Stance Phraseology in Cross-Disciplinary Academic Discourse: the Disciplinary Variation of Writers' Authorial Positioning

Jihua Dong

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Linguistics, The University of Auckland, 2018

Abstract

This thesis investigates the cross-disciplinary features of stance phrases used by agriculture and economics writers to establish their identity and authority in the academic community. This study has three main objectives: investigating the occurrences of stance phrases, their textual distribution, and collocation networks. Using two purposeful-built academic corpora consisting of agriculture and economics research articles, I first adopted a corpus-driven approach to retrieve 2- to 5-grams, and then classified the retrieved phrases functionally by following the metafunctions of Systemic Functional Linguistics (SFL). The cross-corpora comparisons show that the two disciplines share similarities in terms of the frequency ranking and proportional distribution of the four types of stance phrases. The statistical analysis indicates that agriculture researchers tend to use cognitive, attitude, and hedges phrases more frequently, while economics writers employ significantly more reference phrases. This study also identifies some salient syntactic patterns associated with stance phrases in the two corpora. These findings can enrich our understanding of the disciplinary uses and syntactic patterns of stance phrases in the two disciplines.

To explore text colligation of stance phrases, this study employed *Wordskew* software (Barlow, 2016) to extract the text positions of stance phrases at the levels of sentence, paragraph and text in the two academic disciplines. The cross-comparison indicates that stance phrases are distributed unevenly in different text positions, and the two disciplines have notably similar proportional preferences for particular text positions. Significant differences are also identified in the textual colligation of stance phrases distributed at different text positions in the two corpora. These findings suggest that text positions may embody routinized discourse functions result from a process of text priming (Hoey, 2005). Also, the text colligation of stance phrases may reflect the timing when authors opt to project themselves in text-based communication. From a communicative viewpoint, the text colligation of stance phrases may serve as a strategy for academic writers to construct appropriate persona and promote effective communications with readers.

This study also investigates the collocation networks of stance phrases in cross-disciplinary academic discourse from a quantitative perspective. Using GraphColl to identify the collocates of stance phrases, this study depicts the collocation networks of different categories of stance phrases in the two corpora. The findings reveal a complex and intricate collocation network figuring both the within-stance category collocation and the collocation between stance phrases and other semantic domains. Also, significant discipline-specific features are identified in the collocation networks

centered on stance phrases. These findings contribute to our knowledge of how different aspects of stance operate with each other and with their surrounding linguistic domains to assist writers to develop convincing arguments, construct coherent texts and communicate effectively with anticipated readers.

Acknowledgements

I would like to express my sincere gratitude to my main supervisor Dr Louisa Buckingham for her encouragement, support, expertise and enlightenment. She has always helped to plan my study and keep my thesis writing organised. She not only had expertise with regard to my thesis but has also been a mentor in many aspects of my life. The supervision meetings have always been very inspirational, thought-provoking and enjoyable. Her support and guidance were valuable in helping me to develop professionally and intellectually, and prepare me to explore possibilities I never dreamt of before the programme.

My great appreciation also goes to my co-supervisor, Dr Shaofeng Li, who has given me timely support and guidance through my PhD study. He was always ready to listen, and suggest new perspectives for my study, and offered useful approaches to data analysis and suggestions for other aspects of my study. His support and comments were invaluable for the completion of my thesis.

I would also like to express my thanks to people who offered their time, advice and professional insights during different stages of this study. Their kind response to my enquiries and their insightful comments contributed substantially to the interpretation of the quantitative data.

Last but not least, I would also like to thank my husband, Dr Hao Wu, who helped me to double check the data, and offered useful suggestions throughout my study and thesis writing stage. My thanks also go to my beloved son, Franklin Yanhui Wu, for cheering me up and giving me great motivation to complete the thesis. I would also like to express my great thanks to all the other family members, my parents and parents-in-law, brothers and sisters, for their understanding and support.

Table of Contents

List of Figures

List of Tables

Glossary

1. Attitude stance relate to "attitude toward the proposition and convey stance, or assessment of expectations" (Biber et al., 1999, p.966)
2. Collocate refers to "a word which occurs in close proximity to a word under investigation is called a collocate of it" (Sinclair, 1991, p.170).
3. Collocation is concerned with the occurrences of "two or more words within a short space of each other in a text. The unusual measure of proximity is a maximum of four words intervening." (Sinclair, 1991, p.170).
4. Concordance is a list of all the contexts in which a given word or phrase occur in the corpus (Gonzalez-Marquez, 2007, p.39)
5. Cognitive stance, in the present study, is used to relate to writer's reason-oriented perspective of their mental states and activities (Biber et al, 1999) towards entity or proposition in expressing assertions and construing knowledge in the discourse.
6. Evaluation is defined as an attitude towards a person, situation or other entity and is both subjective and located within a societal value-system (Hunston, 1994).
7. Hedges are rhetorical devices used to display due caution, modesty, and humility when making statements and their removal is a major linguistics means of conferring greater certainty to propositions (Hyland, 1998a)
8. Linguistic devices are ways by which a writer expresses his opinion towards certain situation and way by which he makes audience interested in his topic. (web: https://aenga2. wordpress.com/2008/03/18/linguistic-devices-%E2%80%93-the-way-they-are-used-in-articles/)
9. Metadiscourse is defined as "discourse which goes beyond and above the actual content of the basic propositional information being presented, indicating to readers how they may 'organize, classify, interpret, evaluate, and react" (Vande Kopple, 1985, p.83).
10. Phrases refers to "a very general term used to describe the tendency of words, and groups of words, to occur more frequently in some environments than in others." (Hunston, 2011, p.5)
11. Reference stance is defined as "the discourse markers that locate the source of statements in the textual description and direct readers' understanding of the primary message in terms of its content and structure" (Crismore, 1989, p.64).
12. Semantic preference is concerned with the semantic field a word's collocates predominantly belong to (Sinclair 1991).
13. Stance refers to "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, 1999b, p.101).

14. Textual colligation refers to the place in a sentence that a lexical item or group of items prefers or avoids (Hoey, 2005).
15. Collocation network is concept used to depict the collocational relationship between a linguistic item and its context (Phillips, 1989).

List of Acronyms

ARAC- Agriculture Research Article Corpus

AFL- Academic Formulas List

BAWE - British Academic Written English Corpus

BNC - British National Corpus

BRILC - Book Reviews in Linguistics Corpus

COCA-Corpus of Contemporary American English

EAP-English for Academic Purpose

ESP-English for Specific Purpose

ERAC - Economics Research Article Corpus

EASPL- Economics Academic Stance Phrases List

FTW-Formulas Teaching Worth

LL- Log-Likelihood

MI - Mutual Information

MICASE - Michigan Corpus of Academic Spoken English

MICUSP - Michigan Corpus of Upper-level Student Papers

MWC- Multi-word Construction

RA - Research Article

SCF - Stance Classification Framework

SCI- Science Citation Index

SFL-Systemic Functional Linguistics

SSCI -Social Science Citation Index

USAS- UCREL Semantic Analysis System

Chapter 1 Introduction

1.1 General background

Over the past few decades, writers in applied linguistics have paid growing attention to writers' authorial positioning in academic discourse. The way writers position themselves is widely regarded as an essential element of persuasive and argumentative discourse (Crismore & Farnsworth, 1990; Hyland, 1998, 2004) and is "pervasive in all uses of language" (Stubbs, 1996, p.202), because "whenever writers say anything, they encode their point of view towards it" (Martin & White, 2005, p.92). The authorial positioning expressions, therefore, play a pivotal role in the effective communication between writers and readers, particularly in influencing readers' acceptance of the claims delivered in a text.

A substantial number of studies have delved into rhetorical devices expressing writers' authorial positioning, which has been variously termed as evaluation (Hunston, 1994; Hunston, 2011), appraisal (Martin & White, 2005), metadiscourse (Hyland, 1999a; Hyland, 2005b), engagement (Hyland, 2005a), stance (Biber, 2006; Hyland, 2005a), modality (Nuyts, 2001) and voice (Bowden, 1999; Matsuda, 2001; Matsuda & Tardy, 2007). Stance refers to "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, 1999b, p.101). This concept entails a close connection with authors' presence in a text, and reveals an authorial positioning constructed to provide convincing evidence and persuade readers (explicitly or implicitly) to adopt a certain position or point of view (Hunston, 2011). Given that the present study attempts to examine how authors involve themselves in textual construction, the term 'stance' is used in this study.

This study aims to expand studies in stance expressions by examining stance phrases in the genre of written academic discourse. Being perceived as "a class of communicative events that share a recognisable communicative purpose" (Swales, 1990, p.58), academic discourse is recognized as a genre on account of the many similarities found across academic texts with regards to style, form, structure, purpose and intended audience (Hyland, 2009). Interactional in nature, this genre is nevertheless known for its "hidden dialogicality" (Bakhtin, 1984, p.242), in the sense that writers can only negotiate with the reader's anticipated response to a text (Lancaster, 2016).

A considerable number of studies have been conducted to map the discipline-specific uses of stance features in academic discourse (e.g., Bruce, 2014; Cortes, 2008; Hyland, 2005a; Giannoni, 2010; Samraj, 2008). However, few attempts have been made to uncover stance expressions in agriculture and economics academic discourses, and specific stance phrases used in the academic

discourse of the two disciplines remain largely uncharted. Therefore, this study sets out to investigate stance phrases in the two academic disciplines in terms of frequency, textual colligation and collocation networks.

The following section first discusses the theoretical frameworks underpinning this study (Section 1.2), and then provides justifications for the phraseological perspective (Section 1.3) and the disciplinary perspective (Section 1.4). Section 1.5 presents the implication and purpose of this study, and Section 1.6 provides an overview of the thesis structure.

1.2 Theoretical frameworks

1.2.1 Metafunctions of systemic functional linguistics

This study falls under the theoretical framework of Systemic Functional Linguistics (SFL) and Pattern Grammar theory. According to SFL, language is considered as a social semiotic system with the function of meaning-maker. Three metafunctions, namely, ideational, interpersonal and textual, are proposed to explain the semantic organization of language (Halliday, 1994). The ideational function is concerned with how language is built and maintained during writers' or speakers' interaction with the world around. It is composed of the experiential and logical functions. The experiential function concerns the rhetorical devices that enable writers or speakers to make sense of the world, and the use of linguistic expressions which help language users to "construe a theoretical model of their experience" (Halliday, 2003, p.15). The logical function pertains to the rhetorical choices "which set up logical–semantic relationships between one clausal unit and another" (Halliday, 2003, p.17). The logical markers, such *as due to, as a result*, and *in contrast,* allow writers or speakers to demonstrate meaning relationships when joining clauses together. It is worth noting that logical devices not only connect two clausal units in a logical and coherent way, but also show authors' mental processing of the information. The second metafunction, the interpersonal function, relates to the social relationship construed by rhetorical devices. The interpersonal expressions e.g., *it is interesting to* and *it is surprising that* allow language users to intrude into text organization and communicate their attitude or evaluation in order to establish rapport with readers, thus contributing to a creation of social bonds. Textual function is concerned with grammatical devices used to maintain the flow of discourse and to construct coherent and well-organized discourse. The linguistic choices such as *the remainder of the paper* and *as shown in Table1* enable authors to "create coherent text – text that coheres within itself and with the context of situation" (Halliday, 2003, p.17).

1.2.2 Pattern grammar

Pattern grammar (Francis, Hunston & Manning, 1996; Francis, Hunston & Manning, 1998) stems from Sinclair's (1991, p.110) *Idiom Principle* and the open-choice principle (Sinclair, 1991) and results from the systematic empirical observation of recurrent patterns in authentic language corpora.

Opposite to the traditional dichotomy of syntax and lexis, the theory takes the view that there is a close and systematic association between words and particular syntactic patterns. According to pattern grammar, words tend to occur in association with one or more syntactic patterns (Hunston, 2002) and the 'semi-preconstructed phrases' constitute the core unit of linguistic meanings and structures in language (Sinclair, 1991, p.110). By focusing on the syntactic environments of lexis, the theory highlights the relationship between syntactic patterns and word meanings, particularly the synthesis of the grammatical structures and simple words pattern (Hunston & Francis, 2000). In corpus-based studies, the association between pattern and meaning also makes it possible to perform automatic extraction of certain meanings or functions embedded in particular patterns on a large scale (Hunston & Sinclair, 2000, p.82), as has been illustrated by the automatic search for patterns previously (e.g., Groom, 2005; Hunston, 2011).

1.3 Justification for the phraseological perspective

The present study seeks to take formulaic expressions as the point of departure and explore how academic writers use fixed multiword strings to express their stance in academic discourse. In previous literature, various labels have been used to refer to formulaic expressions, such as lexical bundles (Biber & Barbieri, 2007), multi-word constructions (Liu, 2012), collocation (Durrant, 2009), formulaic expressions (Staples, Egbert, Biber & McClair, 2013) and phrases (Hunston, 2011). In the present study, I will use the term 'phrase' as an umbrella term for "a frequently occurring sequence of words which is usually not idiomatic in meaning, and is usually not a complete structural unit" (Biber, Johansson, Leech, Conrad, & Finegan, 1999, p.264).

The justification for the phraseological perspective is as follows. Firstly, it is generally accepted that the meanings of words are usually influenced by collocating words in a "repertoire of multi-word patterns" (Sinclair, 1991, p.108). Thus focusing on phrases helps to decrease or even remove ambiguities that are likely to arise in the individual word-based analysis. Also, phraseological sequences have been identified to be more easily and effectively processed by language users (Conklin & Schmitt, 2008); hence, the formulaic perspective is seen to be of high pedagogical value, particularly in terms of assisting learners' language learning and enhancing their information processing capacity. In addition, the identification of phrases in a text is generally considered an important step in defining the characteristic of the community communication (Firth, 1935/1957, p.180), because "semi-preconstructed phrases" constitute the core unit of linguistic meaning and structure (Sinclair, 1991, p.110). Conversely, the inappropriate use of phrases may expose ones' inexperience in academic writing, which is often regarded as insufficient generic competence by experienced readers (e.g., examiners or reviewers).

1.4 Justification for the disciplinary perspective

Disciplinary variation in academic discourse is another perspective taken in the present study. The rationale for this perspective resides in the assumption that each community may select or prioritize its unique "norms, values, language and ways of communication" (Van Dijk, 1997). As suggested previously, disciplinary variation is a prominent feature of academic discourses (Hyland & Tse, 2005; Koutsantoni, 2004; Shaw, 2004).

When investigating stance expressions in academic discourse, this study focuses on stance phrases in the disciplines of agriculture and economics. The justification for focusing on these two disciplines are discussed as follows.

First of all, agriculture and economics are essential components in the established academic disciplinary system followed by universities, academic journals, and conferences. According to the general academic discipline classification system, both agriculture and economics constitute essential elements of professional sciences, one of the macro-division of academic discipline categories, in parallel with humanities, social sciences, natural sciences, and formal sciences.

In terms of the disciplinary feature, research in the agriculture is considered to be analytical in nature and empirically based on observable experience (Del Saz Rubio, 2011; Martínez, Beck & Panza, 2009). Typically subscribing to an experiment-based paradigm, agricultural writers draw on a wide variety of theories and research methods, and their work aimed at making evidence-based recommendations for agricultural practice. In contrast, the economics discipline possesses both theoretical, or 'pure', and applied strands; research typically follows a quantitative paradigm (Becher & Trower, 2001) and draws on data related to "material and financial goods" (Bazerman, 2010, p.14). In a corpus-based study, Bhatia (2010) identified the nature of economics academic discourse as being more characteristic of a discussion on economic facts and figures, not necessarily entailing "a consistent and true representation of the statistical information" and often "lead[ing] to varying interpretations" (p. 41). Thus, the research paradigm of this discipline is characterized by the construction and testing of models for the purpose of prediction, and the formulation of economic forecasts, which are usually intended to inform and guide governmental bodies and business institutions (Bazerman, 2010, p.281).

Previous studies on disciplinary discourse in agriculture, attention has been drawn to academic wordlists in agriculture (Martínez et al., 2009), the macro-structure and metadiscourse features of the research papers (Del Saz Rubio, 2011), and citation styles (Thompson & Tribble, 2001). In the discipline of economics[1], previous studies have focused on the linguistic or rhetorical features, such

[1] It is necessary to point out that different names have been used in previous studies to refer to this discipline, such as economics, finance, banking and accounting. Although the different names refer to particular subfields, as suggested by

as evaluative stance (Shaw, 2003), metadiscourse markers or metatext pertinent to authors' stance (Dahl, 2004; Khedri, Heng & Ebrahimi, 2013; Lancaster, 2016; Mauranen, 1993), rhetorical metatext structure (Moreno, 2004), hedging stance (Donohue, 2006), rhetorical devices used to accomplish academic persuasion (Bondi, 1999), personal pronouns that manifest authors' presence (Yeo & Ting, 2014), and high-frequency collocations of nouns (Peacock, 2012) (a detailed review of the previous studies appears in Section 2.4.2). Khedri, M., Heng, C. S., & Ebrahimi, S. F. (2013).

Also, in a study charting the academic discipline territories, Durant (2015) identified that agriculture and economics both fall under the general discipline of commerce, a generally broad discipline division in the academic discipline system, in parallel with humanities and social sciences, science and technology, and life sciences. When it comes to the sub-disciplinary categories, economics has "a presence in science and technology" (p.11), while agriculture "appears in life sciences" (p.11). With regard to the nature of these two disciplines, it has been suggested that agriculture is analytical in nature and empirically based on observable experience (Del Saz Rubio, 2011; Martínez et al., 2009).

Despite these discussions and findings in these two disciplines, only a few studies have been carried out to compare the discourse features of the two disciplines, such as the forms of argument, particularly the dialogic and reflexive nature of persuasion in economics academic writing (Bondi, 1999), the macro-structure and metadiscoursal features of research article introductions in agricultural sciences (Del Saz Rubio, 2011), and the argument developing skills in economics and agriculture together with several other disciplines (Nesi & Gardner, 2012). Also, little is known about the differences in the stance-taking practice in the two disciplines. Given that stance can inform us about how academic texts are textually shaped and how writers rhetorically position themselves on research subjects, this study sets out to examine stance formulaic expressions in the RAs of the two academic disciplines.

In addition, agricultural and economic-specific courses have been a particular focus in the teaching practice in university contexts. The need for agricultural and economics-specific English for academic purposes (EAP) courses in higher education, especially for non-native speakers, has been highlighted by Huffman and Evenson (2008, p.193) and Angelil-Carter (1998, p.19) respectively. According to Huffman and Evenson (2008, p.193), an increasing number and percentage of non-native students have enrolled in agricultural science degrees, which boosts the demand for language of science and agriculture-specific training programmes. Similarly, Angelil-Carter (1998) found that both native and non-native students are in great need of economics language and concepts in the discipline-specific courses. Also, students in these disciplines are found to be insufficient in their academic writing skills, and this is especially the case for non-native speakers. For instance, a national

the previous studies economics has been used as a concept covering relatively broader areas. Therefore, this study adopted the title "economics".

survey on economics professors' perceptions of student preparedness indicated that economics degree students are weak in their academic writing skills (Tannera & Cuddb, 1999). Similarly, Conroy, Dailey and Shelley-Tolbert (2000) stressed non-native students' needs of English degree training in the agriculture-specific degree programme (p.73, p.193).

Within academia, in the context of increasing non-native English speakers participating in international scholarly publications in English (Buckingham, 2008; 2014; Flowerdew, 2000), an increasing number of non-native writers in agriculture and economics are also actively involved in public English academic communication. One evidence of this trend can be found from the increasing number of research articles written by non-native speakers published in international journals (i.e., Jiang, 2015; Xie 2016, Zhang, 2015, etc.). This trend also stimulates ever-expanding demands for academic conventions and competence in conducting communication in these two disciplines, among many other. In this light, enabling EFL students and novice writers with the knowledge and competence to communicate adequately in their academic writing is essential for preparing them to be participating competently in academic communications.

The high demand for the agriculture and economics-specific courses is particularly the case in an EFL context, such as China. According to a large-scale survey on Chinese university students' employment jointly conducted by the Central School Department of the Communist Youth League and Peking University's Public Policy Research Institute in 2006, the employment of students majoring in agriculture ranked the highest, with a percentage of 78.38%. Similarly, the same survey also highlights the demand for graduates in the discipline of economics in China. Impelled by this demand, increasing numbers of students are enrolling in agriculture and economics-specific degree programmes. With regard to the academic journals focusing on the two disciplines, 172 core economics journals and 138 core agricultural journals currently exist in China, which indicates that the two disciplines have been accommodated into the major Chinese academic discipline system.

Additionally, the high demand for English for Specific Purposes (ESP) courses has spurred a succession of research projects concentrated on the discipline-targeted research of the academic writing of non-native speakers. For example, Hyland & Milton (1997) compared the expression of doubt and certainty in Chinese students' writing with British students' writing. Their analysis showed that L2 Chinese writers use a limited range of items, display strong commitments and have great difficulty in conveying a precise degree of certainty. Similarly, Hu and Cao (2011) compared the use of hedges and boosters in RA abstracts published in Chinese- and English- medium journals in the discipline of applied linguistics. Their analysis indicates that the abstracts published in English-medium journals include more hedges than those in Chinese-medium journals. They explained Chinese writers' lack of hedges in terms of the culturally preferred rhetorical strategies, the writer's epistemological beliefs, and their lack of English knowledge and English writing conventions.

Similarly, Jiang (2015) compared the nominal stance construction in argumentative essays written by Chinese and American students. The study showed that Chinese students use fewer nominal structures when expressing the event, discourse and cognition types of stance nouns, but they had a preference for attitudinal adjectives and first-person possessives used to modify stance nouns. To sum up, the expanding research in Chinese academic writing highlights the great need for EAP and ESP in the Chinese context.

As suggested by previous studies, stance construction is "an elusive concept" (Mauranen & Bondi, 2003, p.269), thus posing a great challenge to students' academic writing. Also, many students and novice writers have been identified to have encountered great difficulties in contextualizing research needs and constructing stance in their academic writing, such as essays and dissertations (Feak & Swales, 2011). Empirical studies centred on examining stance expressions have also pinpointed students' or novice writers' difficulties in establishing critical stance in writing literature reviews (Bruce, 2014), synthesizing and critiquing in composing literature reviews of doctoral theses (Boote & Beile, 2005), discussing and commenting on research results (Lim, 2010), and using *certainty* and *affect* stance appropriately to express their statements in academic discourse (Hyland & Milton, 1997). In addition, my own experience as a teacher of agriculture/economics-specific EAP courses reinforces the view that students, particularly non-native students, have encountered great difficulty in their position-taking and stance construction in their academic writing practices.

Despite the aforementioned need for discipline-specific academic English and the difficulties in stance construction that students often experienced in the two disciplines, few attempts have focused on agriculture and economics academic discourse, and the recurring stance phrases in the two academic disciplines have not been identified so far. Therefore, the present study aims to address this issue and explore how the authors in the two disciplinary academic discourses employ stance phrases to construct their stance. Specifically, this study identifies the common rhetorical devices used to express authors' presence in the RAs of the academic disciplines of agriculture and economics. It also compares the similar and different conventionalized uses of stance phrases in the two academic disciplines.

1.5 The implications

The study stretches across several fields of applied linguistics—corpus linguistics, English for Academic Purposes (EAP), English for Specific purposes (ESP), discourse analysis, and genre analysis. The findings are intended to deepen our understanding of how disciplinary academic writers project themselves and construct their authorial positioning in academic discourse. Also, with a focus on the discipline-specific uses of stance phrases, this study aims to provide a disciplinary picture of stance phrases and to extend to our knowledge of disciplinary variations in terms of stance

construction.

Theoretically, the findings in this study can provide insight into how writers establish their status and maintain their identity by means of stance phrases in RAs. The disciplinary variations in this study can reveal "the differing norms and epistemologies of specific disciplines, or differing roles of the writers" (Basturkmen, 2014, p.379), and provide a glimpse of social values, conventions, and expectations hidden to many novice writers in the academic community. Considering that identity is often constructed through the use of discourse (Paltridge, 2012, p.38), the findings in the present study can enrich our understanding of authors' identity construction by means of stance linguistic resources in academic writings. In addition, academic discourse often absorbs a variety of elements from the communities that writers engage in (Becher & Trowler, 2001, p.59). Thus, the findings can enrich our understanding of the disciplinary features of the two academic communities under scrutiny.

Pedagogically, the findings can be useful for EAP writing in the field of agriculture and economics and in terms of the following perspectives. First of all, the disciplinary usages of stance phrases identified in this study can be used for class instruction intended to arouse students' and novice writers' stance awareness, assisting them to master the appropriate use of stance expressions in their academic writing. Secondly, the corpus-based extraction obtained a list of the academic stance phrases in the disciplines of agriculture and economics, which can be drawn upon in EAP instruction with an aim to help learners to "construct appropriate contents and allude to shared disciplinary assumptions"(Hyland,1998, p.437). The findings regarding stance uses can also be useful for helping novice writers to grasp the contextualized usage of phrases and to familiarize themselves with appropriate stance expressions strategies. In particular, these findings can be incorporated into developing teaching materials, curriculum design and designing other teaching and learning activities in terms of stance expressions and disciplinary discourse writing.

1.6 An overview of the thesis

The structure of the thesis is organized as follows. Chapter 2 provides an overview of the previous studies in the area of stance expressions, disciplinary variations, corpus-based and corpus-driven studies in stance features, and the studies conducted on textual colligation and collocation network. Chapter 3 presents the methodology involved in corpus-building, stance phrases extraction, the categorization of stance phrases and the statistical tests to compare the stance uses in the two academic disciplines. Chapter 4 first reviews the previous classification of stance expressions and then discusses the framework synthesized and used for analysis in the present study. Chapter 5 first discusses the results of the general frequency of stance uses in the two disciplines, and then compares the disciplinary variation of stance phrases in the two corpora; this is followed by a description of the common syntactic patterns associated with stance phrases in the two disciplines. Moving on to the textual distribution aspect of stance phrases, Chapter 6 presents the methods used to extract textual

colligation of stance phrases, and then reports the findings regarding textual colligation of stance phrases, as well as the disciplinary variation in the textual colligation of stance phrases in the two academic disciplines. With a focus on collocation network of stance phrases, Chapter 7 elaborates on how the collocation network was retrieved, and the findings of the collocation network in terms of within-stance collocation (concerning how stance phrases collocate within the different categories), and the collocation network with other semantic domains (relating to how stance phrases co-occur with other semantic features). The cross-disciplinary variation regarding the collocation networks of stance phrases is also discussed in Chapter 7. The final chapter summarizes the main findings of the present study, and then discusses the implications and applications of these research findings. This is followed by a discussion of the limitations in this study, as well as suggestions and directions for future studies.

Chapter 2 Literature Review

This chapter first provides a general background to the present study, and then examines the previous studies conducted in stance expressions in terms of the grammatical patterns and phraseological perspectives (Section 2.2). Section 2.3 reviews the previous research carried out by using the corpus-based and corpus-driven approach, two major approaches used in this study. Section 2.4 then reviews the studies centered on stance expressions in the strand of disciplinary academic discourse. This is followed by a review of studies on textual colligation and collocation networks in Section 2.5 and 2.6 respectively. The last section of this chapter presents the research questions addressed in the study.

2.1 General background

Previous studies have displayed growing attention to the examination of stance expressions in academic discourse (Bamford & Bondi, 2005; Conrad & Biber, 2000; Freddi, 2005; Hunston, 1993; Hyland, 1999a; Shaw, 2004). This interest coincides with the growing awareness that "language is both the foundation and the instrumentality of the social construction of reality" (Berger & Luckmann, 1966, p.108).

Within the flourishing academic community committed to studying writers' stance expressions in academic texts, the previous focus has mainly centered on the following two fields: oral discourse and written discourse. Papers presenting field work on oral academic discourses account for only a minor portion in comparison with those dealing with written discourses. For instance, Swales (2002) examined attitudinal lexis in the Michigan Corpus of Academic Spoken English (MICASE) to answer the question whether academic speech was more like academic written prose or resembled everyday conversation. The study found that "the stance volume" is amplified in spoken texts in comparison to those in academic prose. Rzewski (2005) investigated the lexis expressing interpersonal engagement and the modality-related expressions related to speakers' certainty in students' dissertation defenses. He identified the role that interpersonal meanings play in structuring a coherent discourse. Similarly, Morton (2009) focused on student architecture presentations (also known as jury, the review) in architecture courses, an academic assessment genre. She analyzed the interpersonal rhetorical strategies valued in this genre. Morton identified this genre as notably characterized by narrative style, metaphorical images, and dynamic grammar.

Meanwhile, a substantial number of studies have concentrated on written discourses in the academic community. The prevalence of such focus may be attributed to the fact that written academic discourse is the most common form to deliver authorial opinions in the process of communicating with peers in the community. Within the written genre, substantial efforts have been made to tease out

the rhetorical devices in a wide range of sub-genres, such as academic discourse (Buchingham, 2008; Conrad, 1996; Cortes, 2008; Hyland, 1996, 2005), letters to editors (Bloch, 2003; Magnet & Carnet, 2006), grant proposals (Pascual & Unger, 2010, p.261), book reviews (Römer, 2008, p.115; Shaw, 2004, p.121).

Among the studies conducted in the strand of academic discourse, research article (henceforward RA) has attracted considerable attention possibly for the following reasons. First of all, RAs constitute a well-established social register of communication in the academic community, comprising one of the predominant forms used by writers to disseminate knowledge, and to win academic recognition and reputation in their own community. In pedagogical practice, RAs are often treated as exemplars for students to emulate (Hyland, 2008a) and have been drawn upon extensively as reading materials in the course of various levels in higher education (Hood, 2010, p.6).

Despite the aforementioned important roles that that RAs play in academic communication, it has been suggested that writers often encounter great difficulties in aligning themselves with the community conventions, especially in the use of stance expressions, on account of its sophisticated composition process and linguistic realizations (Bruce, 2003; Feak & Swales, 2011; Lim, 2010). Therefore, the present study investigates the rhetorical devices used by academic writers to express their stance expressions in RAs in the academic disciplines of agriculture and economics.

2.2 Previous studies on stance expressions

2.2.1 The grammatical patterns of stance expressions

Among the burgeoning studies on stance expressions, a large amount of attention has been attracted to different linguistic features, such as lexical items (Charles, 2003; Hyland & Tse 2009; Tucker, 2003), and syntactic patterns (Dong & Jiang, in press; Groom, 2005; Larsson, 2016; Hewings & Hewings, 2002, p.367; Hyland & Tse 2005). The following section reviews the previous research focusing on the various linguistic forms of stance features.

Firstly, previous studies have concentrated on a wide range of lexical items, such as, evaluative citation verbs (Thompson & Ye, 1991, p.374), stance adverbs (Biber et al., 1999), evaluative lexis (Giannoni, 2010), anaphoric nouns (Charles, 2003), reporting clause (Charles, 2006b), and citation practices (Charles, 2006a). To be more specific, Thompson and Ye (1991) investigated citation verbs conveying evaluation meaning in academic papers. The study identified the common devices used by writers to convey their evaluation and uncovered the importance of context in determining the use of reporting verbs. They also proposed a categorization system comprising the denotation and evaluative potential dimensions. The denotation dimension consists of textual (e.g., *state, write* and *point out*,), mental (e.g., *believe, think* and *consider),* and research (e.g., *measure, calculate* and *find*) in the "author acts" category; while the "writer acts" involve comparing (e.g., *accord with, correspond to*

and *contrast with*) and theorizing (e.g., *explain, account for,* and *support*). These findings are useful for enabling students to choose suitable reporting verbs and construe appropriated evaluative persona in the process of constructing their academic texts. Similarly, Giannoni (2010) analyzed the value-laden lexis and how the value system is encoded in a wide range of academic disciplines. Using corpus-driven methods, he retrieved the lexis and classified them into goodness (pointing out the positive quality of an entity), size (quantifying an entity like "large" or "small" relying on shared subjective evaluation between authors and readers), novelty (related to the concept of being "new"), and relevance (conceptualised as being important and pertinent) according to the semantic meaning.

In line with this, Charles explored the relationship between grammatical patterns and stance expressions with respect to anaphoric nouns (Charles, 2003), reporting clause (Charles, 2006b), and citation practices (Charles, 2006a) in the postgraduate thesis in the disciplines of physical science (materials science) and social science (political science). Her studies identified great disciplinary differences in terms of stance construction and argument development. Likewise, Francis (1994) investigated nominal labels (nouns or noun phrases that can organize discourse and project stance) in persuasive expository texts and differentiated the advance or prospective labels (Hatzitheodorou, 2008) and retrospective labels used to organize written discourse. Advance labels are concerned with labels referring anaphorically (backward), often fulfilling the function of summarizing, amplifying, or rejecting what has been mentioned in the previous texts. In contrast, retrospective labels point forward to information that has yet to be spoken, linking what is currently talked about with what is to be discussed. Francis' study identified the shared perception encoded by the nominal labels in the expository texts. In a recent study, Jiang and Hyland (2017) explored the rhetorical functions of metadiscursive nouns (such as *fact, analysis,* and *belief*) in research articles across six disciplines. Based on a functional classification, they explored the interactive and interactional functions of these nouns, together with their role in establishing authors' stance and constructing a cohesive information flow in the disciplinary academic discourse.

In conjunction with the above-mentioned studies from the perspective of evaluative lexis, another particular focus has been designated to the syntactic pattern associated with evaluation, the introductory *it* pattern in particular. For example, Hewings and Hewings (2002) compared the pattern used by students and professional writers in the research articles of business studies and MBA dissertations, and the result showed that students employ greater and more overt expressions to persuade readers more explicitly and forcefully than professional writers. Similarly, Larsson (2016) compared the use of the introductory *it* pattern in expert and apprentice learner writing, and found striking differences between the writings of students and professional writers, particularly in terms of the preference and functional use of this pattern. From the perspectives of disciplines and registers, Groom (2005) investigated the use of this pattern in research articles and book reviews within the

disciplines of history and literary criticism, and discovered that the pattern displayed notable register- and discipline-specific features. In an effort to explore the cross-register features, Zhang's (2015) comparison of this pattern in academic and non-academic writing showed that academic writing employs this pattern much more frequently than non-academic writing. More recently, Dong and Jiang (in press) compared the use of this pattern in six registers in the BNC and identified notable register-specific variation in the evaluative semantics, the associated clauses, and the agents of evaluated clauses of this pattern. The findings reveal how evaluation is constructed in different registers, as well as the evaluative values held by the community members. As suggested by these studies, the introductory *it* pattern is a salient pattern denoting evaluative meanings in academic writing, and this pattern has been used differently in student and professional writings, in different disciplines, as well as in academic and non-academic writings.

In another exploration on evaluative patterns, Hyland and Tse (2005) examined the evaluative *that* constructions in the abstracts of L2 post-graduate dissertations (both masters and doctoral) and the RAs in six disciplines. The findings indicate that the *that* structure is a prevalent form in the abstract of the six disciplines. The structure could not only allow writers to organize discourse, but also signpost their attitude towards the information in the text. Another evaluative syntactic pattern study carried out by Groom (2005) made a detailed comparison of the grammatical pattern *it v-link ADJ that- and it v-link ADJ to-inf* in two different genres (RAs and book reviews) and two different disciplinary discourses (history and literary criticism). He found that the two patterns displayed notable genre-specific and discipline-specific features, and their use was motivated by contextual practices. That is, *it seems ADJ that/to-inf* was found to be a dominant pattern in RAs; while *it would be ADJ that/to-inf* was popularly used in book reviews. With regard to the disciplinary variation of *it is ADJ* patterns, the RAs in literary criticism contain previously introduced aspects of the writer's interpretation, while RAs in History tend to emphasize the empirical grounding of claims. The analysis of *it seems ADJ* patterns in the review of the two disciplines show that history reviews are more likely to assess the validity of the interpretation of the reviewed book, while literary criticism used *to*-infinitive pattern more frequently and focused on the highly personalized evaluation of autonomous written arguments. The analysis of *It would be ADJ* patterns indicates that history RAs inclines to persuade readers to reject plausible alternative interpretations, and literary RAs are more likely to use the pattern to persuade readers to accept seemingly far-fetched interpretations.

