

Stance phraseology in academic discourse: cross-disciplinary variation in authors' presence

Jihua Dong & Louisa Buckingham

Shandong University (China) & The University of Auckland (New Zealand)
dongjihua@sdu.edu.cn & l.buckingham@auckland.ac.nz

Abstract

This study aims to investigate the stance phrases used by agricultural science and economics scholars to establish their identity and authority in academic discourse. Adopting a corpus-driven approach, this study first retrieved 2- to 5-grams from two self-built corpora consisting of agriculture and economics research articles and then classified them functionally by following the three metafunctions in Systemic Functional Linguistics (SFL). Statistical comparisons show that agriculture researchers use cognitive, attitude, and hedging phrases more frequently, while the economics corpus contains significantly more evidentiality phrases. This paper contributes to the categorization of stance phrases by synthesizing and refining existing stance classifications and incorporating the reason-oriented dimension. The findings can enrich our understanding of disciplinary variation in academic discourse, and can inform academic writing pedagogy with respect to curriculum design and materials development.

Keywords: classification framework, corpus-driven approach, disciplinary variation, cognitive stance, Natural Language ToolKit.

Resumen

La expresión de la posición mediante colocaciones en el discurso académico: variación disciplinar en la presencia de los autores

El objetivo de este artículo es analizar las colocaciones que emplean los investigadores de Ingeniería Agronómica y Economía para marcar su posición (*stance*) como autores, y, así, establecer su identidad y autoridad en el discurso académico. Desde un enfoque dirigido por corpus o *corpus-driven*, se han recuperado en primer lugar los n-gramas de orden 2, 3, 4 y 5 de dos corpus

propios compuestos por artículos científicos de Ingeniería Agronómica y Economía y, posteriormente, se han clasificado con criterios funcionales de acuerdo con las tres metafunciones de la Lingüística Sistémico-Funcional. A través de comparaciones estadísticas, se ha evidenciado que los investigadores de Ingeniería Agronómica utilizan más colocaciones de naturaleza cognitiva, actitudinal y mitigadora con mayor frecuencia, mientras que el corpus de artículos de Economía contiene un número significativamente mayor de colocaciones de carácter evidencial. Asimismo, este artículo presenta una aportación respecto a la categorización de las colocaciones que permiten marcar la posición (*stance*) del autor, ya que sintetiza y mejora las clasificaciones ya existentes e incorpora una dimensión de naturaleza racional dentro del concepto de *stance*. Los resultados obtenidos enriquecen nuestra comprensión de la variación disciplinar en el discurso académico y pueden ser de gran utilidad para la didáctica de la escritura académica en relación con el diseño del currículum y el desarrollo de materiales.

Palabras clave: marco de clasificación, corpus-driven, variación disciplinar, posición cognitiva, *Natural Language ToolKit*.

1. Introduction

Over the past few decades, research articles (RAs) have become one of the most prevalent genres for the scholarly dissemination of knowledge, and a medium through which academics can attain recognition, and build their reputation within the academic community. In teaching practice, RAs are often treated as exemplars for students to emulate (Feak & Swales, 2011; Hyland, 2008) and have been drawn upon extensively as reading materials in the ESP/EAP courses of various levels in higher education (Hood, 2010). Disciplinary variation has become a central focus of ESP studies to meet the local labour market demand for university graduates with specific skill sets (Tao & Gao, 2018), and the needs of an increasing number of academics and students who use English as an additional language for research, publishing and educational purposes.

The interest in disciplinary variation is also grounded upon the assumption that each community may select or prioritize its unique norms, values, and ways of communication (Van Dijk, 1997). As indicated by previous studies (Cortes, 2004; Hewings & Hewings, 2002; Hyland, 2008), writers in different disciplinary communities tend to employ different linguistic features to align themselves with community practice. A central focus in this line of research is stance features in RAs. Stance plays a pivotal role in

facilitating effective communication between writers and anticipated readers and in constructing social relations between writers and readers (Hyland & Jiang, 2016), as it not only enables academic writers to establish their identity and construct an authoritative persona during communication but it also allows authors to engage readers and present their propositions in a persuasive manner.

Substantial efforts have been made to tease out stance features in academic discourse, and a variety of terms have been used, such as evaluation (Hunston & Thompson, 2000; Hunston, 2011; Breeze, 2018), appraisal (Martin & White, 2005), metadiscourse (Hyland, 2005b), engagement (Hyland, 2005a), stance (Biber, 2006; Hyland, 2005a; Di Carlo, 2015), voice (Bowden, 1999; Fogal, 2019; Matsuda, 2001; Matsuda & Tardy, 2007), modality (Coates, 1987; Nuyts, 2001), and evidentiality (Chafe & Nichols, 1986; Wierzbicka, 1994).

Hunston (1994; 2011) and Martin and White (2005) use the terms *evaluation* and *appraisal framework* respectively to refer to an umbrella concept concerning evaluative stance. Evaluative language is seen as expressing “an attitude towards a person, situation or other entity” with the characteristics of subjectivity which are “located within a societal value-system” (Hunston, 1994: 210). By contrast, Martin and White’s framework (2005) encompasses a broader scope, including *engagement* (attributing the source for attitudes and opinions in discourse), *attitude* (centers on the aspects of affect, judgment, and appreciation), and *graduation* (concerned with the gradability of feelings). In addition, Biber et al. (1999), Hyland and Tse (2005), Hyland (2005b), and Simpson-Vlach and Ellis (2010) use the term *evaluation* to refer to a sub-set of linguistic devices within the category of attitude. Also, previous studies have examined particular linguistic elements that can be equated with one of the sub-categories of broader classification schemes (Biber et al., 1999), such as attitude markers in the metadiscourse framework (Hyland & Tse, 2005; Hyland, 2005b) and evaluation in the AFL (Simpson-Vlach & Ellis, 2010). Biber et al. (1999) focus on the emotional attitude (e.g., *surprisingly*, *unfortunately*); Hyland (2004; 2005a) considers this concept in his categories of *boosters* (e.g., *clearly*, *obviously*, *highly*) and *attitude markers* (e.g., *important*, *hopefully*, *remarkable*). Simpson-Vlach and Ellis (2010) concentrate on the value-laden and attitudinal aspect within the category of evaluation, including linguistic markers, e.g., *important role in*, *it is important (to)*, *it is necessary (to)*, *(it) is clear (that)*, and *it is difficult*.

