

# A multi-dimensional contrastive study of English abstracts by native and non-native writers

---

Yan Cao<sup>1</sup> and Richard Xiao<sup>2</sup>

## Abstract

This article takes the multi-dimensional (MD) analysis approach to explore the textual variations between native and non-native English abstracts on the basis of a balanced corpus containing English abstracts written by native English and native Chinese writers from twelve academic disciplines. A total of 47 out of 163 linguistic features are retained after factor analysis, which underlies a seven-dimension framework representing seven communicative functions. The results show that the two types of abstracts demonstrate significant differences in five out of the seven dimensions. To be more specific, native English writers display a more active involvement and commitment in presenting their ideas than Chinese writers. They also use intensifying devices more frequently. In contrast, Chinese writers show stronger preferences for conceptual elaboration, passives and abstract noun phrases no matter whether the two types of data are examined as a whole or whether variations across disciplines are taken into account. The results are discussed in relation to the possible reasons and suggestions for English abstract writing in China. Methodologically, this study innovatively expands on Biber's (1988) MD analytical framework by integrating colligation in addition to grammatical and semantic features.

**Keywords:** academic writing, contrastive analysis, Chinese, MD analysis, research article abstracts

---

<sup>1</sup> College Foreign Language Teaching Department, Ludong University, and the Key Laboratory of Language Resources Development and Application of Shandong Province, China.

<sup>2</sup> Department of Linguistics and English Language, County South, Lancaster University, Lancaster, LA1 4YL, United Kingdom.

*Correspondence to:* Richard Xiao, *e-mail:* r.xiao@lancaster.ac.uk

## 1. Introduction

This paper aims to explore the textual variations between native and non-native English abstracts from a contrastive perspective by using the multi-dimensional (MD) analytical approach on the basis of a sizeable corpus which is composed of two balanced and matching components representing English abstracts written by native English and native Chinese speakers from twelve academic disciplines.

A research article (RA) is written to communicate new knowledge to members of the academic community and persuade them to accept the author's claims (Hyland, 2000). As an important genre in academic discourse, the RA has been the focus of English for Academic Purposes (EAP) studies over the past decades, and has been approached from both macro- and micro perspectives. The macro view is largely concerned with rhetorical patterns. For example, the move-based studies of the Swalesian school have examined the rhetorical organisation across different sections of RAs and the overall structural organisation (e.g., Swales, 2004; Swales and Feak, 2004; and Swales and Luebs, 2002). From a micro perspective, an extensive range of specific linguistic features of RAs have been explored including, for example, voice (Martínez, 2001), tenses (Malcolm, 1987), hedging (Hyland, 1996, 1998), reporting verbs (Thompson and Ye, 1991), personal pronouns (Harwood, 2005), and a combination of a variety of features such as attitudinal stance adjectives, epistemic stance adjectives, self-reference words and *that*-complement clauses (Pho, 2008).

The abstract as the essence of a research article not only has a significant role to play in RA information search, indexing and citation, but also 'serves as an effective tool for readers to master and manage the ever increasing information flow in the scientific community' (Ventola, 1994: 333). In addition, a concisely and accurately presented abstract will make it easier for a non-native writer to be admitted into the research community of their discipline. Hyland (2004) notes that the way that abstracts are written convinces potential readers of the articles that the writers have the professional credibility to discuss their topic as an 'insider', which could, ultimately, encourage the readers to read the related articles. In this sense, a good abstract 'sells' the article (Pho, 2008). As a critical part of the RA and a specific genre in its own right (Hyland, 2004), abstract writing has become a focus of research that has received considerable attention in recent years.

A range of topics have been addressed in previous studies of abstracts, but they have focussed primarily on the macro structure—limited in both the subject areas that are covered and the number of linguistic features that are investigated. In contrast, this study covers twelve disciplines and involves 163 linguistic features. In order to manipulate such a large number of linguistic features effectively, we adopt the multi-feature, MD analysis framework pioneered by Biber (1988) in this study. On the basis of a sizeable corpus composed of two matching components that represent English abstracts written by native English and native Chinese speakers from

twelve academic disciplines, we aim to explore systematically the textual variations between native and non-native English abstracts across different subject areas. Methodologically, this study seeks to expand innovatively on Biber's (1988) original MD framework and Xiao's (2009) extended MD model by integrating colligation into the MD analytical framework.

In the sections that follow, we will first introduce the design and annotation of the corpus used in the study (Section 2). Section 3 contains a review of the MD analysis, discusses the selection of linguistic features, establishes a new MD model of RA abstracts through factor analysis, and interprets the resulting dimensions. The new model provides a useful basis for the contrastive analysis, in Section 4, of English abstracts written by native speakers (NS) and non-native speakers (NNS) as a whole as well as at the level of discipline. Finally, Section 5 concludes the article by summarising our major research findings and discussing the implications of our research.

## **2. Corpus design and corpus annotation**

Given that no existing corpus available to us was suitable for a large-scale MD contrastive study of NS and NNS English abstracts, we decided to design and develop the comparable corpus required for our research. The corpus comprises two subcorpora, one for native English speakers' English abstracts (the NS corpus) and the other for Chinese speakers' English abstracts (the NNS corpus). Both subcorpora cover the same twelve academic disciplines (namely, mathematics, physics, chemistry, biology, mechanics, astronomy, oceanography, earth science, survey, atmospheric science, geology and geography), which will not only enable us to avoid potential disciplinary biases in abstract selection but also makes it possible to explore variations between native and non-native writers as well as across academic disciplines.

To ensure the reliability of our sources of data, we first obtained a list of representative journals by referring to three authoritative Chinese guidebooks, namely, *A Handbook of Core Foreign Journals of Science and Technology* (Dai and Cai, 2004), *Chinese Sci-Tech Journal Citation Reports* (Pan and Ma, 2010) and *A Comprehensive List of Chinese Core Journals* (Ren, 2008). When selecting native English speakers' abstracts, we took account of both the impact factor, which provides a rough measure of the relative importance of an academic journal in a particular field, and the balance of sub-disciplines. We mainly considered the most prestigious journals in each discipline as reflected by their impact factors, assuming that the journals' qualities are positively correlated with the native speakers' professional writing level, and therefore they can be used as a norm against which Chinese samples can be compared. As for the Chinese speakers' abstracts, we decided to select journals with varying impact factors so that they could represent Chinese writers' average writing level. All the

Subcorpus	No. of texts	No. of tokens	Average text length
NNS	6,356	1,053,959	166
NS	4,385	894,084	203
Total	10,741	1,948,043	181

**Table 1:** Corpus size

abstracts were randomly selected and evenly distributed in the chosen journals published in 2007–8 for NS abstracts and in 2008–10 for NNS abstracts. The resulting corpus contains 1,948,043 words, with the NNS and NS components amounting to approximately 1,050,000 and 890,000 words, respectively, as shown in Table 1, which also indicates that NS abstracts are, on average, longer than NNS abstracts.