Overall, previous studies have uncovered a notable association between the structural patterns and semantic properties. Notably, co-text imposes a meaning on a word (Sinclair, 2003) and thus co-text tends to select words that share similar meanings (Hunston & Francis, 2000, p.86). The relationship between semantic sets and patterns, therefore, allows us to view the two concepts in a more integral and dual perspective instead of the traditional dichotomy between the two concepts. In

practice, the recurring patterns could be "parsed to identify common semantic elements" (Fillmore et al., 2003, p.235) and be treated as a search principle to navigate synonyms and antonyms (Renouf, 2009). On the other hand, the common semantic elements could serve as a clue for identifying the similar syntactic patterns or co-text which accompany the use of the vocabulary with similar semantic groups.

2.2.2 Phrases and stance phrases

This section discusses the previous studies related to phrases and stance phrases. The recurring lexical string is viewed as a common feature of English expressions, and writers in linguistics have paid increasing attention to investigating the recursive word strings (e.g., Ädel & Erman, 2012; Biber, Conrad & Cortes, 2004; Cortes & Hardy, 2013; Hyland, 2007, 2008; Gledhill, 2000; Wray, 2002). The use of word strings is often found to be shaped by people's experience in the process of interacting with the world around, and on the other hand, the recurring sequence of words is often characteristic of a situated text (Hoey, 2005). Recurrent lexico-grammatical patterns of academic discourse, therefore, represent the discourse conventions in terms of how the discourse features are embodied by the linguistic realizations.

A plethora of expressions have been used to refer to this notion, such as lexical bundles, phrases, formulaic expressions, chunks, and multiword expressions, to name just a few. Underlying these different labels, writers seem to converge on a similar concept, that is, the combination of word strings which are often found to appear together frequently in a text. In the present study, the term 'phrase' is adopted to refer to "a frequently occurring sequence of words" (Biber et al., 1999, p.264). As previously suggested, phrase is an important component of fluent linguistic production and an essential factor in successful language learning (Cortes, 2004; Hyland, 2007). It has also been identified to play a crucial role in comprehending and producing academic language in disciplinary academic communication (Biber et al., 2004; Hunston, 2002; Hyland, 2012; Römer, 2008).

Since the initial efforts on fixed expressions by Jespersen (1924, p.2), phrases have attracted increasing attention in linguistics. Many linguists have made considerable attempts to define phrases from different perspectives, such as polysystematism (Firth, 1935), language teaching and learning (Pawley & Syder, 1983, p.191), patterned sequences and a relationship with lexis and syntax (Nattinger & DeCarrico, 1992, p.66), and idiomatic nature of language (i.e., the dependent relationship between the vocabulary and the grammatical system) (Sinclair, 1991, p.108).

Essentially, phraseological expressions are generally viewed as a "ubiquitous" (Martínez & Schmitt, 2012, p.300) feature in English use, and an important constituent of the conventions of academic genre. Also, single lexical items regularly appear in combination with particular words and structures; in other words, stance phrases often appears in patterns (Hunston & Francis, 2000, p.51).

Thus, formulaic expressions tend to be a notable characteristic of many discourses (Nattinger & DeCarrico, 1992). From a communicative point of view, Ellis (1996) regarded that the appropriate use of formulaic sequences contributed to attaining fluency in language expressions, while their inappropriate usages are often perceived as a sign of inexperience in academic writing. The recognized importance of formulaic expressions has resulted in a substantial number of studies concentrating on this line (e.g., Biber, Conrad, & Cortes, 2004; Charles, 2013; Cortes, 2004; Hunston, 2011; Hyland, 2008a; Lin, 2009; Sinclair, 1991).

Previous studies in this line have enriched our knowledge of formulaic expressions and have laid the foundation for research into phrases (Biber et al.1999; Halliday, 1993; Hunston, 2002).

A salient trend has focused on identifying formulaic expressions and exploring their functions in disciplinary academic discourse (Biber, Conrad, & Cortes, 2004; Hyland, 2008b; Römer, 2008; Simpson-Vlach & Ellis, 2010). Previous studies have suggested that each discipline is characterized by a particular formulaic sequence (Gledhill, 2000; Hyland, 2008a); and the formulaic expressions, in turn, fulfills different functions in discipline-specific academic discourse (Charles, 2006b; Hewings & Hewings, 2002). Studies in this line mainly have been carried out from the following perspectives: the investigation of one specific discipline, comparisons between two or more disciplines, comparisons between student and professional writers, comparison between writers from different cultural backgrounds, and finally, comparisons between formulaic expressions in different genres or registers.

As regards studies focusing on one specific discipline, a considerable number of studies have been conducted to examine the notable features regarding the disciplinary use of phrases. For example, Gledhill (2000) investigated the phrases in the introduction of RAs in the pharmaceutical sciences. Her study investigated the fixedness and idiosyncratic nature of scientific phrases in this genre, and identified a notable discipline-specific correspondence between lexis and grammar. Marco (2000) also explored collocation frameworks in medical papers, and found that the frameworks like *the ... of, a ... of, and be ... to* enclose restricted sets of lexical items, and that the selection of specific collocates for these frameworks was conditioned by the linguistic conventions of the genre. Similarly, Charles's (2006b) investigation of the phraseological patterning in reporting clauses showed the rhetorical functions of the patterns, as well as the roles the phrases play in assisting learners' information processing.

As for the contrastive studies focusing on the disciplinary variation, substantial attention has been attracted to the comparisons of formulaic expressions between two disciplines and various disciplines. For instance, Cortes (2004) compared the use of phrases in published writings in history and biology and those in students' writing. Based on the four-word lexical bundles in the two disciplines and using

a functional category (Cortes, 2001; Biber, Conrad & Cortes, 2003), she identified notable features in terms of the structure and functions of the four-word strings in the two disciplines. Similarly, Durrant (2009) investigated two-word collocations in the disciplines of arts and humanities, life science, science, engineering. His study identified notable cross-disciplinary differences in the use of the collocations. Focusing on four-word sequences in discipline-specific university students' writing in the British Academic Written English Corpus (BAWE), Durrant (2017) investigated overlapping four-word strings, as well as their similarity and difference. By analyzing the shared and distinctive phrases, he identified the BAWE as comprising of hard (science/technology) and soft (humanities/social sciences) subjects, with life sciences and commerce being intermediate between these two disciplinary groups. His study also discussed the discourse functions of the four-word expressions in these disciplinary categories.

It has also been a notable perspective to compare the formulaic expressions used by student and expert writers in studies centering on formulaic expressions. Previous studies have identified notable linguistic features that can distinguish students' writing from those of expert writers. For instance, Cortes (2004) compared four word lexical bundles in students' writing and published papers in the disciplines of history and biology. The study made a structural and functional distinction of the four-word lexical bundles, such as *the extent to which*, *on the basis of,* and *as a result of*, in the two disciplines. Her comparison revealed notable differences in the frequency and appropriate use of the four-word bundles. In this line, Hyland (2008a) explored four-word bundles between the RAs and doctoral dissertations and masters' theses in the disciplines of electrical engineering, microbiology, business studies, and applied linguistics in terms of forms, structure, and functions. Based on four-word bundles extracted from the corpora, he classified them into research-oriented (location, procedure, quantification, description, topic), text-oriented (transition signals, resultative signals, structuring signals, framing signals), and participant-oriented (stance features and engagements features). His disciplinary comparison of four-word bundles used in different disciplines showed that significantly more research-oriented bundles were used in the hard knowledge fields, probably due to the empirical emphasis. In contrast, the soft disciplines (applied linguistics and business studies) were dominated by text-oriented strings, reflecting that the RAs in soft science usually had more discursive and evaluative patterns of arguments and explicit interpretative persuasion. Hyland also identified great disciplinary variation in students' corpora. Overall, the contrastive perspectives provide evidence for the differing use of the formulaic expressions used by students and expert writers , as well as by non-native speakers and native speakers.

Fourthly, studies in this strand have also concentrated on comparing the formulaic expressions used by non-native and native English speakers. For example, Ädel and Erman (2012) compared the four-word lexical bundles in the writings of Swedish students with that of native speakers. They found

that native speakers had a great variety of lexical bundles, such as unattended 'this' bundles (e.g., *this can be seen; this may be because; this would suggest that*), existential 'there' bundles (i.e., *there appears to be, there has been a*), and hedging bundles (e.g., *or something, sort of, kind of*). Then they compared the use of this pattern with those of the phraseological research tradition in SLA.

A particular focus in this line has also been centered on creating a phrase list for the sake of pedagogical use. For instance, Gardner and Davies (2007) identified the highest frequency phrasal verb patterns from the British National Corpus (BNC) corpus, and analyzed the grammatical structure and semantic function of the phrases. The list of high-frequency verb phrases is of pedagogical value in multi-word research teaching and learning. Likewise, Simpson-Vlach and Ellis (2010) examined frequent recurrent phrases (strings of 3, 4, 5 words) in the corpora of written and spoken language, and created the academic formulas list (AFL) by using statistical approaches and the ratings of experienced EAP teachers, taking into account fixedness, cohesive function and teaching value. The extracted AFL was further categorized by taking into account the predominant pragmatic function into referential, stance, and discourse organizing expressions in both written and spoken academic discourse. A similar effort in Martínez and Schmitt (2012) also generated a phrasal expressions list by extracting the phrases from the BNC, according to the statistical and manual selection criteria. The phrasal expressions list can be drawn upon as a pedagogical instrument for textbook designs and language tests. Meanwhile, Liu's (2012) corpus-based extraction produced a list of phrases, which was referred as multi-word construction (MWC) in his study, from the academic writing sub-corpora of the Corpus of Contemporary American English (COCA) and the BNC. His analysis identified notable features concerning patterns in the phraseological expressions, for example, the pattern ''NP + linguistic action verb (e.g., suggest) + that'' was frequently used in academic discourse. He also made a distinction between American and British English in terms of the phrase usage. For example, phrases like *as long as* and *in general* occur more frequently in American English, while *as far as* and *as a whole* are preferred by the British English.

2.3 Stance expressions from corpus-based and corpus-driven perspectives

It is widely accepted that corpus linguistics has revolutionized traditional linguistics research and fundamentally changed the way data are approached and accessed. The applications of corpus in language analysis are generally regarded as a "newly emerging methodology for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject" (Leech, 1992, p.106). The previous unitization of this tool has suggested that it not only contributes to an efficient access to the large authentic texts, but also facilities data retrieval and quantitative analysis substantially.

Corpora have been applied extensively in the analysis of linguistic items mainly by drawing upon

corpus-based and corpus-driven approaches. Corpus-based analysis departs from predetermined linguistic items and investigates their use based on samples in a corpus. This approach can verify or falsify the linguistic features predetermined prior to the corpus search (Biber, 2012). Corpus-based approaches are characterized by empirical analyses of natural texts, the utilization of large and principled collections of natural texts, and the extensive application of automatic and interactive techniques, etc. Therefore, it has been applied as an effective and reliable approach to investigating linguistic features of discourse, as well as their distribution across various text types (Biber, Conrad, & Reppen, 1998, p.169). The corpus software (such as Antconc, Wordsmith and concordancer) renders it possible to process huge volumes of transcribed texts, and conduct corpus-based search through a corpus. Considerable studies have made use of corpus-based approaches to investigate a particular linguistic feature (e.g., Biber et al., 1999; Conrad & Biber, 2000; Fortanet-Gómez, 2004; Hunston, 2008).

It is generally accepted that corpus-based search enables an efficient search of words and phrases under scrutiny, to identify their detailed features and possible patterns. However, a predetermined selection of words is required while in applying this approach. In cases where the standard for determining the items is not empirically tested, the basis for selecting particular words is unclear. Also, the predetermined searching queries may confine the exploration to only a limited set of linguistic features, thus overlooking the items that do not appear on the list of pre-defined searching queries.

A corpus-driven approach is "more inductive" (Biber, 2012, p.3), as it extracts linguistics items under scrutiny from a corpus. Taking a bottom-up perspective, the corpus-driven approach can determine linguistic items from a corpus for analysis, and thus can "uncover linguistic units that are not detectable using the standard methods of linguistic analysis" (ibid, p.3). This approach has been employed previously, mainly from the perspective of extracting word strings at a pre-set span, such as four-word strings (e.g., Biber, Conrad, & Cortes, 2004; Cortes, 2004; Hyland, 2008a), or 3-5 word strings (e.g., Simpson-Vlach & Ellis, 2010).

However, the use of these two corpus approaches in linguistic analysis varies greatly in terms of tagging of corpus, the automatic search, and the application of quantitative analysis. Some studies have used corpora to search and analyze mainly the frequency of linguistic items, and use statistical tests, such as Log-likelihood test, Chi-square, and mutual information (MI) scores in their analysis (i.e., Ellis, Simpson-Vlach & Maynard, 2008; Simpson-Vlach & Ellis, 2010). In contrast, some studies just used corpora as a data source stored in computers to facilitate access to linguistic items, while no systematic tag or examination of concordance lines was used (e.g., Charles, 2003; Fortanet-Gómez, 2004).

A more advanced use of corpus involves automatic extraction and statistical analysis by using

tagged corpora (e.g., Conrad & Biber, 2000; Gries, 2008). A tag refers to the annotation to mark the specific items, such as grammatical function or part of speech, etc. Currently, automatic tagging has been realized for some items, particularly those can be recognized often by their forms (e.g., part-of-speech (POS) tagging) and some semantic tagging by using for example *Wmatrix* software, but for the complex phenomenon, manual tagging is still required. It is clear that the automatic tagging based on tagged corpus programmes not only facilitates the search but also provides an informed picture regarding the quantitative information and contextualized use of a particular linguistic item.

In the previous studies on stance expressions, corpora have been extensively used (e.g., Charles, 2009; Csomay, 2013; Hu & Cao, 2011; Wharton, 2012). The application of these studies has been mainly conducted by automatic extraction of a certain type of grammatical patterns, such as the introductory *it* pattern (Dong & Jiang, 2019; Groom, 2005; Hewings & Hewings, 2002; Larson, 2016), and the non-continuous word strings (e.g. *a * of* and *are * likely to,* and *it would be * to*) (Römer & O'Donnell, 2010).

However, the identification of stance expressions is "far from straightforward" (Mauranen, 2004, p.209), due to the fact that the lexical items conveying stance meaning are "large and open" (Hunston, 2004, p.157). In addition, stance expression are usually associated with the context in which they appear (Teubert, 2005), so a fully systematic and comprehensive account of stance may require considerable efforts, particularly when it comes to an unexplored disciplinary field or a detailed exploration of items (Hunston, 2011).

Overall, a corpus can provide authentic data to draw upon, and facilitate access to data, thus it is an efficient tool in the discourse-based research. Therefore, these two approaches were employed in this study to address stance phrases in this study: the corpus-driven approach was used in the phrases retrieval, and the corpus-based approach was used when examining the use of stance phrases in terms of frequency, textual colligation and collocation network in the two academic corpora.

2.4 Previous studies on discipline-specific linguistic features

2.4.1 Disciplinary variation of stance expressions

It has generally been established that stance plays an important role in construing a persuasive persona in this academic discourse, as stance reveals how authors present self-positioning in relation to knowledge construction and expected readers (Crismore & Famsworth, 1990; Hyland, 2004). As previously suggested, the disciplinary presence of stance features varies from discipline to discipline (Cortes, 2007; Dressen-Hammouda, 2008; Hardy & Römer, 2013; Hyland, 2004), and are also shaped by the social values and cultural specificities of the academic community (van Dijk, 1997). The identification of the disciplinary uses of stance features, therefore, can reveal conventionalized characteristics of stance positioning and show how academic writers present themselves in relation to

their community members and the constructed texts.

Considerable attention has been attracted to the disciplinary variation in stance uses in cross-disciplinary academic discourse. At a macro-level, studies have adopted a dichotomous division between the soft sciences (e.g., humanity, linguistics, and sociology) and hard sciences (e.g., engineering, medical, and biology). For instance, Hyland (2005a) identified variation in the frequency of stance expressions across eight academic disciplines and found that writers deploy disciplinary-sensitive linguistic resources to facilitate effective communication with anticipated readers. Likewise, Stotesbury's (2003) investigation of abstracts in humanities, social and natural sciences showed that the manifestation of stance expressions of evaluation differs substantially across these disciplines in terms of quantity, type and rhetorical structure. Hyland and Tse (2005) compared the use of phrases containing the evaluative *that* in research article abstracts and L2 masters' and doctoral students' dissertations across six disciplines. Their analysis indicates that more explicit evaluation and interpersonal engagement (e.g., *it is hypothesized that, it is possible that,* and *it seems likely that this*) was used in the soft sciences, while dissertations in the hard sciences and business studies employed more reporting verbs (e.g., *show that, demonstrate that, find that,* and *conclude that*). These studies suggest that writers in the soft sciences are more inclined to employ stance markers to engage readers and convince them to accept their arguments than their peer writers in hard sciences. At the micro-level, cross-disciplinary investigations into the use of stance have identified substantial disciplinary variation and "distinct communicative conventions" (Hyland & Hamp-Lyons, 2002, p.5) in writers' approaches to constructing knowledge and communicating with readers. Such differences were found, for instance, between applied linguistics and psychology (Bruce, 2014), English literature and sociology (Bruce, 2016), politics/international relations and materials science (Charles, 2003), conservation biology and wildlife behaviour (Samraj, 2005).

This disciplinary variation in terms of how stance has been constructed to express authors' attitudes, interpretations and assessment has been examined from the following perspectives: investigating stance feature of one discipline, comparing two or more disciplines, comparisons between student and professional writers, comparisons between writers from different cultural backgrounds, and the comparisons between formulaic expressions in different genres or registers.

Firstly, previous studies have delved into the function and realization of stance features in an individual discipline, such as geology (Dressen, 2003), art history (Tucker, 2003), and economics (Donohue, 2006). Other studies compared the use of stance expressions in a variety of disciplines, such as applied linguistics, biology, business studies, computer science, electrical engineering and public administration (Hyland & Tse, 2005), history and economics (Silver, 2003). Also investigated were: politics, international relations and materials science (Charles, 2003); humanities, social sciences and natural sciences (Stotesbury, 2003), electronics and electrical engineering (Koutsantoni,

2004), history and biology (Cortes, 2004), as well as electrical engineering, microbiology, business studies and applied linguistics (Hyland, 2008a). These studies identified notable expressions used to express authors' stance construction in particular disciplines, as well as the common lexico-grammatical markers and their communicative functions in the discipline-specific academic discourse.

Secondly, the contrastive comparison of stance expressions in different academic disciplines has also attracted substantial attention in applied linguistics. For instance, Hyland (2005a) identified variation in the frequency of stance expressions across eight disciplines and deduced that writers deploy disciplinary-sensitive linguistic resources to facilitate communication with envisaged readers. Similarly, Stotesbury (2003) investigated the expressions of evaluation in abstracts in narrative (humanities or social sciences) and hard science RAs. The results revealed that the abstracts in the humanities, social and natural sciences differ in the manifestation of stance expressions of evaluation, in terms of quantity, type and rhetorical structure. Likewise, Hyland and Tse (2005) compared the use of phrases containing the evaluative *that* in research article abstracts and L2 masters' and doctoral students' dissertations across six disciplines. More explicit evaluation and interpersonal engagement markers (e.g., *it is hypothesized that, it is possible that,* and *it seems likely that this*) were found in the soft sciences, while dissertations in the hard sciences and business studies employed more reporting verbs (e.g., *show that, demonstrate that, find that,* and *conclude that*).

To explore disciplinary features, Hyland (2004) proposed an interactive and interactional model of metadiscourse and investigated the differences in metadiscourse used in doctoral and master dissertations in the disciplines of applied linguistics, public administration, business studies, computer science, electronic engineering, and biology. He also conducted interviews to obtain insiders' perceptions of stance expressions. In addition, he designed a model of interaction, with a distinction between *stance* and *engagement*, by analyzing the published RAs in eight academic disciplines, namely, mechanical engineering, electrical engineering, marketing, philosophy, sociology, applied linguistics, physics, and microbiology. The findings revealed how academic writers in those disciplines employ linguistic markers to express their positions in the socially constructed academic community, to demonstrate their relationship with the contexts, and to engage readers in the discourse-based communications. The study also further discussed the role that metadiscourse plays in exploring rhetorical preferences of academic writers in different discourse communities, and in teaching students in a discipline-specific academic writing instruction.

Dressen (2003) extended the scope by analyzing how geologists construct evaluative evidence by means of implicit persuasive markers. Her analysis identified three functional categories of the implicit evaluative markers, that is, explicit implication in the research account, disguised indications of researcher activity in the field, and the demonstration of research community-based professional expertise. Likewise, Dressen-Hammouda (2008) traced the development of one geology student's

writing after his graduating after his PhD. The study revealed the important role of specialist frames and implicit textual cues in the academic writing process, and concluded that the knowledge and skillful use of "entire semiotic genre chain" (Dressen-Hammouda, 2008, p.249) can prepare students for writing up academic discourse. In addition, Dressen-Hammouda (2014) examined the development of experienced geology writers' voice over the ten-year period after the doctoral dissertation. The study identified a set of indexes concerning geologists' disciplinary novice and practice, and uncovered that writers in the field of geology displayed similar development patterns in their stance construction. The study also suggested that situated genre analysis by taking account of specific sophisticated contexts contributes to dealing with the complexities of voice used by experienced writers to construct their self-representation.

A similar study on authors' stance and identity was carried out by Gillaerts and Van de Velde (2010), who examined the interatcional metadiscourse in RAs abstracts in applied linguistics. A synchronic analysis revealed that RA abstracts and RAs differed greatly in the use of interactional metadiscoursal markers, while a diachronic analysis uncovered notable changes in abstracts development over the past 30 years. These changes were discussed from the perspective of genre, discourse community, and research practice. Likewise, Khedri et al. (2013) compared interactive metadiscourse markers in RAs abstracts in the fields of applied linguistics and economics. The findings revealed that the two disciplines displayed similarities in the interactive metadiscoursal markers, and the abstracts of applied linguistics contains a higher frequency of interactive metadiscourse makers. These findings contribute to our knowledge of how academic writers in the two disciplines construct organized discourse and interact with expected readers. In addition, Koutsantoni (2004) explored appraisal markers used by in RAs in the field of electronic and electrical engineering. Based on a combined taxonomy consisting *attitude markers, certainty markers, emphatics* and *boosters,* Koutsantoni investigated how these disciplinary writers made use of the appraisal resources to position themselves interpersonally and intertextually in their RAs. The results indicate that the appraisal markers assist the disciplinary writers to assert their authority and expertise, gain readers' agreement, and win a community consensus.

Also, Samraj (2008) analyzed the introduction of master's theses in the disciplines of biology, philosophy, and linguistics and carried out interviews with subject specialists. Her study identified disciplinary variation in the introduction of master's theses. The results showed that philosophy students had stronger authorial presence but displayed weaker contextual links to previous studies. Afros and Schryer (2009) investigated the promotional strategies used in the RAs of language and literary studies. The analysis identified the positive and negative evaluation markers employed by writers in language and literary studies to flag their favored and dissenting views. They found that these disciplinary writers used devices, such as *evaluative lexis, coordination, comment clauses,*

personal pronouns, lexical cohesion, and *discourse chunks sequencing* to express their promotion stance. The study also differentiated three types of persuasive strategies, including *logos* (discipline specific arguments), *ethos* (appeals related to trustworthiness and credibility), and *pathos* (appeals derived from a community's values). In addition, the study examined the distribution of different linguistic devices in the abstract, introduction, discussion, and conclusion, and found that the use of those devices varied in different sections of the two disciplinary RAs. To be specific, literary studies contained few overt promotional expressions and tended to downplay their product move (a move used to report main findings or results, the argument) in abstract sections, while the abstract of language studies used the discourse structures like *parallel structure* and linguistic devices like *potential consequences* to promote their articles. In introduction sections, both linguistic and literary studies used evaluative expressions (e.g., *significant* and *fundamental)* to evaluate their work, while linguistic studies employed self-citations to promote their studies. In the discussion section, both linguistics and literary studies quoted and evaluated others' studies to justify or support their own findings. In the conclusion section, literary studies used intensifiers like *absolutely different* to support their claims, while literary writers used more strategies including *lexical cohesion, evaluative lexis, comment clauses,* and *the inclusive pronoun we* to manifest the implication and significance of their studies. Overall, the study found a wide range of discipline-specific promotional markers and the evaluative presence in the disciplines of linguistic and literary studies.

Apart from the aforementioned contrastive studies centered on the discipline-specific features, the comparison between students and expert writers in the same or different disciplines has also attracted considerable attention. Previous studies have identified notable differences between student writings and the writings of expert writers (Hyland & Milton, 1997; Hewings & Hewings, 2002). For example, Hyland and Milton (1997) compared the expressions of doubt and certainty used by Chinese and British students, and found that EFL Chinese students used limited linguistic devices to express doubt and certainty, manifested higher commitment in text description, and had problem in delivering the degree of certainty appropriately. Likewise, Hewings and Hewings (2002) examined the use of a metadiscourse grammatical pattern, anticipatory *it* and extraposed subject like *it is interesting to note*, in published RAs and the dissertations of non-native speakers in the disciplines of business studies and MBA. The findings indicated that student writing used the grammatical pattern much more frequently and expressed greater, more forceful, and more overt persuasion.

In terms of scope, a majority of these aforementioned studies have been concerned with stance expressions in terms of engagement or epistemic modality, and the focus has been centered on two important issues of academic writing: namely, how authors state their propositions and how they define their relations with other members of the community (e.g., through the use of citations). Another main type of study of stance in academic written discourse relates to the characterization of

attitudinal meanings in research writings, especially the values expressed to serve the writer's interpersonal purposes. For example, Thetela (1997) outlined the common types of attitudes expressed in RAs towards two kinds of entity (research-oriented entities and topic-oriented entities); Hood (2006) investigated stance prosodies (the accumulative effect of stances that tend to cluster with each other) and the important values that contribute to authors' successful persuasion.

In addition, authors' presence has also been examined together with other linguistic features, such as move and the macro-structure. For instance, Del Saz Rubio (2011) investigated the metadiscourse features together with the macro-structure in the introduction of agriculture RAs. The findings revealed some commonly used interactive and international devices used to express authors' argumentation in projecting themselves in the context and aligning with the academic conventions. Nathan (2013) investigated authors' stance markers and the rhetorical moves in the business case report. He analyzed case reports in the writings of marketing and marketing management written by native and non-native postgraduates and found common rhetorical devices (i.e., the presence of explicit structure, impersonal style, and business specialism-dependent lexis) and the obligatory rhetorical moves. Lim's (2012) study focused on the move of establishing research niches in the introduction sections of research papers in the field of management. The study uncovered notable features in the linguistic devices, such as contrastive adverbs, adversative prepositional phrases, and pre-modified adjective phrases used to denote the insufficiency in this field. In a more recent study, Liu and Buckingham (2018) examined the interaction between metadiscourse markers and rhetorical moves in the discussion section of applied linguistics RAs. The analysis indicated a close connection between metadiscsourse markers used to construct rhetorical moves in discussion. Liu and Buckingham also identified notable differences in the distribution of the interpersonal and textual markers: more textual markers are found in move with argumentative or persuasive functions, while more interpersonal markers show a high occurrence with sequencing or linking moves.

Taken together, as shown by the considerable number of studies discussed above, studies on stance feature in academic discourse have become increasingly interdisciplinary and cross-disciplinary. As has been suggested, stance features are restricted to particular genres, and thus displays considerable variation across disciplines (Hyland, 2004, 2015; Samraj, 2008). The discipline-specific academic writers members in a differing academic community are found to have developed their own preference for the discipline-specific stance construction in the course of projecting themselves and communicating with the expected readers.

Despite these extensive studies, our understanding of the cross-disciplinary particularities of stance features remains incomplete, with numerous disciplinary domains uncharted. Therefore, this study delved into stance features in the academic disciplines of agriculture and economics. The following section will elaborate studies on these two disciplines, as well as the rationale for examining

these two disciplines.

2.4.2 Studies focusing on agriculture and economics

Among the studies on disciplinary variation of academic discourse, economics and agriculture are two particular disciplines that have been focused on by researchers in applied linguistics. This section reviews the previous studies focusing on the linguistic features in these two disciplines.

(1) Economics

First of all, considerable attention has been attracted to the linguistic features used to convey stance in the discipline of economics. Previous work has concentrated on rhetorical features, such as the argumentative development, move framework in academic writing, and interpersonal stancetaking markers. For instance, Bondi (1999) examined the language variation in the research papers and abstracts, textbooks and newspaper articles in economics from the perspective of systemic linguistics and rhetorical analysis. The findings revealed notable features that economics writers use to develop arguments and persuade writers to accept their arguments. Bondi found that economics writers had a preference for engaging readers directly and positioning them as participants, and used an anonymous third person to maintain a certain distance with readers. She also found that economics writers tended to take a dialogic position to interact with readers by using metadiscourse signals and specific move sequences.

Similarly, Liu and Lim (2014) investigated the evaluative stance by examining how economics writers evaluate their own empirical research in RAs. Making use of the connection between evaluative markers and the Swalesian move-step analytical framework, the study found that self-evaluation is a principal move in economics research reports and the move interacts closely with the evaluative markers in interpreting the results for readers. Apart from the focus on evaluative stance, another aspect of stance, namely hedging in the economics RAs, has also attracted writers' attention. For instance, Donohue (2006) examined the hedging devices used in English economic forecast texts. The study identified that hedging devices perform the following three functions, conversation (concerned with the dialogue that forecasters have with readers, such as "the likelihood is that"), discrimination (how forecasters foreground and back ground predictions, such as 'should be well able to 'may be'), and organisation (the way writers construct a text). The study also further suggested that hedging devices are associated with the concept of 'caution' and 'degree of commitment' and serve as a reductive strategy of 'face-saving'.

Meanwhile, Lancaster (2011) investigated the interpersonal stancetaking markers in argumentative term papers written by upper-level undergraduates majoring in economics. By drawing on the *Engagement* classification within the Appraisal framework (Martin & White, 2005), the study compared the stance markers in high- and low-graded papers written in two distinct undergraduate

courses. The study found that the argumentative texts in economics contain more disclaim expressions, including denial (e.g., *not, never,* and *didn't*) and counter argument (*yet, although,* and *but*) than proclaim markers, including concur (e.g., *naturally, of course and admittedly*), pronouncement (e.g., *I contend, the facts of the matter ar*e), and endorsement (e.g., *the reports demonstrate*). The comparison showed that high-grade papers contain a significantly larger number and a wider range of stance expressions than the low-grade ones. Based on the findings, Lancaster (2011) suggested incorporating the academic stance into the academic writing instruction to enable students to align themselves with academic conventional practices.

The linguistic features in economics have also been examined from a viewpoint of interlanguage comparison. For instance, Mauranen (1993) contrasted the cultural differences in using 'metatext' (a concept similar to 'metadiscourse') between the published research articles in economics written by Anglo-American English writers and Finnish writers. The study found that English writers employed more explicit presence and used more explicit textual rhetorical devices, while Finnish writers used little explicit metalanguage to organize the text and orient readers. Her study indicated that Anglo-American writers had a stronger reader-oriented awareness, more explicit textual features and a positive notion of politeness, while the Finnish writers had a greater tendency to show negative politeness and implicitness in their writing. Similarly, Shaw (2004) compared the evaluations used in published articles in applied economics written by Danish writers in Danish, Danish writers in English, and English writers in English. The study found that Danish writers had less promotional evaluation in research articles than their English peers. In the same vein, Moreno (2004) compared the preferred uses of rhetorical structure 'premise–conclusion sequence intersentential coherence' (PQ metatext), a construct indicating the premise and conclusion relationship between two related semantic units (i.e., *thus, this means that*) in the research articles of business and economics. The study found that both English and Spanish academic writers in these disciplines used more implicit labels of the PQ metatext, while English writers were found to use more non-metalinguistic labels (like an argument, a point, or a statement) than their Spanish counterparts.

In conjunction with the cross-cultural comparison, the linguistic features used in economics have also been compared with those in other disciplines. For instance, Dahl (2004) investigated the manifestation of writers' presence in three languages, English, French and Norwegian within three disciplines, economics, linguistics and medicine. The results indicated that there were more meta-discourse features in the English and Norwegian than their French counterparts, which indicate that English and Norwegian represent writer responsible cultures, while French was representative of a reader responsible culture. The study also found that economics and linguistics had less formalised structures in organizing the research articles, and the writers in these two disciplines were more likely to present their findings through argumentation in texts. Also, Khedri et al. (2013) compared the

interactive metadiscourse markers used in RA abstracts in economics and applied linguistics. The study identified great differences in the interactive metadiscourse markers of the two disciplines. The economics abstracts contained more transitive devices (e.g., *furthermore, equally, in the same way*) and more logical relations (e.g., *however, but*) than those in applied linguistics to facilitate readers' understanding of the statements delivered in the texts. In contrast, writers in applied linguistics were found to use more endophoric markers (e.g., *this study, this paper, this research*) in their text construction.

Likewise, Yeo and Ting (2014) investigated the personal pronouns (like *we, I, you*), one of the components of stance expressions, in lecture introductions in the arts and science disciplines. According to their division, economics is one discipline within the macro discipline category of arts. The study identified the discourse functions of the personal pronouns, namely, activating prior knowledge and giving instructions or announcements in order to enhance students' engagement. The result also showed that science lectures contained more *you* referring to the audience, *we* referring to *I*, and *we* referring to both the lecturers and the audience, while arts lecture introductions used *you* as a generalised concept (e.g., *you see, you know*).

(2) Agriculture

In contrast to the extensive studies on economics, the review of previous literature shows that only a few studies have been conducted to explore linguistic features in agriculture. Previous studies on this academic discipline have mainly concentrated on the linguistics features from the perspective of academic words, rhetorical moves and quoting stance. For instance, Martínez et al. (2009) examined RAs in agriculture and identified academic wordlists with meanings specific to agriculture RAs by comparing them to the Academic Word List (Coxhead, 2000). The study found that the meanings of some words in the agriculture corpus are related to the academic genre and the discipline of agriculture. Some words (e.g., *used, treatment, high, experiments*) from the GSL (General Service List) (West, 1953) are used more frequently and are typical academic vocabulary in general.

Similarly, Del Saz Rubio (2011) analysed the macro-structure and metadiscourse features of the research papers in agriculture academic papers by using Swales's (1990, 2004) Create a Research Space model (CARS) and Hyland's (2005) metadiscourse classification scheme. The analysis showed that the rhetorical patterns used in agriculture resemble the Swales CARS models, and the corpus contained a common combinational pattern of the steps for each move in the CARS models. For example, topic generalizations within move 1 are often combined with the step of "establish the territory"; niche establishment in move 2 is often accompanied by the indication of research gap; occupying the niche in move 3 is often accompanied by outlining the purpose or summarizing the method. With regard to metadiscoursal markers, the study identified the rhetorical moves containing

a high frequency of metadiscoursal markers that are commonly used in agriculture discourse. It was also found that evidential (e.g., *According to X; Z state*), transition markers (e.g., *in addition, but, thus*), and code glosses (e.g., *namely, such as; in other words*) mainly were frequently used in this discipline, and the interactional metadiscourse is mainly expressed by hedges (e.g., *might; perhaps; possibly; about*) and boosters (*in fact; definitely; it is clear that*).

Thompson and Tribble (2001) examined the citation styles, one important component of stance expressions, in the doctoral theses in the field of agricultural botany and agricultural economics in the discipline of agriculture. The study identified the sub-disciplinary variation in terms of the citation practices. To be specific, writers in agricultural botany use more non-integral source, such as *"involves judgements based on textual factors such as the narrative point of view (Fowler 1986; Simpson 1990)",* an example cited from Thompson & Tribble (2001), and identification types (e.g*., "A simulation model has therefore been developed to incorporate all the important features in the population dynamics [Potts 1980],"*an example cited from Thompson & Tribble (2001)); while agricultural economists use more integral naming citations (e.g., *"..For Seers, 'Development is inevitably a normative concept' ... (Seers,1972, p 22),"* an example cited from Thompson & Tribble (2001)). With regard to the distribution of the citations in the different sections of the theses, the study found that there is a high density of citations in introduction and discussion, and low density in the methods and results sections. They also explored the pattern of citation types, and found that the citation pattern "in Author (year)" is much more frequent in agricultural economics theses, while agricultural botany theses contain more "of Author (year)" and "by Author (year)" pattern. From these findings, they inferred that agricultural economics writers are more concerned with making reference to others' texts and concepts, while agricultural botany writers have a preference for other writers' activities and techniques. The study also found that the novice writers in the two sub-disciplines of agriculture use a limited range of citation types, and the study suggested a need to expose students to a wide range of linguistic choices in the EAP instruction.