Some previous explorations of stance have considered it to represent a type of metadiscourse marker (i.e. Hyland, 2005a, 2005b; Hong & Cao, 2014). For instance, according to Hyland's (2005a) model, stance, which was distinguished from engagement, was generally regarded as being composed of the subcategory of *hedges*, *boosters*, *attitude markers* and *self-mentions*, and viewed as a sub-category of metadiscourse markers. The term *engagement*, in contrast, denotes linguistic devices employed by writers to relate to their readers (Hyland, 2001, 2005). Both stance and engagement in this model were considered to be types of interactional metadiscourse markers (used to involve the writer and the reader in the text) in the metadiscourse framework proposed by Hyland (2005b). This framework presents a more differentiated understanding of stance compared with his earlier work that considered stance more generally as "the way that writers project themselves into their texts to communicate their integrity, credibility, involvement, and a relationship to their subject matter and their readers" (Hyland, 1999: 101).

Modality in this study is used to refer to a speaker's commitment and detachment to propositional information (Stubbs, 1996). The modality aspect of authors' projection resembles the notion of hedges (1998a; 2005a), as they both convey authors' presence in terms of certainty, uncertainty, and imprecision. In contrast, evidentiality is concerned with authors' endeavors to signal the source of particular propositions (Chafe & Nichols, 1986) and evidentiality-related linguistic devices to represent a linguistic category "whose primary meaning is the source of information" (Aikhenvald, 2004: 3). As suggested previously, these two concepts are two semantically distinct categories (Greco, 2018).

One strand of stance studies has concentrated on stance phrases, which refers to recurring contiguous word sequences that convey the meaning and function of stance. Previous studies have principally focused on identifying stance phrases and exploring their functions (Biber, Conrad & Cortes, 2004; Hunston, 2011; Hyland, 2008; Römer, 2008; Simpson-Vlach & Ellis, 2010). For instance, Biber et al. (2004) examined four-word sequences in classroom teaching and textbooks and developed a functional taxonomy, including stance expressions, discourse organizers, and referential expressions. They included epistemic (certainty, uncertainty, possibility, and imprecision) and attitudinal (desire, obligation/directive, intention/predication, and ability) within the category of stance, and analyzed them from the perspective of personal and impersonal functions. Likewise, Hyland (2008) investigated four-word bundles in research-focused genres (RAS, doctoral dissertations,

and master's theses) and differentiated stance from engagement using the metadiscourse framework. In this taxonomy, he restricted stance to linguistic devices which convey a writer's attitudes and evaluations (e.g., *are likely to be*, *may be due to*, and *it is possible that*). In a recent attempt to identify the pedagogical utility of formulaic language, Simpson-Vlach and Ellis (2010) compiled the Academic Formulas List (AFL) and treated stance phrases as one of the main formulaic categories. The list was derived from a large-scale corpus-driven study, and differentiated between phrases in spoken and written academic texts. The focus on the phrase as a pedagogically useful unit of analysis reflects the idea that a text is comprised of a "repertoire of multi-word patterns" (Sinclair, 1991: 108). As the meanings of most words are usually influenced by the collocating words, focusing on phrases also contributes to removing some of the ambiguities that are likely to arise in word-based analyses.

One notable trend in stance phrase research has emerged in the examination of their interdisciplinary and cross-disciplinary features (Cortes, 2004; Hyland, 2008; Hyland & Tse, 2005; Koutsantoni, 2004; Stotesbury, 2003). For instance, Cortes (2004) investigated formulaic expressions in published and students' writing in the disciplines of history and biology, and found that biology researchers used more *possibility* stance markers. Likewise, Hyland (2008) examined formulaic expressions in the academic discourse of four disciplines and found that researchers in the hard sciences often minimized their presence, while their peers in the soft sciences displayed a more explicit presence and employed more interpretive expressions. Taken together, these cross-disciplinary studies identified notable variation in the way disciplinary writers use formulaic expressions to convey their stance and communicate with the anticipated readers.

However, our understanding of cross-disciplinary stance-taking features remains incomplete, and Bondi (2007) identified the need for a finer grained examination of disciplinary features to meet the needs of enhancing academic literacy of students from diverse disciplines. This study thus proposes to examine writers' rhetorical use of stance devices in the written discourse of agriculture and economics. Different epistemological assumptions underlie each of these two disciplines and the discursive practices that typify each have an impact on authors' projection in the text and the formulation of their propositions (Bondi, 1999; Del Saz Rubio, 2011; Jiang & Hyland, 2015; Nesi & Gardner, 2012). Research in agriculture is analytical in nature and empirically based on observable experience (Del

Saz Rubio, 2011; Martinez et al., 2009). Research in economics, on the other hand, tends to subscribe to a quantitative paradigm (Becher & Trower, 2001; Bargiela-Chiappini & Nickerson, 2002) and often involves the construction and testing of models, with the purpose of formulating economic forecasts to provide an evidence-informed basis for governmental bodies and business institutions' decision making (Bazerman, 2010: 281). Durrant (2017: 11) identifies agriculture and economics as "central disciplines within the commerce cluster". Within this broad category, he observes that economics displays greater linkages with science and technology, while agriculture is associated with the life sciences.

Previous work has examined the use of stance in each of these disciplines separately (see for instance, Del Saz Rubio, 2011; Martinez et al., 2009; Lancaster, 2014; Mauranen, 1993). A comparative approach can reveal the discipline-specific nature of research texts in these two fields and identify differences in how disciplinary writers rhetorically position themselves. Such findings can have pedagogical relevance, as stance-taking has been deemed "an elusive concept" (Mauranen & Bondi, 2003: 269). Many students and novice writers have been reported to encounter great difficulties in constructing stance in the introduction of papers and dissertations (Feak & Swales, 2011), establishing critical stance in writing literature reviews (Bruce, 2014), synthesizing and critiquing in literature reviews of doctoral theses (Boote & Beile, 2005), discussing and commenting on research results (Lim, 2010), and using *certainty* and *affect* stance appropriately in academic discourse (Hyland & Milton, 1997).

Overall, despite the aforementioned important roles of the two academic disciplines and the difficulties students often encounter with the use of stance, few studies have focused on recurring stance phrases in these two academic disciplines. Therefore, this study investigates the stance-taking practice of disciplinary writers in published academic texts and will address the following research questions: (1) What stance phrases are commonly used in the RAS of agriculture and economics? (2) How can reason-oriented stance expressions be integrated into the existing stance classification frameworks? (3) What disciplinary variation can be found in the use of the stance phrases in the two disciplines?

A secondary purpose of this study is to use the stance phrases retrieved through a corpus-driven approach to sort out the categorization of stance phrases by using the metafunction of SFL (Systemic Functional Linguistics).

In the following section, we elaborate on the justification for sorting out the stance categorization and for using metafunction in SFL.

