2010) register prosody and semantic prosody in business discourse in Nelson (2006).
- 11 The comparative value—one of three evaluative values (*viz.* comparative, subjective, and social)—indicates a situation in which something is compared to the disciplinary, cultural, or social norm (Labov, 1972; Thompson & Hunston, 2000).

### References

- Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 55–76). Amsterdam: Benjamins.
- Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories* (2nd ed.). Milton Keynes: SRHE/Open University Press
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5, 97–116.
- Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text*, 9 (1), 93–124.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Caldas-Coulthard, C.R., & Coulthard, M. (Eds.). (1996). *Texts and Practices: Reading in Critical discourse analysis*. London: Routledge.
- Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35–53). Norwood, NJ: Ablex.
- Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. In D. R.

- Olson, N. Torrance & A. Hildyard (Eds.), *Literacy, language, and learning: The nature and consequences of reading and writing* (pp. 105–123). Cambridge: Cambridge University Press.
- Chafe, W., & Danielewicz, J. (1986). Properties of spoken and written language. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 83–113). New York, NY: Academic Press.
- Charles, M. (2003). “This mystery...”: A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes* 2, 313–326.
- Charles, M. (2004). The construction of stance: A corpus-based investigation of two contrasting disciplines. Unpublished doctoral dissertation, University of Birmingham, Birmingham, UK.
- Charles, M. (2006a). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25(3), 310–331.
- Charles, M. (2006b). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics* 27, 492–518.
- Charles, M. (2009). Stance, interaction and the rhetorical patterns of restrictive adverbs: Discourse roles of *only*, *just*, *simply* and *merely*. In M. Charles, D. Pecorari & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 152–169). London: Continuum.
- Conrad, S., & Biber, D. (2000). Adverbial marking of stance in speech and writing. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 56–73). Oxford: Oxford University Press.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29–62.
- Fang, W. (2009). A corpus-based analysis of the lexical item “international community”. *The Birmingham Journal of Literature and Language*, 2, 52–61.
- Fillmore, C. J. (1977). The case for case reopened. In P. Cole & J. Sadock (Eds.), *Syntax and semantics. vol. & Grammatical relations* (pp. 59–82). New York: Academic Press.
- Fillmore, C. J. (1977). Topics in lexical semantics. In R.W. Cole (Ed.), *Current issues in linguistic theory* (pp. 76–138). Bloomington: Indiana University Press.
- Fillmore, C. J. (1985). Frames and semantics of understanding. *Quaderni di Semantica*, 6, 222–254.
- Firth, J. R. (1957). *Papers in linguistics 1934–1951*. Oxford: Oxford University Press.
- Firth, J. R. (1935). The technique of semantics. *Transactions of the Philological Society*, 7, 36–72.
- Fontenelle, T. (Ed.). (2003). Special issue: FrameNet and frame semantics. *International Journal of Lexicography*, 16(3), 235–50.
- Francis, G., Hunston, S., & Manning, E. (1996). *Collins COBUILD grammar patterns 1: Verbs*. London: Harper Collins.
- Francis, G., Hunston, S., & Manning, E. (1998). *Collins COBUILD grammar patterns 2: Nouns and adjectives*. London: Harper Collins.
- Gries, S. T. (2008). *Quantitative corpus linguistics with R: A practical introduction*. London: Routledge.
- Groom, N. (2007). Phraseology and epistemology in humanities writing. Unpublished doctoral dissertation, University of Birmingham, Birmingham, UK.
- Halliday, M.A.K. (1993). Quantitative studies and probabilities in grammar. In M. Hoey (Ed.), *Data description discourse* (pp. 1–25). London: HarperCollins.
- Hunston, S. (2000). Evaluation and the planes of discourse: Status and value in persuasive texts. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 176–207). Oxford: Oxford University Press.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. (2006). Phraseology and system: A contribution to the debate. In G. Thompson & S. Hunston (Eds.), *System and corpus: Exploring connections* (pp. 55–80). London: Equinox.
- Hunston, S. (2008). Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics*, 13(3), 271–295.
- Hunston, S. (2011). *Corpus approach to evaluation: Phraseology and evaluative language*. New York: Routledge.
- Hunston, S., & Francis, G. (1999). *Pattern grammar: A corpus-driven approach to the lexical grammar of*