2.5 Previous studies on textual colligation

Textual colligation concerns the phenomenon that linguistic items tend to occur at a certain textual position (Hoey, 2004, 2005). Words or phrases are "primed to either occupy or avoid certain recognised discourse positions" (Hoey & Donnell, 2015, p.125), specifically "to occur (or to avoid occurring) in the beginning or end of independently recognised discourse units, e.g. the sentence, the paragraph, the speech turn" (Hoey, 2005, p.115). One notable feature of textual colligation relates to the genre-specific features, as defined by Hoey "textual colligation may only or especially be operative in texts of a particular type of genre or designed for a particular community of users, e.g., academic papers" (Hoey, 2004, p.187). Studies carried in this strand have suggested that the textual positioning

for a lexis/ phrase primed to align with the conventionalized practice of a genre can contribute to the naturalness of a text advocated by Sinclair (Mahlberg & O'Donnell, 2008; O'Donnell et al., 2012).

Corpus analyses focusing on the textual positions of particular linguistic items have mainly centered on the domains of written media texts and academic discourse. Studies in this strand often employed a keyword extraction approach to identifying linguistic units of analysis. For instance, Mahlberg and O'Donnell (2008) investigated the textual colligation of keywords in hard news stories, and identified differences in the collocational patterns of the keyword *fresh* in text initial and non-initial positions in news reporting. Also based on pre-determined keywords, O'Donnell et al. (2012) investigated the continuous and non-continuous patterns that co-occurred with the keyword *fresh.* Expressions located in the text-initial position were found to constitute an essential section of the stories in news and play a central role in the organization of the news story.

Key phrases in newspaper texts have also been identified to display particular distribution patterns across a text. Hoey and O'Donnell (2008), for instance, examined the distribution of key phrases such as *yesterday it was announced that* and the shell noun (or general noun) phrase *a move* and *the move*. They found that this shell noun may occur at some distance from its antecedent, the nucleus, and the intervening text did not need to contain additional information to maintain the clear link with the antecedent. In addition to creating an anaphoric relationship, this word combination simultaneously signaled a shift in topic, or a shift to the next step of argumentation. Relatedly, an analysis of news reporting identified the textual priming function of the phrase *according to a.* Hoey and Donnell (2015) found that this phrase tends to avoid sentence-initial and paragraph-initial positions, that is, the theme position, or may occur in the second half of paragraph initial sentences, following the theme. The features of the textual distributions of these linguistic items indicate how authors use textual position to convey key information and conduct effective communication with readers.

The textual position of phrases has also been examined in academic texts. Typical studies in this line have used a corpus of student papers. In a relatively early study on the textual distribution of phrases, Römer and O'Donnell (2010) examined the distributions of n-grams (fixed phrases) and P-frames (non-contiguous phrases, like *in the * of*) in the positions of sentence, paragraph and text[2], using the Michigan Corpus of Upper-level Student Papers (MICUSP). Considerable variation was found in the position with regard to n-grams and P-frames. Some phrases, such as *it is * that* and *it would be * to* favour sentence-initial position, while some phrases (e.g., *as * in figure*) have a strong preference for sentence-final positions. Some differences were also found with respect to the n-grams and the P-frame at the level of paragraph (e.g., *in order to * the* and *are * likely to)* and text (e.g., *like*

[2] The authors used the following division for their units of analysis: 25% beginning, 50% middle, and 25% end.

it is that* and *it would be * to).*

The concept of textual colligation has also been investigated in the domain of academic writing. Using the British Academic Written English Corpus (BAWE), Thompson (2014) examined the occurrence of single and multiword items at paragraph level. While his results partly supported the existence of textual colligation, this evidence was not uniform. An important contribution of his study was the consideration of the function of key phrases. Discerning phrases such as *this essay* and *one of the* tended to appear in paragraph initial position, Thompson recommended that analyses of textual colligation take into account the semantic or rhetorical function of linguistic units.

As shown in the review of the discipline-specific stance features (Section 2.4.1), academic texts have been found to figure notable disciplinary variation with regard to stance construction. However, an important aspect of stance, namely, textual colligation remains largely unexplored in studies on academic texts. Textual colligation, according to Hoey (2005, p.13), is domain specific. It thus seems reasonable to postulate that variation may be found in the text contexts typical of stance expressions in the two disciplines analyzed in this study (agriculture and economics).

2.6 Previous studies on collocation networks

The co-occurrence of linguistic markers has been a particular research focus in applied linguistics. The exploration of the co-occurrence is motivated by the assumption that there is a complementary relationship between form and meaning, and meanings construction is characterized by different structural configurations (Sinclair, 1991). In this vein, considerable attention has been paid to examining the co-occurrence of different linguistic markers. The knowledge concerning the co-occurrence allows us to explore the recursive patterns and structures of language, and thus the correlation between syntactic patterns and the semantic meanings.

Before introducing the previous studies and discussing the present work, it is necessary to differentiate two similar concepts, i.e., 'collocate', 'collocation'. Collocate refers to "a word which occurs in close proximity to a word under investigation is called a collocate of it" (Sinclair, 1991, p.170). Collocation, on the other hand, is concerned with the occurrences of "two or more words within a short space of each other in a text. The unusual measure of proximity is a maximum of four words intervening." (Sinclair, 1991, p.170).

Another similar concept is collocation networks, which was originally proposed by Phillips (1989) to denote the contextual interconnectedness of lexical units. With a focus on the "lexical connections in discourse" (Brezina et al., 2015, p.142), this concept permits a relatively extensive search of the possible co-selecting continuous and discontinuous collocates of a linguistic item within a pre-defined span around a node word or phrase.

The contextual relationship between linguistic items has been a particular focus in corpus linguistics (Brezina et al., 2015; Gablasova, Brezina, & McEnery, 2017; Hunston, 2011; Xiao & McEnery, 2006). For instance, Xiao and McEnery (2006) examined the collocational relationships and semantic prosody of near-synonyms (i.e., *cause, arouse, lead to, result in/ from, give rise to,* and *bring about*) in English and Chinese. Similarly, Hunston (2011) identified a number of collocates for evaluative words and phrases.

Advancing this line of research, the advent of GraphColl (Brezina et al., 2015) increased the capacity for the quantitative analysis of phrases. This software enables an exploration of the semantic relatedness between one linguistic item and its surrounding linguistic contexts, and measures the degree of relatedness between different linguistic items. With this software, Brezina (2016) investigated the collocation networks of three content words *god*, *love*, and *president* in a Yahoo online forum. Gablasova, et al. (2017) examined the collocational patterns of three grammatical structures, namely, verb + complementation (make + sure/decision/point), adjective + noun (human + beings/rights/nature), and adverb + adjective (vitally/very/really + important), in written and spoken registers. Such studies have provided empirical evidence for the complex collocating relationships between linguistic items and their co-occurring co-texts; however, to date, no such work has been undertaken on a set of linguistic features performing similar communicative functions. Therefore, this study aims to apply this theoretical framework and examine the collocation networks of stance phrases in the disciplinary academic discourse.

2.7 Summary

To summarize, previous literature in this strand has suggested stance entails authors' positioning and constitutes an important rhetorical device in academic discourse to express authors' involvement and manifest their persuasive efforts. Previous studies have yielded fruitful findings regarding stance expressions in cross-disciplinary academic discourse. Functionally, that the appropriate use of stance expressions could show authors' knowledge of the convention in the academic community, as well as serving as an effective means by which authors convey their meanings and deliver their knowledge and arguments. Considering the diversity of disciplinary practices and the complexity of stance expressions, it is not surprising to find that stance expressions in agriculture and economics academic discourse remain underexplored.

2.8 Research questions

This study seeks to investigate stance phrases with respect to their occurrence, syntactic features, textual colligation and their collocation network in the academic disciplines of agriculture and economics, with the aim of uncovering how authors project themselves in cross-disciplinary academic discourse. To be specific, the study sets out to address the following questions:

(1) Type and category of stance phrases

(a) What stance phrases are employed in the agriculture and economics academic corpora?

(b) Re-categorizing the stance framework by the metafunctions of SFL

(2) General use of stance phrases

(a) What is the frequency information of the stance phrases in the agriculture and economics academic corpora?

(b) What disciplinary variation can be found in the occurrences of stance phrases?

(3) Textual colligation of stance phrases

(a) What is the textual colligation of stance phrases in the agriculture and economics corpora?

(b) What disciplinary variation can be found in the textual colligation of stance phrases?

(4) Collocation networks of stance phrases

(a) What are the collocation networks of stance phrases in the agriculture and economics academic corpora?

(b) What disciplinary variation can be found in the collocation networks of stance phrases?

The following chapter discusses the research methods used to address these questions and methods used in building the academic disciplinary corpora, extracting stance phrases from the corpora, and comparing the cross-corpora differences.

Chapter 3 Methodology

This study adopted a combination of corpus-driven and corpus-based approaches in order to obtain a relatively more comprehensive picture of stance phrases in the two academic disciplines under scrutiny. The combination of these two approaches "offers a powerful tool for gaining insight into the way in which texts are constructed and the means by which their purposes are achieved" (Charles, 2009, p.156). This chapter mainly elaborates the corpus-driven approach to extracting stance phrases for analysis, while the specific corpus-based approaches used to examine the different aspects of stance phrases (including the frequency, textual colligation, and collocation network) are described in Chapters 5, 6 and 7 respectively. The structure of this chapter is organized as follows. Section 3.1 spells out the corpora built in this study for analysis. Section 3.2 presents the corpus-driven approach used for stance phrases retrieval, which involves a four-step procedure (including extracting, filtering, manual selection, and collapsing) which are described in detail in Sections 3.2.1 to 3.2.4 respectively). Section 3.3 includes a description of the reliability test used to measure the manual identification and categorization of stance phrases, and the statistical tests (Chi-square, Log-likelihood and effect size) used to compare the uses of stance phrases in the two disciplines. Finally, Section 3.4 provides an overview of the research methods used to address the research questions, and theoretical underpinning for the questions 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 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. The texts were selected from the reading lists which were recommended by 25 and 20 experienced writers in the disciplines of agriculture and economics respectively and intended for 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 seed studies in the discipline of agriculture; and banking studies, corporate economics, investment management, accounting, and public economics within the discipline of economics.

The recommended articles range across 133 journals and 128 journals in agriculture and economics respectively. All the journals are peer-reviewed and indexed by SCI (Science Citation Index) or SSCI (Social Science Citation Index). The specific information regarding the journals and

number of papers selected from the journals are presented in Appendix A. 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 removal of the non-text sections, namely, tables, figures, references, and acknowledgements.

Most of the selected RAs were written by native English speakers while a small number of the texts were written by non-native English speakers. However, it is necessary to point out that the distinction between native and non-native English speakers was not taken into account in this study for the following reasons. Firstly, most of the RAs in the two disciplines were written by more than one author from different cultural backgrounds. As a result of the co-authorship and proofreading system, the interlanguage feature assumed to be shared by nonnative writers may not be notable in RAs. In addition, it is often the case that before being published, the papers often go through a strict scrutiny and peer review procedure. In this sense, the language use of a published RA is shaped by literacy brokers during the writing, revision, peer review and final editing stages (Lillis & Curry, 2010). The writing style of RAs thus reflects a form of language use shared within a community of practice rather than a personal style reflective of an author's linguistic and cultural background.

3.2 Research procedure

This section explicates the four-step procedure followed in extracting stance phrases: data extraction, filtering, collapsing and manual identification and categorization. This procedure is illustrated in Figure 1. The detailed description of the procedure is presented in the following sub-sections.

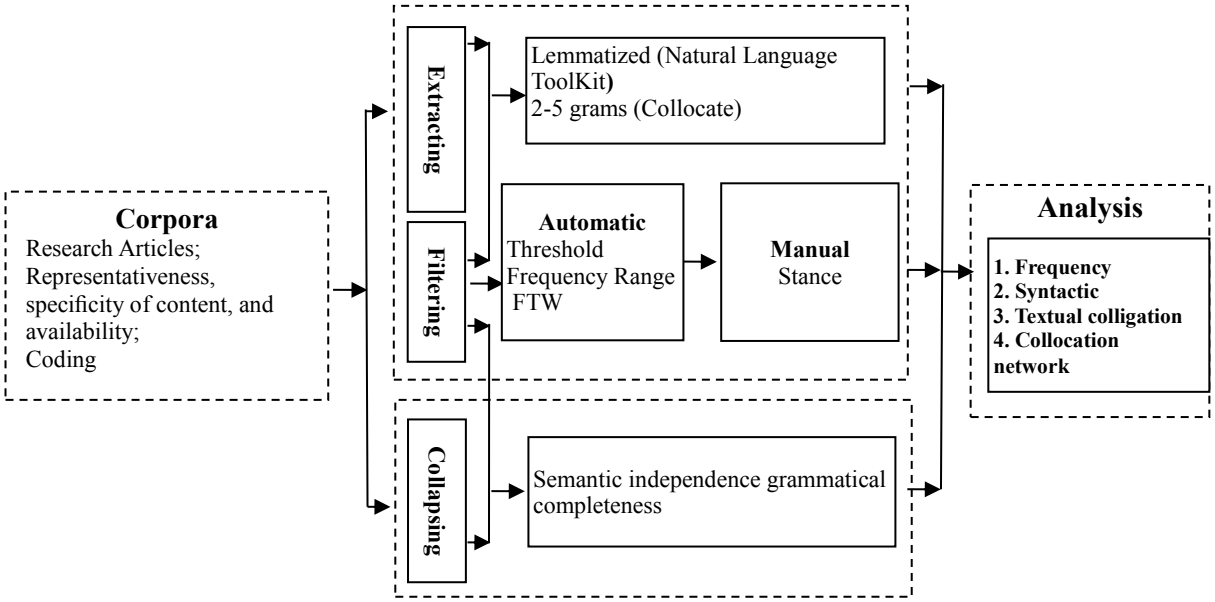

Figure 1 Diagram of the Methodology in the Present Study

3.2.1 Extracting

In retrieving stance expressions, an analysis based on single words is unlikely to lead to satisfactory results (Römer, 2008). Thus, 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 suggested 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 patterns of different forms of a lemma (Hoey, 2005; Stubbs, 2001), lemmatization was carried out in the process of phrase extraction for the following two reasons. Firstly, the phrases retrieved based on the lemmatized texts include some phrases which may otherwise be excluded due to their low frequencies in the concordance list. This is because the lemmatization takes into account the frequency of a particular phrase entry even if some of its grammatical forms do not occur very often in texts. In addition, lemmatization yields a single MI score for a particular stance phrase lemma regardless of the different grammatical forms, instead of different MI scores to each inflectional form. For example, the lemmatized form for *considered as/considering as/ consider as* produced one MI value rather than three. This facilitates the processing of the quantitative data with little loss of information.

The lemmatization was applied to the two corpora at the word-level with the Natural Language ToolKit (Bird, 2006). The lemmatization was conducted in terms of syntactic variants of verbs and nouns; for example, *am, is, are, was, were, and been* were lemmatized into *be*; *refers, referred, referring* were lemmatized into *refer*; and the plural forms of nouns like *articles* were lemmatized into their singular form (i.e., *article)*. In this study, the lemmatization was considered only in the process of retrieving phrases. During the analysis of stance phrases, the original forms of the formulaic stance expressions were considered in order to examine the phrases in context and determine their functional coding.

Then the lemmatized corpora were loaded respectively into Collocate (Barlow, 2004) to retrieve the phrases. For the extraction of phrases, 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 (mutual information) scores. MI score is commonly used to measure the mutual dependence of the words in a word string (Oakes, 1998; Manning & Schütze, 1999). The MI metric reflects the strength of the mutual dependence between the two linguistic items in a word string, and

it gives high scores to fixed phrases (namely, phrases whose constituents have a strong association and appear in a certain order). The higher MI score a word string has, the more likely that it is a fixed phrase, and vice versa. As a popular common metric used to "evaluate the collocational strength of association between words" (Biber, 2009, p.287), MI score has been employed in a number of studies on formulaic sequences (e.g., Ellis et al., 2008; Simpson-Vlach & Ellis, 2010). In this study, this parameter was mainly used in calculating the formulas teaching worth (FTW), a parameter used to rank the stance phrases (the detailed description of this parameter is presented in the section below).

3.2.2 Filtering

The retrieved phrases in the above extraction step were further filtered in this step. The filtering step involves an automatic selection using the parameters, including threshold value for frequency, range, and FTW. The combination was chosen in order to reveal more of the "inherent nature of evaluative items" (Stotesbury, 2003, p.331). In the automatic selection, the threshold for phrase inclusion 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 5%; that is, the phrases selected had to occur in at least 5% of the articles in each corpus, in order to both "guard against idiosyncratic uses by individual authors" (Biber, Conrad & Cortes, 2004, p.376) and encompass a broad range of formulaic expressions to analyze.

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 a parameter in this study to provide evidence for the teaching value of the formulaic expressions. FTW is a new yardstick to measure the teaching value of phrases to select pedagogically useful n-grams. FTW is developed based on stratified random sampling in terms of n-gram length, frequency band, the band of MI score, and the rating by experienced EAP teachers from the perspective of fixedness, cohesive function, and teaching value metric (Simpson-Vlach & Ellis, 2010). Considering that FTW is a metric produced by Simpson-Vlach and Ellis (2010), this study just used it as a parameter, and did not include the process of investigating the judgements of experts on the teaching value. Specifically, I used the established formulas created, namely FTW=β 0.56 MI+ β 0.31 frequency, by weighing MI score and frequency. In this study, the top 5,000 phrases according to FTW rank were included for the analysis, for the purpose of including a relatively large number of samples for analysis.

3.2.3 Manual selection and categorization

Based on the stance phrase obtained after the collapsing step, I carried out a manual selection to

determine whether a phrase is related to a stance meaning. The principle followed in the semantic identification was the definition of stance (Hyland, 1999b, p.101). That is, to consider whether or not the phrases express authors' presence in the textual organization to "communicate their integrity, credibility, involvement, and a relationship to their subject matter and their readers." (Hyland, 1999b, p.101). Then I classified these stance phrases into four functional categories following a synthetic taxonomy composed of cognitive, attitude, hedges and reference stance, based on the metafunctions of SFL (discussed in Section 1.2.1). The reliability test for the manual selection and categorization is presented in Section 3.3.1. a detailed discussion of the stance categories is presented in Section 4.3.

3.2.4 Collapsing

This step involves merging the overlapping phrases obtained in the previous selection step. As it may be assumed, the focus on 2- to 5-grams may result in some overlapping ones, such as *possible to* (2 grams), *is possible to* (3-grams), and *it is possible to* (3-grams), which were further dealt with in this step. The phrase collapsing procedure follows the principles of semantic independence and grammatical completeness for the sake of pedagogical value. While the majority of the stance phrases were treated as what they are, a minority of semantically dependent phrases were merged in favor of the longer semantic independent strings. For example, *possible to* and *is possible to* were merged into the longer strings*, it is possible to*, because the latter carries relatively independent meaning and possesses a higher pedagogical value. As for grammatically different structures, such as, *it is possible that* and *it is possible to,* both were included in the phrases list, because the two phrases are associated with different grammatical patterns and thus it is pedagogically important to distinguish them.

These phrases were analyzed in Chapters 5, 6 and 7 from the perspective of frequency information, textual colligation and the collocation network respectively. (The detailed description of these concepts and the findings are presented in each chapter.)

3.3 Analysis

3.3.1 Reliability test of the manual analysis

As mentioned in Section 3.2.3, manual coding was carried out to determine whether the semantic meaning of a phrase is related to stance meaning and to categorize them functionally. To measure the reliability of the manual selection and classification, a reliability test was conducted. In this step, three experts in linguistics, agriculture, and economics were consulted to determine the meanings and functional categories of the phrases respectively. All nine experts were employed in a research active lecturing positions, experienced writers with doctorate 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 30% of the stance phrases (approximately 60 stance phrases) and their concordance lines showing the contextualized uses of the stance phrases, as

well as the categorization of the stance phrases. 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 related to authors' self- projection; (2) whether they thought the phrases could be subsumed into particular categories.

As indicated by the statistical test of Kendall's coefficient of consistency[3], *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 of the stance phrases' relatedness and functional categorization. (The classification scheme is discussed in detail in Chapter 4.)

3.3.2 Statistical analysis

In analyzing the stance phrases used in the two corpora, both quantitative and qualitative methods were employed. The quantitative analysis involves a statistical comparison of frequency, MI, and FTW value of the stance phrases. Frequency is regarded as an effective indicator in measuring the usefulness of words (e.g., Leech et al. 2001; Martínez & Schmitt, 2012; Nation, 2001). Given that the two corpora differ in size, the frequency of the phrases was normalized in the statistical analysis.

In order to explore disciplinary differences in the use of stance expressions, the raw and normalized frequencies of the stance phrases were compared using the Log-Likelihood (LL) statistic to measure if there is any significant difference in the use of stance phrases in the two corpora. The LL statistic has been commonly employed to assess the statistical significance of differences in words or phrases between different corpora (Botley, 2006; Jurafsky & Martin, 2000; Oakes, 1998; Rayson & Garside, 2000; Simpson-Vlach & Ellis, 2010). In calculating the LL ratio, I employed the Log-likelihood ratio calculator (Xu, 2009) which uses the following equation:

$$\mathrm{LL} = 2 \times \left(Freq1 \times \ln\left(\frac{Freq1}{Exfreq1}\right) + Freq2 \times \ln\left(\frac{Freq2}{Exfreq2}\right)\right)$$

where Freq_i represents the frequency in the *ith* corpora. Exfreq_i represents the expected frequency in the *ith* corpora.

Previous studies have suggested that Chi-square is a useful measure of the statistical differences between two variables (Mu et al., 2015; Ädel, 2014). Given that the frequency data in the present study are categorical data, this study employed Pearson's Chi-square test (Pearson, 1900) to measure the differences in the uses of stance expressions in the two corpora. The following equation was used to calculate the Chi-square in this study.

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

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where χ^2 =Pearson's cumulative test statistic, which asymptotically approaches a χ^2 distribution; O_i= the number of observations of type I; $E_i = N_{p_i}$= the expected (theoretical) frequency of type i, asserted by the null hypothesis that the fraction of type i in the population is p_i; n= the number of cells in the table. (Pearson, 1900, p.157–175)

The reason for taking into account both the LL ratio and the Chi-square score in this study is to contribute to the reliability of the results.

As *p*-values may be affected by sample sizes, this study used an additional statistic, namely effect size, to measure the degree of significant differences in the two disciplines. Effect size measures "the magnitude of the impact of the independent variable on the dependent variable" (Kline, 2004, p.97) and the degree of "the difference between two groups or the relationship between two variables" (Durlak, 2009, p.917). Unlike *p*-values, they are not affected by sample sizes. Therefore, this study used effect size to gain information regarding the magnitude of the significant differences in the use of stance phrases, including their frequency, textual colligation and collocation networks, in the agriculture and economics corpora, and to identify the "size of the difference" (Larson-Hall, 2010, p.114).

As suggested previously, effect size is a robust approach to calculating the degree of difference between variables and has attracted attention from writers in applied linguistics (Mackey & Goo, 2007; Plonsky & Oswald, 2014; Li, 2010). Considering that the data in the study are categorical data (data that could be divided into groups, such as race, sex, age group, and educational level), the study adopted Cohen's *w* (Cohen, 1998, p.157) to calculate the effect size for the Chi-square test. The specific formula for the calculation is presented below:

$$\omega = \sqrt{\frac{\chi^2}{n}}$$

where *n* refers to sample sizes of the phrases in the two corpora. In interpreting the effect size, the study followed Cohen's (1998) magnitude guidelines, that is, a value of 0.1 is considered a small effect, 0.3 a medium effect and 0.5 a large effect.

3.4 An overview of the methods used to address the research questions

This section presents an overview of the methods used to address the research questions. Figure 2 illustrates the research questions, the theoretical frameworks and the methods used to retrieve the different aspects of the stance phrases. The overarching question underpinning this study is how stance phrases are used by academic writers in the disciplines of agriculture and economics. The uses of

stance phrases are examined in terms of frequency, textual colligation (concerning the textual position of the stance phrases), and collocation network (related to the quantitative collocational relationship between the stance phrases with the surrounding contexts). The theoretical framework and the methods used to resolve these research questions are briefly discussed below. The methods used to retrieve the textual colligational and collocation networks of stance phrases are presented in Chapters 6 and 7 respectively.

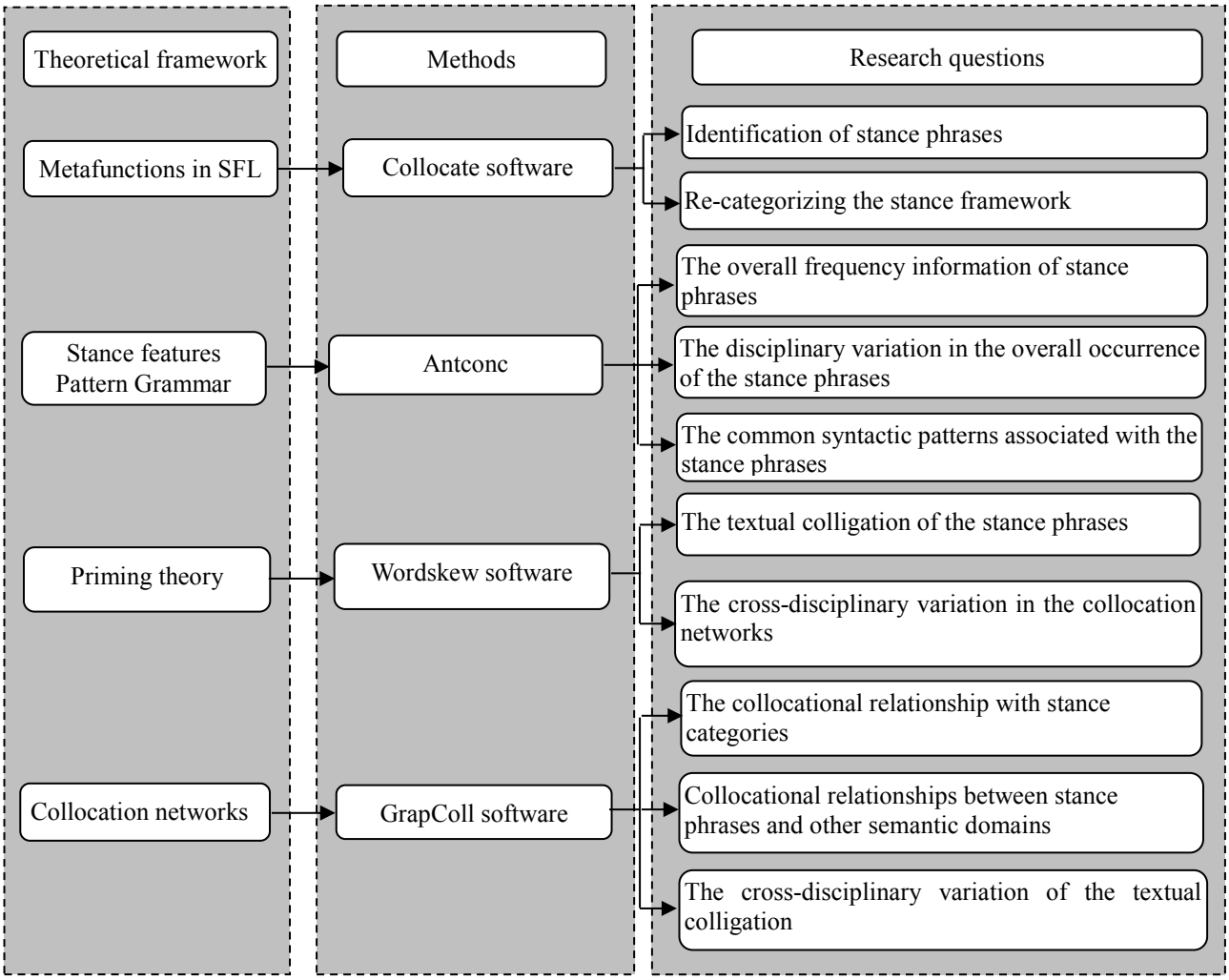

Figure 2 Diagram of the Overall Research Design

The first step involves the identification and categorization of the stance phrases. The identification and retrieval of the stance phrases were carried out using *Collocate* software. Stance phrases were selected by a procedure of filtering with the consideration of frequency, range, and FTW. (The specific information is presented in Section 3.2.2). Following this, I carried out a manual selection to identify the phrases related to stance meaning. Then I carried out a manual categorization of the stance phrases based on the metafunctions of SFL. The reliability of the manual identification and the categorization were measured by using Kendall's coefficient of consistency (described in Section 3.3.2)

Both corpus-driven and corpus-based approaches were employed in the study. Specifically, I used

the corpus-driven approach to retrieve the stance phrases by means of a four-step procedure of extracting, filtering, manual selection and collapsing (described in Section 3.2). The justification of using the bottom-up approach is presented as follows. First of all, there is no consensus reached regarding classification schemes, and the current classification schemes are derived from certain genres or disciplines. Additionally, most writers (i.e., Biber et al, 2004; Hyland, 2008a) acknowledged that their classification schemes are not exhaustive and may vary when applying to different corpora. Thus, it may affect the coding or retrieval of the stance expressions in the different study. The third reason relates to the following two advantages of the corpus-driven approach. Firstly, applying this approach helps to avoid overlooking some unique features that would otherwise not show up in the current classification system, as the analysis of stance phrases shown in many examples have not been discussed in the previous classification schemes. These include, such as *with respect to, provide insight* into, etc. Also, this approach specifies and determines the exact stance phrases used in this study, thus facilitating the automatic corpus-based retrieval and analysis of stance phrases in subsequent chapters.

When categorizing the functional taxonomy of the extracted stance phrases, a top-down approach was employed by referring to the ideational, interpersonal and textual function from the SFL framework. The specific analysis of categories was carried out by mapping the semantic meanings of each stance phrase onto the metafunctions of SFL.

When addressing research question 1, which concerns the similarities and differences in the use of similar linguistic devices by writers in the same discipline, I compared the frequencies of the stance phrases used in the two corpora by using LL, Chi-square and effect size statistics (This is described in Chapter 5).

To address research question 2, which concerns the textual colligation of the stance phrases, I separately examined the textual colligation of the stance phrases extracted in Chapter 3 at the level of sentence, paragraph and text in the two disciplinary RAs and compared the discipline-specific features of the textual colligation in the two academic disciplines. This analytical perspective is underpinned by Hoey's (2005) priming theory.

The software *Wordskew* (Barlow, 2016) was used to extract the textual colligation of the stance phrases. The variation in the textual colligation of stance phrases used in the two corpora are presented in Chapter 6. The specific procedures are presented in Section 6.5.

This study also attempted to examine the collocation network of stance phrases (research question 3) by examining the collocational relationship between stance phrases and their co-texts. The theoretical basis for collocation networks is Phillips (1989), who emphasized the contextual interconnectedness of lexical units. The justification for the contextual relationship of stance phrases is as follows. Firstly, it is argued that stance expressions are often found to rely on the context

(Stotesbury, 2003). In addition, context is generally regarded as a "paramount" factor in stance analysis for its close connection with disciplinary cultural values (Partington, Duguid, & Taylor, 2013, p.52). As previously suggested, the contextual feature of collocational relationship reflects "sets of taken-for-granted values, attitudes and values of behaving, which are articulated through and inferenced by current practices among a group of people in a given text" (Becher & Trowler 2001, p.23). Hence, the disciplinary practice is influenced by the conventionalized practice held by the academic community members. In that sense, the value of a collocational relationship resides in displaying the authentic use of the linguistic devices in context, which contributes to a deeper understanding of language use in a specific context (Flowerdew, 1993; Mudraya, 2006; Thurston & Candlin, 1998; Trebits, 2009).

The collocational relationship of stance phrases enables an in-depth and quantitative understanding of how stance phrases function and enables an examination of the use of stance phrases in agriculture and economics academic discourses. When addressing this aspect of stance phrases, the software *Graphcoll* (Brezina et al., 2015) was used to retrieve the collocates of stance phrases. The specific procedure is presented in Section 7.2.

Chapter 4 The classification scheme of stance phrases

This chapter first presents a review of the previous classification schemes (Section 4.1). It then provides a justification for proposing the new classification framework synthesizing the classification scheme based on the metafunctions of SFL (Systemic Functional Linguistics) (Section 4.2). Section 4.3 elaborates on the classification scheme used in this study.

4.1 Previous stance classification schemes

Previous research has yielded several elaborated stance classification frameworks. For instance, evaluation (Hunston, 1994, 2011; Hunston & Thompson, 2000) and stance (Biber et al., 1999) are derived from the Corpus Linguistics perspective; the appraisal framework (Martin & White, 2005) is based on the interpersonal function within SFL; metadiscourse (Hyland, 2004; 2005b) draws on the distinction between interactive and interaction made by Thompson (2001); modality is rooted in SFL; and AFL (academic formulas list) is obtained by a corpus-driven approach based on the MICASE. This section provides a brief overview of the previous theoretical accounts and classification schemes of stance and other relevant terms (Section 4.1.1-4.1.7), before introducing the justification for the analysis, and discussing the classification framework developed and used for analysis in this study.

4.1.1 Evaluation

Evaluation is defined as an attitude towards a person, situation or other entity and is both subjective and located within a societal value-system (Hunston, 1994). It has been generally viewed as an essential constituent of the community ideology or social values underlying the texts (Hunston & Thompson, 2000, p.6). The evaluation framework proposed by Hunston and Francis (2000, p.100) consists of the following four categories: good-bad, certainty, expectedness and importance. According to this theory, evaluation generally performs the following function of evaluation: 1) expressing opinions (referring to identifying the writers' idea aimed to express, which could reflect the value system of that person and their community); 2) maintaining relationships (concerned with the establishment of and maintaining relations between writer and reader), and 3) organizing discourse (involving how the evaluation have been expressed through the text structure (Hunston & Thompson, 2000).

Showing a close link with the three metafunctions, namely, ideational, interpersonal, and textual functions within the framework of SFL, the framework identifies the typical features and functions of evaluation and explores the implicit evaluative expressions in different discourses. One notable feature of evaluation, according to Hunston (2011), is the property of being "highly implicit" (Hunston, 2011,

p.21), as the act of evaluation often takes place in a subtle way rather than just manifesting any recognizable evaluative expressions. For example, the expression "the clothes made in France" often conveys the evaluative connotation that the clothes are of high quality, and the interpretation of this expression relies on writer/speaker and readers/hearers' shared assumption that the clothes made in France are of higher quality. Thus, the evaluation in the statement is constructed implicitly by making use of no semantically explicit evaluative expressions.

Hunston views evaluation as a "cumulative" concept (Hunston, 2011, p.3), and highly dependent on the context. The interpretation of evaluation entails more than "simply locating those forms", but also our understanding of the evaluative expressions (i.e., classifying instances of evaluation according to a scientifically grounded framework), as well as an in-depth analysis of the co-texts of a construed evaluation (Hunston, 2007b).

4.1.2 Stance

Biber et al. (1999) classified stance into epistemic, attitude and style stance categories and examined their use in the registers of conversation, fiction, news and academic discourse. This classification scheme includes: (1) epistemic adverbs indicating certainty and doubt *(*e.g., *certainly, undoubtedly,* and *maybe),* actuality and reality *(*e.g., *in fact, actually,* and *for a fact),* source of knowledge *(*e.g., *evidently, according to,* and *as X reports/notes),* limitation *(*e.g., *in most cases,* and *typically),* viewpoint or perspective *(*e.g., *in our view, from our viewpoint,* and *to my knowledge),* imprecision*(* e.g., *sort of, kind of,* and *so to speak)*; (2) attitude conveys attitudes, feelings, value judgments, or expectations, and is composed of expectation *(*e.g., *surprisingly, as you might expect,* and *predictably),* evaluation *(*e.g., *conveniently, wisely,* and *sensibly)* and importance *(*e.g., *even more importantly, importantly,* and *significantly)*, and (3) style delivers authors' comment on the manner of speaking, including *honestly, truthfully*, and *if I may say so.* Their comparison of register-specific stance-taking showed that conversation register contains more stance adverbs than academic and news registers.