2. Theoretical framework

Previous research has yielded several stance classification frameworks. For instance, Hunston (2011) identifies the functions of evaluation as status, value, and relevance, which correspond to the three moves of an evaluative act, namely identifying and classifying an object to be evaluated, ascribing a value to that object, and identifying the significance of the information. Martin and White's (2005) appraisal framework consists of engagement, attitude, and graduation, which are concerned with how authors construe their authorial persona during the construction of the text. Based on the data extracted from conversation, academic, and news registers, Biber et al. (1999) divide stance into three sub-types, namely epistemic, attitudinal, and style. In addition, Simpson-Vlach and Ellis (2010) classify stance into hedges, epistemic, obligation and directive, ability and possibility, evaluation, and intention/volition and prediction on the basis of high-frequency phrases extracted from the Michigan Corpus of Academic English (MICASE). Hyland (2005a) views stance from an interactional perspective and categorizes stance features into hedges, boosters, attitude markers, and self-mention. These classification schemes have enriched our understanding of the characteristics and functions of stance phrases and laid the foundation for empirical and theoretical research centered on stance theory.

Nevertheless, a close analysis indicates that most of the above stance classification schemes mainly focus on the aspects of attitude (related to feelings or values), hedges (conveying certainty or estimation), and evidentiality (indicating the source of a proposition). It is surprising that a further important dimension of stance, authors' reason-oriented presence, which contributes to the authors' construction of their persuasive persona, has not received due attention in the previous classification schemes.

The reason-oriented dimension of stance has attracted growing attention from researchers recently. For instance, Swales (2002) investigated the common cognitive verbs co-occurring with the subject *I*, such as *I think*, *I guess*, *I hope*, *I believe*, and *I suppose* in MICASE, and found that cognitive verbs constituted an important linguistic feature in spoken genre. In the AFL compiled by Simpson-Vlach and Ellis (2010), emphasis was placed on the

knowledge-oriented aspect of stance by acknowledging that “epistemic stance phrases have to do with knowledge claims or demonstrations, expressions of certainty or uncertainty, beliefs, thoughts, or reports of claims by others” (506) and included cognitive phrases, such as *assumed to be*, *considered as*, *determined by*, *argued that*, and *have shown that*, in this category. Aull and Lancaster (2014) viewed logical expressions (*of course*, *however*) as one essential part of stance in their analysis of the writing of first-year university students and that of upper-level undergraduate students and published academics. In addition, Lancaster (2016) included relational markers (*however*, *but*, *nevertheless*) in his analysis of undergraduates’ stance-taking practice, under the title of disclaim markers, and found that this aspect of self-projection constitutes an important element of academic writing. Taken together, the cognitive dimension represents how authors project their mental processing of knowledge, and works as an essential linguistic device for writers to “negotiat[e] meanings with the reader” (Lancaster, 2016: 19). Considering the importance of this dimension in constructing authors’ persuasive persona and the endeavors that have been made in the above-mentioned literature, it is surprising that this dimension has not been incorporated into the common analytical framework of stance studies.

Therefore, this study attempts to incorporate the reason-oriented dimension of authors’ presence into the stance classification framework in accordance with the theoretical underpinning of this study, namely the three SFL metafunctions: ideational, interpersonal, and textual (Halliday, 1994). According to SFL, language is a social semiotic system organized in relation to the aforementioned three metafunctions. The ideational function is concerned with how language is used in relation to the experiential world. It is composed of experiential and logical functions: the first covers the rhetorical devices that construct meanings around the world; and the second is relevant to the linguistic forms that build “logical–semantic relationships between one clausal unit and another” (Halliday, 2003: 17). The interpersonal function refers to the social relationship construed by the language resource for the purpose of building the relationship in communication. The textual function relates to the syntactic systems used to construct coherent and well-organized discourse. Drawing on this framework, Hunston and Thompson (2000) undertook a detailed analysis of evaluation and identified that evaluation performs the following functions: (1) expressing opinions, reflecting the value system of authors and their community; (2) maintaining relationships, concerned with building and maintaining relations between

writer and reader; (3) organizing discourse, concerned with the use of linguistic devices to structure a text.

3. Corpus and methods

3.1. Corpora

The corpora were compiled from the electronic versions of empirical RAS in agriculture and economics, as empirical RAS play a dominant role in disseminating research findings in these academic disciplines. In order to ensure the papers selected were representative of empirical RAS in the two disciplines, we took into account the corpus design criteria suggested by Biber (1993), i.e., sample size, range of text types, and the range of linguistic distributions.

To ensure the range of text types, this study focused on the texts taken from the reading lists which were prepared by experienced scholars in the disciplines of agriculture (n=25) and economics (n=20) to the master's candidates under their supervision. The papers selected mainly represent empirical studies in the fields of horticulture, crop genetics and breeding studies, crop cultivation studies, plant resources studies, and seeds studies in the discipline of agriculture; and banking studies, corporate economics, investment management, accounting, and public economics within the discipline of economics. We achieved a satisfactory distribution range by choosing the phrases that occurred in at least 10% of the texts in each corpus, as discussed in Section 3.2.2.

The recommended articles were taken from a wide range of different journals in agriculture (n=133) and economics (n=128). All the journals are peer-reviewed and SCI or SSCI indexed. The agriculture research article corpus (ARAC) consisted of 372 RAS, with 1,669,396 tokens, and the economics research article corpus (ERAC) was composed of 283 RAS, with 1,352,973 tokens, following the removal of the non-text sections, namely tables, figures, references, and acknowledgements.

3.2. Research procedures

3.2.1. Extraction

In retrieving stance phrases, this study focused on recurring n-grams involving 2- to 5-word sequences, as the word sequences in this range are

more frequent (i.e., Biber et al., 1999; Ellis, Simpson-Vlach & Maynard, 2008; Simpson-Vlach & Ellis, 2010). Although 2-word sequences were excluded in some previous phrases studies (i.e., Simpson-Vlach & Ellis, 2010) to ensure the manageability of the data, it has been reported that bigrams (like *tend to* and *seem to*) play an important role in conveying authors' stance (Biber et al., 1999; Hyland, 2005a); hence bigrams were included in this study despite the data-processing demands involved.

The two corpora were first lemmatized. Although it is argued that lemmatization may hide the collocation differences among different forms of a lemma (Cortes, 2008; Hoey, 2005), lemmatization was carried out in the process of phrase extraction for the following two reasons. Firstly, the phrases retrieved based on the lemmatized texts allow us to recognize “the different inflectional forms of the same lemma as the same word type” (Lu, 2014: 91) and analyse them as a unified form when appropriate, thus helping identify some phrases which may otherwise be excluded due to their low frequencies in the concordance list, and so can allow for a more comprehensive inclusion of the phrases examined in this study. In addition, the lemmatization facilitates the quantitative analysis of the phrases examined as it yields one identical raw frequency and one mutual information (MI) score for a phrase lemma regardless of its different inflectional forms.