Each abstract in the corpus is marked up with metadata information about its provenance (i.e., whether it is written by a NS or a NNS) and discipline. In addition, all the textual data in the corpus is automatically annotated with word class and semantic information by using the web-based Wmatrix system that combines the CLAWS part-of-speech tagger and the USAS semantic tagger (Rayson, 2003, 2008). The tagset used for part-of-speech tagging in this study is CLAWS C7, which has consistently achieved 96 to 97 percent accuracy and can provide detailed word-class categories. The USAS semantic tagger has a multi-tier structure with twenty-one major discourse fields, subdivided into 232 category labels and containing nearly 37,000 words and over 16,000 multi-word units (Archer *et al.*, 2002). Such detailed word class and semantic tags, and combinations of the two, have facilitated the retrieval of an extensive range of linguistic features used in establishing the MD model in this study (see Section 3.2).

### 3. Multi-dimensional analysis of RA abstracts

This section first provides a brief introduction to the MD analysis approach, which was pioneered by Biber (1988), and is followed by a discussion of our initial selection of linguistic features that are included in the factor analysis. We will then present the new MD model of RA abstracts established in this study and interpret each of the seven dimensions on the basis of the positively and negatively loaded linguistic features.

#### 3.1 An introduction to MD analysis

Originally developed by Biber (1988) to identify and compare the systematic variation between spoken and written registers, MD analysis has been extended in both methodology and research domains over the past

twenty-five years. The approach is primarily concerned with the identification and interpretation of systematic register variation as revealed by the underlying patterns of linguistic features. With comprehensive linguistic features and registers of different domains involved, 'it gives a formal status to the notion of linguistic co-occurrence' (Biber, 2004: 16), assuming that statistical co-occurrence patterns reflect underlying shared communicative functions. The comprehensive view of MD analysis demonstrates that it is functional in nature. The diverse variation studies that have been enabled and inspired by MD analysis facilitate a better understanding of the interaction between linguistic styles and register variation. On the one hand, linguistic features reflect 'the discourse functions of such features in relation to the situational characteristics of each register' (Biber, 2004: 15); on the other hand, register difference also helps to explain the patterns of variation for a specific text type.

The MD approach has been applicable to a diverse range of types of research into language variation, and addresses the issues of written and spoken language, including those that are synchronic (Biber, 1991; Biber and Finegan 1994a; Conrad, 1994; and Reppen, 1994) and diachronic (Biber and Finegan, 1989, 1992; and Conrad and Biber, 2001), and contrastive analyses of different languages (Biber, 1995), ESP studies (Biber *et al.*, 1998), text-type analyses (Biber, 1989; Biber and Finegan, 1994b; Grieve, 2011; Jones, 2007; and Watson, 1994), and world Englishes (Xiao, 2009). However, the MD approach has probably made its most significant impact in the area of academic discourse studies, providing a complementary approach to the traditional rhetorical and linguistic analyses of academic discourse. The relevant research includes both synchronic and diachronic studies in areas such as university spoken and written registers (Biber, 2003; Biber, *et al.*, 2004; and Csomay, 2005, 2007), and specific academic disciplines (Biber and Finegan, 1994b; and Conrad, 1996, 2001). In this study, we follow the steps outlined in Biber (1988) and Xiao (2009) and further include categories of colligation as linguistic features to be examined in an attempt to provide a stronger interpretation of register variation in RA abstracts.

### **3.2 Selection of linguistic features for factor analysis**

Biber (1988) bases his initial model of underlying dimensions on the co-occurring patterns of sixty-seven grammatical features. His recent work also includes semantic features of word classes together with part of their lexico-grammatical information in his new model, with 129 linguistic features included for factor analysis (Biber, 2007; and Biber *et al.*, 2004). Xiao (2009) enhances the MD model by incorporating a more comprehensive set of semantic categories using the annotation results of the Wmatrix system. A total of 141 linguistic features are used in the study of world Englishes, language variation across different registers and world English

varieties. Crossley and Louwse (2007) introduce bigrams into the MD analytical framework, demonstrating its strength for classifying spoken and written registers. Since we did not know, before the model was established, which linguistic features would be sufficiently strong and significant, we followed Biber's (1995) suggestion that as many features as possible should be included, initially, and at the lowest possible level of groupings. Hence, our initial selection included the features used in previous MD studies, but, in addition, we also introduced into the MD framework the notion of colligation, which was motivated by Sinclair's (1991) concept of co-selection as a central mechanism for describing language in use.

Sinclair's (1991) model of extended lexical units 'reconciles the paradigmatic and syntagmatic dimensions of choice at each choice point' (Sinclair, 2004: 141), asserting that 'form and meaning cannot be separated because they are the same thing' (Sinclair, 2004: 139). Highly patterned construction in language is considered as a major finding of modern corpus linguistic research over the past four decades (Römer, 2009: 140). It has been the focus of a range of corpus-based studies employing different terminologies, (e.g., pattern, collocation, colligation, multi-word units, lexical bundles, n-gram, construction, among others), but all emphasise the inter-dependence of form and meaning (Biber, 2006; Biber *et al.*, 1999, 2004; Hoey, 2005; Hunston and Francis, 2000; Hyland, 2008; and Goldberg, 2006). Crossley and Louwse (2007) classify registers using the frequency of bigrams shared among nine spoken and two written corpora, the findings of which demonstrate that the phrasal units and grammatical constructions can function as a powerful approach to MD analysis. Indeed, as Gries *et al.* (2011) observe, 'a pure n-gram-based approach can be used as an initial, computationally cheap, way of classifying corpus registers that produces useful results.'

According to Gries *et al.* (2011), colligation patterns in the form of 3-grams are tested to have the greatest discriminatory power and a great potential in pattern research. In comparison with bigrams, 3-grams and 4-grams are preferable because longer word clusters contain more information than shorter lexical bundles, and they are more distinguishable because they require writers to be more highly productive in terms of their language. However, in our study, 5-grams are also excluded because they are relatively rare in a corpus of moderate size. We decided to focus mainly on 3-grams and employed those patterns identified by our previous study of colligation (Li and Cao, 2013). We first based our selection on the categories of parts of speech, (namely, nouns, verbs, adjectives, adverbs and prepositions), with the verbs further divided into five subcategories (i.e., past participle, predicate verb, verb *be*, *-ing* form and infinitive). Then we selected from each group the most frequently used 3-grams, using the main word classes as the node. We also included a few bigrams and 4-grams in the list because they show stronger collocational tendencies than 3-grams. For example, the bigram 'verb + adverb' in the adverb category shows a closer relationship than its longer sequences, and is, therefore, included. Finally, forty n-grams from five

categories (see Appendix A) were selected for inclusion on our feature list for factor analysis.