- English*. Amsterdam: John Benjamins.
- Hunston, S., & Sinclair, J. (2000). A local grammar of evaluation. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 74–101). Oxford: Oxford University Press.
- Hunston, S., & Thompson, G. (2000). *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman.
- Hyland, K. (2009). Corpus informed discourse analysis: The case study of academic engagement. In M. Charles, D. Pecorari & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 110–128). London: Continuum.
- Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25, 156–176.
- Krishnamurthy, R. (Ed.). (2004). *English collocation studies: The OSTI report*. London and New York: Continuum.
- Labov, W. (1972). *Language in the inner city*. Philadelphia, PA: University of Pennsylvania.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 157–176). Amsterdam: John Benjamins.
- Martin, J. R., & Rose, D. (2007). *Working with discourse: Meaning beyond the clause* (2nd ed.). London: Continuum.
- Martin, J. R., & White, P. (2005). *The language of evaluation: Appraisal in English*. London: Palgrave.
- Mason, O. (2000). *Programming for corpus linguistics: How to do text analysis with java*. Edinburgh: Edinburgh University Press.
- Mason, O., & Hunston, S. (2004). The automatic recognition of verb patterns: A feasibility study. *International Journal of Corpus Linguistics*, 9(2), 253–270.
- Matthiessen, C. (2006). Frequency profiles of some basic grammatical systems: An interim report. In G. Thompson & S. Hunston (Eds.), *System and corpus: Exploring connections* (pp. 103–142). London: Equinox.
- Moon, R. (1998). *Fixed expressions and idioms in English: A corpus based approach*. Oxford: Clarendon Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nelson, M. (2006). Semantic associations in business English: A corpus-based analysis. *English for Specific Purposes*, 25(2), 217–234.
- Nishina, Y. (2010). Evaluative meaning and disciplinary values: A corpus-based study of adjective patterns in research articles in applied linguistics and business studies. Doctoral dissertation, University of Birmingham, Birmingham, UK.
- O'Halloran, K. (2007). Critical discourse analysis and the corpus-informed interpretation of metaphor at the register level. *Applied Linguistics*, 28(1), 1–24.
- O'Halloran, K. (2010). Investigating metaphor and ideology in hard news stories. In S. Hunston & D. Oakey (Eds.), *Introducing applied linguistics: Concepts and skills* (pp. 100–107). New York: Routledge.
- Palmer, H. & A. S. Hornby. (1933). *The second interim report on English collocations*. Tokyo: Kaitakusha.
- Partington, A. (2004). Utterly content in each other's company: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9(1), 131–156.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). London: Longman.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. M. (1996). The search for units of meaning. *Textus*, 9(1), 75–106.

- Sinclair, J. M. (1998). The lexical item. In E. Weigand (Ed.), *Contrastive lexical semantics* (pp. 1-24). Amsterdam: Benjamins.
- Sinclair, J. M. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.
- Sinclair, J. M. (Ed.). (1987). *Collins COBUILD English language dictionary* (1st ed.). London: Williams Collins Sons & Co Ltd.
- Sinclair, J. M. (Ed.). (2006). *Collins COBUILD advanced learner's English dictionary* (5th ed.). Glasgow: HarperCollins.
- Stubbs, M. (1995). Corpus evidence for norms of lexical collocation. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 245-256). Oxford: Oxford University Press.
- Stubbs, M. (2001). *Words and phrases*. Oxford: Blackwell.
- Tagg, C. (2009). A corpus linguistics study of SMS text messaging. Unpublished doctoral dissertation, University of Birmingham, Birmingham, UK.
- Teubert, W. (2003). Writing, hermeneutics, and corpus linguistics. *Logos and Language*, 4, 1-17.
- Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10(1), 1-13.
- Teubert, W. (2008). Sinclair, pattern grammar and the question of *hatred*. *International Journal of Corpus Linguistics*, 12, 223-247.
- Teubert, W., & Cermakova, A. (2007). *Corpus linguistics: A short introduction*. London and New York: Continuum.
- Thompson, G., & Hunston, S. (2000). Evaluation: An introduction. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 1-27). Oxford: Oxford University Press.