The lemmatization was applied to the two corpora at the word-level with the Natural Language ToolKit (Bird, 2006). In this study, the lemmatization was considered only in the process of retrieving phrases. During the analysis of the stance phrases, the original forms of the formulaic stance phrases were considered in order to examine the phrases in context and determine their functional coding.

The lemmatized corpora were each uploaded to Collocate (Barlow, 2004) to retrieve the phrases. When the phrases were extracted, the range of 2- to 5-grams was selected in the search field of the software. The software produced a long list of phrases, accompanied by their respective frequencies and MI scores. MI score is often used as a parameter for comparative information (Simpson-Vlach & Ellis, 2010), thus no threshold score was applied in this study.

3.2.2. Filtering

In identifying the phrases to be included in this study, both automatic and manual selections were conducted. The combination was chosen in order to

reveal more of the “inherent nature of evaluative items” (Stotesbury, 2003: 331). In the automatic selection, the threshold for the inclusion of phrases was set following the general practice of corpus-based studies (i.e., Biber et al., 1999; Simpson-Vlach & Ellis, 2010). That is, the frequency level used as a cut-off was 10 instances per million words, a relatively low end of the threshold value used by corpus linguists, with the purpose of retrieving relatively less restricted data. The distributional range of the phrases was 10%; that is, the phrases selected had to occur in at least 10% of the articles in each corpus.

Given that one of the main purposes of this study is to identify the phrases with pedagogical value in the two disciplines, the formula teaching worth (FTW) metric (Simpson-Vlach & Ellis, 2010) was also employed as one parameter in this study to provide evidence for the teaching value of the formulaic expressions. According to Simpson-Vlach and Ellis (2010), this parameter “enables a reliable and valid rank ordering of the formulas” (496), thus FTW was just used as an index for the value of instructional purposes, but could “not provide a threshold cutoff score” (496). So no cut-off score was applied to this parameter in this study in accordance with Simpson-Vlach and Ellis’s (2010) suggested practice.

3.2.3. Collapsing

Although filtering from the perspective of frequency, range, and FTW score yielded a relatively short list of pedagogically useful stance phrases, there still existed a number of overlapping phrases, which were further dealt with through the ‘collapsing’ stage. This entailed grouping together phrases which could be subsumed within the same phrase entry, such as *interesting to*, *[be] interesting*, and *it [be] interesting to*. The formulaic expressions within one entry were then merged into one phrase by following the principles of semantic completeness and structural independence. To be specific, semantically incomplete phrases were collapsed in favor of semantically complete strings. For example, *in terms* and *terms of* were merged into the word string, *in terms of*, because the latter is relatively complete in meaning and possesses higher pedagogical value. The structural dependent strings, such as *[be] possible*, *[be] possible to*, *possible to*, and *it [be] possible* were trimmed, because these strings do not possess as high structural independence and pedagogical importance as *it [be] possible to*. It is necessary to point out this step was only applied to overlapping phrases, while the original forms of non-overlapping phrases were retained as retrieved from the corpora.

3.3. Analysis

In this step, we first determined whether the semantic meaning of a phrase, obtained from the above-mentioned corpus-driven approach, is related to stance meaning. Then we categorized the phrases by judging the semantic meaning in accordance with the metafunction of SFL. In instances where the meanings and functions were not clear, three experts in linguistics, agriculture, and economics were consulted to determine the meanings and functional categories of the phrases respectively. All nine experts were experienced academics with doctoral degrees and a number of publications in their research domains. In rating the phrases, the informants were first provided with training on the stance classification procedure. They were then provided with fifty samples randomly selected from the concordance lines containing both the co-text and contexts, together with the descriptions and examples of stance categories. They were then asked to rate the phrases on a scale of 1 (no agreement) to 10 (complete agreement) from the following two perspectives: (1) whether they thought the phrases delivered authorial projection in the texts; (2) whether they thought the phrases could be subsumed into particular categories. As indicated by the statistic test of Kendall's coefficient of consistency¹, W value was 0.91 ($p < 0.05$) for the first rating, and 0.87 ($p < 0.05$) for the second rating. This indicates a high degree of consensus and reliability on the informants' judgment on the stance phrases' relatedness and functional categorization. (The classification scheme is discussed in detail in Section 4).

As regards the normalized frequency, two main approaches have been previously used to measure the normalized frequency of phrases, namely against the number of tokens in the respective corpus (Biber, Cornrad & Cortes, 2004; Hyland 2008; Simpson-Vlach & Ellis, 2010) and against the number of texts in a corpus (Cortes, 2004). Although there may be some controversy in the approaches used for the measure, this study followed the general practice and measured the normalized frequency against the number of tokens in the corpora.

In order to explore the specific use of stance phrases in the two corpora, the frequency of the categories of stance phrases was further analyzed using the log-likelihood (LL) statistic. This statistic test has been commonly employed to assess the statistical significance of differences in words or phrases between different corpora (Botley, 2006; Jiang, 2015; Oakes, 1998; Rayson & Garside, 2000; Simpson-Vlach & Ellis, 2010). Since the LL statistic does not

assume that data are normally distributed, it is considered to be more reliable in text analysis than statistics based on assumed normal distributions as “normal distribution overestimates the significance” (Dunning, 1993: 65).

4. Results and discussion

This section presents the stance categorization scheme (SCF) that was constructed after the analysis of the phrases retrieved in the above steps (Section 4.1), and the statistical comparison of the stance phrases used in the two corpora (Section 4.2).

4.1 The hybrid model of stance features

As mentioned in Section 2, previous stance frameworks, such as Biber et al. (1999), Hyland (2005a, 2005b), Hunston (2011), and Simpson-Vlach and Ellis (2010) unanimously converged on the following aspects, namely attitude (emotion or evaluation-oriented aspect, related to feelings or values), hedges (modality-oriented aspect, conveying certainty or estimation), and evidentiality (evidentially-oriented aspect, indicating the source for a proposition). Based on the analysis of the stance phrases retrieved from the corpus, the present study introduced reason-oriented stance, termed ‘cognitive stance’, into the stance classification scheme as an independent stance category, alongside the other three aspects of stance, namely attitude, modality, and evidentiality. The descriptions and examples of stance phrase categories are presented in Table 1.