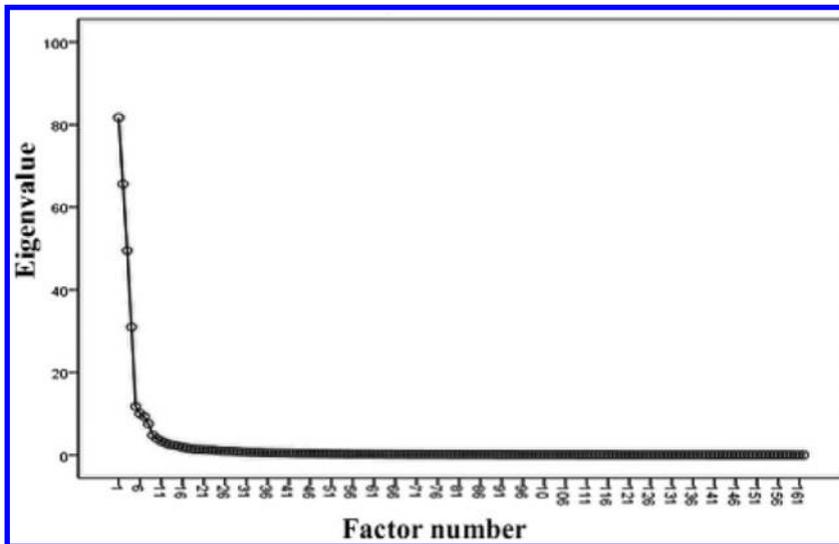
Our feature list includes a total of 163 linguistic features, which can be divided into three categories—that is to say, grammatical features, semantic features and colligation patterns. Linguistic features that rarely occur in both subcorpora were dropped and small groups with relatively low frequencies were aggregated into larger ones. For example, the seven subcategories of degree adverbs were combined into one category. We used the search algorithms developed by McEnery *et al.* (2006) together with the additional search patterns designed in our project to retrieve the required linguistic features, with the help of WordSmith 6.0 (Scott, 2011), and PatCount (Liang and Xiong, 2008)—where necessary, in combination with the grammatical and semantic tags produced by Wmatrix. In our study, raw frequencies were normalised to a common base of 100 words, instead of 1,000 words as in Biber (1988), because an abstract typically contains a few hundred words and artificially reducing or enlarging the base of normalisation may distort the picture. Finally, the results were ready for factor analysis using SPSS (v. 19.0), as described in the following section.

### 3.3 The factorial structure and interpretations of dimensions

Factor analysis is commonly used in linguistic research to group a large number of linguistic features into small sets of correlated features which are known as factors or dimensions, representing the underlying communicative functions. In this study, we have followed the same factor extraction method used in Biber (1988), with factors extracted using Principal Axis Factoring and rotated using Promax. A total of forty-seven linguistic features, with an absolute value of loading greater than 0.30 on a factor, are finally retained in our MD model after factor analysis. By examining both the covariance matrix and the scree plot (see Figure 1) yielded in factor analysis, we find a seven-dimension factorial structure to be optimal for the subsequent interpretations of the extracted factors, which account for 79.6 percent of the total variance after rotation.

In comparison with previous MD studies, a distinguishing feature of our new model lies in the co-occurrence of items in the same word class within one dimension. For example, Dimension 1 is primarily concerned with the use of adverbs while Dimension 3 is exclusively associated with adjectives. The following interpretative labels are proposed, on the basis of co-occurring linguistic features, for the dimensions of the factorial structure in our new model:

- Dimension 1: focussing and intensification
- Dimension 2: active involvement and interaction
- Dimension 3: explicit conceptualisation of methodology



**Figure 1:** The scree plot of factor analysis

- Dimension 4: conceptual elaboration
- Dimension 5: formal, abstract and impersonal style
- Dimension 6: textual cohesion
- Dimension 7: informational density

In Sections 3.3.1 to 3.3.7, we will interpret the communicative functions of each dimension with reference to the co-occurring features loaded on to each factor.

### 3.3.1 Dimension 1: focussing and intensification

The first dimension is primarily concerned with various forms and functions of adverbs (see Table 2). The most salient feature, ‘other adverbs’, refers to all adverbs other than intensifying adverbs, hedges, amplifiers, downtoners as well as place adverbials and time adverbials, while adverbs of degree and measurement, boosters and general emphatics are all linked to intensification. Such adverbs are often used in abstracts to modify verbs to indicate degrees, manners and authors’ stances. Structures with split auxiliaries such as split infinitives, where the auxiliaries and the lexical verbs are typically separated by adverbs of focus, time, manner and degree, are frequently associated with a following focus which may involve ‘intensification or comparison’ (Hasselgård, 2010: 110). Hence, this dimension can be aptly labelled as ‘focussing and intensification’.

	Features	Weight	Mean	Std. dev.
1	other adverbs	1.024	2.59	1.73
2	adv: degree	0.539	0.49	0.62
3	adv: measurement	0.518	0.50	0.61
4	adv+V	0.491	0.44	0.52
5	adv: boosters	0.409	0.23	0.39
6	V+adv	0.346	0.17	0.31
7	general emphatics	0.327	0.15	0.35
8	split auxiliaries	0.305	0.29	0.48

**Table 2:** Factor 1 – focussing and intensification

	Features	Weight	Mean	Std. dev.
1	pronouns	0.758	1.98	1.51
2	reladel <sup>3</sup>	0.745	0.86	0.98
3	first-person pronouns	0.679	0.45	0.88
4	<i>we</i> +V+N/adj/adv/Art	0.541	0.13	0.38
5	<i>that</i> -clauses as verb complements	0.436	0.48	0.57
6	<i>we</i> +V+ <i>that</i>	0.348	0.03	0.13
7	private verbs	0.336	1.01	0.86
8	demonstratives	0.331	0.63	0.74
9	V+N/adj/adv/Art +N	0.307	0.41	0.59
10	pronoun <i>it</i>	0.302	8.00	3.45

**Table 3:** Factor 2 – active involvement and interaction

### 3.3.2 Dimension 2: active involvement and interaction

The linguistic features loaded on Dimension 2 essentially fall into two groups (see Table 3). The first group includes various types of pronouns (first-person pronouns, demonstratives, pronoun *it* and other pronouns), *that*-deletion and private verbs, which are all features positively loaded on the first dimension in Biber's (1988) original MD model, (i.e., 'informational *versus* involved interaction'), and indicate an involved and interactive style. These features naturally co-occur with colligation patterns in the second group (e.g., '*we* + verb' followed by a noun or *that*-clause), which are frequently used in academic writing to show the author's active involvement or evaluation and stance. This dimension is labelled, 'active involvement and interaction'.

	Features	Weight	Mean	Std. dev.
1	nouns: mental object	0.823	1.87	1.70
2	verbs: using	0.359	0.33	0.47
3	verbs: getting and giving; possession	0.353	1.01	0.85
4	science and technology in general	0.346	2.12	1.82
5	nouns: evaluation	0.305	1.28	1.15
6	verbs: speech acts	0.304	0.68	0.69
7	(preposition)	-0.535	13.79	3.15
8	<i>be</i> as main verb	-0.342	1.29	1.07

**Table 4:** Factor 3 – explicit conceptualisation of methodology

### 3.3.3 Dimension 3: explicit conceptualisation of methodology

In Dimension 3 (see Table 4), nouns of mental object, verbs of using and verbs of getting, giving and possession express abstract concepts, particularly the means and methods used in research. It is not surprising that they tend to co-occur with lexis of science and technology in academic prose. Nouns of evaluation are a positive loading on this factor because scientific writing in essence deals with problems of various kinds (cf. Csomay, 2005) while verbs of speech acts and communication are essential in making explicit statements or arguments in scientific writing. Note that ‘*be* as main verb’, which is a feature for involvement in Biber’s (1988) model, is a negative loading on this factor, because the conceptualisation of methodology is typically abstract and informationally dense. Through combined use of these method-orientated features, writers can engage readers in understanding and following their research process. Therefore, we refer to this dimension as ‘explicit conceptualisation of methodology’.