Based on this framework, Conrad and Biber (2000) examined the use of stance adverbials in three registers (conversation, academic and news), and identified notable grammatical features of stance adverbials, including adverbs e.g., *perhaps*, prepositional phrases e.g., *in most cases*, and subordinate clauses e.g., *I think*. Their study also compared the use of three types of stance expressions and identified the register-specific variation of stance expressions in terms of the typical meanings and grammatical forms.

4.1.3 Metadiscourse

Metadiscourse is defined as "discourse which goes beyond and above the actual content of the basic propositional information being presented, indicating to readers how they may 'organize, classify, interpret, evaluate, and react" (Vande Kopple, 1985, p.83). Using the term *metadiscourse*, Crismore (1989), Vande Kopple and Crismore (1990), Hyland (2005b; 2008a) and Hyland and Tse (2005) examined the use of metadiscourse markers in a variety of genres[4]. For instance, Hyland (2004) compared the frequency of stance expressions in academic discourse of various academic disciplines and identified discipline-specific features in metadisocurse markers.

As regards the categorization of metadiscourse, Hyland and Tse (2005) introduced a model of metadiscourse consisting of two dimensions, namely, *interactive and interactional* dimensions. The *interactive* dimension relates to linguistic makers that "set out an argument to explicitly establish the writers' preferred interpretations" (p.168), and *interactional* concerns markers that "involve readers in the argument by alerting them to the authors' perspective towards both propositional information and readers themselves" (p.168). Hyland (2005a) further proposed a classification scheme of metadiscourse, and defined it as being composed of stance *(hedges, boosters, attitude markers,* and *self-mentions*) and engagement (*reader pronoun, appeals to shared knowledge, directives and questions,* and *personal asides*).

The interactional and interactive aspects of metadiscourse in this framework have also been explored extensively in previous studies. According to Thompson and Thetela (1995), *interactive* resource is concerned with linguistic markers used to manage information flow in texts and used to guide readers through text reading; while *interactional* focuses on involving readers in developing argument in a text. Drawing on the distinction between these two resources, Thompson (2001) examined the interactive and interactional resources in academic texts written by novice writers. His study showed how to improve the written drafts by using interactional resources and she suggested raising students' awareness by using the *interactive* and *interactional* resources in their academic writing. From a theoretical point of view, Sinclair (2004) distinguished the concept between the *interactive* plane (concerns the continuous negotiation between participants) and *autonomous* plane (relates to a developing record of experience). By adopting a dynamic perspective, he elaborated on

[4] The distinction between register and genre has not reached consensus, one of the most notable attempts is made within SFL, which views genre and register on different "semiotic planes" (Martin, 1985). Genre is regarded as a social process by which writers or speakers use language in a sequential structure, while register is concerned with the "expression-plan" of genre (Martin, 1985). Drawing on the distinction, Biber and Conrad (2009) elaborated the difference between the two concepts and further proposed that register and genre are different approaches and perspectives in text varieties analysis, rather than the text types. Specifically, genre perspective focuses on the structural organization features (like the rhetorical organization), while the register perspective concentrates on the typical linguistic characteristics used in particular text varieties. Given that the main purpose of the present study is to examine the linguistic devices used by writers in the academic discourse, so the term register is used in this study.

how the *interactive* and *autonomous* planes of language interact with the real world.

One notable feature of the metadiscourse framework is the emphasis on the feature of interaction, and also shows how linguistic features have been used to organize text structures. The two-dimensional model of metadiscourse provides a dual view from the perspective of both the author and reader, which is of value for exploring the interaction between writers and readers.

4.1.4 Academic formulas list

The Academic Formulas List (AFL) is a list of word strings extracted from corpora of written and spoken language by Simpson-Vlach and Ellis (2010). In the study, Simpson-Vlach and Ellis (2010) extracted high-frequency phraseological expressions from the Michigan Corpus of Academic Spoken and Written English and provided 131 high-frequent 3-5 words phrases. Based on the predominant pragmatic functions, they classified the phrases into Referential expressions, stance expressions and discourse organizing expressions. The stance expressions in the AFL were further categorized to include *hedges, epistemic, obligation and directive, ability and possibility, evaluation and intention/volition,* and *prediction*.

Taking a pedagogical perspective, AFL provides a list of the phrases extracted from the corpora, and made a distinction between written and spoken phrases in academic domains. The functional categorization and the stance phrases can be useful for the EAP curriculum design, class activities in EAP contexts.

However, a close look at the classification scheme reveals some overlapping phrases. For example, *it is possible ([that/to]), (it) is not possible (to)* were classified under the category of *expressions of ability and possibility,* whereas a similar phrase *it is impossible to* was categorized as *evaluation,* a different category from the ability category that the phrase *it is possible to* belongs to. One of the possible reasons for the difference may be attributed to the overlapping principles followed in categorizing the stance phrases. For example, the category *expressions of ability and possibility (most likely to, it is possible)* may overlap with *hedges* (such as *likely to* and *appear that*). Therefore, the category *ability and possibility* seems to overlap with *hedges*, as they both perform a similar function and contain some expressions of similar meanings. The overlapping classification may pose some difficulties when using the classification schemes to categorize the stance expressions. Hence, a scientific guiding principle for classification can provide the basis for systematic classification frameworks, thus minimizing the overlaps that might be caused by multiple criteria. In this light, the present study employed the metafunctions of the SFL framework as a criterion for stance classifying. (The specific classification scheme used in this study is discussed in detail in Section 4.3).

4.1.5 Modality

In the SFL tradition, Halliday (1994) proposed the distinction between *Modalization* (relation to probability and usuality) and *Modulation* (obligation and inclination). He discussed modality and attitude meaning in relation to interpersonal meaning, but that category also includes other areas (such as mood). In this line, Bybee and Fleischman (1995, p.25) distinguished between modality with the two main subcategories of epistemic (*probability, possibility, certainty*, etc.) and deontic (*obligation, permission, ability,* etc.) modality and evaluation. Epistemic stance was developed from the concept of *modalization* in Halliday's terms (Conrad & Biber, 2000). Similarly, Stubbs (1996, p.208) also focused on the modality of a speaker's commitment and detachment to propositional information. 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.

This concept also differentiates the degree of probability associated with a proposition and the degree of obligation or inclination associated with a proposal, and is identified to be realized in a continuum of truth, e.g., *certainly-probably-possibly*, *required to-supposed to-allowed to* and *will-would- may-might.* The theory of *modality* specifies the degree of commitment that writers intend to manifest towards the proposition or proposal, thus laying the foundation for the examination of hedging expressions.

4.1.6 Appraisal framework

Appraisal framework is proposed to "flag the existence of a wide array of resources that are used to negotiate group identity and so co-operate with appraisal theory and negotiation in the realization of tenor relationships" (Martin & White, 2005, p.34). Martin and White (2005) described a taxonomy of the evaluation construction system, including *attitude* (related to people's feelings, such as emotional reactions, judgment of behavior and evaluation of objects), *engagement* (is concerned with pointing out a source for attitudes and deals with people's voices and opinions), and *graduation* (conveying how writers modify the strength of their attitude/engagement). Among the three categories, attitude is further subsumed into *affect* (refers to expressions of emotion), *judgment* (moral assessment of behavior), and *appreciation* (aesthetic assessment). Engagement is composed of *heterogloss* (including *contract* e.g., *X demonstrated that*, and *expand* e.g., *X is claiming that)* and *monogloss* (using authors' own voices instead of quoting other voices and viewpoints).

With respect to the comparison between the categories with other classification schemes, the *engagement* category in the appraisal framework is a concept similar to evidentiality (attributing the source of information in a statement) (Boye & Harder, 2009; Charles, 2003), as they both point out the source for a statement or an argument, whereas, this concept differs from evidentiality. Specifically,

the *engagement* category in the appraisal framework entails a social perspective as it centers on the interaction between writers and readers, while the evidentiality is influenced by the philosophical truth-functional orientation which views statement or information as supposedly grounded in logical truth values or evidence, which can be realized by expressions such as *I see that*, *reportedly*, and *obviously*. However, it is necessary to point out that engagement has also been used in Hyland's (2005) stance classification scheme, with a different meaning. In contrast to the notion of attributing or specifying the source of a certain statement delivered by the *engagement* in the appraisal framework, *engagement* in Hyland's (2005) definition is used to show the connections with readers to engage readers in text description, including reader pronouns like *we,* personal asides (authors interrupting to offer comments often inserted in brackets) and appeals to shared knowledge e.g., *we know*, directives e.g., *consider, imagine,* and questions e.g., *what do these two have in common one may ask?*

Built on the interpersonal metafunction within the SFL framework, the appraisal framework views stance as "dialogically directed towards aligning the addressee into a community of shared values and belief" (Martin & White, 2005, p.95). It is necessary to note that both the appraisal framework (Martin & White, 2005) and the stance and engagement framework (Hyland & Tse 2005) draw on the interpersonal function, and attach great importance to the interaction between writers and readers. However, the appraisal framework (Martin & White, 2005) concerns interpersonal function as a holistic concept, and theorizes expressions that fulfill the interpersonal function; while Hyland and Tse (2005) employ the interactive and interactional resources as two differing criteria for the categorization system.

Overall, the interpersonal function constitutes an important aspect of evaluation, as it relates to one of the most importance functions of evaluation, that is, building and maintaining relations (Thompson & Hunston, 2000, p.8). This perspective contributes to revealing the interactive feature of discourse and revealing how writers and readers communicate by making use of evaluative expressions. However, as suggested by Halliday and Hasan (1989, p.45), only some evaluation fulfills an interpersonal function, while many other evaluation acts are connected with the ideational function and textual function. Therefore, the categorization of stance merely built on the interpersonal function may overlook some other aspects of stance features.

4.1.7 Summary of the previous frameworks

To sum up, the aforementioned frameworks and previous research on stance suggest that writers have generally reached consensus on the following points. Firstly, stance closely entails authors' self-projection in the process of delivering knowledge and organizing texts. Three dimensions of stance, namely, evaluative stance, hedging stance, and referential stance, have attracted due attention and been extensively investigated in the previous studies. Evaluative stance is regarded as "ways of feeling"

(Martin & White, 2005, p.42); hedging stance is a means to "withhold complete commitment to a proposition" (Hyland, 2005a, p.178), and reference stance shows authors' way of attributing the source of information and organizing the texts (Crismore, 1989; Hyland, 2003; 2005a). Considering that stance is closely related to authors' degree of commitment to a proposition, the exploration of stance expressions can reveal authors' particular approaches to constructing knowledge and communicating information. Also, the knowledge of stance expressions allows readers to keep track of "how knowledge is made" (Hunston, 1993a, p.58) in the text organization. In addition, stance embodies a "shared ideologies assumption" by the community members (Hunston, 2011, p.13; Martin &White, p.67), as subjective stance act "takes place within a social and ideological framework" (Hunston, 2011). In this light, the examination of stance features could reveal the value shared by the community members, thus bringing to light the conventionalized community practice and the shared ideology within community.

Secondly, as indicated by the previous studies, stance is characterized by writers' interaction with readers in the course of communicating knowledge via text organization. The construing of authorial stance allows writers to build and maintain relations (Thompson & Hunston, 2000, p.8). Highlighting the interactive feature, Sinclair (2004) proposed the concept of *interactive* plane and maintained that writing involves continuous negotiation between participants, i.e., the writers and readers. Additionally, Martin and White (2005) drew upon the interpersonal function of appraisal framework and perceived that evaluation fulfills a function of establishing a relationship with readers. Likewise, based on the distinction between interactional and interactive resources, Hyland and Tse (2005) propose the metadiscourse framework, and Hyland (2005a) distinguishes between stance and engagement to highlight the interactive aspect of authors' self-presence in discourse-based communication with readers. The interactive feature of stance expressions has also been emphasized by the stance triangle (Du Bois, 2007), which underlines the role that stance plays in the alignment between writers and readers.

Thirdly, stance encompasses a wide range of open-ended linguistic markers such as adjectives (Hunston, 2004), adverbs (Biber et al., 1999; Conrad & Biber, 2000), phrases (Simpson-Vlach & Ellis, 2010), and salient syntactic patterns (Hewings & Hewings, 2002; Groom, 2005; Hunston & Sinclair, 2000). Riloff and Wiebe (2003) point out that "subjective language can be exhibited by a staggering variety of words and phrases" (p.105).

Fourthly, apart from the diverse linguistic forms, stance is also generally accepted to be "heavily dependent on context" (Hunston, 2011, p.13). Stance entails "a deeper understanding of the discourse as a whole" (Hunston, 2007b, p.28) rather than simply locating linguistic forms. Also, the interpretation of evaluation tends to rely on readers' understanding of the context and their shared knowledge of the concept to be evaluated. The reliance on context can also be illustrated by the

observation that evaluative stance is often presented in an implicit way (Hunston, 2011, p.3). According to Hunston (2011), stance expressions are not always straightforward but in some cases expressed subtly, which needs to be interpreted by readers' comprehension of the context of the stance devices.

With regard to the stance classification categorization, previous classification schemes are found to share the following similarities. First of all, previous writers place great emphasis on the evaluative aspect of stance. As indicated in the classification schemes mentioned in Section 4.1, most classification schemes include the category of evaluative stance. The importance of evaluative stance can also be illustrated by several writers' using the evaluative aspect as the title of their classification, such as evaluation (Hunston, 1994; 2011) and the appraisal framework (Martin &White, 2005). For instance, Hunston (1994; 2011) and Martin and White (2005) use the headings of *evaluation* and *appraisal framework* respectively to refer to an umbrella concept concerning evaluative stance. Hunston (1994) defines evaluation as "an attitude towards a person, situation or other entity" with the characteristics of subjectivity which are "located within a societal value-system" (p.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 their framework. In contrast, Biber et al (1999), Hyland and Tse (2005); Hyland (2005b), and Simpson-Vlach and Ellis (2010) use the term of evaluation to refer to a concept within the attitude category. Also, writers often investigate the concept with different headings, such as attitude stance in the stance classification scheme (Biber et al, 1999), attitude markers in the metadiscourse scheme (Hyland & Tse, 2005; Hyland, 2005b) and evaluation in the AFL (Simpson-Vlach & Ellis, 2010). In terms of scope, Biber et al. (1999) focuses on the emotional attitude e.g., *surprisingly* and *unfortunately*; Hyland (2004; 2005a) considers the concept in his category of *booster* (e.g., clearly, obviously, and highly) and *attitude makers* (e.g., *important, hopefully*, and *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.*

Secondly, the hedging aspect of stance has been extensively explored, though under varying headings, such as *hedges* (Hyland, 1998), *epistemic modality* (Nuyts, 2001) or *epistemic* (Biber et al, 1999). In the previous categorization schemes, hedges have been included in the stance classification under the heading of *epistemic stance* (Biber et al, 1999), *hedges* in metadiscourse scheme (Hyland, 2005b), and *hedges* in AFL (Simpson-Vlach & Ellis, 2010). With respect to the scope, Hyland (2004; 2005a) uses *hedges* to refer to expressions that mitigate authors' presence, while Biber et al. (1999) employ the concept of epistemic stance (including *certainty, reality, source of knowledge, limitation,*

viewpoint, and *imprecision*); Martin and White (2005) employ the term *entertain* (e.g., *possibly, I think,* and *it seems*) to deliver possible positions which make "dialogic space for possibilities" (Martin & White, 2005, p.104)*;* while Simpson-Vlach and Ellis (2010) use the term *possibility*.

Thirdly, the referential aspect has also attracted interest from previous writers, and different headings have been employed, such as "evidentiality" (Chafe & Nichols, 1986) and reference (Crismore, 1989; Hyland, 2003; 2005a). In the previous classifications, the concept has generally been treated as one category of the broad concept of stance. For instance, it has been included as a kind of metadiscourse markers (Hyland & Tse, 2005), under the heading of *endophoric markers* (e.g., *see fig.2*)*, evidentials* (like *according to*)*,* and *code glosses* (like *see namely*) within interactive resources*,* and *self-mentions* (e.g., *we, my*) within interactional resources, as well as *engagement* (positioning the voice of authors and other recourses) within the appraisal framework (Martin & White, 2005). In terms of the scope, the concept of referential stance has a different denotation: Hyland and Tse (2005), for example, subsume them into *endophoric markers* (attributing the source to other parts of the text), *self-mentions* (attributing the source to authors or the present study), and evidentials (attributing to an extra-source other than the authors or the present study) (Hyland & Tse, 2005). In contrast, Martin and White (2005) distinguish between monogloss (making no reference to other voices and viewpoints) and heterogloss (e.g., i*n my view, everyone knows,* by pointing out the source to invoke or allow for dialogistic alternatives) following the criterion concerning the source of information.

These similarities aside, the aforementioned frameworks are also found to have the following differences. Firstly, apart from the three major aspects of stance (evaluative stance, hedging stance, and referential stance) mentioned above, previous classification frameworks are seen to have their unique perspectives on a particular aspect of stance features. For example, Hyland (2004) elaborates on the interactive resources, including transitions (*in addition* and *thus*), frame markers (*to conclude* and *my purpose is to*), endophoric markers (*note above* and *in section 2*), evidentials (*according to,* and *X states*), and code gloss (*namely* and *such as*). Meanwhile, Biber et al. (1999) focus on the manner of speaking, including expressions such as *honestly, frankly*, and *simply*. In the appraisal framework, Martin and White (2005) elaborate on the graduation aspect of appraisal from the perspective of force (including intensification e.g., *slightly sad, very sad* and quantification e.g., *small problem* and *large problem*), and focus (including sharpen e.g., *a true father*, and soften e.g., *an apology of sorts*). Whereas, Simpson-Vlach and Ellis (2010) examine the knowledge-oriented aspect of the stance, and include phrases such as *be considered as, is determined by, be argued that* in their stance explanation.

The divergence in the previous classification schemes may be attributed to the following two reasons. Firstly, most of the above classification approaches are grounded in an inspection of stance expressions of different texts (Hunston, 2011, p.22). For instance, Huston's evaluation taxonomy centers on value-laden expressions in empirical RAs; Biber's work focuses on stance adverbs retrieved

from three genres in conversation, academic papers, and newspapers; Hyland's metadiscourse framework involves expressions extracted from published RAs and postgraduate dissertations, while the stance phrases in the AFL retrieved by Simpson-Vlach and Ellis (2010) are based on the spoken and written academic discourse. The appraisal framework (Martin & White, 2005, p.161) centers on newspaper reports. In contrast, the classification framework on modality (Halliday, 1994) adopts an intuition-based approach and explores the concept from a theoretical point of view.

Secondly, previous classification schemes have focused on different aspects of stance expressions. For instance, Hunston (2011) centers on the evaluative feature of stance; Biber et al. (1999) and Biber et al. (2004) focus on the attitudinal and style aspect of stance; Martin and White (2005) place a strong emphasis on the interpersonal dimension of stance; Hyland (2004; 2005a) has a focus on the interactional and interactive dimensions of stance, and finally Simpson-Vlach and Ellis (2010) concentrate on the knowledge aspect of stance.

4.2 Justification for the framework used in the present study

Underpinned by the aforementioned classification schemes, the present study proposes a Stance Classification Framework (SCF), which is elaborated in Section 4.3. The metafunctions of SFL were employed as the classifying theoretical underpinning for SCF with an aim to obtain a relatively systematic stance classification framework. This following section presents the rationale for proposing SCF.

According to the extended definition proposed by Hyland (1999a, p.101), stance delivers writers' authorial positioning in text construction to "communicate their integrity, credibility, involvement, and a relationship to their subject matter and their readers". The analysis of the previous classification schemes, as discussed in Section 4.1, shows that the six current popular stance categorizations converge in the following three aspects of stance features: attitudinal stance (authors' emotion or affection-oriented presence), hedging stance (expressions to downplay authors' role), or referential stance (markers signaling the source of the opinion or statement or textual organization markers). However, another important aspect of stance, which delivers authors' reason-oriented presence and constructs a persuasive persona, e.g., *due to* and *as a result of*, has been overlooked in these classification schemes.

It is necessary to make a distinction between reason and emotion before discussing the previous literature on the reason-oriented aspect of stance. Reason and emotion are generally accepted to deal with distinct aspects of authors' stance. According to Collins Cobuild English dictionary (Sinclair, 1995), reason is concerned with "the ability that people have to think and to make sensible judgments" (p.1280); emotion relates to "a feeling such as happiness, love, fear, anger, or hatred, which can be caused by the situation that you are in or the people you are with" (p.502). Philosophers tend to view

the two concepts as two antagonistic aspects. For instance, Plato viewed emotion and reason as two horses pulling us in opposite directions. Reason guides people to make logical and rational judgment and decision, while emotion directs to an opposite direction. According to Macmurray (1935), the reason and emotion-oriented aspects are "in the eternal nature of things distinct and opposite" (p.16). Based on the distinction between the two concepts further, in his theory of Moral Sentiments (1759), Adam Smith believed that it is human's passion or emotion that rules the morality development instead of reason, while reason plays an essential role in discovering and supporting the institutions which "directed the passions to universally beneficial ends" (Sharpe,1996, p.63).

Although it is beyond the scope of the present study to elaborate the specific relationship between the two concepts, the above discussions indicate that reason and emotion address two distinct aspects of authors' self-presence in text construction. Reason is concerned with people's mental processing of the information by following the logical rules while emotion addresses people's intuitive feelings.

As suggested by previous linguistic analysis, the reason-oriented aspect of stance has attracted growing attention from writers 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" (p.506) and included cognitive phrases, such as *assumed to be, considered as, determined by, argued that, have shown that,* in this category. Furthermore, Hyland (2012) recently resumed the knowledge-oriented aspect of stance in his extended definition (Hyland, 1999b) and asserted that expressions (like *in terms of the*) delivered explicitly interpretive presence and persuasive endeavors, and he also identified that cognitive verbs often collocate with personal subjects like *I think that* and *we suspect that.* Similarly, Aull and Lancaster (2014) view the logical expressions conveying concede/counter formulations (e.g., of course, however) in the stance analysis of the writing of first-year university students and that of upper-level undergraduate students and published writers. In addition, Lancaster (2016) included some of the cognitive markers denoting 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, cognitive represents how authors project their mental processing of knowledge, and works as an essential linguistic device for writers to "negotiate meanings with the reader" (Lancaster, 2016, p.16). 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, drawing upon the traditional distinction between reason and emotion, this study introduced the reason-oriented stance, embodied as cognitive stance, into the stance classification scheme and treated it as an independent stance category, and differentiate it from the other three aspects of stance, namely attitude, modality, and reference.

Taking into account the different aspects of the stance features, this study attempts to synthesize the existing stance classification and incorporate the reason-oriented dimension of authors' presence into the SCF classification framework, in order to capture a relatively more comprehensive picture of stance. The hybrid SCF classification is underpinned by the three metafunctions, namely ideational, interpersonal, and textual functions within SFL (Halliday, 1994). The following section elaborates on the SCF by referring to the stance phrases extracted from the agriculture and economics academic corpora. The correspondence of this framework and the metafunctions of SFL is also discussed below.

4.3 The stance classification framework proposed in the present study

The SCF used in the present study, as illustrated in Figure 3, is composed of cognitive, attitude, hedges and reference stance, in which subcategories are also made for the purpose of providing a deeper and more thorough picture of the stance features of the two academic disciplines. The four categories and their subcategories are discussed in detail in Section 4.3.1, 4.3.2, 4.3.3, and 4.3.4 respectively. The examples used to illustrate the categories of the stance phrases are based on the data extracted in the two academic disciplinary corpora under scrutiny in this study.

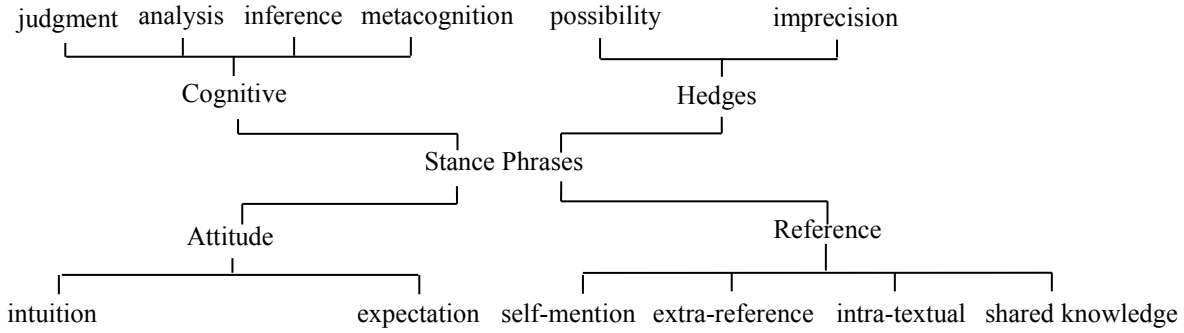

Figure 3 The Stance Classification Framework

4.3.1 Cognitive stance expressions

Cognitive stance expressions are mainly used to "formalize the logic of discourse and clarify the principles of reasoning" (Rescher, 2005, p.1). Their use reflects writer's reason-oriented presence in expressing assertions and constructing knowledge, mainly in the form of logical relation (*in addition, because of, explained by*) and mental processing act (*take into account, assume that, consider that*) towards entity or proposition. The cognitive stance expressions reflect authors' efforts in situating themselves in a framework of norms and ideas (Martin, 2006) and play a critical role in establishing the research territory and construing knowledge. Cognitive stance phrases principally serve a purpose

of "establishing a warrant for research" (Hood, 2010, p.30) and "creating rational appeals" (Hyland, 2005b, p.75), and contribute to augmenting the validity of the knowledge statement and thus constructing a persuasive persona.

This dimension shares a strong relationship with the ideational function, particularly the logical function within SFL. According to the logical metafunction of 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 expressions are concerned with authors' reasoning and speculation in raising assertions. This sub-category is composed of *result shows that, the evidence indicates that, it is demonstrated that, etc.* The expressions often signpost a conclusive statement established on evidence or data following logical induction or deduction. Thus such expressions contribute to boosting the validity of the conclusion and the soundness of their inference, and fulfilling the persuasiveness function of academic discourse.

Metacognition stance expressions concentrate 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, etc.* Generally, the cognitive markers demonstrate authors' reason-oriented acts. The employment of this type of expression therefore functions to justify their statements, provide "research warrant" (Hood, 2010, p.39) and "create rational appeals" (Hyland, 2005b, p.230).

Essentially, cognitive stance characterizes authors' cognitive presence and processing in the process of knowledge construction. However, it is surprising to see that such a category has seldom been proposed as an independent category, albeit having been occasionally dealt with in previous literature (i.e., Hyland, 2012; Simpson-Vlach & Ellis, 2010; Swales, 2002). One possible reason is that most of the previous schemes are underpinned by the interpersonal function, as above mentioned, and focus on the explicit stance expressions and authors' direct evaluation. Cognitive stance, derived from the ideational function, allows us not only to ascertain the implicit rhetorical devices that indicate authors' presence but also to distinguish author's rational projection from the emotional ones. The inclusion of this dimension thus can add an insight into how authors construct their authorial positioning from a reason-oriented perspective.

4.3.2 Attitude stance expressions

In contrast with this reason-oriented projection delivered by cognitive stance, attitude stance represents authors' emotion-oriented involvement. This aspect is concerned with the recurring

expressions that clearly manifest authors' feelings in presenting opinions or propositions (Biber et al., 1999; Hunston, 2011; Martin & White, 2005). The attitudinal stance phrases are composed of evaluative markers (*it is important, play an essential role*), and the feeling-oriented expressions (*it is interesting to*, *it is surprising that*). Corresponding to the interpersonal metafunction, attitude stance phrases arouse readers' emotional reactions and thus enhance their engagement in processing the knowledge or assertions delivered in the discourse, by means of the following two types of subcategories (*intuition* and *expectation*).

Specifically, intuition expressions are pertinent to authors' subjective stance which functions to arouse readers' curiosity or interests and steer their attention to the statements. The intuition expressions consist of *it is important to/that, it is necessary to, and it is noteworthy to*, etc. The intuition expressions are usually employed to convey authors' stance in terms of importance, necessity and significance, etc. towards entity, phenomenon or proposition. Those phrases reflect authors' efforts to share their interpretation and stance with readers, thus contributing great value in realizing effective communication with readers.

Expectation stance involves authors' anticipation or expectation towards a certain phenomenon or proposition, and includes *it is recommended that, should be, must be,* etc. Attitude expressions convey personal stance, comment, interpretation, and expectations. Such expressions allow authors to flag their explicit emotion-oriented commitment in interaction with readers via academic discourse.

With respect to the grammatical forms associated with this category, adjectival or adverbial phrases take an overwhelming proportion of the authors' evaluative stance. This finding is supported by Francis (1994) who found that adjectives indicate an evaluation or judgment, likelihood, clarity, necessity, significance, goodness or badness. Hyland and Tse (2005) also view that attitude as expressed throughout a text by using subordination, comparatives, and progressive particles. Additionally, the verb phrase is also found to constitute a notable structure in presenting authors' stance, e.g., *play a role, contribute to, need to,* and *shed light on/upon*, etc., which work as idioms to assess or evaluate an entity or proposition. Although insufficiently addressed in the previous literature, verb phrases are found to play an active role in constructing evaluative stance and covertly showing the author's presence and engagement in the text organization.

4.3.3 Hedges stance expressions

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*) assist writers to establish a close rapport with readers by softening the tone of statement claims. Hedge stance phrases not only reduce the risk of being criticized for proposing new statements, but also allow writers to assume "dialogic tones" (Bakhtin, 1981, p.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, p.23) and is a process to reduce the strength of a statement (Zuck & Zuck, 1986).

The major theoretical underpinning of hedges in the study is also the interpersonal function in SFL by downplaying authors' presence, the hedging function of which is to establish a rapport and communicate the ideas to the expected readers. This line of study is seen to be extensively researched (Hu & Cao, 2011; Hyland, 1996; Varttala, 1999), and the extensive and in-depth findings by Hyland (1998; 2005a) have enriched our understanding of this hedging dimension of authorial' presence.

The following two types of linguistic devices are often used to mitigate authorial' presence in statement construction. Firstly, possibility is used to present authors' uncertainty and mitigate their tone when claiming new assertions or raising different propositions. It includes phrases such as *it seems that, it is likely to, tend to,* etc. The second type, imprecision hedges, relates to the linguistics forms which expresses an imprecise estimation, for example, *at least, at most, to some extent, in some cases, etc.* Hedges demonstrate authors' caution in scientific research, which thus enhances the reliability of the assertion and the credibility of the hedged claims. Their use indicates that authors often tailor the amount of information to fit what the assumed readers need to know, rather than express more than what is required. In this light, hedge expressions contribute to striking a compromise between the nature of reality and the acceptability of a claim. By manifesting author's rigor in proposing claims or reporting data, hedging stance expressions could serve as a strategy to avoid the risk of being criticized by adopting a compromising tone and mitigating the force used to express their claims.

4.3.4 Reference stance expressions

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

The justification for including this category in the classification scheme is presented as follows. Firstly, in terms of the definition, this study considers stance as an umbrella term, which includes a broad scope of authors' presence in a text, namely cognitive, emotional, modality and referential aspects. In contrast, most previous studies, such as Biber et al (1999) and Hyland (2005a, 2008), treat

stance as a narrow concept and only include the evaluative and modality aspect of stance in their categorization.

The second reason relates to the theoretical framework used in this study, namely the metafunction of SFL, which is composed of ideational, interpersonal and textual functions. Among the three functions, the textual function underpins the referential category in this framework. The textual function is concerned with grammatical devices used to maintain discourse flow and to construct coherent and well-organized discourse. The linguistic choices such as *the remainder of the paper* and *as shown in Table 1* enable authors to "create coherent text – text that coheres within itself and with the context of situation" (Halliday, 2003, p.17). By pointing to a particular textual position, these expressions indicate authors' intervention in the text in creating or strengthening intra-textual connections, and enhancing coherence.

The third reason derives from an insight by from Martin and White (2005) and Hyland (2005a). Specifically, Martin and White (2005) consider 'textual voice' as an important form of authors' presence and that its use plays an essential role in the information flow (Martin & White, 2005, p. 2). In the Appraisal Framework, they categorize expressions such as *X said, X believes …, according to X, in X's view (distance) X claims that*, as a type of textual voice. Meanwhile, Hyland's (2005) stance categorization includes self-mention and shared knowledge, which enables authors to establish a connection between a source and the author, or the author's discourse self.

Undoubtedly, this type of expression explicitly performs a textual function in that they connect the text into a coherent and cohesive whole. By means of this type of expression, writers are able to express an authorial positioning of constructing a coherent text and purposefully guiding readers through their text-based communication.

This category mainly corresponds to the rhetorical devices that direct the primary source of particular statements. The analysis of the extracted phrases reveals that the following types of rhetorical devices, namely, self-mention, extra-reference, intra-textual reference and shared knowledge, are commonly used in the corpora to refer to the source of the statements or assertions.

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, etc.* Such phrases can initiate a knowledge-knower structure, in which writer is represented as a knower to provide further indication of the relative strength of social or cognitive relations in the discourse (Hood, 2010, p.172).Therefore, the use these markers play a key role in orientating readers' attention to focus on authors' involvement and contributions in the text construction.

In contrast, extra-reference stance expressions, such as *according to, by other authors,* etc., direct

readers' attention to statements from an outer source, rather than authors' own efforts. Such stance expressions often reflect the multi-voiced or heteroglossic nature of statements (Martin & White, 2005, p.37) as it uses others voices to express authors opinions, and traces the statement source to previous studies or writers. Extra-reference phrases allow writers to support their propositions by referring to an authority and thereby gaining support for their position (Hood, 2010, p.85).

The intra-textual markers allow writers to construct coherent text and show the link between different sections of a text (Vande Kopple & Crismore, 1990).The use of this types of reference stance markers reflects authors' attempts to facilitate readers' comprehension by constructing a coherent and cohesive textual organization, thus fulfilling the textual and interpersonal metafunctions.

Apart from the two kinds of reference categories, in some cases authors direct readers' attention away from a particular piece of information or knowledge to another section or sentence in the textual organization, often by using intra-textual stance. 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, see figure, etc.* Such expressions function to link an argument or proposition with an 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). In this light, the use of intra-textual formulas contributes to constructing a coherent and cohesive textual organization. The formulas can also be viewed as a "discourse strategy to encourage reading" (Hood, 2004:75); their usage reflects authors' attempts to facilitate readers' comprehension of the information delivered in the text organization.

Shared knowledge concerns linguistic markers conveying commomsense 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, etc.* Such expressions allow writers to present their work with assurance (Hyland, 2005a). They encourage the need to interpret the propositions with greater confidence and thus function as implicit boosters to support the authority and reliability of the information. In addition, shared knowledge can save writer's labor of repeating familiar knowledge assumed to be known by the expected readers in their academic community. Therefore, it can serve as a technique to "build consensus with the disciplinary community by evoking shared knowledge and expectations" (Charles et al., 2009, p.157).

4.3.5 Overlapping stance phrases

Although the use of the metafunctions in the stance classification scheme has effectively reduced the degree of overlap found in multicriteria, some overlapping phrases nevertheless occur due to the polysemous meanings of stance phrases. The analysis of concordance lines reveals that the following

four phrases, namely, *contribute to*, *according to, account for*, and *close to*, deliver two meanings and accordingly perform two functions in the cases identified in the two corpora. For instance, the phrase *account for* expresses the meaning of *take up* or *constitute*, as shown in Example (1), thus being subsumed into the judgment function within the cognitive stance. As in Example (2), the phrase means *explaining*, performs the function of analysis in cognitive stance. (Phrases with multiple functions are specified in Appendix B.)