Category	SFL metafunction	Descriptions	Examples
Cognition	Ideational function	Linguistic markers related to writers' reason-oriented evaluation of a proposition when expressing assertions and construing knowledge in discourse	<i>in contrast, due to, imply that, take into account, etc.</i>
Evidentiality	Ideational/ Textual function	Expressions used to point out the source for certain statements or link to a certain textual position in construing the text flow.	<i>other authors, it is well-known that, this paper, as shown in, etc.</i>
Attitude	Interpersonal function	Expressions conveying writers' affective or evaluative presence towards or assessment of a proposition.	<i>it is interesting to, it is essential that, play an important role, etc.</i>
Hedges	Interpersonal function	Rhetorical devices expressing writers' degree of caution, modesty, and humility when making statements.	<i>appear to, seem to, may be, to some degree, etc.</i>

Table 1. The functional classification of stance phrases.

Cognitive stance phrases are mainly used to “formalize the logic of discourse, and clarify the principles of reasoning” (Rescher, 2005: 1). Their use reflects the writer’s reason-oriented presence in expressing assertions and constructing knowledge, mainly in the form of logical relation (*in addition, because of, and explained by*) and mental processing act (*take into account, assume that, and consider that*) towards entity or proposition. As suggested previously, the cognitive phrases displayed a notable co-occurrence with evidentiality expressions (Dong & Buckingham, 2018a). This collocational relationship enables writers to “ground their analytical behaviour in the evidence or source introduced by reference markers” (Dong & Buckingham, 2018a: 124).

This dimension shares a strong relationship with the ideational function, particularly the logical function within SFL. According to the logical metafunction of the ideational function within SFL, cognitive phrases fall into the following types, namely judgment, analysis, inference, and metacognition stance. Judgment stance covers the knowledge-building expressions in terms of definition, connotation, denotation, interpretation, ability, function, characteristics, and classification, etc. Inference stance is concerned with authors’ reasoning and speculation when making assertions. This sub-category is composed of *show that, indicate that, and it is demonstrated that*, etc. The expressions often signpost a conclusive statement informed by evidence or data following logical induction or deduction. Metacognitive stance concentrates on authors’ mental processing in the form of understanding, interpreting, and reasonable analysis of the knowledge or

statements. This category includes expressions such as *it is noted that*, *it is considered that*, and *take into account*.

In contrast with this reason-oriented projection delivered by cognitive stance, attitude stance represents authors' evaluation-oriented involvement. This aspect is concerned with the recurring expressions that manifest authors' sentiments in presenting opinions or propositions (Biber et al., 1999; Hunston, 2011; Martin & White, 2005). Attitudinal stance phrases comprise evaluative markers (*it is important*, *play an essential role*), and formulations expressing a subjective impression (*it is interesting to*, *it is surprising that*). Corresponding to the interpersonal metafunction, attitude stance phrases appeal to their emotionally-oriented engagement in processing knowledge or assertions.

Hedges are devices used to display due caution, modesty, and prudence in their statements construction (Hyland, 1998). Similar to attitude stance, hedging markers (*seem to*, *it is likely*) also correspond to the interpersonal function as they assist writers to establish a close rapport with readers by softening the tone of a claim. Hedging stance phrases not only reduce the risk of being criticized for proposing new statements, but also allow writers to assume "dialogic tones" (Bakhtin, 1981: 294) and moderate the strength of their opinions. This type of rhetorical device is regarded as a "manipulative non-direct sentence strategy of saying less than one means" (Hübler, 1983: 23) and a process to reduce the strength of a statement (Zuck & Zuck, 1986).

Evidentiality stance refers to the discourse markers that locate the sources of statements and direct readers' attention to particular information in the textual description (Crismore, 1990). The sources of evidentiality are seen to be directed to self-mention (*this study*, *our paper*), extra-reference (*previous studies*, *previously reported*), shared knowledge (*it is generally accepted that*, *it is well known that*) and intra-textual (*as described above*, *shown in Table 1*). By attributing a statement to a certain source, evidentiality stance establishes an 'evidentiality' persona (Boye & Harder, 2009; Chafe & Nichols, 1986; Wierzbicka, 1994) and also contributes to constructing a coherent and cohesive text and a multi-voiced or heteroglossia feature of academic discourse (Martin & White, 2005).

Specifically, self-mention expressions are concerned with the rhetorical devices directing the primary source of particular statements to the study, paper or author, including *in the present study*, *in this paper*, and *to our knowledge*.

Such phrases can initiate a knowledge-knower structure, in which the writer is represented as a knower to provide further indication of the relative strength of social or cognitive relations in the discourse (Hood, 2010: 172). In contrast, extra-reference stance phrases, such as *according to* or *by other authors*, direct readers' attention to statements from an outer source. Such stance phrases often reflect the multi-voiced or heteroglossic nature of statements (Martin & White, 2005: 37) as it uses others' voices to express authors' opinions, and traces the statement source to previous studies or writers.

Intra-textual expressions relate to the rhetorical devices that can direct information into a specific position in the textual organization, including *as shown in table*, and *see figure*. Such expressions function to link an argument or proposition with evidence indicated in a figure or table, or a source mentioned in other sections. By doing so, these expressions create the link between one element and other parts of the text (Vande Kopple & Crismore, 1990). Shared knowledge concerns linguistic markers conveying commonsense knowledge assumed to be shared by readers within a certain discipline (Hyland, 2005a). The phrases include *it is well known that*, *it is clear that*, and *it is widely accepted that*.

4.2. The cross-disciplinary comparison of stance phrases

Table 2 presents the descriptive and log-likelihood statistics of the stance phrases used in the two disciplines. As can be seen, cognitive phrases are most frequently used, and account for more than half of the total stance phrases in the two corpora.

The cross-disciplinary comparison shows that the agriculture corpus contains significantly more cognitive phrases ($LL=144.01$, $p=0.00$). By employing reason-based cognitive stance, writers are able to shed light on their reasoning process with respect to the proposition delivered in the subsequent clause. Example (1) shows a case in which the inference verb phrase, *suggest that*, is used to make an explicit statement related to knowledge construction.

- (1) The results *suggest that* changes in SOC contents induced by the conversion from fields to grassland or vice versa are primarily caused by changes in the mineral associated SOC pool. (ARAC#058:2)

Categories	Raw frequency		Normalized frequency (per million)		Percentage		LL	Sig
	Agr.	Econ.	Agr.	Econ.	Agr.	Econ.		
Cognitive	19204	13636	11503.56	10078.55	56.11%	55.22%	144.01	0.00
Hedges	7916	5200	4741.83	3843.39	23.13%	21.06%	140.27	0.00
Evidentiality	3484	3496	2086.98	2583.94	10.18%	14.16%	-79.47	0.00
Attitude	3619	2362	2167.85	1745.79	10.57%	9.57%	67.91	0.00
Total stance	34223	24694	20500.23	18251.66	100%	100%	194.72	0.00

Note: Sig stands for p-value; The LL ratio is significant at the 0.05 level.

Table 2. Descriptive and log-likelihood statistics for stance phrases categories in agriculture and economics corpora.