### 3.3.4 Dimension 4: conceptual elaboration

Dimension 4 builds on a range of features related to adjectives and their colligation patterns (see Table 5), particularly adjectives denoting measurement, physical attributes, importance and evaluation. The most salient feature loaded on this dimension is attributive adjectives, which ‘allow scientists to successfully describe, clarify, and qualify additional information about scientific phenomena or entities [thus providing] a more conceptual description of referents’ (Biber *et al.*, 2007: 91). Noun phrases with heavy pre-modification by adjectives and nouns, as reflected in the colligation patterns, are usually long technical terms in

	Features	Weight	Mean	Std. dev.
1	attributive adjectives	0.923	8.64	3.01
2	adj: measurement	0.522	1.75	1.30
3	adj+adj+N	0.452	0.61	0.74
4	adj: physical attributes	0.396	0.98	1.06
5	adj+N+N	0.374	1.06	0.93
6	prep+art+adj+N	0.368	0.71	0.75
7	adj: importance	0.330	0.77	0.76
8	adj: evaluation	0.319	0.76	0.77

**Table 5:** Factor 4 – conceptual elaboration

scientific writing, and thus result in high informational density (Biber *et al.*, 2007). As Biber *et al.* (2006: 14) observe, ‘The reliance on nouns and complex noun phrases results in a style of text with dense informational content packed into relatively few words.’ It is clear that this dimension is essentially concerned with ‘conceptual elaboration’ and is thus labelled.

### 3.3.5 Dimension 5: formal, abstract and impersonal style

The linguistic features used in computing the factor score of Dimension 5 (see Table 6) include passives (agentless and *by*-passives, and the past participle structure) and verbs that semantically denote general and abstract terms relating to being and existence, which pair perfectly well because of their common abstractness. Passive constructions obscure identification of the agent, which in turn highlights the role of the corresponding patient. This pattern also helps to achieve discourse cohesion in terms of theme and maintains an objective tone. In academic writing, many verbs frequently occur in their passive forms, used either to describe the scientific methodology and research procedure, or to report research findings. It is well documented in the literature that English passives express abstract information in a formal and impersonal style (e.g., Biber, 1988; Biber *et al.*, 1999; and Xiao *et al.*, 2006). Hence, this dimension can reasonably be labelled as ‘formal, abstract and impersonal style’.

### 3.3.6 Dimension 6: textual cohesion

Dimension 6 is based on linguistic features associated with prepositions (see Table 7), including their colligation patterns, in addition to other grammatical

	Features	Weight	Mean	Std. dev.
1	agentless passives	0.899	2.16	1.57
2	verbs: being	0.689	2.45	1.36
3	<i>be</i> +past participle+preposition	0.531	0.56	0.62
4	<i>by</i> -passives	0.405	0.27	0.51
5	(V+preposition)	0.304	1.58	0.99

**Table 6:** Factor 5 – formal, abstract and impersonal style

	Features	Weight	Mean	Std. dev.
1	preposition	0.851	13.79	3.15
2	grammatical bin	0.661	22.28	7.65
3	<i>the</i> +N+ <i>of</i>	0.480	8.00	3.45
4	prep+art+N+ <i>of</i>	0.480	0.42	0.53
5	V+preposition	0.410	1.58	0.99
6	(prep+art+adj+N)	0.303	0.71	0.75

**Table 7:** Factor 6 – textual cohesion

words such as articles and conjunctions which, together with prepositions, are visually labelled by Wmatrix as ‘grammatical bin’. Such words primarily serve a grammatical, rather than a semantic, function to express relational information. The frequent use of sequences of prepositional phrases is one of the features of academic writing (Gillet, 2013). In the English abstracts in our corpus, the most common preposition, *of*, accounts for 31 percent of the total, with the ten most frequently used prepositions (*of*, *in*, *to*, *for*, *with*, *by*, *from*, *on*, *at* and *as*) accounting for 86.7 percent. The colligations ‘*the* + noun + *of*’ and ‘preposition + article + noun + *of*’ are both very common patterns in English, while ‘verb + preposition’ is mostly a formulaic expression that is almost fixed structurally and semantically. Clearly, this dimension is characterised by the highly frequent use of prepositional word clusters, including formulaic expressions, which all contribute to textual cohesion (Xiao, 2011), and this accounts for the name we have given to this dimension.

### 3.3.7 Dimension 7: informational density

The features loaded on Dimension 7 are largely those that are informationally dense (see Table 8). Nominalisation is a kind of grammatical metaphor that is commonly used in academic writing to provide densely compacted abstract

	Features	Weight	Mean	Std. dev.
1	nominalisation	1.078	8.00	3.45
2	mean word length	0.503	5.34	0.45
3	nouns: affect	0.384	2.05	1.51
4	N+co-ordinating conj+N	0.366	0.52	0.65
5	phrasal co-ordination	0.305	1.11	1.12

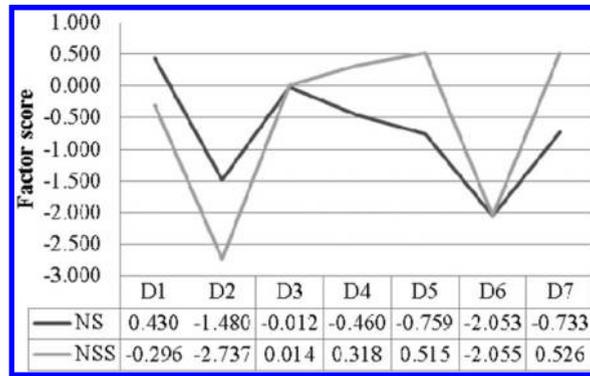
**Table 8:** Factor 7 – informational density

information, while greater mean word length and phrasal co-ordination, particularly conjoined nouns, also add to information density (cf. Biber, 1988: 227). Nouns of affect are general or abstract terms that either denote change or the propensity for change, or indicate a causal relationship or lack of it (Archer *et al.*, 2002), suggesting that the frequent use of nouns of this semantic category signals a high informational load in discourse. As can be seen, all of the positive loadings on this dimension relate to information density.

While features such as nominalisations, attributive adjectives, prepositions and passives tend to co-occur on one dimension in many factor analyses (e.g., Biber, 1988), reflecting the information-orientated characteristics, they are distinguished as different dimensions in this study, having their own emphases and distinctive communicative functions. To be specific, Dimension 4 mainly focusses on the compact conceptual elaboration; Dimension 5 highlights the style of formality, objectivity and abstractness; while Dimensions 6 and 7 are linked to textual cohesion and informational density, respectively. With regard to the possible reason, we consider that most of Biber's studies are concerned with two markedly distinct genres, spoken *versus* written English, while our sources of data are both abstracts with very similar styles of academic writing.