(1) The first component ***accounts for*** more than 80% of variance in the five series; it is used as the financial liberalization index (FLI) in our paper. (FRAC#014:5)

(2) Different phenomena may ***account for*** this fact: (i) Evolutionary rates do not differ because generation times or recombination rates are similar among the interacting species (Gandon and Michalakis, 2002). (ii) Migration, the patterns of adaptation or maladaptation are highly variable over time because of the stochastic nature of the arrival of new favorable alleles. (ARAC#111:2)

It is interesting to note that the overlap is caused by polysemy of the phrases and the examination of the context when the phrases occur allows us to determine which meaning is intended. For example, *according to*, if the object is the researcher, study, or other sources of information, as in (3), the phrase is used to express an inferential meaning by pointing out the source for a statement, thus it can be classified into extra reference within the reference stance. On the contrary, if the phrase is followed by principles, methods, rules, etc., it means *following* or *agreeing with*, and fulfills the function of analysis, one category in the cognitive stance.

(3) ***According to*** Magid et al. (1999), microorganisms lose some of their ability to degrade complex substrates during desiccation. They partly regained that activity upon rewetting, but not to the extent maintained by microorganisms in CWC conditions. (ARAC#125:1)

(4) Thus ***according to*** the tenets of RDT a firm must respond to the external environment (deal with contingencies) by managing interorganizational relations in order to acquire and maintain resources (tangible and intangible) crucial to the firm's ability to compete (i.e., survive) within a given market. (FRAC#42:1)

When it comes to the analysis, the occurrences of the overlapping stance phrases performing two functions were calculated separately in order to differentiate the use of the phrases within different categories and investigate the general features of different types of stance phrases.

4.3.6 Summary

In this chapter, I categorized the most frequent 2- to 5-grams stance expressions extracted from two self-built academic corpora into cognitive, evaluation, hedges, and reference stance, based on the metafunctions within SFL. Cognitive stance shows writer's reason-oriented presence; attitude stance presents authors' emotion-oriented commitment; hedges stance concerns authors' endeavor in downplaying their presence, while reference relates to authors' projection by attributing statements to

different sources. The four types of stance expressions reflect four different aspects of authors' endeavors in construing disciplinary knowledge and maintaining the balance between reason-oriented and emotion-oriented involvement, boosting and downplaying, expressing personal commitment and detachment, reasonable analysis, and a cautious tone.

Overall, based on the metafunctions of SFL, the SCF framework develops previous stance classification schemes in the following manner. Firstly, the framework supplements the classification schemes by adding a reason-oriented dimension of stance, namely, cognitive stance, and treats it in parallel with the other three dimensions of stance (the attitudinal, hedging and reference stance). Given that cognitive stance entails authors' mental processing of the information and constitutes an essential part of their projection, the inclusion of cognitive stance allows us to examine authors' reason-oriented commitment.

The use of cognitive stance entails authors' logical and rational analysis in the course of processing information, which reflects the reliability of the knowledge and opinions delivered in the text. By deploying cognitive resources, authors make evident that their processing of information or knowledge conforms to the norms and procedure in analyzing and explaining the world, such as defining, interpreting, comparing, explaining, and inferring. The use of cognitive stance thus reveals not only how authors mentally process the information by a logical approach, but also how they persuade readers effectively while communicating information to readers.

Additionally, cognitive stance expressions can "signal the reasoning involved in connecting the two clauses" (Martin, 1985, p.18). The use of such expressions not only serves as the connective between clauses to help the arguments flow in an organized manner, but it also enables writers to make the idea explicit to readers and help them to keep track of the development of the information. In this light, cognitive stance entails the interaction between clauses, authors and information, and authors and readers. Therefore, the examination of this concept can enrich our knowledge of authors' particular ways of constructing writer-reader relationships, delivering an argument, and communicating with readers by means of text organization.

Secondly, the classification framework adheres to the criterion of authorial' stance, which not only expands our understanding of authors' role in constructing knowledge and organizing text but also reduces the overlaps that may be caused by multiple perspectives. For one thing, the examination from the viewpoint of authors' projection concerns the ways by which authors process the information and construct knowledge in the process of interacting with readers. The elements pertained to authors' involvement, such as authors' mental processing of the information, their feelings or attitudes towards the statements, their position in presenting new ideas, and their ways of organizing knowledge and ideas, are all embedded in the stance expressions. Given that academic discourse embodies the

interaction between authors' stance and objective knowledge, the use of authors' stance allows information to be presented in a personalized manner, which thus contributes to facilitating readers' processing and understanding of the knowledge. The perspective of the stance classification presented here contributes to an exploration of the role that writers play in the text organization, as well as their relationship with both the information delivered and with the expected readers.

This perspective allows us to examine the concept in a relatively systematic way, thereby minimizing the overlaps caused by multiple perspectives. One example of using more than one perspective is the metadiscourse classification framework (Hyland & Tse, 2005). Arguably, the interactional and interactive perspective contributes to our knowledge of how writers interact with readers via academic discourse. However, the use of the dual criteria may cause some overlapping, as discussed in Section 4.1.3.

It is necessary to point out that the perspective adopted in the present study also captures the interaction between writers and readers, as the use of the four categories of stance entails an underlying interaction between writers and readers. For example, cognitive stance expresses authors' logical processing, which serves as a persuasive device.

The use of the attitude can enhance readers' engagement or participation in the negotiation of knowledge. Similarly, hedging stance allows writers to take a mitigated position when expressing ideas. Reference stance guides readers through the texts and identifies the source of the statements and helps to create linkage between different sections of a text.

In addition, the classification of SCF is underpinned by the metafunctions of SFL, which allows us to examine the concept of stance in a systematic and consistent way. As mentioned in Section 1.2.1, Halliday (1978) considers that language is a system of social semiotics; its operation is shaped and organized in relation to three metafunctions, namely, ideational, interpersonal and textual functions. The three metafunctions address different aspects of language use respectively, and "contribute jointly to the description of linguistic units" (Hasan, 2014, p.26).

Following the metafunctions of SFL, SCF classifies the stance into its four categories: cognitive, attitude, hedging, and reference. Cognitive stance builds on the ideational function, the logical function in particular. With respect to the interpersonal function, SCF includes two aspects of stance, attitude, and hedges; the former manifests author's attitudinal projection with an aim to enhance readers' engagement in the text, while the hedging stance downplays autshors' presence towards some statements to ensure an effective acceptance of the opinions conveyed. Referential stance is based on textual function and shows how authors identify the source of particular opinions or statements, as well creating a cohesive link between different sections of a text. Overall, this theoretical basis not only reduces the overlap caused by multiple criteria, but also helps to identify a dimension of stance

that has been overlooked in previous classification schemes, namely the cognitive stance (as discussed in Section 4.3.1).

Chapter 5 Comparison of stance phrases in agriculture and economics corpora

This chapter addresses research question 1, concerning the frequency of stance phrases in the agriculture and economics corpora. The frequencies of the stance phrases were retrieved by *Collocate* software, as described in Chapter 3. The frequencies of the stance phrases were normalized to a million words (in line with Biber, Conrad, & Reppen, 1998), and Log-likelihood and Chi-square statistical tests were employed to measure if there were significant differences in the occurrences of the stance phrases used in two academic disciplinary corpora, an effect size statistic was used to measure the degree of significance. (The detailed procedure was presented in section 3.3.2). Section 5.1 presents the overall frequency of the categories and compares the disciplinary variation in the two corpora. Section 5.2 presents a comparison of subcategories of stance phrases in detail. Then Section 5.3 discusses the common syntactic patterns associated with the stance phrases in the analysis of two corpora.

5.1 General comparison of stance phrases categories

This section presents the overall similarities and differences with respect to the four types of stance phrases in the two academic disciplines. The comparison of the overall frequency of the four stance phrases categories in the two disciplines is presented in Table 1. The first two columns show the raw frequency of the four types of stance phrases used in the two corpora. Columns 3 and 4 present the normalized frequency per million words in agriculture and economics corpora respectively. Columns 5 and 6 list the percentage of the four types of stance phrases used in each corpus. Columns 7 and 8 report the normalized mean frequency of the stance categories in the two corpora. In order to find whether there is a statistical difference in the use of stance categories, the study presented the results of the LL score (Column 9) and Chi-square (Column 10), and the sig value (Column 11). The effect size is presented in Column 12.

As shown in Table 1, the two disciplines shared similarities in terms of the ranking and percentage of each category. To be specific, cognitive stance phrases were most frequently used, accounting for more than half of the total stance phrases in the two corpora, with an occurrence of 11,503.56 and 10,078.55 per million words in agriculture and economics respectively. Such preference for using cognitive stance reflects these disciplinary writers' efforts to justify their arguments by making use of the logical markers conveyed by cognitive stance. Examples (5) and (6) illustrate two cases where cognitive phrases, *indicate that* and *attributed to*, were used to manifest writers' cognitive presence in form of making a reference and giving an explanation.

(5) Satisfactory results of testing the GMDH regressions against data from NRCS studies of soil quality ***indicate***

that soil survey databases can be used to project changes in water retention caused by differences in soil management. (ARAC#058:2)

(6) The first model uses logistic regression to test whether ranking improvements can be ***attributed to*** a journal's increased global expertise. (ERAC#056:3)

Hedge phrases were the second most frequently employed stance category, constituting about one fifth of the total stance phrases in the two corpora (24.71% and 21.97% in agriculture and economics respectively). According to Hyland (1996), hedging expressions "allow writers to anticipate possible opposition to claims by expressing statements with precision, caution, and diplomatic deference to the views of colleagues" (p.433). The high occurrence of hedging phrases thus suggests that writers in the two disciplines are more likely to hold a prudent attitude to developing their arguments and organizing their texts. It may be the case that the disciplinary writers often anticipate readers' attitudes towards their statements in the process of writing up their academic articles, and they are probably aware that the statements may be criticized or denied when presented to readers. Therefore, in order to avoid the possible negative response from the expected readers, writers tend to make use of hedge stances to downplay their presence and display their prudence to ensure the statements can be effectively communicated to the readers. It is also interesting to note that such expressions are often associated with a proposition which relates to their study or an interpretive statement. As shown in Examples (7) and (8) where writers in agriculture and economics academic discourse utilize the hedging markers, *tend to* and *is likely* to respectively to express caution in constructing a statement.

(7) If roots are unable to penetrate soil aggregates they will ***tend to*** proliferate in the soil cracks sometimes giving a clustered root distribution. (ARAC#111:2)

(8) In less competitive markets, credit unions may be more ***likely to*** provide credit to businesses since it is less available from banks and may be more profitable. (ERAC#077:1)

Reference stance phrases were the third most common category in each corpus, with a percentage of 10.87% and 14.77% in the agriculture and economics corpora respectively. Examples (9) and (10) show two cases where writers in the two disciplines use referential markers to point to a source. The use of these expressions shows that academic writers' efforts to present organized and coherent holistic discourse. By signaling the source with reference stance phrases, writers in the two disciplines identify the source for certain statements, guide readers through the discussion, and steer them to another part of the text (Hyland, 2005b, p.51).

(9) ***According to*** Bal¢zs et al. (1996). S. myopaeformis has become a significant pest and this can be attributed to changes in apple production technology. (ARAC#103:2)

(10) For example, the arithmetical expectation of consumption growth ***as in*** [1] does not necessarily mean that it encapsulates all information about consumers' expectations and their behaviour. (ERAC#003:2)

Table 1 The Occurrences of Stance Phrases in Agriculture and Economics Corpora

Stance	Raw frequency		Normalized frequency (per million)		Percentage		Mean		LL	Chi-square	Sig	Effect size
	Agr	Fin	Agr	Fin	Agr	Fin	Agr	Fin				
cognitive	19204	13636	11503.56	10078.55	59.94%	57.61%	197.12	167.45	144.01	141.20	0.00	0.08
attitude	1433	1339	863.19	989.67	4.47%	5.66%	74.67	76.9	13.98	14.06	0.00	0.09
hedges	7916	5200	4741.83	3843.39	24.71%	21.97%	163.51	132.53	140.27	139.61	0.00	0.13
reference	3484	3496	2015.10	2558.07	10.87%	14.77%	51.67	48.27	79.47	80.10	0.00	0.13
total	32037	23671	19123.68	17469.68	100.00%	100.00%	486.97	425.15	116.93	118.70	0.00	0.06

Note: The following abbreviations are used: Agr (agriculture), Fin (economics).
Sig stands for p-value; The LL ratio and Chi-Square scores are significant at the 0.05 level.

However, in the two corpora, the proportions of attitude stance phrases account for only about 5%, showing that the writers in the two disciplines had a limited presence in terms of flagging their emotion-oriented presence for the purpose of hooking readers' attention and enhancing their engagement. Examples (11) and (12) show two cases where agriculture and economic writers use attitude stance phrases, *it is interesting to* and *It may be surprising* to engage readers to the subsequent statements.

(11) ***It is interesting to*** note that the correlation for Vereecken is numerically the lowest. (ARAC#006:1)

(12) ***It may be surprising*** to know the rate of interest has not turned out a significant factor in determining loan repayment by sample households. (ERAC#018:2)

As the attitude stance is generally regarded as a presence associated with writers' subjective presence (Thetela, 1997), the low percentage in the use of the attitude stance phrases shows that writers in the two disciplines are less prone to present their subjective presence in comparison with the rational commitment conveyed by the cognitive stance.

In conjunction with the similarities mentioned above, the analysis indicates that the two corpora displayed significant differences in the frequency of the four types of stance phrases. Specifically, as shown in Table 1, the statistical analysis shows that the agriculture corpus has a significantly higher frequency of cognitive stance phrases (LL=144.01; Chi-square=141.20; p=0.00), which suggests that agriculture writers tend to express reason-oriented presence more frequently in the textual description. The frequency of attitude stance was significantly higher than that in the economics corpus (LL=13.98; Chi-square=14.06; p=0.00), indicating that agriculture writers were more likely to manifest their emotion-oriented presence in RAs. Likewise, agriculture writers employed more hedging phrases (LL=140.27; Chi-square=139.61; p=0.00) to downplay their involvement in expressing assertions. Whereas, economics writers seemed to use reference phrases more frequently (LL=79.47; Chi-square=80.10; p=0.00), which indicate a tendency to support assertions with evidence.

As shown by the effect size statistic, the statistical difference between the stance phrases used in the two corpora are small (w=0.08, 0.09, 0.13, and 0.13 for cognitive, attitude, hedges, and reference stance phrases respectively). The detailed comparison concerning the subcategories of each type of stance and the reasons for the disciplinary variation are discussed in Section 5.2.

5.2 Detailed comparison of subcategories of stance phrases

5.2.1 Cognitive stance phrases

Table 2 presents the statistic comparison of the cognitive stance phrases in the corpora of agriculture and economics. The first two columns show the raw frequency of the four types of cognitive stance

phrases extracted from the two corpora. Columns 3 and 4 present the normalized frequency (per million words) of the four types of stance phrases in agriculture and economics corpora respectively. Columns 5 and 6 list the percentage of each category in the cognitive stance used in each corpus. Columns 7 and 8 report the normalized mean frequency of the cognitive stance subcategories in the two corpora. In order to identify whether there is a statistical difference in the use of cognitive stance categories, I report the findings of the LL score (Column 9) and Chi-square (Column 10), and the sig value (Column 1). The effect size for the Chi-Square is presented in Column 12.

As indicated in Table 2, the two disciplines share similarities in terms of the frequency of the subcategories usages within cognitive stance phrases. Firstly, the two disciplines present an identical ranking order in terms of the use of the subcategories, that is, analysis phrases ranked in the first place, followed by judgment, inference and metacognition phrases. The congruent order shows that writers in the two disciplines shared similar patterns concerning authors' involvement in the different forms of cognitive stance. In addition, a roughly similar proportion was found in terms of the percentage of each subcategory, that is, analysis stance, as illustrated in Example (13), accounted for about one half; judgment as in (14) accounted for about one third, inference in (15) accounting for slightly above 10%, and metacognition in (16) taking up the smallest proportion with a percentage of 2.05% and 5.62% respectively in agriculture and economics respectively. It is interesting to note that analysis, judgment and inference phrases occupied an overwhelmingly dominant percentage, with 97.95% and 94.38% in agriculture and economics corpora respectively. The prominent use of the three subcategories of stance phrases indicates that these disciplinary writers have a tendency to present significantly more knowledge-oriented logic analysis than subjective mental processing acts. Logic is generally accepted as an indispensable tool to construct knowledge and justify authors' arguments.

(13) Neither approach may be an ideal model for gradual shifts in climate over decades-to-centuries. However, the results described above are largely ***consistent with*** studies of long-term trends at individual sites. (ARAC#003:1)

(14) This paper examines the local properties of perfect foresight equilibrium of a economics constrained economy featuring two classes of infinitely-lived agents with heterogeneous general preferences. It is primarily ***concerned with*** the conceivability of endogenous fluctuations for large plausible capital-labor elasticities of substitution. (FRAC#021:1)

(15) *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)

(16) Finally, following Berger et al. (2000), we ***take into account*** the home country of foreign banks in order to test the limited global advantage hypothesis. (FRAC#056:3)

Apart from the similarities, the cross-corpora comparison also shows significant differences in the use of the analysis, inference and metacognition phrases in the two disciplines. As shown in Table 2, analysis stance phrases are found to be significantly more frequent in the agriculture corpus

(LL=311.92; Chi-square=309.78; p=0.00). This shows that agriculture writers tend to display their analytical thinking, a presence conveyed by analysis stance, towards statements more frequently. Similarly, inference phrases are found to be significantly more frequently used in agriculture corpus, indicating that agriculture writers are more inclined to use inference markers to interpret or generalize propositions in their academic discourse. In contrast, metacognition phrases are more significantly preferred in the economics corpus (LL=-214.15; Chi-square=214.04; p=0.00), reflecting that economics writers are likely to display more explicit presence of their mental processing act in their discourse construction. With regard to the magnitude of significance, the effect size statistic shows the disciplinary differences in the use of the analysis, judgment, and inference phrases in the two corpora are small (w = 0.17, 0.02, and 0.04 for analysis, judgment, and inference respectively). In contrast, the degree in the difference in the case metacognition phrase is large (w = 0.52).

The differing degree of involvement in the two corpora could be explained by the research practices in the two academic disciplines. Analysis stance phrases are significantly more frequently used in agriculture. Also, the higher frequency of inference phrases in the agriculture corpus indicates agriculture papers tend to require a higher presence of writers' inferential presence in knowledge construction and argument development. In contrast, the economics texts are seen to contain significantly more phrases related to authors' metacognitive presence. This shows that economics writers have more explicit self-presence in the form of authors' cognitive processing in the course of organizing the texts.

Table 2 The Frequency of Cognitive Stance Phrases in Agriculture and Economics Corpora

Cognitive	Raw frequency		Normalized frequency (per million)		Percentage		Mean		LL	Sig	Chi-square	Sig	Effect size
	Agr	Fin	Agr	Fin	Agr	Fin	Agr	Fin					
analysis	10625	6541	6364.58	4834.54	55.33%	47.97%	442.44	335.95	311.92	0.00	309.78	0.00	0.17
judgment	5802	4526	3475.51	3345.23	30.21%	33.19%	34.97	28.16	3.72	0.05	3.72	0.05	0.02
inference	2384	1802	1428.06	1331.88	12.41%	13.22%	677.40	549.06	5.00	0.03	5.00	0.03	0.04
metacognition	393	767	235.41	566.90	2.05%	5.62%	223.90	255.81	214.15	0.00	214.04	0.00	0.52
Cognitive total	19204	13636	11503.56	10078.55	100.00%	100.00%	197.12	167.45	144.01	0.00	141.20	0.00	0.08

Note: The following abbreviations are used: Agr (agriculture), Fin (economics).
Sig stands for p-value; The LL ratio and Chi-Square scores are significant at the 0.05 level.

5.2.2 Attitude stance phrases

Table 3 presents the statistical comparison of attitude stance phrases in the corpora of agriculture and economics. The first two columns show the raw frequency of the four types of attitude stance phrases extracted from the two corpora. Columns 3 and 4 present their normalized frequency per million words in agriculture and economics corpus respectively. Columns 5 and 6 list the percentage of each category in the total attitude stance in each corpus. Columns 7 and 8 report the normalized mean frequency of the attitude stance subcategories in the two corpora. In order to identify whether is a statistical difference in the use of attitude stance categories, the study presented the findings of the LL score (Column 9) and Chi-square (Column 10), and the sig value (Column 11). The effect size for the Chi-Square is presented in Column 12.

In contrast with the overwhelming proportion of cognitive stance, the total frequency of attitude stance phrases accounts for the lowest percentage, slightly above 10%. As shown in Table 3, RAs in the two corpora shared the following similarities with respect to the proportion of each subcategory: first of all, *expectation* phrases as in (18) ranked second, with 23.16% and 37.55% in agriculture and economics, respectively, followed by *intuition* stance as in (17), with about 16.44% and 19.14% in the agriculture and economics respectively. As for the magnitude of significance in the use of the three types of attitude stance phrases, the result shows that the difference in using *intuition* and *expectation* is small. The result also shows that the *expectation* phrases constitute the second-most favored strategy in sharing authors' attitude, and this indicates that the writers seemed to be cautious in expressing their intuition-based attitude in academic articles.

(17) Moreover, keeping in view developmental problems like burgeoning population, growing food demand, shrinking natural resources, ***it is necessary*** *to* sustain the production of crop yield as well as soil health in an eco-friendly way. (ARAC#119:2)

(18) This study ***is expected to*** result in recognizing the impact of pollution due to these pesticides and fungicides in the marine environment, using D. faba as a bioindicator. (ARAC#030:2)

Table 3 The Frequency of Attitude Stance Phrases in Agriculture and Economics Corpora

Attitude	Raw frequency		Normalized frequency (per million)		Percentage		Mean		LL	Sig	Chi-square	Sig	Effect size
	Agr	Fin	Agr	Fin	Agr	Fin	Agr	Fin					
intuition	595	452	361.21	334.08	41.52%	33.76%	32.84	22.27	1.08	0.30	1.077	0.30	0.04
expectation	838	887	501.98	655.59	58.48%	66.24%	41.83	54.63	-30.7	0.00	30.92	0.00	0.16
total	1433	1339	863.19	989.67	100.00%	100.00%	74.67	76.9	13.98	0.00	14.06	0.00	0.09

Note: The following abbreviations are used: Agr (agriculture), Fin (economics).
Sig stands for p-value; The LL ratio and Chi-Square scores are significant at the 0.05 level.

The analysis shows that the agriculture corpus contains more *intuition* phrases, suggesting that agriculture writers are more inclined to present their intuitive evaluation when reporting research findings in constructing their academic discourse. The difference may be explained by the research paradigms subscribed to by these two disciplines. Agriculture studies seemed to employ a large amount of first-hand data which were collected by writers, who then interpret the dataset and comment on it. In contrast, economics may regard it as a convention to report findings and refrain from too much intuitive interpretation and comment.

On the other hand, the statistical analysis of *expectation* phrases shows that the economics corpus contains a higher frequency of *expectation* phrases than the agriculture corpus. The result indicates that writers in economics are more likely to express their involvement in academic discourse.

5.2.3 Hedges stance phrases

While presenting their reason-oriented projection by cognitive stance and displaying emotion-oriented presence with attitude stance, authors in the two disciplines were also found to mitigate their commitment in manifesting some assertions by means of hedges.

Table 4 presents the statistic comparison of the hedging stance phrases in the corpora of agriculture and economics. The first two columns show the raw frequency of the four types of hedging stance phrases extracted from the two corpora. Columns 3 and 4 present their normalized frequency per million words in agriculture and economics corpus respectively. Columns 5 and 6 list the percentage of each hedging stance in each corpus. Columns 7 and 8 report the normalized mean frequency of the hedging stance subcategories in the two corpora. In order to identify whether is a statistical difference in the use of hedging stance categories, the study presented the findings of the LL score (Column 9) and Chi-square (Column 10), and the sig value (Column 11). The effect size for the Chi-Square is presented in Column 12.

As for the percentage of the subcategory of hedges stance phrases, agriculture and economics writers are found to display a surprisingly similar proportion of possibility as in (19) and imprecision hedges as in (20), with about 35% and 65% respectively. The similar proportion in using the possibility and imprecision hedges indicates that the writers in two disciplines share a similar epistemological pattern when mitigating their presence by hedges makers. To be more specific, the dominant use of imprecision stance suggests the academic writers in the two disciplines tend to express more approximations or conditions for certain statements to "make things fuzzier" (Lakoff, 1973, p.471). In the two corpora, about one third of hedges are used to express authors' uncertainty and tentativeness as a way of showing authors' caution in expressing claims or statements. The use of hedges contributes to establishing a "conceptual framework within which a particular world-view or state of affairs is qualified" (Hoey, 1997, p.41)

(19) 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)

(20) The theory helps to explain how (though constrained by the external environment) organizations strive to reduce environmental interdependence and uncertainty through strategies that, ***at least*** partially, enact their environment. (FRAC#042:1)

The cross-corpora comparison shows that the agriculture corpus consists of significantly more hedge phrases than the economics corpus, suggesting that agriculture writers tend to display a higher degree of mitigated presence. As shown in Table 4, both possibility and imprecision hedges, the two subcategories, are significantly more frequently used in the agriculture corpus than that in the economics corpus. However, as indicated by the value of effect size, the degrees of the difference in using the two subcategories of hedge stances are small (w = 0.12 and 0.13 for possibility and imprecision subcategories respectively).

Table 4 The Frequency of the Hedges Stance Phrases in Agriculture and Economics Corpora

Hedges	Raw frequency		Normalized frequency (per million)		Percentage		Mean		LL	Sig	Chi-square	Sig	Effect size
	Agr	Fin	Agr	Fin	Agr	Fin	Agr	Fin					
possibility	2725	1804	1632.33	1333.36	34.42%	34.69%	102.02	83.34	44.96	0.00	44.64	0.00	0.12
imprecision	5191	3396	3109.51	2510.03	65.58%	65.31%	239.19	193.08	95.41	0.00	94.80	0.00	0.13
hedges	7916	5200	4741.83	3843.39	100.00%	100.00%	163.51	132.53	140.27	0.00	139.61	0.00	0.13

Note: The following abbreviations are used: Agr (agriculture), Fin (economics).
Sig stands for p-value; The LL ratio and Chi-Square scores are significant at the 0.05 level.

Such differences may be attributed to the different research approaches adopted in the two disciplines. Studies in agriculture RAs are often carried out in natural environments or labs, thus many natural factors, such as rainfall, temperature, latitude, and locations, may affect the research results to some degree. These experimental conditions may help agriculture researchers develop a cautious attitude in reporting and interpreting their research findings, due to many potential confounding variables. The academic papers of geologists, who share similar experience of interacting with nature, are also found to use a high frequency of hedges as "a rhetoric of understatement" (Dressen, 2003, p.278). In contrast, economics studies are more likely to employ particular models or tools to solve the economics problems. Therefore, the findings are less subject to the changes than those in the agriculture studies, which thus enable a more confident tone in economics' reporting of their research findings.

5.2.4 Reference stance phrases

Table 5 presents the statistic comparison of the reference stance phrases in the corpora of agriculture and economics. The first two columns show the raw frequency of the four types of reference stance phrases extracted from the two corpora. Columns 3 and 4 present their normalized frequency per million words in agriculture and economics corpus respectively. Columns 5 and 6 list the percentage of each reference subcategory in each corpus. Columns 7 and 8 report the normalized mean frequency of the stance categories in the two corpora. In order to identify whether is a statistical difference in the use of reference stance categories, the study presented the findings of the LL score (Column 9) and Chi-square (Column 10), and the sig value (Column 11). The effect size for the Chi-Square is presented in Column 12.

The analysis of reference stance phrases, shown in Table 5, shows that writers in the two corpora share the following similarities in terms of the frequency ranking and the percentage of the different subcategories. First of all, both agriculture and economics writers are inclined to flag their explicit engagement by *self-mention* as in (21), with 44.86% and 61.10% of the total reference phrases in agriculture and economics respectively. Also, they are found to both employ fewer *extra-reference* phrases as shown in (22), with 11.65% and 5.78% in agriculture and economics respectively. Considering that extra-reference phrases are mainly used to point claims to sources other than the study, the percentage difference between *self-mention* and *extra-reference* stance indicates that the academic writers tend to present more of their own propositions than quoting from other statements. The authors are also found to use *intra-textual* phrases, as in (23), as their second most frequently strategy, with the percentage of 42.19% and 31.86% of the reference phrases in the corpora of agriculture and economics RAs respectively. This shows that writers have a preference for making use of the intra-textual phrases to orient readers throughout the text description by communicating to readers "where they have been in the text and where are they going" (Johns, 1997, p.135) and to

construct a discursive, coherent and well-organized information flow. Finally, academic writers in the two disciplines are found to both employ a similar low frequency of shared knowledge phrases as in (24), with 1.29% and 1.26% respectively in the two corpora, to present their common knowledge in their own disciplines.

The effect size statistic shows a differing degree of significance in the use of the subcategories of reference stance phrases. Specifically, the effect size of *self-mention* and *extra-reference* are medium ($w = 0.32$ and 0.35 respectively), while the magnitude of the significant differences in *shared knowledge* and *intra-textual* are small, with a value of 0.11 and 0.04 respectively.

(21) Climate system feedbacks resulting from phenological shifts influencing the seasonal course (cf. Lee et al., 2011) of surface roughness length have not, ***to our knowledge,*** been fully quantified. (ARAC#003:2)

(22) ***It is widely accepted that*** well-functioning financial markets depend on sound legal and government institutions. (FRAC#074:1)

(23) However, the benefit owing to the smaller-size molecules is limited, ***as can be seen*** from Fig. 4b, in which significant amounts of ketones and sulfoxides were generated from both asphalt and lignin molecules when the temperature is raised. (ARAC#125:3)

(24) Finally, ***previous studies*** revealed that firms have an incentive to manage earnings during the time when a firm pursues a seasoned equity offering (SEO) (Teoh et al., 1998), so we control for this effect by including a variable that indicates whether a firm pursues an SEO in a particular year. (FRAC#054:2)

Table 5 The Frequency of Reference Stance Phrases in Agriculture and Economics Corpora

Reference	Raw frequency		Normalized frequency (per million)		Percentage		Mean		LL	Chi-square	Sig	Effect size
	Agr	Fin	Agr	Fin	Agr	Fin	Agr	Fin				
self-mention	1563	2136	936.27	1578.75	44.86%	61.10%	117.03	197.34	250.31	252.36	0.00	0.32
shared knowledge	45	44	26.96	32.52	1.29%	1.26%	2.25	2.71	0.78	0.79	0.38	0.11
intra-textual	1470	1114	880.56	823.37	42.19%	31.86%	62.90	28.39	2.87	2.86	0.09	0.04
extra-reference	406	202	171.32	123.43	11.65%	5.78%	34.26	30.86	33.62	35.76	0.00	0.35
total	3484	3496	2015.10	2558.07	100.00%	100%	51.67	48.27	79.47	80.10	0.00	0.13

Note: The following abbreviations are used: Agr (agriculture), Fin (economics).
Sig stands for p-value; The LL ratio and Chi-Square scores are significant at the 0.05 level.

The cross-corpus comparison shows that the two academic corpora display significant differences in the usage of reference phrases. To be more specific, *extra-reference* phrases are significantly more frequently used in the agriculture corpus, suggesting that agriculture writers have a higher tendency to assert or present propositions by appealing to authority to make their statements more convincing and persuasive. In contrast, the writers in economics contain a high frequency of *self-mention* phrases, indicating that economics writers are more likely to manifest their own involvement in their text construction. One possible reason for such frequent occurrence of self-mention stance in the economics corpus is that most of the research materials are based on publicly accessible databases, such as, the annual government report, national survey, which are often analyzed and processed by computer or mathematics models. Hence, the writers tend to exert a relatively higher degree of control in the research process than their agriculture peers. The high occurrence of *self-mention* can also indicate that it may be a conventional practice for economics writers to present their explicit self-presence in reporting their findings and organizing their academic discourse.

5.3 Syntactic patterns of stance phrases

Aside from the statistical similarities and differences in the use of the stance phrases, the present study also identified some salient syntactic patterns associated with the stance phrases. The analysis of the stance phrases reveals notable syntactic patterns associated with the stance phrases. (The following section presents the patterns in terms of each stance category.)

The cognitive stance phrases, particularly inference and metacognition phrases, are found to be associated with the following syntactic patterns. Firstly, as can be found from the concordance lines, more than two thirds of inference stance expressions are collocated with *evidence, result* or *it* as subjects, i.e., *the results indicate that*. The structure foregrounds the evidence or objective structure for certain statements or claims delivered in the clauses. The usage of this syntactic pattern supports Halliday's (1994, p.342) view that most of the cognitive phrases are associated with the clause, working as a structural unit expressing a particular range of ideational meanings. Secondly, the metacognition stance phrase is identified to be notably associated with the introductory *it* pattern (namely, *it is verb (pp) tha*t, i.e., *it is noted that, it is argued that,* and *it is considered that*), including eight out of thirteen phrase entries. By integrating the impersonal grammatical structure, denoted by *it,* and the subjective mental behavior verb, e.g., *consider*, the syntactic pattern functions to downplay the author's subjective intrusion, thus leaving an impression of objectivity and contributing to a persuasive function of academic discourse.

Secondly, within the use of attitude stance phrases, writers in the two disciplines seem to favor the grammatical pattern *it is adj. to/that*, i.e., *it is interesting to/that, it is important to/that,* and *it is difficult to/that.* Slightly more than one quarter of the *intuition* stance phrases entries are found to

contain this structural pattern. The balance between the objective impersonal structure, represented by *it*, and the subjective intuition attitude, for example, *interesting,* could work as a "rubber-glove" (Johns, 1997). In other words, by allowing authors to depersonalize their opinions and mitigate their subjective intrusion, the pattern allows them to maintain a certain distance from ideas expressed and thereby fulfill the requirement of objectivity and persuasion required in academic written texts.

Similar to the grammatical patterns in attitude stance phrases, the impersonal structure, *it is adj. to/that,* i.e., *it is possible to,* and *it is likely that,* also constitutes a notable structure in hedge phrases. In the phrases retrieved in this study, 7 out of the 16 hedge phrase entries were found to contain this structure. The use of this structure is in line with the function of hedges, that is, to minimize the presence of authors in the statements. In addition, verb phrases, such as *tend to* and *seem to,* with a percentage of 25% of the phrase entries, are also found to be frequently used among the hedges phrases to express tentativeness and downplay authors' presence. Adverbial phrases, such as, *to some extent, to date, at least,* and *in general* also constitute a salient pattern in the stance phrases, particularly in expressing imprecision hedges stance. Most of the imprecision phrases, with about 92.3% (12 out of the 13 phrases), are adverbial phrases. By indicating the degree and condition for a certain statement or proposition, those patterns function to lessen the force of tone in delivering the proposed statements. Their use also reflects authors' rigorous attitude and prudent approach to reporting findings and putting forward new or different propositions.

As far as reference stance phrases are concerned, it was observed that self-mention phrases are often associated with evidence-related subjects. As shown in the phrases list, about 87.5% (seven out of eight) self-mention phrases entries are associated with the evidence-based expressions, such as *the results of the present study, our data.* This combination of the self-mention and the evidence-related expressions functions to underscore writers' contribution and manifest their involvement. Also, such combination is found to have an inclination to occur at the beginning of sentences in the corpora. Such foregrounded positions may contribute to highlighting the findings obtained in their study, and direct readers' attention to the assumed valuable information, such as main contribution and significance (Hyland & Tse, 2005).

5.4. Summary

The comparison between the agriculture and economics papers shows that the two disciplines not only share academic conventions but also display some disciplinary variation in stance phrases usage. Specifically, they both attach great importance to presenting reason-oriented projection by cognitive stance, mitigating their presence with hedging phrases, enabling readers to grasp the source of statements by employing reference phrases, and sharing their evaluation and comment with attitude phrases. The two academic corpora are also found to include more reason-oriented than emotion-

oriented stance phrases as well as more objective than subjective stance phrases. When it comes to the disciplinary difference, agriculture writers use analysis and inference cognitive stance more frequently to present their analytical and speculative processing of the information, more hedges to mitigate their commitment in particular statements, more extra-reference stance to point out the source to win support for some statements. In contrast, economics academic papers are more inclined to make use of *self-mention, metacognition* and *expectation* stance phrases. This reflects that economics writers are inclined to display their explicit engagement in reporting findings and asserting claims.