Hedging phrases are the second most frequently employed type, accounting for about one-fifth of the total cases. This suggests that these disciplinary writers have an inclination to mitigate their presence and exercise prudence in making strong claims. The mitigation of a claim also has the effect of attenuating readers’ potential opposition to the proposition. This effort in expressing caution may also be perceived, by some experienced readers, as a form of manifesting authors’ contribution or the originality of particular propositions, because the hedged component signals the authors’ speculative position.

As shown in Table 2, the occurrences of hedging phrases are significantly higher in the agriculture corpus, showing that agriculture researchers are more likely to assume a cautious attitude when constructing their texts. This can be seen in (2), in which *it is likely that* expresses the writer’s circumspection with respect to the reported finding.

- (2) WLCC treatments had 3 years of winter legume cover crop input and *it is likely that* there was a significant buildup of both labile C and N in these systems compared to the NCC systems. (ARAC#009:2)

The third most frequent type of stance phrase in the two corpora is evidentiality stance. These phrases enable writers to identify the source of certain statements, guide readers through the elaboration of an argument, and steer them to a specific point in the text (Hyland, 2005b). The statistical analysis shows that the economics corpus contains more evidentiality phrases, and in particular, self-mention phrases. This strong self-presence may be attributed to the research paradigms economics research subscribes to. Such research usually requires the construction of models that can be used for purposes of prediction, which thus enable economists to exercise a relatively higher degree of objective control over the research procedure compared to agricultural scientists. Example (3) shows a case where the self-

mention marker is used to initiate the research focus of this study, namely *investigating whether any significant difference exists in the default and operating risk of government-owned banks*.

- (3) *This study* investigates whether any significant difference exists in the default and operating risk of government-owned banks with respect to private banks. (ERAC#074:3)

The analysis of the attitude stance phrases in the two corpora shows that most of the attitudinal presence is expressed in the form of an evaluation, as in (4) which contains the evaluative phrase *play an important role*. The high frequency of this type of marker indicates that these disciplinary texts tend to appeal to readers' appreciation of the value of a statement. Notably, the occurrence of this dimension is found to be much less frequent than the cognitive dimension, indicating that researchers in the two disciplines have a tendency to express greater rational than emotional commitment in text construction.

- (4) Fungal hyphae and polysaccharides of microbial origin *play an important role* in soil aggregation. (ARAC#46:2)

The frequency of attitude stance is significantly higher in agriculture academic discourse ($LL=67.91$, $p=0.00$), indicating that agriculture researchers are more likely to show their evaluation-oriented presence in RAS. One of the possible reasons is that agriculture researchers tend to employ a larger amount of first-hand data, which are commonly observed or collected in their fieldwork or laboratory analysis. Thus, attitude stance phrases allow agriculture researchers to reveal their personal involvement in research procedures and potential subjectivities in the interpretation of findings. In contrast, the less frequent use of attitude phrases may be attributed to the model-based reasoning in economics discourse identified by Bondi (2006), which is "either developed through scenarios and simulations or tested with empirical data or simulations" (59). This feature thus permits little space for researchers in economics to express their subjective presence in their academic presence.

5. Conclusion

This study has presented an integrated stance classification framework by exploring the rhetorical use of stance phrases, and examining the disciplinary

variation of the stance phrases in the academic disciplines of agriculture and economics. To this end, this study extracted the most frequent 2- to 5-grams from two self-built corpora consisting of RAS in agriculture and economics and classified them functionally into cognitive, evaluation, hedges, and evidentiality stance by following the metafunctions in SFL. Drawing upon the extended definition given by Hyland (1999), this framework brings together previous classification schemes and extends stance classification by including the cognitive dimension, which denotes the reason-oriented dimension, as an independent category in the classification scheme.

The results show that these two disciplines (agriculture and economics) share a substantial number of academic conventions and also display significant disciplinary variation in their use of stance phrases. The findings suggest that academics tend to present a differing degree of stance phrases to align themselves with the community conventions of their respective disciplines. These differences in the use of stance phrases are in line with Biber et al. (2004) and Hyland (2008), who found considerable disciplinary variation in academic writing. By “instantiating the aspects of community’s values” (Hyland, 1999: 115), stance phrases align writers with academic conventions, and contribute to writer’s ability to present information as knowledge to the respective disciplinary community. The appropriate use of stance phrases thus can assist researchers to conform to community expectations (Hunston, 1994).

Pedagogically, the discipline-specific stance phrases and their specific use in RAS are of value to ESP/EAP practitioners and students or novice researchers in the two fields. The findings can be operationalized as strategies for students and novice researchers to develop their generic knowledge of disciplinary conventions. In particular, the findings can be drawn upon by practitioners in classroom or workshop-based training to assist novice writers in these disciplines to gain the skills and confidence in employing the appropriate type of stance markers in alignment with discipline-specific practices. The stance phrases retrieved in this study can serve as the searching items or teaching resources for corpus-based teaching (Dong & Buckingham, 2018b) and data-driven learning in academic writing instruction.

It should be noted that this study only scratches the surface of stance expressions by examining 2- to 5-grams, and thus excludes single words (such as *perhaps* and *maybe*), which also express stance. Finally, the focus of this study was limited to empirical RAS in two academic disciplines; an exploration of stance features in a wider range of disciplines and in studies

which employ different research paradigms would enable a more comprehensive view of stance features in academic discourse.

Acknowledgements

The authors gratefully acknowledge the support received from the National Education Examinations Authority and British Council English Assessment Research Grant, Young Taishan Scholars Program, and MOE (Ministry of Education in China) Project of Humanities and Social Sciences (No. 18YJCZH190).

Article history:

Received 1 October 2018

Received in revised form 6 April 2020

Accepted 30 April 2020

References

- Aikhenvald, A.Y. (2004). *Evidentiality*. Oxford: oup.
- Aull, L.L. & Z. Lancaster (2014). "Linguistic markers of stance in early and advanced academic writing: A corpus-based comparison". *Written Communication* 31(2): 151-183.
- Bakhtin, M. (1981). *The Dialogic Imagination: Four Essays by M.M. Bakhtin*. Austin: University of Texas Press.
- Bargiela-Chiappini, F. & C. Nickerson (2002). "Business discourse: Old debates, new horizons". *IRAL - International Review of Applied Linguistics in Language Teaching* 40: 273-286.
- Barlow, M. (2004). *Collocate 1.0*. Collocation Extraction Software. Houston, TX: Athelstan.
- Bazerman, C. (2010). *The Informed Writer: Using Sources in the Disciplines*, 5th ed. Fort Collins: The WAC Clearinghouse.
- Bhatia, V.K. (2010). "Interdiscursivity in professional communication". *Discourse & Communication* 4(1): 32-50.
- Biber, D. (1993). "Representativeness in corpus design". *Literary and Linguistic Computing* 8(4): 243-257.
- Biber, D. (2006). "Stance in spoken and written university registers". *Journal of English for Academic Purposes* 5(2): 97-116.
- Biber, D. (2009). "A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing". *International Journal of Corpus Linguistics* 14(3): 275-311.
- Biber, D. (2010). "Corpus-based and corpus-driven analyses of language variation and use" in B. Heine & H. Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 159-191. Oxford: oup.
- Biber, D., S. Conrad & V. Cortes (2004). "If you look at...: Lexical bundles in university teaching and textbooks". *Applied Linguistics* 25(3): 371-405.
- Biber, D., S. Johansson, G. Leech, S. Conrad, E. Finegan & R. Quirk (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Bird, S. (2006). "NLTK: The natural language toolkit" in J. Curran (ed.), *COLING-ACL '06: Proceedings of the COLING/ACL on Interactive presentation sessions*, 69-72. Stroudsburg, PA: Association for Computational Linguistics.
- Bondi, M. (1999). *English across Genres: Language Variation in the Discourse of Economics*. Modena: Edizioni Il Fiorino.
- Bondi, M. (2006). "A case in point: Signals of narrative development in Business and Economics" in K. Hyland & M. Bondi (eds.), *Academic Discourse across Disciplines*, 49-74. Bern: Peter Lang.
- Bondi, M. (2007). "Authority and expert voices in the discourse of history" in K. Fløttum (ed.),

- Language and Discipline Perspectives on Academic Discourse*, 66-88. Newcastle: Cambridge Scholars Publishing.
- Boote, D.N. & P. Beile (2005). "Scholars before researchers: On the centrality of the dissertation literature review in research preparation". *Educational Researcher* 34(6): 3-15.
- Botley, S.P. (2006). "Indirect anaphora: Testing the limits of corpus-based linguistics". *International Journal of Corpus Linguistics* 11(1): 73-112.
- Bowden, D. (1999). *The Mythology of Voice*. Portsmouth: Heinemann.
- Boye, K. & P. Harder (2009). "Evidentiality: Linguistic categories and grammaticalization". *Functions of Language* 16(1): 9-43.
- Breeze, R. (2018). "Researching evaluative discourse in Annual Reports using semantic tagging". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 35: 41-66.
- Bruce, I. (2014). "Expressing criticality in the literature review in research article introductions in applied linguistics and psychology". *English for Specific Purposes* 36: 85-96.
- Chafe, W.L. & J. Nichols (eds.) (1986). *Evidentiality: The Linguistic Coding of Epistemology*. Norwood: Ablex.
- 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(4): 313-326.
- Coates, J. (1987). "Epistemic modality and spoken discourse". *Transactions of the Philological Society* 85(1): 110-131.
- Cortes, V. (2004). "Lexical bundles in published and student disciplinary writing: Examples from history and biology". *English for Specific Purposes* 23(4): 397-423.
- Crismore, A. (1990). "Metadiscourse and discourse processes: Interactions and issues". *Discourse Processes* 13(2): 191-205.
- De Haan, F. (1999). "Evidentiality and epistemic modality: Setting boundaries". *Southwest Journal of Linguistics* 18(1): 83-101.
- Del Saz Rubio, M.M. (2011). "A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of agricultural sciences". *English for Specific Purposes* 30(4): 258-271.
- Di Carlo, G.S. (2015). "Stance in TED talks: Strategic use of subjective adjectives in online popularisation". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 29: 201-222.
- Dong, J. & L. Buckingham (2018a). "The collocation networks of stance phrases". *Journal of English for Academic Purposes* 36: 119-131.
- Dong, J. & L. Buckingham (2018b). "The textual colligation of stance phraseology in cross-disciplinary academic discourse: The timing of authors' self-projection". *International Journal of Corpus Linguistics* 23(4): 408-436.
- Dunning, T. (1993). "Accurate methods for the statistics of surprise and coincidence". *Computational Linguistics* 19(1): 61-74.
- Durrant, P. (2017). "Lexical bundles and disciplinary variation in university students' writing: Mapping the territories". *Applied Linguistics* 38(2): 165-193.
- Ellis, N.C., R. Simpson-Vlach & C. Maynard (2008). "Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL". *TESOL Quarterly* 42(3): 375-396.
- Feak, C.B. & J.M. Swales (2009). *Telling a Research Story: Writing a Literature Review*. Ann Arbor: University of Michigan Press.
- Feak, C.B. & J.M. Swales (2011). *Creating contexts: Writing Introductions across Genres*. Ann Arbor: University of Michigan Press.
- Fogal, G.G. (2019). "Tracking microgenetic changes in authorial voice development from a Complexity Theory Perspective". *Applied Linguistics* 40(3): 432-455.
- Greco, P. (2018). "Evidentiality and epistemic modality in witness testimony in the context of Italian criminal trials". *Journal of Pragmatics* 128: 128-136.
- Halliday, M.A.K. (1994). *Functional Grammar*. London: Edward Arnold.
- Halliday, M.A.K. (2003). "On the 'architecture' of human language" in M.A.K. Halliday, *On Language and Linguistics*, 1-29. London/ New York: Continuum.
- Hewings, M. & A. Hewings (2002). "'It is interesting to note that...': a comparative study of anticipatory 'it' in student and published writing". *English for Specific Purposes* 21(4): 367-383.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hong, H. & F. Cao (2014). "Interactional