#### 4. NS and NNS abstracts in contrast

Following the three formulae for computing factor scores as proposed in McEnery *et al.* (2006: 303), we used the normalised frequency of each linguistic feature in each text together with the mean frequency and standard deviation of the feature to compute the factor score of a given linguistic feature in each text. The factor scores of these linguistic features were then used to compute the factor score of a text and finally of the NS and NNS categories of RA abstracts. Computer programs were written to do the complex and time-consuming computations in batches. With the factor scores obtained, we can now compare the overall dimensional variations of



**Figure 2:** Factor scores of NS and NNS abstracts along the seven dimensions

NS and NNS abstracts, followed by more fine-grained comparisons across disciplines, discussing possible reasons for the observed similarities and differences.

Figure 2 provides an overall comparison of the mean factor scores of NS and NNS abstracts along the seven dimensions. Differences can be found in Dimensions 1, 2, 4, 5 and 7 by direct observation of the figure, while Dimensions 3 and 6 display almost the same factor scores, with NS and NNS scores nearly overlapping in the figure. We used *t*-tests to establish the statistical significance of differences between the two types of abstracts along each dimension. A probability value  $p < 0.05$  is regarded as statistically significant. The *t*-tests indicate that there are significant differences between NS and NNS abstracts in Dimensions 1, 2, 4, 5 and 7, all with  $p < 0.001$ , while Dimensions 3 ( $p = 0.709$ ) and 6 ( $p = 0.971$ ) show no significant differences.

#### 4.1 NS and NNS abstracts in contrast along Dimension 1

Overall NS abstracts show a significantly greater factor score than NNS abstracts in Dimension 1 (with a mean difference of 0.726653,  $t = 7.961$ ,  $p < 0.001$ ). A difference that is similarly significant is also observed in all disciplines barring astronomy where there is a significant difference (in astronomy, the dimension score of NNS is significantly higher than NS while the differences in the disciplines of geography, physics, mechanics and survey are not significant).

In comparison with native English authors, Chinese writers appear to be less focussed and less confident in using intensifying devices to emphasise the points that they make. This finding is in line with previous reports on the discrepancies between NS and NNS in using subcategories of

adverbs such as intensifiers (Liang, 2003; and Zhang, 2008a) and frequency adverbs (Wen and Ding, 2004). These studies relate the inadequate use of adverbs to the lack of effective input and Chinese writers' heavy reliance on the most frequently used adverbs and avoidance of others, as well as the absence of equivalent adverbs in Chinese. This suggests that there is a need to acquaint Chinese authors with the proper use of adverbs in EAP writing.

#### 4.2 NS and NNS abstracts in contrast along Dimension 2

NS English writers generally show a substantially higher factor score than their Chinese counterparts in Dimension 2 (mean difference = 1.257074,  $t = 12.119$ ,  $p < 0.001$ ). Eight out of the twelve disciplines covered in our corpus demonstrate a significant difference (no significant difference is found in earth science, geology, oceanography and physics), with seven of them (i.e., atmospheric science, astronomy, biology, chemistry, geography, mathematics and mechanics, barring survey) showing a significantly higher score for NS abstracts. It is not surprising that both sides have negative factor scores in this dimension because academic writing, as Biber (1988) suggests, is less focussed on interaction than speech. Despite its academic tone, co-occurring features like first-person pronouns, private verbs and the colligation pattern '*we + verb + that*', are a powerful combination used to make generalisations (Biber, 2000), or to interact with and gain acceptance from the readers (Hyland, 2001).

As our results indicate, native English writers demonstrate a more active involvement and commitment in their proposals, results and conclusions than NNS writers do. They also have a stronger preference for an interactive style in their writing. Our finding is in accord with previous observations of this difference (Ding, 2009; Liang, 2003; and Zhang, 2008b). This is attributable both to the influence of Chinese authors' native language writing conventions and to the deliberate omission of first-person pronouns by many Chinese scholars and publishers – especially in scientific writing, for the purpose of maintaining objectivity, which is widely accepted by Chinese academic communities.

#### 4.3 NS and NNS abstracts in contrast along Dimension 3

The overall contrast between NS and NNS abstracts in Dimension 3 is not significant (mean difference = -0.025369,  $t = -0.373$ ,  $p = 0.709$ ), but there are variations across different disciplines, possibly because of the different conventions in various disciplines. For example, in atmospheric science, biology, chemistry, geography, geology and survey, the factor score of this dimension is significantly greater in NNS abstracts whereas in earth science,

mathematics, mechanics and oceanography, it is significantly higher in the NS data.

#### 4.4 NS and NNS abstracts in contrast along Dimension 4

In Dimension 4, native English authors' abstracts as a whole display a significantly lower factor score than their Chinese counterparts (mean difference =  $-0.777681$ ,  $t = -9.164$ ,  $p < 0.001$ ). Chinese authors' preference for conceptual elaboration is reflected in various disciplines. The differences between NS and NNS abstracts in six out of the twelve disciplines covered in our corpus (namely, atmospheric science, biology, chemistry, mechanics, physics and survey) are not significant. However, amongst the other six disciplines, NS abstracts display a significantly lower factor score in this dimension in five of them (i.e., astronomy, earth science, geography, geology and mathematics), with oceanography alone showing a significantly higher score for NS abstracts.

Chinese writers' stronger preference for conceptual elaboration is also evidenced by their more frequent use of modified noun phrases: 'adjective + adjective + noun' and 'adjective + noun + noun', reflecting a tendency towards idea unit integration and expansion. Chinese writers' greater propensity to use pre-modification is probably a result of their first-language transfer, because in Chinese, unlike English, nouns can only take pre-modifiers.

However, highly packed conceptual elaboration may increase comprehension difficulties amongst readers, slowing down their speed to filter and index information. Recently, many Chinese journals have realised the problem and suggested their authors avoid using a long sequence of noun phrases with several adjective or nominal pre-modifiers (see, for example, the journal of *Acta Geodaetica et Cartographica Sinica*, Volume 2, 2010). It is essential for NNS abstracts to shift from informationally dense writing to a more readable style with simpler sentence patterns.

#### 4.5 NS and NNS abstracts in contrast along Dimension 5

The factor score of Dimension 5 is significantly lower in NS than NNS abstracts (mean difference =  $-1.274042$ ,  $t = -23.688$ ,  $p < 0.001$ ). This is a surprising contrast that becomes even more striking given that cross-linguistic differences between English and Chinese in the use of passives – passives are over ten times as frequent in English as in Chinese (Xiao *et al.*, 2006) – would lead to the natural expectation of a more frequent use of passives in NS than NNS abstracts. Nevertheless, our data unmistakably shows that, in spite of our expectations, NNS abstracts in our corpus make more frequent use of passives than NS abstracts, no matter

whether the two types of data are examined as a whole, as noted above, or variations in disciplines are taken into account. In terms of disciplinary variations, with the exceptions of atmospheric science, geography and survey, where no significant differences between NS and NNS data are observed, NNS Chinese writers make significantly more frequent use of such formal, abstract and impersonal style markers in all of the other disciplines covered in our corpus.