In terms of syntactic patterns, impersonal structures are favorably used in the two disciplines to display authors' prudence and objectivity in the process of expressing their subjective attitude or mental processing act towards information or knowledge. By "instantiating the aspects of community's values" (Hyland, 1999b), the stance phrases not only help writers to grasp the community strategies required as a community member, but also to orientate readers' focus to the key points conveyed in the papers. The conventionalized appropriate use of the stance phrases in the discourse may reveal the writers' efforts to "fit in the community by measuring up the expectations and yards of the community" (Hunston, 1993).

Overall, the analysis of stance phrases identified in the present chapter enriches our understandings of how authors position themselves by means of the four aspects of self-presence, and shows how writers interact with the expected readers in the academic communities. The following chapter investigates stance phrases in terms of their textual position based on Hoey's (2005) priming theory.

Chapter 6 The textual colligation of stance phrases in the two academic disciplines

In this chapter, I address research question 2, namely, "(a) What is the textual colligation of stance phrases in the agriculture and economics corpora? (b) What disciplinary variation can be found in the textual colligation of stance phrases?"

To do this, I first present the background for the studies carried out from this perspective (Section 6.1), then I discuss the theoretical framework of textual colligation in Section 6.2. Section 6.3 describes the methods used to retrieve the textual colligation of stance phrases and the approaches of the textual division. The results concerning the textual colligation of each category of stance phrases are presented in Section 6.4. Section 6.5 discusses these findings in detail.

6.1 Introduction

A key aspect of recurring linguistic features is the association between words or phrases and their textual positions. Lexical items, according to Hoey (2005, p.13) "are primed to occur in or avoid, certain positions within the discourse". An analysis of textual colligation, the term Hoey (2005) uses to denote such priming, explores the textual position of linguistic markers in relation to textual structures, and may also examine the interaction between the textual position and discourse functions.

Previous studies on textual colligation have mainly centered on particular words or phrases (e.g., Hoey & O'Donnell, 2008; Mahlberg, 2009; O'Donnell et al., 2012). While insights from such studies have enriched our understanding of the textual colligation of particular linguistic features, most studies have employed the key word procedure (Scott, 1997)[5] to identify words or phrases to be examined. This approach privileges the relative frequency of lexical items over their semantic or functional contribution to the text. Comparatively little attention, however, has been paid to the textual colligation of linguistic items typical of one particular semantic group. This is perhaps unsurprising, given that textual colligation is a relatively recent focus of research.

Therefore, this chapter investigated the textual colligation of stance phrases, based on the purpose-built corpus of research articles from the disciplines of the agricultural sciences and economics. (The specific information of the corpora is available in Section 3.1). With the objective of providing a cross-disciplinary focus, this chapter compared the disciplinary differences in terms of the stance phrases used in the two corpora. Textual colligation, according to Hoey (2005, p.13), is domain

[5]Key word procedure: When identifying the salient lexical items for textual analysis, previous studies mainly (i.e., Hoey & O'Donnell, 2008, 2015) followed Scott (1997); that is, they employed the key word function provided by Wordsmith Tools (Scott, 1996). This function involves computing word lists from a reference corpus and the corpus under scrutiny, comparing the frequency of individual words in the two lists. Statistical analyses, such as chi-square, are used to calculate the keyness.

specific. It thus seems reasonable to postulate that variation may be found in the text contexts typical of stance expressions in the two disciplines analysed in this thesis (agriculture and economics).

6.2 Textual colligation

According to the lexical priming theory proposed by Hoey (2005), words or phrases are "primed to occur in, or avoid, certain positions within the discourse; these are its textual colligations" (p.13). One of the key components of lexical priming is textual colligation. This refers to the phenomenon that words or phrases are primed to either occupy or avoid certain recognised discourse positions" (Hoey & Donnell, 2015, p.125), specifically "to occur (or to avoid occurring) in the beginning or end of independently recognised discourse units, e.g.,the sentence, the paragraph, the speech turn" (Hoey, 2005, p.115). The theory focuses on the distributional aspect of linguistic units of analysis in texts and facilitates insights into the position of target structures in particular discourse types.

With regard to the linguistic unit of analysis receptive to lexical priming, Hoey (2005) refers in general terms to 'lexical items'. Previous studies (Hoey, 2005; Hoey & O'Donnell, 2008; 2015) have demonstrated that the priming procedure occurs at the level of both single words (i.e., *consequence*, *move*) and phrases (i.e., *in winter, in the winter, according to, the move, it was announced yesterday*). This raises a question as to the boundary or the scope of 'lexical items' within priming theory.

Divergence is found in the identification of the basic unit of analysis in discourse. Biber et al. (1999) employ the term 'lexical bundle' for recurring sequences of three or more words. Previous studies (Biber et al., 2004; Hyland, 2008a) have shown that this approach can retrieve lexical, grammatical and functional patterns. However, a potential issue with the retrieval of lexical bundles is that short word sequences are often not favoured when retrieved in longer strings. For example, two-word sequences (like *tend to*), when retrieved at a pre-set length of a four-word bundle, often do not occur as frequently as they do in the 2-word sequence retrieval, and thus may not be identified as a salient feature or may not even pass the threshold criteria. N-grams have also been used previously (i.e., O'Donnell et al, 2012; Simpson-Vlach & Ellis, 2010) to refer to linguistic units of various length, usually comprising consecutive words; non-consecutive constructions, however, also constitute a noteworthy unit of analysis, as shown by Römer (2011) and Römer and O'Donnell (2010).

6. 3 Methods

This section presents the retrieval of the textual colligation of stance phrases at the level of sentence, paragraph and whole text, as well as the statistical comparison of the textual colligation of the stance phrases in the two corpora.

6.3.1 Retrieval of stance phrases and their textual colligation

To identify the textual colligation of these stance phrases, this study employed *WordSkew* (Barlow, 2016). *Wordskew* is a text analysis tool to identify the particular position of words or phrases in terms of their distribution at sentence, paragraph, and text levels. The advantage of this software is that it can specify the units of discourse and it enables automatic identification of the textual positions of predetermined search items.

When searching the textual colligations of the stance phrases, the two corpora were loaded to the software respectively and then the searching of each phrase was carried out at sentence, paragraph and text level. In determining the initial, medial and final positions of sentence and text, I specified the percentage in sentence and textual distribution (as elaborated in Section 3.2). With regard to the distribution at paragraph level, I adopted the absolute slot. In practice, I set the 'absolute slot'[6] to be "1 Other #" in the software, where "1" refers to the first sentence, "#" the last sentence, and "other" the rest of a paragraph.

6.3.2 Analysis

The text division in terms of initial, medial and final position at the level of sentence, paragraph and text mainly followed the approach used by Römer and O'Donnell (2010), and Thompson (2014). Table 6 displays the divisions used with respect to the initial, medial and final positions at the level of sentence, paragraph and text. At sentence level, a percentage division was adopted. That is, if a stance phrase is in the first 25% of a sentence, it is determined to be in sentence-initial position, sentence-medial position if it belongs to the middle 50% of a sentence, and in sentence-final position if it is in the last 25%.

At the paragraph level, I used the approach of absolute slots. That is, if a stance phrase is in the first sentence of a paragraph, it is defined as being in the paragraph-initial position, and paragraph-final if it is in the last sentence; otherwise, it is defined as being in the paragraph-medial position. With regard to the segmentation of the complete texts, a text-initial position is defined to be in the first 15% percentage of a text based on word counts; the text-medial position is in the middle 70% of the text, and the last 15% is defined the text-final position.

[6] According to Barlow (2016), absolute slot refers to the first linguistic units, such as a word/phrase, sentence, or paragraph in a sentence, paragraph, or text respectively. This division method divides the textual position of particular linguistic items by examining whether the item in the first or last slot in a sentence. Here, with reference to the paragraph, the first absolute slot refers to the first sentence in a paragraph, and the last absolute slot refers to the last sentence in a paragraph.

Table 6 Sentence, Paragraph and Text-level Divisions

Textual position	Initial	Medial	Final	Previous practice
Sentence	First 25%	Middle 50%	Final 25%	Römer & O'Donnell (2010) Hoey & O'Donnell (2015)
Paragraph	1st sentence	Remaining sentences	Last sentence	Römer & O'Donnell (2010) Thompson, 2014
Text	First 15%	Middle 75%	Final 15%	Römer & O'Donnell (2010)

When examining the disciplinary variation in the occurrence of stance phrases in different textual positions, I performed a chi-square statistical test, using raw frequencies and the corpus sizes. This chapter also employed effect size calculations in order to report on the degree of difference between variables, in this case, the textual colligation of stance phrases in the agriculture and economics corpora. In interpreting the effect size, the chapter followed Cohen's (1988) magnitude guidelines, that is, a value of 0.1 and below is considered a small effect, 0.3 a medium effect, and 0.5 and above a large effect.

6.4 Results

This section presents the textual positions of the four stance categories: cognitive, attitude, hedges and reference phrases, and the functions that these phrases exercise in the text. A statistical comparison is conducted in terms of the occurrences of the different types of stance phrases in the different textual positions of the agriculture and economics corpora. The similarities and differences are discussed from the perspective of the different academic practices in the two disciplines.

6.4.1 Cognitive stance phrases

The proportional distribution of cognitive stance phrases is presented in Figure 4. Clearly, cognitive stance phrases in agriculture and economics corpora display notable similarities in terms of sentence, paragraph and textual distribution. The sentential distribution analysis shows that about half of the cognitive stance phrases occur at sentence-medial positions, and substantially more cognitive phrases occur at sentence-initial than in the final positions.

The analysis of the paragraph position shows that cognitive stance phrases overwhelmingly occur in the paragraph-medial position. The percentage of cognitive phrases occurring in the paragraph-initial and final positions of the economics corpus is slightly higher than those in the agriculture corpus.

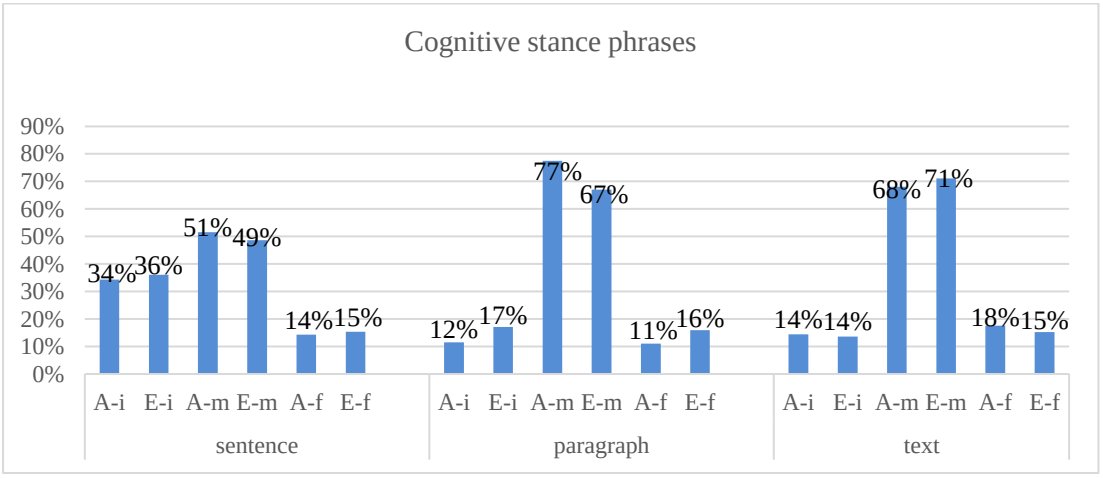

Figure 4 The Proportional Distribution of Cognitive Phrases in Different Textual Positions

Note: A=Agriculture, E=Economics, i = initial, m = medial, f = final

Similarly, with respect to the textual position, the text-medial position is favoured in both corpora. The percentage of stance phrases distributed in the text-final positions is found to be slightly higher than those that occur in the text-initial positions.

Table 7 displays the results of the statistical analysis of the occurrences of textual colligation of the cognitive phrases in the two disciplines. The occurrence of phrases in sentence-initial, medial and final positions of the agriculture corpus is significantly greater than those in the economic corpus ($p = 0.00$, $p = 0.00$, and $p = 0.04$ respectively[7]). The effect sizes are relatively small, however. This signals that while differences exist in sentence position of the cognitive stance phrases, the degree of significance is small.

Table 7 Cross-corpus Comparison of Cognitive Phrases

Text unit	Textual position	Norm freq. A	Norm freq. E	X^2	Sig		Effect size
sentence	initial	3942.10	3633.13	18.79	0.00	***	0.04
	medial	5922.68	4903.05	142.80	0.00	***	0.09
	final	1638.78	1542.38	4.36	0.04	*	0.03
paragraph	initial	1324.57	1725.36	79.89	0.00	***	0.13
	medial	8911.62	6753.08	441.66	0.00	***	0.14
	final	1267.35	1600.10	58.47	0.00	***	0.12
text	initial	1655.34	1368.63	40.16	0.00	***	0.09
	medial	7834.50	7171.01	44.00	0.00	***	0.04
	final	2013.71	1538.92	93.81	0.00	***	0.13

Note: A=Agriculture, E=Economics, X^2 = chi-square

In the paragraph initial and final positions, the occurrences of the cognitive stance phrases are

[7] The p-values are all below the threshold level of 0.05.

significantly higher in the economics corpus than those in the agriculture corpus, while the degree of the significant differences are at medium level (w = 0.13 for the paragraph-initial; w = 0.12 for the paragraph-final positions). In contrast, the agriculture corpus contains significantly more cognitive phrases in the paragraph-medial positions (p = 0.00), although the effect size is small (w = 0.14). With respect to the distribution at the text level, the agriculture corpus contains significantly more phrases in the text initial (p = 0.00, w = 0.09), medial (p = 0.00, w = 0.04), and final positions (p = 0.00, w = 0.13). Effect sizes here are also small.

6.4.2 Attitude stance phrases

Figure 5 presents the proportional distribution of attitude stance phrases with regard to their different textual positions. More than half of the attitude stance phrases occur in sentence- medial position. With regard to the comparison between sentence-initial and final positions, the analysis shows that the proportion of attitude phrases occurring in sentence-initial position is much higher than that in sentence-final positions in the two disciplines.

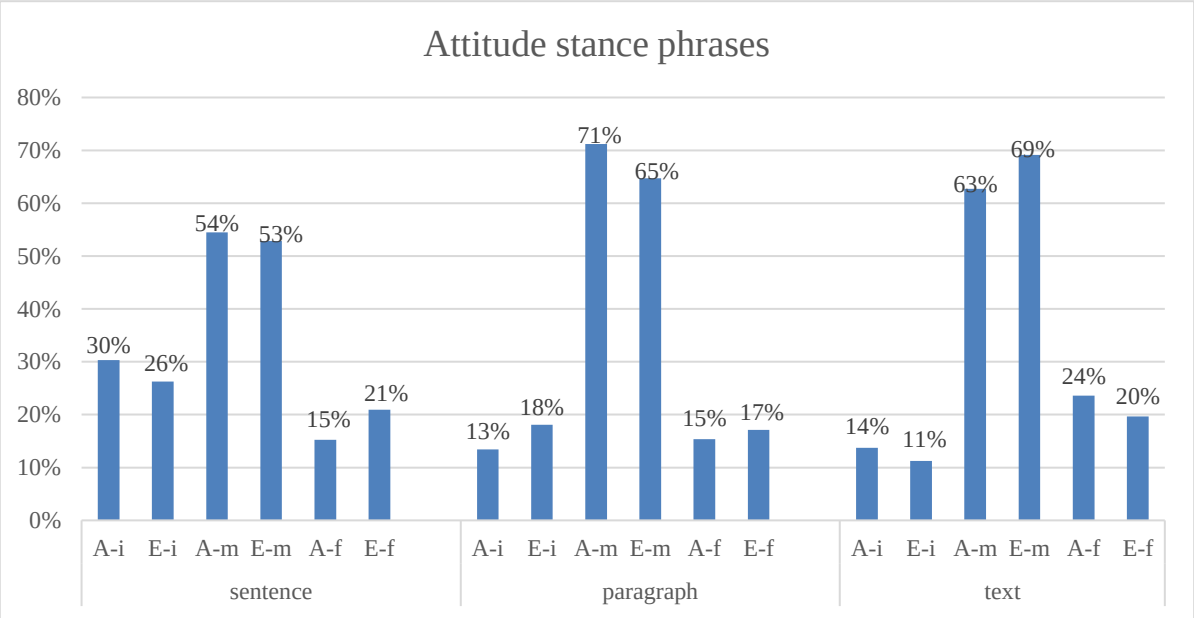

Figure 5 The Proportional Distribution of Attitude Phrases in Different Textual Positions

Note: A=Agriculture, E=Economics, i = initial, m = medial, f = final

The analysis of the paragraph positions shows an overwhelmingly higher percentage of attitude phrases in the paragraph medial position in the two disciplines. With respect to the textual position of the attitude stance, the analysis shows that the largest percentage of attitude phrases occur in the text medial position. The second favourable textual position is the text final position, with an average percentage of 23.6% and 19.6% in agriculture and economics corpora respectively. Least favoured is the text initial position, with just over 10% of attitude phrases occurring here (13.7% and 11.2% in the agriculture and economics respectively).

Table 8 Cross-corpus Comparison of Attitude Phrases

Text unit	Position	Norm freq. A	Norm freq. E	X^2	Sig.		Effect size
sentence	initial	657.47	457.98	52.34	0.00	***	0.17
	medial	1180.51	922.21	46.85	0.00	***	0.12
	final	329.87	365.60	2.77	0.10		0.05
paragraph	initial	291.45	316.29	1.50	0.22		0.04
	medial	1543.70	1130.23	94.21	0.00	***	0.15
	final	332.69	299.28	2.58	0.11		0.05
text	initial	297.10	196.12	30.43	0.00	***	0.20
	medial	1359.56	1206.75	13.53	0.00	***	0.06
	final	511.18	342.92	48.43	0.00	***	0.19

Note: A=Agriculture, E=Economics, X^2 = chi-square

Table 8 displays the Chi-square and effect size of the attitude stance phrases distributed in different textual positions. As can be seen, there are significant differences in the textual colligation of the attitude phrases. Specifically, significantly more attitude phrases are distributed in sentence-initial position (p = 0.00, w = 0.17) and sentence-medial (p = 0.00, w = 0.12) in the agriculture academic corpus than in the economics corpus. Also, although a higher incidence of occurrence of attitude phrases is found in economics corpus than that in the agriculture corpus, no significant difference is observed between the two corpora in sentence-final positions.

With regard to the paragraph position, the statistical analysis shows that significantly more attitude phrases occur in the paragraph-medial position in the agriculture corpus (p = 0.00; w = 0.15). Analysis of the textual positions showed that the agriculture corpus contains significantly more phrases which occur in the text-initial (p = 0.00, w = 0.20), medial (p = 0.00, w = 0.06), and final positions (p = 0.00; w = 0.19) than in the economics corpus.

6.4.3 Hedges stance phrases

The proportional distribution of hedges is presented in Figure 6. As can be seen, more than half of the hedges phrases occur in sentence-medial position, with the percentages about 54.4% and 54.3% respectively. The second favoured position is sentence-initial, while only about 13.0% and 17.2% in the agriculture and economics corpora occur in sentence-final positions. In contrast, the instance of hedge phrases in sentence-final position is slightly higher in the economics than in the agriculture corpus.

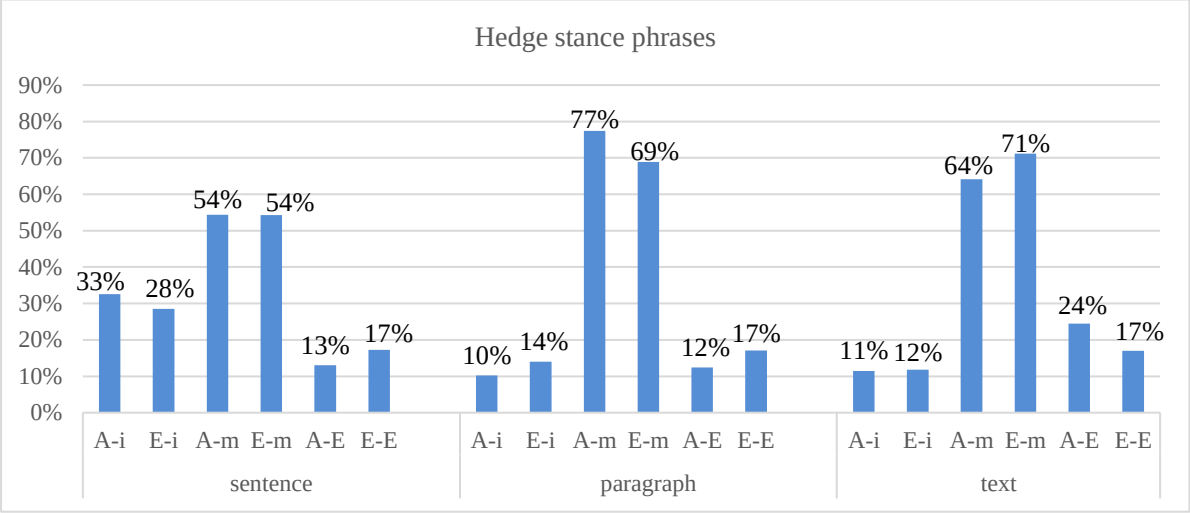

Figure 6 The Proportional Distribution of Hedges Phrases in Different Textual Positions

Note: A=Agriculture, E=Economics, i = initial, m = medial, f = final

The paragraph level analysis shows that the two disciplines have a clear preference for paragraph medial positions. With regard to the final and initial positions, writers in the disciplines express proportionally slightly more hedged expressions in the paragraph final than in the beginning positions. Similarly, the text analysis shows that the text medial position contains a substantial percentage of hedging phrases in both agriculture and economics corpora.

Table 9 Cross-corpus Comparison of Hedge Phrases

Text unit	Textual position	Norm freq. A	Norm freq. E	X^2	Sig.		Effect size
sentence	initial	1545.24	1094.43	113.20	0.00	***	0.17
	medial	2578.98	2086.90	76.72	0.00	***	0.10
	final	617.61	662.06	2.34	0.13		0.03
paragraph	initial	482.88	539.23	4.74	0.03	*	0.06
	medial	3668.74	2648.25	243.19	0.00	***	0.16
	final	590.19	655.93	5.19	0.02	*	0.05
text	initial	542.49	453.24	11.95	0.00	***	0.09
	medial	3039.21	2735.45	23.85	0.00	***	0.05
	final	1160.13	654.70	204.59	0.00	***	0.27

Note: A=Agriculture, E=Economics, X^2 = chi-square

The statistical analysis (in Table 9) shows significant differences in the hedges phrases occurring in the different textual positions of the two corpora. Specifically, the occurrence of phrases in sentence-initial and medial positions are significantly higher in the agriculture corpus than those in the economics corpus ($p = 0.00$, $p = 0.00$). In contrast, the economics corpus contains more hedges phrases in sentence-final positions ($p = 0.13$). The effect size is small in all instances, however.

The paragraph analysis shows that the economics corpus contains significantly more hedges

phrases in the paragraph-initial (p = 0.30). In contrast, the agriculture corpus employs significantly more hedges phrases in the paragraph-medial positions (p = 0.00; w = 0.16). Again, effect sizes are small.

Regarding the distribution of hedges phrases in the whole text, the agriculture corpus employs significantly more hedges phrases in the initial, medial and final positions. Although the effect sizes are relatively small in the first two cases (w = 0.09 and w = 0.05 respectively), the effect size of the text final positions borders on medium level (w = 0.27), indicating a stronger degree of significance in this instance.

6.4.4 Reference stance phrases

Figure 7 presents the proportional distribution of the reference phrases in both corpora. With regard to the distribution at sentence level, more than one half of the reference phrases occur in sentence-initial position. The analysis of the paragraph-level distribution shows that the two disciplines have a substantial proportion of phrases in the paragraph medial position. Of note are the differences in the proportional distribution of phrases across the paragraph in the two disciplines. While the agriculture corpus contains proportionally more reference phrases in the paragraph medial positions, the economics corpus contains a higher proportion of reference phrases in the paragraph initial and paragraph final positions. With respect to the text-level position, an overwhelmingly higher number of reference phrases occur in the text medial position. Also, a slightly higher percentage of reference phrases are distributed in the text final position than the text-initial position in the two disciplines.

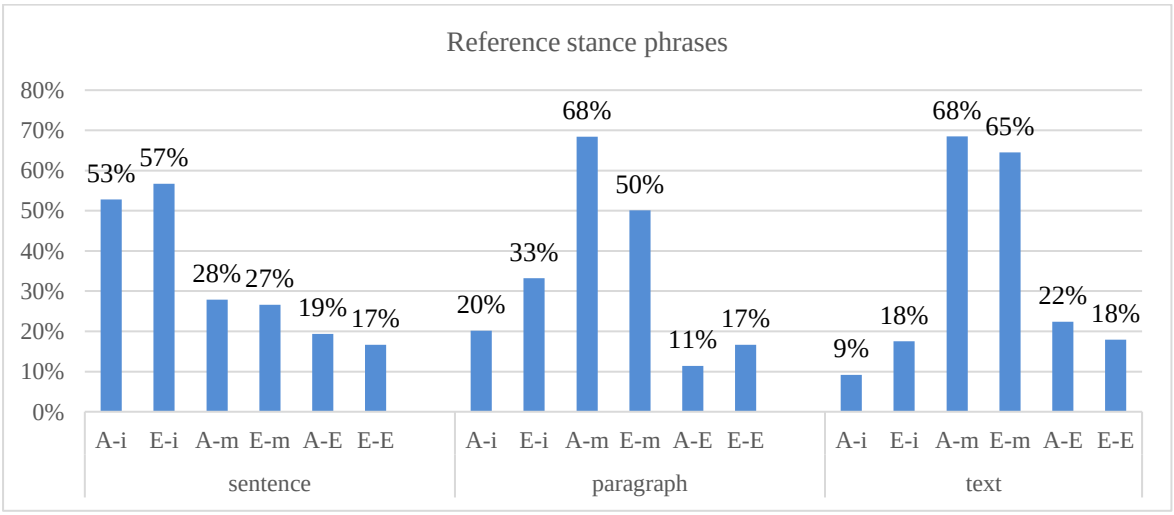

Figure 7 The Proportional Distribution of Reference Phrases in Different Textual Positions

Note: A=Agriculture, E=Economics, i = initial, m = medial, f = final

The result of statistical analysis (see Table 10) shows that the reference phrases have the following two disciplinary differences in terms of their textual colligation. Firstly, the economics corpus contains

significantly more reference phrases occurring in sentence-initial (p = 0.00, w = 0.11) and medial positions (p = 0.04; w = 0.05), although the effect sizes are small. In the paragraph positions, the economics corpus contains significantly more reference phrases in the paragraph initial (p = 0.00) and final (p = 0.00) positions than those in the agriculture corpus. In this case, the effect sizes are at medium level, w = 0.37 and 0.31 respectively. In contrast, the agriculture corpus contains significantly more reference stance phrases in the paragraph-medial positions (p = 0.02), although the effect size is small.

Finally, the textual position analysis shows that significantly more reference phrases occur in the text initial positions (p = 0.00), and the effect size is relatively large (w = 0.43). While the economics corpus display more reference phrases in the text medial position than the agriculture corpus (p = 0.00; w = 0.09), the effect size is again small. No significant difference is found in the occurrence of reference phrases at the text final position.

Table 10 Cross-corpus Comparison of Reference Phrases

Text unit	Textual position	Norm freq. A	Norm freq. E	X^2	Sig.		Effect size
sentence	initial	1063.64	1449.84	90.27	0.00	***	0.16
	medial	561.70	681.36	17.38	0.00	***	0.10
	final	389.76	426.87	2.55	0.11		0.05
paragraph	initial	405.95	849.97	243.56	0.00	***	0.37
	medial	1379.38	1281.85	5.37	0.02	*	0.04
	final	229.77	426.24	90.73	0.00	***	0.31
text	initial	184.66	448.18	171.51	0.00	***	0.43
	medial	1379.77	1651.05	36.79	0.00	***	0.09
	final	450.67	458.83	0.12	0.73		0.01

Note: A=Agriculture, E=Economics, X^2 =chi-square

6.5 Discussion

This analysis has shown that stance phrases are distributed unevenly in the two corpora in terms of textual positions at the level of sentence, paragraph, and text. The result consolidates the finding reached in previous studies that linguistic features are not evenly distributed (O'Donnell et al., 2012; Mahlberg, 2009).

At sentence level, reference phrases tend to occur in sentence-initial position. Deploying this marker at this position allows writers to treat the referential markers, such as *this study*, as the point of departure for the statement delivered subsequently, thus highlighting "a competent scholarly identity and gaining accreditation for research claims" (Hyland, 2001, p.232). This textual position also foregrounds authors' self-representation and their active engagement, as in Example (25), where

the self-representation marker, *in this study*, is preceded by a statement conveying the focus of the study, i.e., *testing potential mechanisms for the dispersal of Joshua tree seeds*. Similarly, Example (26) begins with an initial summary statement which reviews the main content of the paper.

(25) ***In this study, we*** tested potential mechanisms for the dispersal of Joshua tree seeds and seed fate with a combination of field experiments. (ARAC#66-2: sentence-initial, paragraph-initial and text-medial)

(26) <conclusion> ***This paper*** deals with the effect of financial infrastructures on economic growth, with financial intermediaries as local monopolies due to horizontal differentiation. (ERAC#65-1: sentence-initial, paragraph-initial and text-final)

Cognitive phrases are identified to have a tendency to occur at sentence-medial positions. As mentioned in Section 4.3.1, this type of expression concerns authors' reason-oriented presence by means of logic markers. By arranging this type of stance in sentence-medial positions, writers are able to persuade readers implicitly. It is interesting to note that sentence-initial position is also preferred compared to sentence-final position. In Example (27), the cognitive verb phrase, *demonstrates that*, is used to make generalisation based on the evidential results (*our findings*) and strengthen the credibility of the following statement. Notably, the cognitive verb is found to have a high association with evidential expressions, such as *our findings, the result,* and *this evidence*.

(27) ***Our findings*** also ***demonstrate that*** the effect of FRQ on investment efficiency decreases with shorter maturities, suggesting a substitutive role of FRQ and shorter maturities in reducing information asymmetries and monitoring managerial behavior to limit expropriation of creditors and minority shareholders. (ERAC#56-2: sentence-initial, paragraph-final and text-final)

Hedges and attitude phrases, on the other hand, have a preference for sentence-medial positions. Both of these types are associated with authors' subjective presence; hedging phrases tend to mitigate their presence as in (28), while an evaluative value (*essential*) is credited to the topic, *access to public equity,* in Example (29). The preference for sentence-medial position indicates that writers are more likely to present their subjective presence in less notable sentential positions, which may assist writers in influencing readers' acceptance of a proposition.

(28) Given that EM fungi are almost entirely reliant on trees for their C, factors altering C allocation in trees ***are likely to*** alter patterns of EM abundance, diversity, and ultimately belowground functioning. (ARAC#111-2: sentence-medial, paragraph-medial, and text-medial)

(29) Furthermore, access to public equity may be ***essential for*** most young high-tech firms, who typically have limited access to debt and small or negative internal finance. (ERAC#20-1: sentence-medial, paragraph-medial, and text-medial)

At the level of paragraph and text, writers in both disciplines tend to use a high percentage of stance phrases in the medial position. One reason may be that the medial position accounts for a substantial section of the text, according to the division used in this study. This may also reflect the tendency for writers to place stance expressions in a paragraph non-salient position.

The comparison between the initial and final positions at the level of paragraph and text, however, shows that hedging phrases occur more frequently in the paragraph- and text-final than the initial position. A close examination reveals that the paragraph-initial and medial positions are favoured for reports of findings or statements of observation, while in the paragraph final position authors tend to interpret the findings through the use of hedging phrases. According to priming theory, the textual environment of linguistic items often results from writers' frequent encounter with the textual positioning of linguistic features (Hoey, 2005, p.130). Therefore, the positive priming of hedging phrases for the paragraph- and text-final reflects a process in which writers internalise this textual association through their repeated exposure to this frequent textual use. In other words, this frequent exposure primes writers to consciously or unconsciously align their use of stance markers to conventionalised community practices by deploying the hedging markers in these final positions.

It has also been found that reference phrases have a higher tendency to occur in the paragraph-initial than final positions. Initiating a paragraph with reference markers allows writers to shift readers' attention to a different focus. Example (30) shows a typical case, in which the self-mention expression (*in the present study, we**)*** signifies a transition from a preceding paragraph, talking of the previous studies on *the oriental fruit moth Cydia molesta*, to the research focus of the present study, *establishing the 3D glomerular map for males and females of C. molesta.*

(30) ***In the present study, we*** have established the 3D glomerular map for males and females of C. molesta as a tool for future functional studies. (ARAC# 27-1: sentence-initial, paragraph-initial and text-initial)

At the text level, slightly more attitude phrases occur in the text-initial (in particular the abstract and introduction sections) than final position. This position is used to establish the importance of research topic, as in (31), and to discuss or promote research findings, as in (32). The attitude stance used in the textual final position, particularly in the conclusion section, is mainly for the purpose of promoting the study or research findings, as in (33).

(31) Below-ground processes ***play a key role*** in the global carbon (C) cycle because they regulate storage of large quantities of C, and are potentially very sensitive to direct and indirect effects of elevated CO2 and temperature. (ARAC#102-2: sentence-initial, paragraph-medial and text-initial)

(32) These findings ***offer insights into*** the evolution of the nature of partnerships and the power over resources. (ERAC#85-2: sentence-initial, paragraph-medial and text-initial)

(33) Such empirical results ***are important*** because they ***provide insight into*** how enacting changes in governance structures and the design of contracts might improve investment performance and thereby either lower the fiscal burden to taxpayers or perhaps increase the benefits to beneficiaries. (ARAC#62-2: sentence-medial, paragraph-final and text-final)

In contrast, reference and attitude phrases occur more frequently in the text-final than initial position. The text-final position contains an interpretation and discussion of findings and implications,

and these rhetorical functions may require writers to draw on reference expressions, self-mention markers in particular, to refer to their study and boost their findings. Example (34) illustrates writers' use of self-mention markers to compare their findings with those from earlier studies, while in Example (35), the self-mention marker (*this study*) and the booster phrase (*demonstrates that*) work in tandem to introduce a generalized account of the findings.

(34) ***This finding*** is in line with Levine et al. (2000), who show that there is a strong connection between private sector credit and economic growth. (ERAC#56-3: sentence-initial, paragraph-medial and text-final)

(35) ***This study*** also ***demonstrates that*** structural characteristics of shrub layer, such as cover or height, were also important determinants of abundance and richness. (ARAC#51-2: sentence-initial, paragraph-initial and text-final)

Overall, the different types of stance phrases are primed to appear at certain textual positions. Hoey (2005) considers that the textual position of linguistic items is a consequence of frequent exposure to their contextual positions. Functionally, the textual positioning may assist in constructing the writer's position vis-a-vis the content of a text and readers. That is, writers employ stance phrases in appropriate textual positions to aid the information flow and facilitate readers' cognitive processing of the information.

The disciplinary comparison, on the other hand, reveals that the two disciplines share similarities in terms of the proportional distribution. That is, these academic disciplines tend to associate stance phrases of the same type with particular textual positions. The findings support the central tenet of lexical priming theory that certain words or phrases are primed to occur at certain recognised textual positions (Hoey & O'Donnell, 2015). The similar textual distribution of stance phrases suggests that authors in the two disciplines are exposed to similar priming influences.

Meanwhile, great disciplinary variation has been identified in the association between stance phrases and particular textual positions at the levels of sentence, paragraph and text, which may correspond to their disciplinary practices. For instance, the agriculture corpus contains significantly more attitude and hedges in sentence-initial and medial positions. The theme position, embedded in sentence-initial position, allows writers to "make their evaluative position the starting point of the message" (Hyland & Tse, 2005, p.61), and imbues the subsequent proposition expressed in sentence with the writer's position or attitude. Similarly, the agriculture corpus also contains significantly more attitude and hedging phrases in sentence non-salient position, namely, sentence-medial position. In contrast, the economics corpus contained significantly more reference stance phrases in sentence-initial and medial position, reflecting the fact that the economic writers tend to depart from making reference to their own study, or an extra-resource, shared knowledge or other parts of the text organisation while developing their ideas.