- metadiscourse in young EFL learner writing: A corpus-based study". *International Journal of Corpus Linguistics* 19(2): 201-224.
- Hood, S. (2010). *Appraising Research: Evaluation in Academic Writing*. London: Palgrave.
- Hunston, S. & G. Thompson (2000). *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: OUP.
- Hunston, S. (2011). *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. New York: Routledge.
- Hyland, K. & J. Milton (1997). "Qualification and certainty in L1 and L2 students' writing". *Journal of Second Language Writing* 6(2): 183-205.
- Hyland, K. & J.K. Jiang. (2016). "Change of attitude? A diachronic study of stance". *Written Communication* 33(3): 251-274.
- Hyland, K. & P. Tse (2005). "Hooking the reader: A corpus study of evaluative that in abstracts". *English for Specific Purposes* 24(2): 123-139.
- Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam/ Philadelphia: John Benjamins.
- Hyland, K. (1999). "Disciplinary discourses: Writer stance in research articles" in C. Candlin & K. Hyland (eds.), *Writing: Texts, Processes and Practices*, 99-121. London: Longman.
- Hyland, K. (2003). "Self-citation and self-reference: Credibility and promotion in academic publication". *Journal of the American Society for Information Science and Technology* 54(3): 251-259.
- Hyland, K. (2005a). "Stance and engagement: A model of interaction in academic discourse". *Discourse studies* 7(2): 173-192.
- Hyland, K. (2005b). *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
- Hyland, K. (2008). "As can be seen: Lexical bundles and disciplinary variation". *English for Specific Purposes* 27(1): 4-21.
- Jiang, F.K. (2015). "Nominal stance construction in L1 and L2 students' writing". *Journal of English for Academic Purposes* 20: 90-102.
- Jiang, F.K. & K. Hyland, (2015). "'The fact that': Stance nouns in disciplinary writing". *Discourse Studies* 17(5), 529-550.
- Johns, A. M. (1997). *Text, Role, and Context: Developing Academic Literacies*. Cambridge: CUP.
- Koutsantoni, D. (2004). "Attitude, certainty and allusions to common knowledge in scientific research articles". *Journal of English for Academic Purposes* 3(2): 163-182.
- Kraska-Miller, M. (2013). *Nonparametric Statistics for Social and Behavioral Sciences*. Boca Raton: CRC Press.
- Lakoff, G. (1973). "Hedges: A study in meaning criteria and the logic of fuzzy concepts". *Journal of Philosophical Logic* 2(4): 458-508.
- Lancaster, Z. (2014). "Exploring valued patterns of stance in upper-level student writing in the disciplines". *Written Communication* 31(1): 27-57.
- Lancaster, Z. (2016). "Expressing stance in undergraduate writing: Discipline-specific and general qualities". *Journal of English for Academic Purposes* 23: 16-30.
- Lim, J.M.H. (2010). "Commenting on research results in applied linguistics and education: A comparative genre-based investigation". *Journal of English for Academic Purposes* 9(4): 280-294.
- Lu, X. (2014). *Computational Methods for Corpus Annotation and Analysis*. New York: Springer.
- Manning, C.D. & H. Schütze (1999). *Foundations of Statistical Natural Language Processing*. Massachusetts: MIT press.
- Martin, J.R. & P.R.R. White (2005). *The Language of Evaluation: Appraisal in English*. New York: Palgrave Macmillan.
- Martin, W.M. (2006). *Theories of Judgment: Psychology, Logic, Phenomenology*. Cambridge: CUP.
- Martinez, I.A., S.C. Beck & C.B. Panza (2009). "Academic vocabulary in agriculture research articles: A corpus-based study". *English for Specific Purposes* 28(3): 183-198.
- Matsuda, P.K. & C.M. Tardy (2007). "Voice in academic writing: The rhetorical construction of author identity in blind manuscript review". *English for Specific Purposes* 26(2): 235-249.
- Matsuda, P.K. (2001). "Voice in Japanese written discourse: Implications for second language writing". *Journal of Second Language Writing* 10(1): 35-53.
- Mauranen, A. & M. Bondi (2003). "Evaluative language use in academic discourse". *Journal of English for Academic Purposes* 2(4): 269-271.
- Mauranen, A. (1993). "Contrastive ESP rhetoric: Metatext in Finnish-English economics texts". *English for Specific Purposes* 12: 3-22.
- Nesi, H. & S. Gardner (2012). *Genres across the Disciplines: Student Writing in Higher Education*. Cambridge: CUP.

- Nuyts, J. (2001). *Epistemic Modality, Language, and Conceptualization: A Cognitive-pragmatic Perspective*. Amsterdam: John Benjamins.
- Oakes, M.P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Rayson, P. & R. Garside (2000). "Comparing corpora using frequency profiling" in *WCC'00: Proceedings of the Workshop on Comparing Corpora*, 1-6. Stroudsburg, PA: Association for Computational Linguistics.
- Rescher, N. (2005). *Epistemic Logic: A Survey of the Logic of Knowledge*. Pittsburgh: University of Pittsburgh Press.
- Römer, U. (2008). "Identification impossible? A corpus approach to realisations of evaluative meaning in academic writing". *Functions of Language* 15(1): 115-130.
- Simpson-Vlach, R. & N.C. Ellis (2010). "An academic formulas list: New methods in phraseology research". *Applied Linguistics* 31(4): 487-512.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: OUP.
- Stotesbury, H. (2003). "Evaluation in research article abstracts in the narrative and hard sciences". *Journal of English for Academic Purposes* 2(4): 327-341.
- Stubbs, M. (1996). *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture*. Oxford: Blackwell.
- Swales, J.M. & C.B. Feak (2000). *English in Today's Research World: A Writing Guide*. Ann Arbor: University of Michigan Press.
- Swales, J.M. (2002). "Integrated and fragmented worlds: EAP materials and corpus linguistics" in J. Flowerdew (ed.), *Academic Discourse*, 150-164. London: Longman.
- Tao, J.T. & X.A. Gao (2018). "Identity constructions of ESP teachers in a Chinese university". *English for Specific Purposes* 49: 1-13.
- Tucker, P. (2003). "Evaluation in the art-historical research article". *Journal of English for Academic Purposes* 2(4): 291-312.
- Van Dijk, T.A. (1997). "Discourse as interaction in society" in T.A. van Dijk (ed.), *Discourse as Social Interaction*, 1-37. Vol. 2. London: SAGE.
- Wierzbicka, A. (1994). "Semantics and epistemology: The meaning of 'evidentials' in a cross-linguistic perspective". *Language Sciences* 16(1): 81-137.
- Zare, J., A. Eslami-Rasekh & A. Dabaghi (2017). "Importance marking: Giving reference to the highlighted discourse in academic lectures". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 33: 213-234.
- Zuck, J.G. & L.V. Zuck (1986). "Hedging in newswriting" in A.M. Cornu, J. Vanparijs & M. Delahaye (eds.), *Beads or Bracelets: How Do We Approach LSP: Selected Papers from the Fifth European Symposium on LSP*, 172-181. Oxford: OUP.

Jihua Dong is Professor, Qilu Young Scholar, and Taishan Young Scholar in the School of Foreign Languages and Literature at Shandong University, China. Her research interests include corpus linguistics, corpus-based teaching, academic writing, and discourse analysis. She has previously published in *English for Specific Purposes*, *International Journal of Corpus Linguistics*, *Journal of English for Academic Purposes*, and *System*, among others.

Louisa Buckingham is a Senior Lecturer in Applied Language Studies and Linguistics at the University of Auckland. She has a broad range of research interests which include the use of corpus linguistics as a methodology to explore discourse. She has published in various journals including *TESOL Quarterly*, *System*, *Journal of English for Academic Purposes*, and *English for Specific Purposes*.

NOTES

¹ The specific information on how to conduct this test is available in Kraska-Miller (2013).