Chinese authors' preference for the passive voice in their academic writing has been observed in previous studies (e.g., He, 2003; and Teng and Tan, 2004). Despite the unexpected result from a cross-linguistic perspective, this interesting observation may be accounted for. First, as Xiao (2007) demonstrates, the overuse of passives is a type of 'error' with passive use in L1 Chinese learners' English inter-language, which is possibly a result of training transfer, (i.e., excessive passive training in classroom instructions in China). Secondly, despite the western tradition of using passive constructions in scientific writing, directness and conciseness have become the expected style that is typical of the academic circle in recent decades. Writers are now encouraged to turn the passive voice into a more direct style, and to use first-person pronouns and strong verbs where appropriate (see Bem, 2004; and Booth, 1993). With little attention given to this stylistic shift, many Chinese scientific journals still include in their guidance for abstract writing the avoidance of first-person pronouns and the active voice (see, for example, the journal of Spectroscopy and Spectral Analysis, Volume 31, 2012). Therefore, we think that both the Chinese tradition of supplying instructions to authors and the conflicting requirements of journals contribute to the observed difference in using the passive voice in NS and NNS abstracts.

#### **4.6 NS and NNS abstracts in contrast along Dimension 6**

In Dimension 6, the overall difference between NS and NNS abstracts is not significant (mean difference = 0.002002,  $t = 0.037$ ,  $p = 0.971$ ), nor are the differences observed in the disciplines of earth science, geography, geology, mechanics and physics, in spite of some significant differences in other disciplines: that is, NS data has a greater dimension score in atmospheric science, biology, oceanography and survey but a lower score in astronomy, chemistry and mathematics. Biology is a discipline that shows a significantly higher score for NS abstracts.

The results indicate that NNS abstracts demonstrate little difference concerning Chinese writers' ability to deal with textual cohesion in spite of disciplinary variations. The overall results are hardly surprising, given that the abstracts written by both NS and NNS authors are formally published materials, which, usually, have undergone rigorous copy-editing and proofreading by professionals (albeit that they are subject to the different

stylistic preferences of disciplines and journals). As a result, textual cohesion is achieved in both types of abstracts.

#### **4.7 NS and NNS abstracts in contrast along Dimension 7**

The overall difference between NS and NNS abstracts is statistically significant, with the abstracts written by Chinese authors showing a greater factor score in this dimension. The same is observed in all disciplines barring biology, which shows no significant difference. This suggests that Chinese authors tend to write more informationally dense abstracts than their English counterparts.

This is surprising to most researchers who may presume that NS writers should have been more skilful in dealing with the complex converting process of nominalisation than their Chinese counterparts due to cross-linguistic differences between English and Chinese. To our knowledge, there have been few studies that have compared NS and NNS writers' use of nominalisation in their academic writing. Nominalisation has been acknowledged and advocated as the most remarkable characteristic of English for science and technology in China (Liu, 1998; and Qin, 2001). However, a sentence using too many instances of nominalisation would make it difficult for readers to comprehend, which is contrary to the style that is currently preferred: being direct, concise and readable. As a result, authors are advised, by some scholars (e.g., Williams, 2000; and Zeiger, 2000) and international academic communities (see, for example, CBE Style Manual Committee, 1994), to use verbs to avoid intricate nominalisation wherever possible.

### **5. Conclusion**

This paper has sought to explore the textual variations between native and non-native English abstracts from a contrastive perspective using the MD approach. Further comparisons have also been made at the level of discipline.

In our MD study, factor analysis reduced our initial selection of 163 linguistic features to 47 with significant loadings, which are loaded on seven dimensions in our new model of RA abstracts, namely, (1) focussing and intensification; (2) active involvement and interaction; (3) explicit conceptualisation of methodology; (4) conceptual elaboration; (5) formal, abstract and impersonal style; (6) textual cohesion; and (7) informational density. Our contrastive analysis of abstracts written by native and non-native authors shows that the two differ significantly along five dimensions other than Dimensions 3 and 6. Both similarities and differences between NS and NNS abstracts have been found in various disciplines.

With respect to the communicative function along each dimension, native English writers demonstrate a more active involvement and commitment in presenting their ideas than NNS writers. They also use intensifying devices more frequently to emphasise the points they make—thus showing a more confident writing style. In contrast, Chinese writers are found to demonstrate a stronger preference for conceptual elaboration through heavier reliance on passives and compact abstract noun phrases no matter whether the NS and NNS abstracts are considered as a whole or whether disciplinary variations are taken into account.

A variety of reasons are put forward in this article to account for the divergence of NNS abstracts from NS abstracts including—for example, the transfer of native language writing, improper classroom instructions, widely accepted conventions by academic communities, and conflicting requirements of individual journals. We would like to suggest, on the basis of our research findings, that there is a need for Chinese writers to be fully aware of the currently preferred writing style of international journals, to follow the writing norms of specific disciplines, and to make their writing more direct and readable. On the other hand, publishers in China should be cautious when providing writing guidance to authors because rigorously defined requirements or an over emphasis on one aspect of language use may lead to fossilisation of language, and, thus, runs the risk of taking us from one extreme to the other (see Qian and Wang, 2010).

The great number of abstracts sampled, the wide range of disciplines covered, the large number of linguistic features investigated, the rigorous statistic measures taken, and the depth of analysis performed in this study have enabled us to provide a more realistic and accurate account of Chinese writers' English scientific abstracts in relation to native English norms. Methodologically, this study has innovatively expanded Biber (1988) and Xiao's (2009) MD models by integrating colligation into the multi-dimensional analytical framework, which, in combination with grammatical and semantic features, has helped the MD approach to offer even stronger interpretations of the discourse functions of dimensions based on closer integration of form and meaning in language variation research. In addition, the research presented in this article has also helped to extend the MD approach from vastly different genres to those of a similar kind.

We hope that these original aspects of this research will become welcome new developments of the MD approach on the twenty-fifth anniversary of Biber's (1988) pioneering multi-feature/multi-dimensional analytical framework of register variation.

### **Acknowledgments**

This research is funded by China's Ministry of Education under its Social Science Research Scheme (Grant Reference: 10YJC740007).

**Appendix A: Colligation patterns on the feature list**

Categories	Colligation patterns
Noun (N)	art+N+of, <i>a/an</i> +adj+of, N+and/or+N, N+N+N, adj+N+N, adj+adj+N
Preposition (prep)	prep+art+N+of, <i>of</i> +art+adj+N, prep+art+adj+N
Verb: past participle (PP)	<i>be</i> +PP+prep, <i>be</i> +PP+present participle, PP+inf, third-person pron+ <i>be</i> +PP, adv+PP+prep, adv+PP+N, N+PP+prep
Verb: as predicate (V)	N+N+V, adj+N+V, N+V+ <i>that</i> , V+art/adj/adv/N+N, V+prep, <i>we</i> +V+ <i>that</i> , <i>we</i> +V+art/adj/adv/N
Verb: <i>be</i> ( <i>be</i> )	third-person pron+ <i>be</i> +adj, N+ <i>be</i> +adj, <i>be</i> +adj+prep, <i>be</i> +inf, <i>be</i> +adv+adj
Verb: <i>-ing</i> form ( <i>V-ing</i> )	N+V- <i>ing</i> +prep, N+prep+V- <i>ing</i> , PP+prep+V- <i>ing</i> , N+V- <i>ing</i> +N
Verb: infinitive (inf)	N+inf, third-person pronoun+ <i>be</i> +adj+inf
Adjective (adj)	adv+adj+N, V+adv+adj, <i>be</i> +adj+prep
Adverb (adv)	adv+V, V+adv, V+prepositional adv