The cross-disciplinary comparison at the paragraph level shows that the agriculture corpus

contains significantly more attitude and self-mention phrases in the paragraph-medial positions than the economics corpus. The result indicates that agriculture writers tend to incorporate more hedging phrases at the paragraph-initial positions, and more attitudinal and explicit self-presence in paragraph non-salient positions, thus indicating a less notable self-projection persona in the text organisation. In contrast, significantly more stance phrases (cognitive, hedges and reference in particular) occur in the salient paragraph-initial positions of the economics corpus. This indicates that the economic writers may choose to embody more of the three types of stance phrases at the paragraph salient positions to impress readers and thus achieve an acceptance from the readers of the information delivered in the texts.

As far as the distribution at the text level is concerned, the analysis shows that the agriculture corpus displays significantly more of each of the four categories in text-medial positions. This indicates that agriculture writers tend express their subjective presence in text non-salient positions. In contrast, the economics corpus contains more self-mention in the text-initial position (e.g., abstract and introduction). The beginning of the text has been generally accepted as an important textual position (Swales & Feak, 2004). According to Mauranen (1993, p.11), "[t]he orientating aspect of metatext is particularly important in initial positions, such as introductions". The text-final position, on the other hand, is "the last chance a writer has to convince [his or] her reader[s]" (Ädel, 2006, p.125). This position is found to contain more attitude markers in the economics corpus, showing that economic writers are more inclined to make use of this textual final position to promote their research findings.

The use of *Wordskew* (Barlow, 2016) was found to be an effective tool to identify the textual positions at sentence, paragraph, and text levels. It provides an efficient way to quantify the textual position of particular linguistic features, and contributes to visualising the distribution of particular linguistic features across texts.

Hyland (2011) points out that academic writing, in essence, "initiate[s] a writer-reader interaction" (p. 194) through which writers construct knowledge, develop arguments, and persuade readers of the validity of propositions. In this process, the textual arrangement of the linguistic units displays a temporal dimension. In other words, the textual position, despite being framed as a linear-spatial arrangement in the written form, reflects the timing with which linguistic units are deployed to achieve certain communicative purposes.

As this study has demonstrated, different types of stance phrases are seen to display particular priming for certain textual positions. The textual positioning of authors' presence is characteristic of all three discourse positions examined in this study. This suggests that textual position of stance markers is not a random practice, but embodies authors' intentional arrangement of linguistic

resources to achieve particular communicative functions. Employing linguistic markers with the appropriate timing arguably indicates skill and experience in professional writing and awareness of the potential impact of one's writing on the reader. Overall, the appropriate timing of stance-taking facilitates the coherent organization of discourse, enables a nuanced interpretation of a study's findings, and also contributes to the construction of the writer as a credible persona.

6.6 Summary

This chapter extends previous work on textual positions by examining the textual distribution of stance phrases in the academic discourse of agriculture and economics. The study examined the distribution of stance phrases operating at sentence, paragraph and text level in academic discourse, and explored the existence of disciplinary variation with respect to textual positions. The results indicate that writers in the two disciplines tend to associate stance phrases of the same type with similar textual positions. Thus, textual colligation seems to be a characteristic of linguistic markers which perform similar discourse functions. The findings not only enhance our understanding of the interaction between the textual colligation of stance phrases and their 'local textual functions' (Mahlberg, 2009, p.265), but they also reveal considerable variation between the two academic disciplines in the frequency of stance phrases in particular textual positions.

The textual positions may reflect the timing with which authors choose to project themselves in the process of communicating with readers. Thus, it may be inferred that effective writing depends not only on the type and frequency of stance phrases, but also the appropriate timing of the usage of these phrases in academic discourse. The findings from this study may thus contribute to developing Hoey's (2005) priming theory. The findings regarding the textual position of the stance phrases consolidate Hoey's premise that certain expressions favour or avoid particular textual positions. In addition, this study reveals that the phrases of a particular function tend to share some positional similarities with regard to their distribution in sentence, paragraph and the whole text.

Although a strong association between the stance phrases and textual positions has been identified in the disciplines examined in this study, the present study only focuses on two disciplines using published studies of an empirical nature. Results from other disciplines may be different, and the use of stance phrases in research reflecting a qualitative paradigm has yet to be explored. Pedagogically, insights into the textual colligation of stance phrases could be of value for EFL students and novice writers in the two disciplines in their discipline-specific academic writing practice.

In the following chapter, I explore how stance phrases interact with their surrounding environment from the following two perspectives: how the four categories of stance phrases interact with each other, and how stance phrases collocate with other semantic domains, namely the non-stance phrases categories.

Chapter 7 The collocation networks of stance phrases

This chapter addresses the third research question, concerning the collocation network of stance phrases and the disciplinary-specific nature of these relationships in the two corpora. Section 7.1 provides a general background for the studies concentrated on the collocation network of linguistic items. Section 7.2 presents the methods used to retrieve collocation network of the stance phrases and the statistical tests used to investigate the cross-disciplinary variation of the collocation network of stance phrases. Section 7.3 discusses the collocation network of within-stance collocational patterns, while Section 7.4 presents the collocation patterns of stance phrases with other semantic domains. The investigation of the disciplinary variation of collocation networks is presented in Section 7.5.

7.1 Introduction

The interaction of linguistic items with their surrounding environment is an ongoing focus in applied linguistics, and increasing attention has been devoted to the study of a phenomenon known as 'collocation networks' (Brezina, Mcenery, & Wattam, 2015; Gablasova, Brezina, & McEnery, 2017; Hunston, 2002, 2011; Partington, 2004). A collocation network not only displays the co-occurrence of targeted linguistic items with their surrounding contexts, but also reveals the extent to which different linguistic items in a network relate to one other. This approach calculates collocational strength by means of a quantitative measurement (i.e., mutual information, log-likelihood, and z score) within a certain span. The results are displayed using a collocation network, which depicts the collocational relationship between a linguistic item and its context. An example of the collocation network of *seem to* and its different inflectional forms (e.g., *seemed to*, and *seems to*) is illustrated in Figure 8. This visualisation provides a rich picture of the complex collocation networks of the phrase. For example, the collocation network in Figure 8 shows a strong collocation between *seem to* and the evaluative expression *important*, a co-selection pattern may escape observation in a long list of concordance lines.

Previous studies have explored the collocational relationships of linguistic items from one of two perspectives. One prevalent focus has been on the linear relationship between co-occurring linguistic items (e.g., Hunston 2007a; Partington 2004). This entails selecting a linguistic item, often by consulting a high-frequency list, and investigating its relationship with the linguistic contexts in a linear fashion. More recently, writers have investigated the semantic relationship between a linguistic item and its co-occurring context from a hierarchical perspective by identifying how linguistic items or units (e.g., lexis, phrases, and clauses) are inter-related in a text (see, for instance, Hunston, 2008, p.273). This line of research has unveiled a complex and intricate linkage existing between linguistic

items and their surrounding contexts (e.g., Brezina et al., 2015; Gablasova, Brezina & McEnery, 2017).

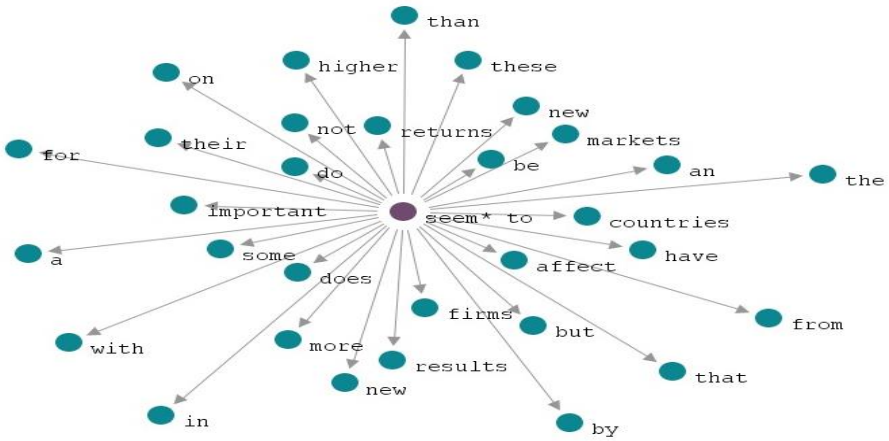

Figure 8 Collocation Network of the Phrase Seem* to [MI (>3), R5-L5][7]

However, previous studies have been frequently confined to a limited number of words or phrases and have merely examined collocational relationships from a descriptive perspective (Hunston, 2007b; 2008). A limitation of this approach is the difficulty of attaining "a more complete account" of semantic relatedness (Hunston, 2008, p.277). A focus on the collocational relationships of a set of phrases performing similar functions could arguably enable broader generalizations with regard to semantic relatedness. Despite this importance, advances in this line of research have been modest to date, possibly on account of the technical difficulties involved in retrieving the collocates of multiple phrases.

A recent development in corpus software, namely the release of GraphColl, has made it possible to investigate the contexts of stance features by examining both the immediate and discontinuous collocates from a quantitative perspective. A large-scale examination of the quantitative semantic relationship, namely the collocational strength of a linguistic item with its surrounding environment, can facilitate insight into "latent patterning" (Sinclair & Coulthard, 1975, p.125), a concept concerned with the language use patterns which are not easily observable or identifiable in manual searches, and which may therefore be neglected in research which relies on an intuitive analysis of linguistic forms.

As discussed in Section 4.3, stance expressions convey authors' self-projection into a text in the form of cognitive involvement and involve interaction with envisaged readers through the formulation of evaluative, hedging, and referential cues. As such, they thus "constitute an essential aspect of the interpersonal in discourse" (Mauranen & Bondi, 2003, p.269). Previous work has investigated the type and frequency of stance expressions (Biber, Conrad, & Cortes, 2004; Conrad & Biber, 2000; Hyland, 2005a; Jiang & Hyland, 2016) and their occurring contexts (Hunston, 2002, 2011).

While the existence of disciplinary variation in the occurrence of stance expressions has been previously identified (Hu & Cao, 2015; Hyland, 2004, 2005; Johns & Swales, 2002), little is known

about the collocational preferences of these expressions in different academic disciplines. Therefore, this chapter ascertain the extent to which the collocation network of stance expressions may be considered disciplinary specific.

Specifically, I investigate the collocation network of stance phrases by describing the within-stance collocational relationship (that is, the co-occurrence of the different categories of stance phrases). Employing a corpus of research articles from the disciplines of agriculture and economics, this chapter addressed the following questions:

(1) What are the collocational relationships between different categories of stance phrases?

(1a) To what extent is there disciplinary variation in the collocational relationships identified?

(2) What semantic domains do stance phrases typically collocate with?

(2a) To what extent is there disciplinary variation in the collocational relationships identified?

7.2 Methods

7.2.1 Collocate retrieval

In identifying the collocates of stance phrases, this study used the software Graphcoll (Brezina et al., 2015) to retrieve the collocates of the stance phrases. The extraction of the collocates for a node word or phrases contributes to identifying the "meaning connections in text and discourse that may otherwise pass unnoticed" (Brezina et al., 2015, p. 41). When depicting the relationship of words and surrounding contexts, the software allows users to select a particular statistical test among various measures, such as the dimensional Deltap P, MI, and T test, etc. In this study, I used the parameter of both MI and frequency to measure the collocational relatedness of the linguistic items.

The current version of GraphColl does not permit a search for a set of words or for phrases in batches; for this reason, I first coded the stance phrases by assigning a label to each stance category[8]. For example*, seem to, it is likely that* were labelled as 'hedging phrases', and *it is interesting to* and *plays an important role* were given a category marker of 'attitude phrases'. These category labels subsequently enabled us to use GraphColl to search for the category labels instead of the specific phrases, and to identify the collocation networks of each of the four stance categories.

A collocational span of five words on each side of the node was adopted, following the practice in Brezina et al. (2015) and McEnery (2006). This extended span allows us to investigate both

[8] Considering that the search of a group of words or phrases has not been previously undertaken with GraphColl, I developed a coding system to replace the group of stance phrases in the stance framework. For instance, C&& was used to represent the category of the cognitive stance phrase, A&& stands for the attitude stance, H&& is used to refer to the hedges stance, and R&& for the reference stance phrases. In order to ensure that the coding did not overlap with the words or other contents in the corpora, the coding system was checked against the two corpora and it was found that the system used do not exist in the corpus. I also checked the entire corpora to ensure theses codes do not overlap with any items in the corpora.

immediate and non-immediate collocates of stance phrases (Sinclair et al., 2004: 42), and should thus provide an adequate window for the analysis of semantic relationships.

When filtering the collocates, I took into account both MI and normalised frequency. The MI metric has been commonly used as a criterion to measure collocational strength (e.g., Simpson-Valch & Ellis, 2010; Gablasova et al., 2017). However, the use of this parameter has raised some concerns. For example, Biber (2009) points out that the MI value disfavours word sequences with high occurrences. That is, if the individual words in a word string have comparatively low frequency, the word string tends to have a high MI value; and vice versa. Therefore, I set the threshold values by considering both MI and frequency: MI above 3 (following Brezina et al., 2015 and Gablasova et al., 2017), and frequency above 10 instances per million words (in line with Biber et al., 1999), to ensure sufficient data for the analysis of each type.

7.2.2 Analysis

In the analysis, I employed Wmatrix (Rayson, 2008) to conduct semantic tagging and analysis. The WMatrix's categorization system comprises 21 major discourse fields at the top level, which are further divided into subcategories at a second level and third level. For example, the semantic group 'A. general and abstract terms' (a top-level category) is broken down into 'A1 general', 'A2 affect', 'A3 being', etc. at the second level; some of these are further broken down into A2.1 Affect: modify, change and A 2.2 affect: cause/connected at the third level. To ensure sufficiently nuanced results were retained, while also keeping the semantic domains to a manageable number, I categorized the data according to the second semantic categorisation level, e.g., 'A5 Evaluation' in the UCREL Semantic Analysis System (USAS).

For the entries tagged with multiple labels, their concordance lines were examined closely. For instance, *table* was labelled as "Furniture and household fittings", "written communication-related expressions", and "mathematical terms" by Wmatrix. The examination of the concordance lines assured that all instances consisting of *table* in the stance collocation network of the two academic corpora referred to the semantic category of "written communication-related expressions" (e.g., *in Table 1*). Similarly, I checked the multiple taggers ascribed to *have,* and adopted the corresponding one by referring to the co-text. Specifically, the label of "possession" was adopted for cases like *have higher concentration, have the flexibility*; and "function word[9]" was kept for the instances like *have been reported*, and *have been established*. Acronyms were also tagged by referring to their original full meanings, for example *CBI* (stands for *Confederation of British Industry),* which refers to an association, was labelled as "groups and affiliation", while *CEO (*short for *Chief Executive Officer)*

[9] 'Grammatical bin' is the term used in USAS. Here, we adjust the term to 'function word' to align with the commonly used linguistic term and make the meaning of this category clearer.

was tagged as “people”.

For this analysis, I compared the frequencies of the two types of collating patterns, namely, the within-stance phrases collocation (i.e., the collocating patterns of the four types of stance phrases), and the collocation patterns between stance phrases and other semantic groups in the two corpora. A Chi-square analysis was performed to establish whether significant differences exist with regard to the within-stance collocation and their collocation with other semantic domains. Similar to Chapters 5 and 6, this chapter also used effect size statistic to measure the degree of the significant differences in the collocation network of stance phrases in the two academic corpora. The scale of effect was interpreted following Cohen (1988): ≤ 0.1 small effect, 0.1- 0.5 medium effect, and ≥ 0.5 large effect.

7.3 Collocation network: the collocation within different categories of stance phrases

In identifying the collocation network, a stepwise procedure was adopted in the GraphColl search by selecting the initial node, namely one particular category of stance phrase (e.g., cognitive phrases), and the other three categories (e.g., attitude, hedging and reference phrases) were operationalised as the second, third and fourth nodes respectively. In this step, I searched the stance categories one by one in GraphColl: starting with cognitive phrase, then attitude, followed by hedging and reference stance phrases respectively. Since the relationships between the stance phrases are symmetrical, the order in which these four categories are searched does not matter in terms of the collocation networks of the four types of stance phrases.

The software then produced collocation networks for each of the four categories of stance phrases (see Figure 9). This section presents and discusses the collocation network of within-stance collocational patterns; the collocation patterns of stance phrases with other semantic domains are presented in Section 7.4.

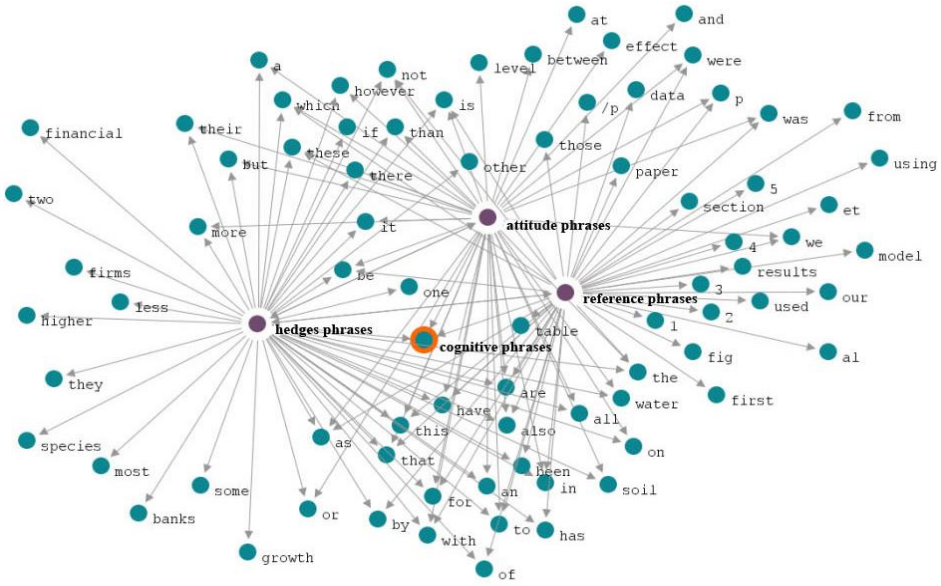

Figure 9 Collocation Network of the Different Categories of Stance Phrases [MI (>3), R5-L5 [8]]

Figure 9 depicts the collocational relationship between the four categories of stance phrases: cognitive, attitude, hedges and reference phrases, as well as their correlation with other linguistic items. In this figure, the four large dots or nodes in the centre represent the four stance categories, and the smaller surrounding dots represent the collocates of each stance category. The length of lines between the stance nodes and each collocate indicates the strength of the collocational bond. That is, the shorter the line between two collocating items, the greater the collocation strength, and vice versa. The positioning of the stance nodes in relation to one another is random in this figure. Table 11 displays the relative strength of the collocational bond between two types of stance phrases in terms of frequency and MI value. When ranking the strength of these collocational patterns, I used MI value, as a secondary parameter in conjunction with the normalised frequency score, due to the aforementioned concerns regarding MI (outlined in Section 3.2.1).

Table 11 The Collocation Pattern within the Four Types of Stance Phrases

	Cognitive			Attitude			Hedges			Reference		
Category	MI	Freq	Norm Freq	MI	Freq	Norm Freq	MI	Freq	Norm Freq	MI	Freq	Norm Freq
Cognitive	-	-	-	3.53	1029	34.27	3.86	1531	51.00	3.67	1227	40.87
Attitude	3.53	1029	34.27	3.84	172	5.73	3.66	190	6.33	-	-	-
Hedges	3.84	1531	51.00	3.66	190	6.33	3.46	208	6.93	3.15	146	4.86
Reference	3.67	1227	40.87	-	-	-	3.15	146	4.86	4.21	287	9.56

Note: '-' stands for non-collocating patterns identified in the searches.

Cognitive phrases collocate most strongly with hedging phrases (MI=3.86 and norm freq=51 per million). This category expresses authors' logic-oriented approach to formulating arguments through processes such as explaining, inferring, and inducing, etc. The co-selection with hedges phrases indicates that authors in the two academic disciplines tend to employ a cautious tone in their logical analysis, when using phrases such as *giving explanation, make comparison* or *reaching* conclusion, an approach that may better "gain ratification for claims from a powerful peer groups" (Hyland, 1996, p.433). In example (36), the author combines the hedging resource *may be* with the cognitive behaviour *explain*. In so doing, the author opens up a negotiating space where he/she can accommodate the reader's perspective in explaining the '*negative and neutral microbial biomass responses'*.

(36) Negative and neutral microbial biomass responses ***may be explained by*** increased turnover rates if soil water content is high (Hungateet al., 1997; Arnone & Bohlen, 1998). [ARAC#102:3][10]

The second strongest collocational relationship occurs between cognitive and reference (MI=3.67; norm freq=40.87). Reference denotes the evidentially-aspect of stance by signalling the

[2] For the examples in Section 5, one stance phrase is shown in **bold**, and the second is presented in ***bold, italicised and underlined.*** Sources of examples are presented in bracket.

source of particular propositions (Chafe & Nichols, 1986). Its co-occurrence with cognitive stance entails an evidentiality-attributing process whereby academic writers ground their analytical behaviour in the evidence or source introduced by the reference markers. This practice of specifying the evidential source for a cognitive processing allows writers to strengthen the nature of their; essentially, subjective engagement with a proposition (implicit in words such as *imply, infer*) by referring to an ostensibly neutral reference (such as *result, table*).

In example (37), the reference phrase (*results*) denotes an objective evidential source, and the co-occurring cognitive phrase (*imply that*) conveys the writers' logical processing based on the source. The binding of these two linguistic resources allows the writer to support his/her cognitive processing with an ostensibly objective reference and provide authorial endorsement for the arguments delivered in the subsequent clause.

(37) **Our results *imply that*** only at intermediate and high viral doses are there effects of mixed infections on the response of the larval host. </p></discussion> [ARAC#76:2]

As shown in Table 11, cognitive phrases also collocate frequently with attitude phrases (MI=3.53; norm freq=34.27). This finding is in line with Bruce's (2016) view on the co-occurrence of attitude markers and relation-related expressions, a term that corresponds to the cognitive category in this study. This collocating structure, which entails projecting the affective aspect of stance prior to the cognitive phrase, contributes to priming readers to share authors' emotive response towards the cognitive actions. This can be seen in example (38); the attitude phrase *it is important to* prompts a particular interpretation of the cognitive activity, expressed by '*note'*.

(38) Since this period is a baseline, ***it is important to note that*** the 19 too-big-to-fail banks had not been explicitly named at this point, but we contend that the market was able to accurately infer which banks were likely considered too big to fail. [FRAC#56:2]

Strong collocational relationships were also found between examples from the same stance category, especially within-attitude, within-hedges and within-reference stance phrases. Of these, the co-occurrences of the reference phrases are the most frequent (MI=4.21; norm freq=9.56). The co-occurrence of these expressions contributes evidential weight to the claims, and lends structure to the textual organisation. In example (39), the two reference phrases operate in tandem to establish a link between two distinct sections of the text.

(39) We can test the model that we developed ***in the previous section*** using the market data ***described in Section 6.*** [ERAC #56: 5]

The co-occurrence of attitude phrases is also quite noticeable in the network (MI=3.84; freq=5.73). The co-occurrence of two attitude phrases appeared to mutually enforce the positive semantic value. This can be seen in example (40); here the attitudinal markers, *interesting to* function to stimulate engagement with the following proposition, while *contribute to* gives *this* a credit of being

conducive *to the debt policy and debt management theories.*

(40) However, it could be ***interesting to*** see the extent to which this can ***contribute to*** the debt policy and debt management theories. </conclusion> [FRAC#76:3]

Similarly, hedging phrases frequently co-occur (MI=3.46, and norm freq=6.93), mostly in a parallel structure to enable a certain economy of style. For example, the co-occurring two hedges in (41), *tend to* and *likely to*, are used to modify two parallel and relatively interdependent aspects of the topic: 'be stable over time' and 'move in relative unison'.

(41) Book income and taxable income ***tend to*** be stable over time and are ***likely to*** move in relative unison unless the firm makes a different strategic choice or new information is reflected in one, but not both, of the measures. </p> [ERAC#2: 1]

Interestingly, hedges are found to collocate with all categories of stance phrases. The collocation of this category with attitude phrases (MI = 3.66; norm freq=6.33) shows writers' endeavours to mitigate their attitudinal subjective involvement. For instance, the hedging expression *may be* in (42) nuances the following attitudinal expression *interesting to* by introducing an element of modesty. By expressing a proposition in a relatively low-risk manner, the writer accommodates possible alternative views.

(42) Because the correlation between organic carbon content and bulk density is relatively low (Fig. 1e), ***it may be interesting to*** try using both those properties along with texture as the water retention predictors. </p> [ARAC#58:2]

The juxtaposition of hedges and reference phrases (MI=3.15; norm freq=4.86) permits the writer to assume a more nuanced position with respect to the cited sources (see example 43). The perception that the writer is exercising a degree of caution in the interpretation of sources contributes to maintaining the readers' confidence in the writer's approach (Hyland 1996).

(43) ***According to*** these findings, there ***seems to*** be a positive interaction among carob (legume family), Aleppo pine (pine family) and Kermes oak (beech family). [ARAC# 45:2]

7.4 Collocational relationships between stance phrases and other semantic domains

Table 12 displays the collocation relationships between stance phrases and other semantic domains, namely non-stance phrases expressions, produced with the assistance of Wmatrix. Such collocational patterns reveal important information regarding how stance phrases operate with other semantic domains. For reasons of economy, the 18 semantic domains identified here, according to the WMatrix's second level categorization system, were conflated into the following three semantic categories: topic-oriented expressions, involvement-oriented expressions, and general expressions. The general category of 'topic-oriented collocates' functions as an umbrella term for 'farming and horticulture' and 'environment' etc; 'involvement-oriented expressions' comprise expressions denoting linguistic actions, states or process, psychological actions, states and process, etc; and

'general expressions' contain those pertinent to general actions or entities, number and measurement, and function words, etc. The specific findings regarding the semantic domains (including the average and total MI, raw frequency and normalised frequency) which collocate with stance phrases are presented in Table 12, and discussed in the following subsections. The calculation of the average MI value, frequency and normalized frequency of all members in each category are displayed here to enable a generalized statement of the collocational relationship.

Table 12 The Collocation between Stance Phrases and Other Semantic Domains

General category	Semantic domains	Average MI	Average Freq.	Average Norm freq. (per million)	Total MI	Total Freq.	Total Norm freq.
Topic-oriented expressions	Farming and horticulture	7.16	93.23	56.51	489.60	6666	4010.99
	Science and technology	3.91	37.13	23.8	747.59	7953	5389.80
	Substances and materials	3.76	64.4	38.97	583.37	9982	6039.65
	Social actions, states and processes	3.76	39.96	27.08	747.59	7953	5389.80
	Environment	3.76	47.37	29.46	323.34	4074	2533.15
	Money and commerce in industry	3.66	68.14	50.15	921.98	17171	12636.67
	Life and living things	3.6	58.71	35.17	360.22	5871	3516.84
Involvement-oriented expressions	Government	3.58	50.27	36.85	39.40	553	405.37
	Linguistic actions, states or process	4.14	47.86	31.17	396.97	4595	2992.47
	Psychological actions, states and process	3.92	45.45	30.46	1434.70	16634	11146.90
	Modality	3.85	74.74	49.26	223.08	4335	2857.32
	Evaluation	3.72	46.38	31.09	661.87	8255	5533.30
General expressions	General actions or entities	3.76	46.17	30	2261.54	27792	18059.88
	Number and measurement	3.72	57.99	38.21	1761.94	27487	18112.32
	Function words	3.69	336.37	221.87	1743.07	159104	104943.05
	Degree	3.92	52.36	34.64	301.84	4032	2667.55
	Location	3.7	42.16	28.54	447.57	5101	3453.18
	Time	3.63	40.47	25.83	585.01	6516	4158.04

7.4.1 Topic-oriented expressions

As shown in Table 12, the collocation network of stance phrases contains a substantial number of topic-related expressions, including 'farming and horticulture' (MI=7.16; norm freq=93.23), science and technology' (MI=3.91; norm freq=37.13), 'substances and materials' (MI=3.76; norm freq=64.40).

Two points may be raised regarding the semantic relationship. Firstly, there is a close interaction between topic-related phrases and stance phrases. For instance, *biomass production,* in Example (44), is governed by *in terms of,* which specifies the perspective from which the author makes the statement that *plants are affected.* Similarly, example (45) shows that the cognitive stance phrase, *such as*, is used to exemplify *the level of creditor rights and enforcement*, which falls under the semantic domain

of ‘social actions, states and processes’.

(44) If that was the case, we would expect differences in the multiplication of the root-feeders as a function of the plant population. On the other hand, plants would be differently affected, ***in terms of biomass production***, by the root-feeders. [ARAC #111: 2]

(45) Second, the role of debt maturity and FRQ may differ according to institutional features, ***such as*** the level of ***creditor rights and enforcement***, so these results may not be generalized to other contexts. [ERAC #56:4]

(46) It ***appears that REIT returns*** are largely independent with respect to shocks in the other assets ¨C neither direct real estate nor stock market shocks appear to be driving REIT market performance. [ERAC #84: 1]

Secondly, the stance phrase is also seen directly related to a proposition containing topic-related words, as in (46). Here, the topic-related expression, *REIT returns*, functions as the subject of the proposition which is governed by *it appears that.* By allocating the hedging marker to the position preceding the proposition, the author is able to downplay the argument that he/she constructs regarding *REIT returns*, thus enhancing the communication entailed in this pattern.

7.4.2 Involvement-oriented expressions

The collocation network also includes a substantial number of semantic domains that represent the following four forms of authors’ involvement, including (1) ‘modality’ *(*e.g., *may, likely)* (MI=3.85; norm freq=74.74); (2) ‘psychological actions, states and process’ (MI=3.92; norm freq=45.45), consisting of ‘mental actions’ (e.g., *estimate, interpret, assume)* and ‘mental objects’ (e.g*., hypothesis, opinion*); (3) ‘evaluation’ *(*e.g., *good, important)* (MI=3.72; norm freq=46.38), and (4) ‘linguistic actions, states or process’ (MI=4.14; norm freq=47.86), composed of words denoting author’s linguistic behaviours, like *describe*, *report* and *present*, and text-related expressions, like *article*, *table*, and *figure.*

These words are used to modify the stance phrases from different perspectives; for instance, in (47), the modality-related word, *perhaps,* is combined with a cognitive phrase, *due to*, to express the author’s cautious tone in developing the justification for the argument developed in the subsequent clause. In other words, the modality marker operates in conjunction with stance phrases to construct a semblance of prudence in the articulation of an argument.

(47) ***Perhaps due to*** the magnitude of the questions, the bits and pieces that are found in the literature seem inconsistent or contradictory. </p> [ERAC #62: 1]

Similarly, in example (48), the mental actions, *consider* and *examine,* are preceded by the cognitive phrase, *in order to*, to show the purpose of the study, and *it is interesting to*, which serves to encourage readers’ attention to the following proposition.

(48) ***In order to consider*** in turn the two explanations described above, ***it is interesting to examine*** our finding for other factors associated with equity premia. [ERAC #18: 2]

The co-occurrence of stance phrases and 'linguistic actions, states or process' contributes to locating the source for a particular information or proposition, thereby specifying the "speech act-relation between speaker and addressees" (Lemke, 1995, p. 40). For example, in (49), the reference stance phrase, *as shown in*, co-occurs with the text-related structure, *table 1*, which entails a conventionalized manner of signalling an information source.

(49) ***As shown in Table*** 1, in 2006, the industry collectively held 121,546 business loans with an NMBLB amount of $22.5 billion. [FRAC # 77:2]

7.4.3 General expressions

Number and measurement-related words occur relatively frequently in the collocation network of stance phrases (MI=3.72; norm freq=57.99). This semantic category includes specific numbers (cardinal, like *three*, and ordinal numbers, like *fifth*), quantifiers (like *each, several, some, many, all*), and measurement (*parameter, scale, t-test*). Measurement-related words connote the means by which writers gauge or evaluate a particular subject, like *weight, scale, level,* and *ratio.*

Most number- and quantifier-related words appear in the determiner position and often function as a modifier for the topic-related words or phrases. For example, in (50), the number-related word, *more*, quantifies the head noun (*water),* while the measurement-related word, *level,* in example (51)*,* functions as the head noun of the phrase *a certain level of institutional quality*. In both these cases, the number and measurement phrases introduce a proposition that is grammatically governed by the stance phrases, *suggest that* and *may be*.

(50) Similar cumulative ET that year between mulched and bare soil surfaces but significantly higher LAI ***suggest that more*** water use in the mulched treatment was partitioned into transpiration and consequently plant growth rather than evaporation from the soil. [ARAC #125: 2]

(51) It ***may be*** that a certain ***level*** of institutional quality has to be attained before financial development can have any impact on growth. [FRAC #56: 1]

In the collocation network, I also identify a large number of general actions or entities which co-occur with stance phrases (MI=3.76; norm freq=46.17), including actions or behaviours (e.g., *control, obtain, select, provide, process*), and general objects (e.g., *sample, component, instruments*). This semantic relatedness shows authors tend to manifest their attitude towards the behaviour or entity denoted by the semantic domain concerning general actions or entities. For example, in (52), the hedge phrase, *likely to,* is used to mitigate the actions denoted by the verb phrase, *to produce new metabolites;* while example (53) contains a hedging phrase, *may be*, which again mitigates the following noun phrase, *the best possible control*.

(52) New members of this prolific genus were also identified, which is important as these new members are more ***likely to produce*** new metabolites. [ARAC #83: 2]

(53) My own behaviour before a tax change ***may be*** the best possible ***control*** for assessing the tax effects on my

behaviour of a tax experiment. [ERAC #52: 2]

Another component appearing in the collocation network of stance phrases concerns degree-related expressions (e.g., *greatly, mainly, almost,* and *partially)* (MI=3.92; norm freq=52.36). As regards the semantic relationship, some degree expressions are used to scale the degree of authors' stance, as in (54). Here, *very* functions to intensify the possibility (headed by *likely that*) for the statement conveyed in the subsequent clause. In contrast, the degree marker, *greatly,* in (55), is used to bolster the action *vary,* rather than the stance phrase *show that.*

(54) It is ***very likely that*** water affects the reaction via surface sorption/ modification processes. [ARAC #131:2]

(55) They also ***show that*** readership varies ***greatly*** across continents, with the Journal of Finance being ranked first in every geographical region. </p> [FRAC #56:3]

There are also numerous instances of the semantic domains concerning 'time' (MI=3.63; norm freq=40.47), and 'location' (MI=3.7; norm freq=42.16). The time-related category encompasses expressions showing year, month, date, or time adverbial or adjective (*now* or *early*), while 'location' includes a region, and a specific location (*lab*, *field, Europe*). Similar to the 'number and measurement', these two semantic domains occurring in the textual proximity of the stance phrases are seen to provide detailed temporal or spatial information for particular topic-oriented expressions or propositions, as in (56) and (57), which are semantically and grammatically bound to the stance phrases (*suggest to* and *tend to*)

(56) This ***suggests that*** only some of the ***current*** vegetation fingerprints are seen at the permafrost table. [ARAC #58:2]

(57) This increased risk aversion may cause investors to perceive foreign investments (especially in emerging ***markets***, where political risks ***tend to*** be greater) as more risky and may cause them to withdraw money from these markets. [FRAC #93:1]

Function words occur frequently in the environment of stance phrases (norm freq=336.37), but with a relatively low MI value (MI=3.69). As suggested by Biber (2009:288), function words should not be just treated as the grammatical by-products, as they capture some hidden textual features. In this study, function words are seen to have the following two types of semantic relationship with the stance phrases. Firstly, they assist grammatically to construct the syntactic structure for the stance phrase, such as *to*, in example (58), supplementing *take into account*. Secondly, as shown by the collocational patterns, they tend to comprise another construct which may be governed by the stance phrase (*take into account).*

(58) Despite contributions to understanding real estate price dynamics at the regional level, none of the above studies have extended their conceptual framework ***to take into account the*** impact ***of*** international trade on the urban

real estate prices. (ERAC: #14:1)[11]

The co-occurrence of function words not only reveals the syntactic environments where the stance phrases occur, but also provides a clue as to the grammatical environments of an adjacent phrase which interacts with the stance phrases at the phrasal level.