**References**

- Archer, D., A. Wilson and P. Rayson. 2002. Introduction to the USAS Category System. Unpublished technical report. Lancaster: UCREL, Lancaster University. Accessed 6 February 2012, at: <http://ucrel.lancs.ac.uk/usas/usas%20guide.pdf>
- Bem, D. 2004. 'Writing the empirical journal article' in J. Darley, M. Zanna and H. Roediger III (eds) *The Complete Academic: A Career Guide*, pp. 185–219. (Second edition.) Washington, District of Columbia: American Psychological Association.
- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. 1989. 'A typology of English texts', *Linguistics* 27 (1), pp. 3–43.
- Biber, D. 1991. 'Oral and literate characteristics of selected primary school reading materials', *Text* 11 (1), pp. 73–96.
- Biber, D. 1995. *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge: Cambridge University Press.
- Biber, D. 2000. 'Corpus-based analysis of grammar: variability in the form and use of English complement clauses' in M. Bilger (ed.) *Corpus:*

- Méthodologie et Applications Linguistiques, pp. 224–36. Paris and Perpignan: Honoré Champion and Presses Universitaires de Perpignan.
- Biber, D. 2003. 'Variation among university spoken and written registers: a new multi-dimensional analysis' in C. Meyer and P. Leistyna (eds) *Corpus Analysis: Language Structure and Language Use*, pp. 47–70. Amsterdam: Rodopi.
- Biber, D. 2004. 'Conversation text types: a multi-dimensional analysis' in G. Purnelle, C. Fairon and A. Dister (eds) *Le Poids des Mots: Proceedings of the 7th International Conference on the Statistical Analysis of Textual Data*, pp. 15–34. Louvain: Presses Universitaires de Louvain.
- Biber, D. 2006. 'Stance in spoken and written university registers', *Journal of English for Academic Purposes* 5 (2), pp. 97–116.
- Biber, D. 2007. 'On the complexity of discourse complexity: a multi-dimensional analysis' in S. Conrad and D. Biber (eds) *Variation in English: Multi-dimensional Studies*, pp. 215–40. London: Longman.
- Biber, D. and E. Finegan. 1989. 'Drift and the evolution of English style: a history of three genres', *Language* 65 (3), pp. 487–517.
- Biber, D. and E. Finegan. 1992. 'The linguistic evolution of five written and speech-based English genres from the seventeenth to the twentieth centuries' in M. Rissanen, O. Ihalainen, T. Nevalainen and I. Taavitsainen (eds) *History of Englishes: New Methods and Interpretations in Historical Linguistics*, pp. 688–704. Berlin: Mouton.
- Biber, D. and E. Finegan. 1994a. 'Multi-dimensional analyses of authors' styles: some case studies from the eighteenth century' in D. Ross and D. Brink (eds) *Research in Humanities Computing* 3, pp. 3–17. Oxford: Oxford University Press.
- Biber, D. and E. Finegan. 1994b. 'Intra-textual variation within medical research articles' in N. Oostdijk and P. de Haan (eds) *Corpus-based Research into Language*, pp. 201–22. Amsterdam: Rodopi.
- Biber, D., U. Connor and T. Upton. 2007. *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins.
- Biber, D., S. Conrad and V. Cortes. 2004. 'If you look at...: lexical bundles in university teaching and textbooks', *Applied Linguistics* 25 (3), pp. 371–405.
- Biber, D., S. Conrad and R. Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D., E. Csomay, J. Jones and C. Keck. 2004. 'A corpus linguistic investigation of vocabulary-based discourse units in university registers' in U. Connor and T. Upton (eds) *Applied Corpus Linguistics: A Multi-dimensional Perspective*, pp. 53–72. Amsterdam: Rodopi.

- Biber, D., M. Davies, J. Jones and N. Tracy-Ventura. 2006. 'Spoken and written register variation in Spanish: a multi-dimensional analysis', *Corpora* 1 (1), pp. 1–37.
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.
- Booth, V. 1993. *Communicating in Science: Writing a Scientific Paper and Speaking at Scientific Meetings*. (Second edition.) Cambridge: Cambridge University Press.
- CBE Style Manual Committee. 1994. *Scientific Style and Format: CBE Manual for Authors, Editors and Publishers*. (Sixth edition.) New York: Cambridge University Press.
- Conrad, S. 1994. 'Variation in academic writing: textbooks and research articles across disciplines', paper presented at the annual conference of the American Association of Applied Linguistics. Baltimore, MD, March 1994.
- Conrad, S. 1996. *Academic Discourse in Two Disciplines: Professional Writing and Student Development in Biology and History*. Unpublished PhD thesis. Northern Arizona University.
- Conrad, S. 2001. 'Variation among disciplinary texts: a comparison of textbooks and journal articles in biology and history' in S. Conrad and D. Biber (eds) *Variation in English: Multi-dimensional Studies*, pp. 94–107. Harlow: Pearson.
- Conrad, S. and D. Biber. 2001. *Variation in English: Multi-dimensional Studies*. Harlow: Pearson.
- Crossley, S. and M. Louwse. 2007. 'Multi-dimensional register classification using bigrams', *International Journal of Corpus Linguistics* 12 (4), pp. 453–78.
- Csomay, E. 2005. 'Linguistic variation within university classroom talk: a corpus-based perspective', *Linguistics and Education* 15 (3), pp. 243–74.
- Csomay, E. 2007. 'Vocabulary-based discourse units in university class sessions' in D. Biber, U. Connor and T. Upton (eds) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*, pp. 213–38. Amsterdam: John Benjamins.
- Dai, L. and R. Cai. 2004. *Guowai Keji Hexin Qikan Shouce (A Handbook of Core Foreign Journals of Science and Technology)*. Beijing: Peking University Press.
- Ding, M. 2009. 'Yixue lunwen Yingyu zhaiyao zhong diyi rencheng daici shiyong zhi fenxi (Analysis of the use of first personal pronouns in English abstracts of medical research articles)', *Acta Editologica* 6, pp. 500–2.

- Gillet, A. 2013. *Using English for Academic Purposes: A Guide for Students in Higher Education*. Accessed 6 February 2013, at: <http://www.uefap.com/>
- Goldberg, A. 2006. *Constructions at Work*. Oxford: Oxford University Press.
- Gries, StTh, J. Newman and C. Shaoul. 2011. 'N-grams and the clustering of registers', *ELR Journal*, 5 (1). Accessed 6 February 2013, at: <http://ejournals.org.uk/ELR/article/2011/1>
- Grieve, J. 2011. 'A regional dialect survey of contraction rate in written Standard American English', *International Journal of Corpus Linguistics* 16 (4), pp. 514–46.
- Harwood, N. 2005. "'We do not seem to have a theory... the theory I present here attempts to fill this gap": inclusive and exclusive pronouns in academic writing', *Applied Linguistics* 26 (3), pp. 343–75.
- Hasselgård, H. 2010. *Adjunct Adverbials in English*. Cambridge: Cambridge University Press.
- He, R. 2003. 'Zhong wai keji qikan Yingwen zhaiyao bijiao: yutai fenbu he "tou zhong jiao qing" ju de shiyong pinlü (A comparison between Chinese and foreign scientific and technical periodicals: Voice distribution and frequency of use of "front-weight" sentences)', *Shanghai Journal of Translators for Science and Technology* 1 (1), pp. 16–18.
- Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hunston, S. and G. Francis. 2000. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Hyland, K. 1996. 'Talking to the academy: forms of hedging in science research articles', *Written Communication* 13 (2), pp. 251–81.
- Hyland, K. 1998. *Hedging in Scientific RAs*. Amsterdam: John Benjamins.
- 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), pp. 207–26.
- Hyland, K. 2004. *Disciplinary Discourses: Social Interaction in Academic Genres*. Ann Arbor: The University of Michigan Press.
- Hyland, K. 2008. "'As can be seen": lexical bundles and disciplinary variation', *English for Specific Purposes* 27 (1), pp. 4–21.
- Jones, J. 2007. 'Vocabulary-based discourse units in biology research articles' in D. Biber, U. Connor and T. Upton (eds) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*, pp. 175–212. Amsterdam: John Benjamins.

- Li, P. and Y. Cao. 2013. *Keji Qikan Yingwen Zhaiyao Xiezuotonglun* (Writing Essentials of English Abstracts of Scientific and Technical Journals). Beijing: Higher Education Press.
- Liang, M. 2003. 'A study of intensifiers in Chinese learners' speech production', proceedings of the 2003 International Conference on Corpus Linguistics. 25–27 October 2003. Shanghai.
- Liang, M. and W. Xiong. 2008. 'Wenben fenxi gongju PatCount zai waiyu jiaoxue yu yanjiu zhong de yingyong (Applications of PatCount in foreign language teaching and research)', *Computer-assisted Foreign Language Education* 5, pp. 1–11.
- Liu, M. 1998. *Wenti yu Fanyi (Style and Translation)*. Beijing: China Translation and Publishing Corporation.
- McEnery, T., R. Xiao and Y. Tono. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. London and New York: Routledge.
- Malcolm, L. 1987. 'What rules govern tense usage in scientific articles?', *English for Specific Purposes* 6 (1), pp. 31–43.
- Martínez, I. 2001. 'Impersonality in the research article as revealed by analysis of the transitivity structure', *English for Specific Purposes* 20 (3), pp. 227–47.
- Pan, Y. and Z. Ma. 2010. *Zhongguo Keji Qikan Yinzheng Baogao* (Chinese Sci-Tech Journal Citation Reports). Beijing: Scientific and Technical Documentation Press.
- Pho, P. 2008. 'Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance', *Discourse Studies* 10 (2), pp. 231–50.
- Qian, E. and X. Wang. 2010. 'Keji lunwen zhong Yingwen zhaiyao de rencheng yu yutai wenti (Some aspects on personal pronouns and voices in abstracts of scientific writing)', *Acta Editologica* 4, pp. 319–21.
- Qin, X. 2001. *Yingyu Yuti Wenti Yaolue* (Registers in English and Essentials of Style). Shanghai: Shanghai Foreign Language Education Press.
- Rayson, P. 2003. *Matrix: A Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison*. Unpublished PhD thesis. Lancaster University. Available online, at: <http://ucrel.lancs.ac.uk/people/paul/publications/phd2003.pdf>
- Rayson, P. 2008. 'From key words to key semantic domains', *International Journal of Corpus Linguistics* 13 (4), pp. 519–49.
- Ren, Q. 2008. *Zhongwen Hexin Qikan Mulu Zonglan* (A Comprehensive List of Chinese Core Journals). Beijing: Peking University Press.
- Reppen, R. 1994. *Variation in Elementary Student Writing*. Unpublished PhD thesis. Northern Arizona University.

- Römer, U. 2009. 'The inseparability of lexis and grammar: corpus linguistic perspectives', *Annual Review of Cognitive Linguistics* 7 (1), pp. 140–62.
- Scott, M. 2011. *WordSmith Tools 6.0*. Oxford: Oxford University Press.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Swales, J. 2004. *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press.
- Swales, J. and C. Feak. 2004. *Academic Writing for Graduate Students*. (Second edition.) Ann Arbor, Michigan: University of Michigan Press.
- Swales, J. and M. Luebs. 2002. 'Genre analysis and the advanced second language writer' in E. Barton and G. Stygall (eds) *Discourse Studies in Composition*, pp. 135–54. Cresskill, New Jersey: Hampton Press.
- Teng, Z. and W. Tan. 2004. 'Yingwen zhaiyao de shitai, yutai wenti (Tenses and voices in English abstract writing)', *Chinese Science and Technology Translators Journal* 1 (1), pp. 5–7.
- Thompson, G. and Y. Ye. 1991. 'Evaluation in the reporting verbs used in academic papers', *Applied Linguistics* 12 (4), pp. 365–82.
- Ventola, E. 1994. 'Abstracts as an object of linguistic study' in S. Čmejrková, F. Daneš and E. Havlová (eds) *Writing vs. Speaking: Language, Text, Discourse, Communication*, pp. 333–52. Tübingen: Gunter Narr.
- Watson, G. 1994. 'A multidimensional analysis of style in Mudrooroo Nyoongah's prose works', *Text* 14 (2), pp. 239–85.
- Wen, Q. and Y. Ding. 2004. 'Zhongguo Yingyu zhuan ye xuesheng shiyong pinlü fuci de tedian (A corpus-based analysis of frequency adverbs used by Chinese English majors)', *Modern Foreign Language* 27 (2), pp. 150–6.
- Williams, J. 2000. *Style: Ten Lessons in Clarity and Grace*. London: Longman.
- Xiao, R. 2007. 'What can SLA learn from contrastive corpus linguistics? The case of passive constructions in Chinese learner English', *Indonesian Journal of English Language Teaching* 3 (2), pp. 1–19.
- Xiao, R. 2009. 'Multidimensional analysis and the study of world Englishes', *World English* 28 (4), pp. 421–50.
- Xiao, R. 2011. 'Word clusters and reformulation markers in Chinese and English: implications for translation universal hypotheses', *Languages in Contrast* 11 (2), pp. 145–71.

- Xiao, R., T. McEnery and Y. Qian. 2006. 'Passive constructions in English and Chinese: a corpus-based contrastive study', *Languages in Contrast* 6 (1), pp. 109–49.
- Zeiger, M. 2000. *Essentials of Writing Biomedical Research Papers*. (Second edition.) New York: McGraw-Hill Companies.
- Zhang, X. 2008a. 'Zhongguo fei Yingyu zhuan ye EFL xuexizhe qianghuayu shiyong yanjiu (A corpus-based investigation of the use of intensifiers among Chinese college EFL learners)', *Foreign Language Learning Theory and Practice* 2008 (4), pp. 26–33.
- Zhang, M. 2008b. 'Zhong wai zhaiyao diyi rencheng daici yongfa de bijiao yanjiu (A comparative study of first person pronouns in abstracts in China and English speaking countries)', *Shanghai Journal of Translators* 2008 (2), pp. 31–6.