7.5. Cross-disciplinary comparison

This section presents the results of our investigation into the disciplinary variation of collocation networks. I consider both the collocational network of stance phrases with respect to their within-stance collocation and their collocation with other semantic domains. Table 13 presents the cross-disciplinary comparison of the collocating patterns within the different aspects of stance phrases. As can be seen, the co-selection patterns involving hedges and cognitive stance phrases (p=0.00; w = 0.1) and reference and cognitive stance phrases (p=0.00, w=0.17) (as illustrated in example 1 and 2 respectively) are significantly more frequent in the agriculture corpus. Additionally, the MI values of these two co-occurring patterns in the agriculture corpus are notably higher than those in the economics corpus. This shows that agriculture research texts are more likely to employ a cautious tone and strengthen propositions through the inclusion of explicit references.

Table 13 The Cross-disciplinary Comparison of within-stance Collocating Patterns (per million)

Collocating patterns	Agr MI	Agr norm freq.	Eco MI	Eco norm freq.	Chi-square	Sig	Effect size
Attitude and attitude	3.81	51.52	3.83	63.56	1.91	0.17	0.13
Attitude and hedges	3.78	63.50	3.48	62.09	0.02	0.88	0.01
Cognitive and attitude	3.50	322.87	3.48	362.17	3.39	0.07	0.07
Hedges and cognitive	4.03	542.11	3.57	462.68	9.31	0.00	0.10
Hedges and attitude	3.78	63.50	3.48	62.09	0.02	0.88	0.01
Hedges and hedges	3.42	62.30	3.47	76.87	2.31	0.13	0.13
Reference and cognitive	3.74	457.05	3.52	342.95	23.98	0.00	0.17
Reference and reference	3.88	80.27	4.65	113.08	8.48	0.00	0.21
Reference and hedges	3.26	53.91	3.01	41.39	2.43	0.12	0.16

In contrast, the co-selecting pattern involving reference and reference phrases (p=0.00, w=0.21) is significantly more frequent in the economics corpus, and the collocational strength is much higher in this corpus (MI=4.65 and MI=3.88 in economics and agriculture respectively). This preference

[11] In the examples in Section 6, stance phrases are shown in **bold**, and the collocating semantic domains are *italicised and underlined*. Sources of examples are given in brackets.

shows that economics research articles are more inclined to create linkages between different sections of a text, as seen in example (39).

With regard to the collocation of stance phrases and other linguistic domains (shown in Table 14), the two disciplines display the following discipline-specific variation. Firstly, they differ significantly in the cases with topic-related words co-occurring with stance phrases. The agriculture corpus has a strong preference for discipline-specific words, such as 'farming and horticulture' and 'life and living things', which appear exclusively in the collocation network of stance phrases in the agriculture corpus. Likewise, the collocates related to 'substances and materials' and 'environment' are significantly more frequent in the agriculture corpus, with a large effect size (p=0.00; w=1.05 and p=0.00; w=0.73 for the two categories respectively). In contrast, the collocation network in the economics corpus is heavily populated with economics-related words from categories such as 'money and commerce in industry' (p=0.00; w=1.29), and 'government' (p=0.00; w=1.20). Notably, among all the significant differences identified with topic-oriented expressions, six differences were found to have a large effect size (w>0.5).

Table 14 The Cross-disciplinary Comparison of the Collocation with other Semantic Categories

General category	Semantic domains	Agr norm Freq.	Eco norm Freq.	Chi-Square	Sig (p)	Effect size
Topic-oriented expressions	Farming and horticulture	3916.39	0	5310.25	0.00	1.16
	Life and living things	3516.84	0	4767.45	0.00	1.16
	Substances and materials	5721.83	317.82	6629.80	0.00	1.05
	Environment	2043.85	489.29	1341.62	0.00	0.73
	Money and commerce in industry	23.36	12403.06	20695.03	0.00	1.29
	Government	14.38	390.99	579.42	0.00	1.20
Involvement-oriented expressions	Science and technology	2537.45	1342.23	534.21	0.00	0.37
	Social actions, states and processes	2088.18	3301.62	419.27	0.00	0.28
	Evaluation	2429.02	3104.28	125.10	0.00	0.15
	Psychological actions, states and process	4906.57	6240.33	242.89	0.00	0.15
	Modality	1482.57	1374.75	6.07	0.01	0.05
	Linguistic actions, states or process	1726.37	1266.10	104.29	0.00	0.19
General expressions	Function words	54568.84	50264.86	277.58	0.00	0.05
	Number and measurement	9422.57	8689.75	44.53	0.00	0.05
	General actions or entities	10705.67	7332.00	933.57	0.00	0.23
	Time	2813.59	1344.45	749.78	0.00	0.42
	Location	1355.58	2097.60	244.21	0.00	0.27
	Degree	1336.41	1331.14	0.02	0.90	0.00

Additionally, the seemingly discipline-neutral categories, such as ‘science and technology’ and ‘social actions, states and processes’, display considerable disciplinary variation. That is, ‘science and technology’ occurs predominantly in the network of agriculture stance phrases (p=0.00; w=0.37), while ‘social actions, states and processes’ occurs significantly more often in the economics corpus (p=0.00; w=0.28). This shows a differing emphasis on the topic-related words co-selected by the stance phrases in the two corpora.

Similarly, disciplinary variation is a noticeable feature in the co-selection of stance phrases and involvement-related expressions, including ‘evaluation’, ‘modality’, ‘psychological actions, states and process’, and ‘linguistic actions, states or process’. Specifically, the economic corpus contains a higher occurrence of the semantic domains concerning ‘evaluation’ (p=0.00; w=0.15) and ‘psychological actions, states and process’ (p=0.00; w=0.15); whereas the agriculture corpus shows a preference for the modality-related words (p=0.01; w=0.05), and ‘linguistic actions, states or process’ (p=0.01; w=0.19). In essence, these semantic categories generally correspond to the four dimensions of stance phrases under scrutiny in this study, that is, ‘psychological actions, states and process’ corresponds to cognitive phrases, ‘evaluation’ to attitudinal phrases, ‘modality’ to hedging phrases, and the ‘linguistic behaviour’ within ‘linguistic actions, states or process’ echoes reference phrases. Such co-occurrence, therefore, represents an interesting pattern of co-selection, which involves different aspects of stance at the level of phrase and lexis. From a cross-disciplinary perspective, these co-selection patterns vary significantly from agriculture to economics: in the economics corpus significantly more expressions concerning ‘evaluation’ and ‘psychological actions, states and processes’ co-occur with stance phrases, while the agriculture corpus employs more ‘modality’ and ‘linguistic description’-related words in the proximity of stance phrases.

A surprising statistical difference is also identified with general words occurring in collocation network of stance phrases, including ‘general actions or entities’ (p=0.00; w=0.23), ‘function words’ (p=0.00; w=0.05), and ‘number and measurement’ (p=0.00; w=0.05). All of these three semantic categories appear more frequently in the proximity of stance phrases in the agriculture corpus, but with relatively low effect sizes as indicated by the w value. This shows disciplinary variation also exists in the use of general words to construct the co-texts of stance phrases.

Finally, disciplinary variation is also identified with ‘time’- and ‘location’-related words in the collocation of stance phrases. The agriculture corpus displays a significant preference for ‘time’-related expressions in the network of stance phrases (p=0.01; w=0.42), while the stance phrases in the economics corpus collocate more frequently with ‘location’-related expressions (p=0.01; w=0.27). This suggests that the two disciplines prioritize different types of detail in the vicinity of stance phrases.

7.6 Summary

This chapter investigates the collocation network between stance phrases and their surrounding environment, as well as their variation in two disciplinary academic discourse. Mapping out the collocation networks helps to identify how different dimensions of stance phrases operate in tandem and how they co-occur with other semantic domains to achieve particular textual functions. Generally, in the collocation network of stance phrases, there is a complex and intricate connection between different categories of stance phrases. Also, the stance phrases share a close link with other semantic domains, which directly or indirectly interact with the stance phrases to frame the information and project authors' posture in the course of communicating information to readers. As Lemke (1998, p.34) points out, the semantic relationships "mirror and help to constitute the sociological relationships", thus the complex network can inform us on how different types of stance phrases cooperate with each other and co-select with other semantic domains to construct arguments in the interaction with the envisaged readers through the academic discourse.

The cross-disciplinary comparison uncovers the discipline-specific features regarding the collocation network of stance phrases. The two disciplines are found to have significant differences with regard to the three within-stance phrases collocation patterns, namely cognitive and hedges, cognitive and reference, reference and reference, and with regard to most collocating patterns between stance and other semantic domains. This indicates that the collocating patterns are not just shaped by but are also constitutive of conventionalized communicative norms in the respective academic communities. In other words, writers in different research communities seem to draw upon different collocating patterns when summoning linguistic resources to craft their persona and compose texts which align with disciplinary conventions.

GraphColl proves to be an efficient tool in retrieving collocates for the node words or phrases in this study. However, the analysis shows that the most significant relationships between different linguistic units occur at phrasal level, that is, one phrase often directly relates to another phrase. In view of this, it would be advantageous to be able to employ this analytical tool in future to explore the collocation patterns between formulaic expressions.

The identification of the collocation network of stance phrases regarding their within-stance phrases and collocation with other semantic domains provide a quantitative perspective on how stance phrases operate with their surrounding contexts. However, these findings are based on a corpus of published research with an empirical paradigm; it would thus be of value to expand study in this line to explore the collocation network in studies employing a different research paradigm and other genres to supplement our knowledge of the collocation networks of stance phrases.

Chapter 8 Conclusion

This chapter first reviews the research objectives and research questions of this thesis, and then summarizes the main findings obtained. This is followed by a discussion of the implications and applications of these research findings (Section 8.2). Section 8.3 reflects on the limitations of this study and Section 8.4 provides suggestions and directions for future studies on corpus-based stance expressions and cross-disciplinary academic discourse.

8.1 Introduction

The main objective of this thesis is to examine the disciplinary variation of stance phrases in the disciplines of agriculture and economics. To achieve this objective, I first built two corpora consisting of the published research articles of these two disciplines, and then followed a corpus-driven approach to extract and analyse the stance phrases. A cross-disciplinary comparison was conducted to examine the stance phrases in terms of the frequency, textual colligation and collocation network, with the aim of answering the following four research questions.

(1) Type and category of stance phrases

(a) What stance phrases are employed in the agriculture and economics academic corpora?

(b) Re-categorizing the stance framework by the metafunctions of SFL

(2) General use of stance phrases

(a) What is the frequency information of the stance phrases in the agriculture and economics academic corpora?

(b) What disciplinary variation can be found in the occurrences of stance phrases?

(3) Textual colligation of stance phrases

(a) What is the textual colligation of stance phrases in the agriculture and economics corpora?

(b) What disciplinary variation can be found in the textual colligation of stance phrases?

(4) Collocation networks of stance phrases

(a) What are the collocation networks of stance phrases in the agriculture and economics academic corpora?

(b) What disciplinary variation can be found in the collocation networks of stance phrases?

8.1.1 Research question 1: the type and category of stance phrases

To answer the first research question, a corpus-driven approach was first adopted to extract the most frequent 2- to 5-grams stance phrases in the agriculture and economics academic corpora by using the *Collocate* software. This bottom-up approach produced a relatively more comprehensive inclusion of

stance phrases used in the two corpora after the filtering by using the parameters of FTW, MI and normalised frequency. The extracted stance phrases allowed me to re-examine the current well-established stance frameworks, such as Biber et al. (1999), Martin and White's (2005), Hunston (2011), Simpson-Vlach and Ellis (2010). The analysis showed that most of these schemes converge in the following three aspects of stance features: attitude (related to feelings or values), hedges (conveying certainty or estimation), and reference (indicating the source of a proposition). However, a category which concerns authors' reason-oriented stance had been previously overlooked in these classification schemes. It is also of note that most of the above-mentioned classification schemes are based on the interpersonal function and textual functions from the metafunctions of SFL (Systemic Functional Linguistics), while the logical function, a function that is closely related to author's reason-oriented stance, had been previously neglected in these schemes. In light of the importance of this aspect in constructing authors' persuasive stance, this study draws on the logical metafunction of SFL, and incorporates the reason-oriented aspect of authors' stance, titled as 'cognitive stance' in this study, into the stance classification scheme. Also, in conjunction with the three key aspects of stance features mentioned above, this study proposes a hybrid stance classification scheme, composed of attitude (related to feelings or values), hedges (conveying certainty or estimation), reference (indicating the source of a proposition), and cognitive stance (related to the reason-oriented expressions).

Functionally, stance phrases allow writers to align with the academic conventions, augment persuasion and validity of the constructed knowledge, and enhance the effectiveness of the discipline-specific academic communication. The use of stance phrases thus can assist writers to conform to community expectations (Hunston, 1994) and align with discipline-specific conventions.

8.1.2 Research question 2: the general use of stance phrases

To address the second research question, this study used the aforementioned classification scheme, and compared statistically the occurrences of the four categories of stance phrases and identified notable disciplinary similarities and variation in the stance phrases used by the disciplinary writers. Specifically, cognitive stance phrases are more frequently used in the two corpora, which shows the disciplinary authors' preference for reason-oriented stance construction. The cross-corpora comparison also shows that agricultural writers have a significantly higher use of the cognitive-oriented category, indicating their explicit endeavour to construct reason-oriented persuasive presence in their text construction. The published texts in the two disciplines are also found to contain a substantial number of hedging phrases, suggesting writers' inclination to downplay their presence and exercise caution in making strong claims. In terms of the disciplinary variation, hedging phrases occur significantly more frequently in the agriculture corpus, implying that agriculture writers are more likely to take on a cautious attitude when constructing their academic discourse texts.

Reference stance phrases are identified as the third most frequent category in the two corpora. This shows that writers in the two corpora have a strong inclination to provide an evidential source (e.g., *according to previous studies*) and provide cross-references to construct a coherent text (*as shown in Table 1*). The disciplinary variation analysis shows that the economics corpus contains significantly more reference phrases. The occurrence of attitude stance phrases is least frequent in these two corpora, and most of the attitudinal presence is made through an evaluative lens, e.g., *play an important role,* while only a few entail emotion-related attitudinal phrases (e.g., *it is surprising, and it is interesting*). The cross-corpora comparison shows agriculture texts contain significantly more attitude stance phrases than the economics texts.

The answering of this research question and the categorization of stance phrases shares great similarities with previous frequency-based corpus-driven formulaic language or lexical bundles (i.e., Biber et al., 2004; Cortes & Hardy, 2013). Below is brief comparison between the present study and those mentioned above.

A regards the research approaches, similar to the above two studies, this study adopted a corpus-driven approach, that is, extracting all the phrases from a corpus and then classifying the phrases retrieved into different categories. The large sample and the relatively exhaustive extraction of the phrases can permit a relatively more comprehensive examination of this concept. But the above two studies examined the four-word lexical bundles, and includes *stance, discourse organizer and referential expressions* in their analysis. In contrast, this study focused on exclusively the phrases performing the stance function. This specified focus allows me to conduct a more thorough and in-depth examination of the concept.

In terms of the research focus, similar to the above two studies, this study took a phrasal perspective. While the above two studies just focused on four-word phrases, this study investigated a wider range of phrases, including 2 to 5 grams (namely the phrases containing 2 -5 words). This extended inclusion thus allows me to have a broad inclusion of stance expressions, rather than just four-word strings.

With regard to the scope of the central concept (stance), Biber, Conrad and Cortes (2004), and Cortes (2013) consider stance as a narrow concept, and include only expressions of evaluative and hedging expressions in their classification schemes. In contrast, this study treats stance as an umbrella term, and includes a broader scope, including emotion-oriented presence (which has been widely studied previously), reason-oriented presence (which was developed and added in this study), modality aspect (which has been widely studied previously) and the textual-presence (with three subcategories have been widely studied previously).

In terms of the classification criteria, the classification criteria of Biber, Conrad and Cortes (2004),

and Cortes and Hardy (2013) are based on intuition, without referring to a particular framework. In contrast, this study used the metafunctions of SFL to inform the categorization, this thus allow me to inspect this concept more systematically.

8.1.3 Research question 3: the textual colligation of stance phrases

Research question 3 extends the scope of stance studies from the well-researched areas involving frequency and disciplinary differences to a recently developed perspective, namely the textual colligation, which concerns the textual position of stance phrases. According to Hoey and Donnell (2015, p.125), linguistic items are primed to "either occupy or avoid certain recognised discourse positions", so they may occur or avoid occurring "in the beginning or end of independently recognised discourse units, e.g., the sentence, the paragraph, the speech turn" (Hoey, 2005, p.115). To address the textual colligation of stance phrases, I drew upon priming theory (Hoey, 2005) and employed the *Wordskew* software (Barlow, 2016) to extract the textual position of the stance phrases at the level of sentence, paragraph and whole texts. With regard to the text division in terms of initial, medial and final position text, I adopted approaches used by Römer and O'Donnell (2010), and Thompson (2014) to analyse the textual colligation of the stance phrases. In addition, I conducted a cross-disciplinary comparison to identify if there were significant differences in terms of the textual colligation of stance phrases, that is, if the textual positions of the stance phrases are distributed significantly differently in the two corpora.

The results show that the four types of stance phrases are unevenly distributed at the level of sentence, paragraph, and text. Cognitive, hedges, and attitude phrases have a tendency to occur in sentence-medial positions, while reference phrases are inclined to appear in sentence-initial positions. This shows that the writers in the two disciplines tend to employ more cognitive, hedges, and attitude phrases in sentential non-salient positions, but express their referential presence at the point of departure of a sentence. At the level of paragraph and text, all four types of stance phrases are found to occur in the medial position. This is likely to be because the medial position occupies a significant portion in the text division method adopted in this study. It is interesting to note that there is a higher occurrence of hedging phrases in the paragraph-and text- final than the initial position. This shows that these disciplinary writers tend to employ hedging markers at paragraph- and text- final positions than the initial positions. In contrast, reference phrases tend to occupy the paragraph-initial instead of the final positions, indicating that writers in the two disciplines tend to initiate a new paragraph by expressing their reference stance and resorting to evidential supporting sources.

From a communicative point of view, the textual colligation of stance phrases may represent a strategic choice made by disciplinary writers to ensure effective communication with the expected readers. In particular, writers may use these textual positions of stance phrases to maintain the

information flow in their text organisation, as well as to facilitate readers' cognitive processing of the information.

8.1.4 Research question 4: the collocation networks of stance phrases

The fourth research question is concerned with the collocation networks of stance phrases. To address this question, I retrieved the collocates of the stance phrases and quantified the collocates of the stance phrases in the two corpora. Most of the previous collocate-related studies mainly adopt a descriptive approach by listing the common collocates. However, this study used the software *GraphColl* to extract the collocating stance expressions and measure the quantitative relationships between stance phrases and the surrounding semantic environment. When examining the collocation network of stance phrases, the following two respective aspects are addressed, namely the within-stance phrase collocation and the collocation between stance phrases with other semantic domains.

To examine the collocation relationship within the four categories of stance phrases, I took a stepwise procedure in GraphColl search by selecting the initial node, one particular category of stance phrase (e.g., cognitive phrases), and treated the other three categories as the second, third and fourth nodes respectively. This produced a collocation network graph containing each of the four categories of stance phrases. The analysis of the stance phrases within the collocation network shows a strong collocation between cognitive phrases and hedging phrases (like *may be explained by*). This collocational pattern indicates that academic writers tend to formulate their cognitive-oriented expressions in a cautious fashion. The co-selection between cognitive and reference (like *our results imply that*) is identified as the second strongest collocational pattern, which indicates that disciplinary writers tend to associate their reason-oriented analysis with evidential or source giving expressions conveyed by reference phrases. The analysis also uncovers a notable collocational relationship between cognitive phrases and attitude phrases (like *it is interesting to note*). This indicates writers' inclination to assign an affective expression to accompany the cognitive presence, as a means of manifesting writers' affective involvement towards their cognitive behaviour.

This study also examined the collocational relationship between stance phrases and non-stance phrase expressions in order to investigate how stance phrases operate with their surrounding co-texts. The analysis of the collocating patterns show that stance phrases normally collocate with 18 semantic domains, which can be generalised into the following three macro semantic categories: topic-oriented expressions (including farming and horticulture, science and technology, substances and materials, social actions, states and processes, environment, money and commerce in industry, life and living things, and government,), involvement-oriented expressions (linguistic actions, states or process, psychological actions, states and process, modality, evaluation), and general expressions (general actions or entities, number and measurement, function words, degree, location, time). This finding

reveals the collocation network centered around stance phrases, which thus provides insight into how stance phrases work in conjunction with the different semantic types to construct a text and deliver the information to the expected readers.

The disciplinary comparison shows that the agriculture corpus contains significantly more collocating patterns of hedges and cognitive stance phrases and more patterns of reference and cognitive stance phrases. This indicates that agriculture writers tend to employ a more cautious tone and explicit references when constructing their reason-oriented cognitive expressions. In contrast, the economics texts were found to contain significantly more co-selecting patterns of reference and reference phrases, reflecting economics writers' preference for creating linkages between different sections of a text, as shown in Example (39).

As regards the disciplinary variation of the collocation patterns of stance phrases and other linguistic domains, a significant variation was identified in the co-selection of stance phrases and the topic-related expressions: the agriculture corpus contains a salient collocation of stance phrases with words denoting 'farming and horticulture', 'life and living things', 'substances and materials', and 'environment'; while economics texts comprise a high frequency of stance phrases with economics-related expressions, such as 'money and commerce in industry' and 'government'. It is also interesting to note that the seemingly discipline-neutral categories, such as 'science and technology' and 'social actions, states and processes', are found to be collocated differently with stance phrases in the two corpora. Specifically, the stance phrases in the agriculture corpus co-occur more frequently with 'science and technology', while the economics corpus contains significantly more co-selecting patterns between stance phrases and 'social actions, states and processes'. These differences may indicate a differing disciplinary emphasis in these disciplinary discourse: agriculture texts contain science-related features, while the economics texts contain social science features in the proximity of authors' stance construction.

Stance phrases are also found to co-occur significantly differently with involvement-related expressions (e.g., 'evaluation' and 'modality') in the two corpora. The stance phrases in the economic corpus have a higher co-occurrence with expressions related to 'evaluation' and 'psychological actions, states and process'; whereas the agriculture corpus contains more stance phrases collocating with words expressing 'modality' and 'linguistic actions, states or process'. Finally, the agriculture corpus contains a high occurrence of general words, including 'general actions or entities', 'function words' and 'number and measurement', appearing in the proximity of stance phrases.

Overall, the complex network of stance phrases provides insight into how stance phrases interact with their surrounding semantic environment to assist writers to construct a persuasive text. The disciplinary variation in the collocating patterns of stance phrases shows that these disciplinary

academic writers employ different collocating strategies to craft their persona and compose texts in alignment with the discipline-specific communicative conventions.

8.2 Implications and applications of this research

8.2.1 Theoretical contributions

This study makes three main theoretical contributions to our understanding of stance expressions in disciplinary academic discourse.

First of all, this study investigates the different aspects of stance phrases, namely the frequency, the textual colligation, and the collocation network. The findings obtained can contribute to our knowledge with respect to how often stance phrases are used, where these phrases often occur, and how they interact with the surrounding texts. The latter two perspectives, namely the textual colligation and the collocation network of stance expressions have been inadequately explored in previous studies. These findings thus supplement the current picture of stance features in terms of these two aspects in the academic discourse.

The second theoretical contribution concerns the categorization of stance expressions. Drawing upon the metafunctions of SFL, this study first synthesized the previous classification schemes, attitude (emotion-oriented aspect, related to feelings or values), hedges (modality-oriented aspect, conveying certainty or estimation), and reference (evidentially-oriented aspect), which correspond to the interpersonal and textual metafuctions. Also, based on the ideational function, the logical function in particular, this study introduced a new category concerned the reason-oriented aspect to the stance classification scheme. This synthesized classification scheme is found to be a good fit when applied to the stance phrases retrieved by a corpus-driven approach in the academic discourse. This hybrid stance phrase framework contributes to a systematic categorization of stance phrases from a metafunctional perspective, and thus can provide a useful reference for future studies of stance features.

The third theoretical contribution concerns the disciplinary variation in stance construction. The notable disciplinary differences identified in this study, concerning how agriculture and economics writers express and construct their disciplinary stance and authorial positioning, can add to the current literature of the disciplinary features. The findings are also useful for exploring the common research paradigms that these disciplines often subscribe to, and thus can be drawn upon by future research which aims to identify the disciplinary conventions and chart the disciplinary characteristics. That said, it is also necessary to bear in mind that there are some similarities in terms of the stance expressions used by the disciplinary writers, which are still of value to explore if these similarities are shared by other disciplines.

In terms of the focus, this study takes an extended perspective and examines stance with regard to their frequency, textual colligation and collocation network. Previous corpus-driven and corpus-based studies of stance expressions have mainly focused on frequency information, while their features regarding textual colligation have not been explored to the best of my knowledge. The analysis of the textual colligation of stance phrases suggests that this aspect represents the timing that authors decide to position themselves, and that textual colligation of stance phrases represent a strategical deployment for academic writers to achieve effective communication.

The collocation network aspect of stance phrases advances studies on stance features from a quantitative perspective. The complex collocation network depicted by *GraphColl* can expand our understanding of the complex collocational relationship between stance phrases and their surrounding environments, and thus contributes to a better understanding of stance construction in the disciplinary context.

8.2.2 Methodological implications

Methodologically, this study makes three methodological contributions to corpus-driven and corpus-based analyses.

Firstly, I used lemmatisation when retrieving the phrases for analysis. Previous studies on phrase/collocation extraction often treated the phrases of the same lemma with different inflectional forms as different phrases and used a frequency threshold (like 10 per million words) to decide on the phrases to be included for analysis. This approach may exclude some phrases that have a lower frequency for each of the different inflection forms (like *seems that, seem that, seemed that*), but the total frequency of a lemma phrase can be frequent enough to pass the frequency selection. To solve this issue, I used lemmatisation in the process of phrase extraction to assist in achieving relatively more comprehensive stance phrases and avoid excluding stance phrases with low frequencies for their different inflectional forms but with high total frequency for the lemma phrase. The lemmatisation also helps to obtain a single mutual information (MI) score for a phrase lemma regardless of its different inflectional forms, and this facilitates the phrase selection.

Another methodological contribution of this study is the employment of formula teaching worth (FTW). Previous corpus-based filtering mainly used the parameter of normalised frequency, or MI (mutual information), while the teaching value of a phrase was not well considered. This study used a criterion of FTW developed by Simpson-Vlach and Ellis (2010), which contributes to identifying the phrases of high pedagogical value.

In terms of statistical analysis, this study used effect size, in conjunction with the Log-likelihood and Chi-square tests, to measure the statistical difference, as p-values may be affected by sample sizes. The effect size statistics provide information with regard to the magnitude of the significant difference

between two variables, thus showing an additional perspective regarding the degree of significant difference in stance features of the two disciplines.

This study focuses on two relatively new perspectives, textual colligation and collocation network, in conjunction with frequency information. These extended perspectives may contribute to expanding the scope of corpus linguistics.

Also, this study employed two recently developed software programmes *Wordskew* and *GraphColl* to explore textual positions of stance phrases and the quantitative relationship between stance phrases and surrounding semantic environment. The two corpus tools were found to be useful for investigating the textual colligation and collocation networks of stance expressions, which thus can be drawn upon by writers who intend to explore textual colligation and quantitative collocation of particular linguistic items.

8.2.3 Pedagogical applications

In terms of the pedagogical applications, the findings in this study can be useful in the following three respects. Firstly, the findings regarding the stance phrases obtained in this study can be operationalised as teaching materials and be incorporated into class instruction and curriculum design for EAP/ESP teaching, particularly to guide EFL/ESL student writers in the two disciplines to express an appropriate tone and construct a conventionalised stance in their academic writings.

The findings regarding the textual colligation of stance phrases can be drawn upon in class instruction to raise students' textual colligation awareness and help them to allocate stance markers to suitable textual positions. The findings concerning the collocation network can be used to assist learners to construct stance in relation to their surrounding contexts, thereby establishing their stance in a conventionalized way.

With respect to the findings on collocation network (research question 3), the pedagogical value resides in helping novice and student writers to gain knowledge and awareness of utilizing the appropriate collocational patterns to express an appropriate tone when projecting themselves into the text. For instance, the collocational patterning of stance phrases can be incorporated into teaching materials to draw students' attention to typical stance phrase combinations (e.g., *may be explained by, it is interesting to note that,* and *it may be interesting to*). It can also be useful to undertake guided exercises during classroom instruction to foster students' awareness of the use of stance phrases in appropriate contexts, that is, in combination with other linguistic units or in particular textual positions. A more advanced task would require students to use stance phrases to fulfil certain purposes when drafting a particular text. For instance, the task may instruct students to employ either a confident or a cautious tone when relaying their findings, a stipulation which is intended to guide students in their choice of hedges.

This study can be pedagogically useful for workshop-based publication training for novice writers. In particularly, the findings in this study can be drawn upon by practitioners to assist novice writers in these disciplines to gain the knowledge of employing the right amount of stance markers, placing them at the suitable textual position, and accompanying them with suitable within-stance categories and other semantic domains.

Another possible application avenue for the findings is in machine learning, especially in terms of the textual position and collocation network of stance phrases. The findings may also be useful to improve the current grammar check software, particularly in terms of checking and correcting EFL writers' errors from the perspective of textual colligation and the possible collocation network.

8.3 Limitations and suggestions for future work

It should be noted that this study only scratches the surface of stance expressions from the aspect of 2- to 5-grams, which might exclude some single words with stance meaning such as *perhaps* and *maybe*, as well as the stance expressions with a diversity of forms. It is also possible that stance functions can be performed by non-frequent phrases, which may escape the filtering criterial adopted in this study. Therefore, this study only presents the use of stance phrases that occur most frequently in the two corpora.

In addition, the focus of this study was limited to empirical RAs in two academic disciplines. Thus, an exploration of stance features in a wider range of disciplines would enable a more comprehensive view of stance features in academic discourse. A focus on different research paradigms (quantitative and qualitative) may also contribute to revealing the stance construction features in academic discourse. It is also necessary to point out that this study did not distinguish the research articles produced by native and non-native writers, as this study mainly aimed at exploring disciplinary variation of the stance features. For future studies, it may be of interest to explore differences in the use of stance by writers of different language or cultural backgrounds.

Another limitation is that the focus on the corpus data may limit our understanding to the corpus data, while the writer's perception of their use, and why they are used particular ways would need to be explored through other methods, such as interview or ethnographic approaches.

8.4 Suggestions for future studies

In light of the above-mentioned limitation, future studies may consider advancing the studies in this strand from the following perspectives.

Firstly, stance is a complex concept and tends to be represented by a wide range of linguistic devices, such as evaluative adjectives (Biber 2004; Hunston 2004), adverbs (*clearly, certainly*) (Biber et al., 1999; Degaetano-Ortlieb 2015), and extraposed patterns (Charles, 2006; Groom, 2005; Hewings

& Hewings, 2002; Larson, 2016). Therefore, it would be of high value to explore a broad range of linguistic forms related to stance features.

Also, this study proposes to integrate a new dimension, namely cognitive stance concerning the reason-oriented aspect, into the classification scheme of stance. Further studies are still needed to examine how reason-oriented stance is constructed in other academic disciplines, in different genres, and even in different academic writing cultures.

According to the longitudinal exploration of the discipline-specific writing in different disciplines carried out by Dressen-Hammouda (2014) and Hyland and Jiang (2017, 2018), disciplinary writers' stance construction is a dynamic process and exhibits an interesting changing pattern. Therefore, future studies may consider tracing the dynamic development of authors' stance and exploring the dynamic features of authors' stance construction.

In terms of the disciplinary variation, this study focuses on the stance features in two academic disciplinary tribes. Although notable findings have been obtained in these two disciplines, many other disciplines yet remain to be charted in terms of how authors construct their stance and communicate with their expected readers. This is especially true for some recently emerging disciplines, like bioinformatics, quantum biology, computational social science, cognitive economics, etc. Future work may also consider supplementing the picture of stance construction features by exploring how it is constructed in other genres, as well as in different culture groups. Further comparative studies are needed to explore the differences between EFL/ESL student writing and that of native students and professional writers, in order to identify the gap and common mistakes in EFL/ESL students' and novice writers' writing. This knowledge can help to design appropriate instructional approaches to assist students and novice writers to improve their academic writing.

Another promising direction for the stance studies is to examine the textual colligation and collocation network of stance expressions. The current exploration of stance has mainly focused on the frequency information, while the aspects concerning the textual colligation and collocation network of this concept remain to be explored. For the purpose of a better understanding of stance features from different perspectives, the knowledge concerning the textual colligation and collocation network may be worth exploring for future studies on stance. Also, the two approaches can be applied to the exploration of other linguistic items, as well as their disciplinary features and genre variation. These two approaches have contributed to broadening the type of analysis that can be done in corpus linguistics.

Finally, although disciplinary information regarding the use of stance markers has been identified, much remains to be done before this knowledge can be incorporated into materials and classroom teaching procedures. Future investigation may consider exploring possible pedagogical approaches to

integrating discipline-specific stance expressions into EAP/ESP teaching, academic writing instruction or workshop training in order to prepare students and novice writers for academic writing for publication purposes.

References

Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Philadelphia: John Benjamins.

Ädel, A. (2014). Selecting quantitative data for qualitative analysis: A case study connecting a lexicogrammatical pattern to rhetorical moves. *Journal of English for Academic Purposes*, *16*, 68–80. https://doi.org/10.1016/j.jeap.2014.09.001

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81-92.

Afros, E., & Schryer, C. F. (2009). Promotional (meta) discourse in research articles in language and literary studies. *English for Specific Purposes, 28*(1), 58-68.

Angelil-Carter, S. (1998). *Access to success: Academic literacy in higher education*. Juta and Company Ltd.

Aull, L.L., & Lancaster, Z. (2014). Linguistic Markers of Stance in Early and Advanced Academic Writing: A Corpus-based Comparison. *Written Communictaion*, *31*(2), 151-183.

Bakhtin, M. (1981). *The dialogic imagination*. (M. Holquist, Ed.; C. Emerson & M. Holquist, Trans.). Austin, TX: University of Texas Press.

Bamford, J., & Bondi, M. (2005). *Dialogue within discourse communities: Metadiscursive perspectives on academic genres* (Vol. 28): Walter de Gruyter.

Barlow, M. (2004). Collocate 1.0. *Collocation Extraction Software.Houston TX: Athelstan.*

Barlow, M. (2016). *WordSkew* : Linking corpus data and discourse structure. *International Journal of Corpus Linguistics*, *21*(1), 105–115. https://doi.org/10.1075/ijcl.21.1.05bar

Barnbrook, G. (1996). *Language and computers: A practical introduction to the computer analysis of language / Geoff Barnbrook*. Edinburgh: Edinburgh: Edinburgh University Press 1996 (2004 printing).

Basturkmen, H. (2014). Replication Research in Comparative Genre Analysis in English for Academic Purposes. *Language Teaching: Surveys and Studies, 47*, 377-386.

Basturkmen, H., & Bitchener, J. (2005). The text and beyond: exploring the expectations of the academic community for the discussion of results section in master's theses. *New Zealand Studies in Applied Linguistics*, *11*(1), 1-19.

Bakhtin, M. (1984). Problems of Dostoevsky's Poetics. Edited and Translated by Caryl Emerson. Introduction by Wayne C. Booth. *Minneapolis: University of Minnesota Press. (Theory and History of Literature 8)*.

Bazerman, C. (2010). *The informed writer:* Using Sources in the Disciplines. Houghton Mifflin Company. Fort Collins, Colorado

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. McGraw-Hill International.

Berger, P., & Luckmann, T. (1966). The social construction of knowledge: A treatise in the sociology of knowledge. *Open Road Media: Soho, NY, USA*.

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. (2004). Historical patterns for the grammatical marking of stance: A cross-register comparison. *Journal of Historical Pragmatics*, *5*(1), 107-136.

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. (2012). Corpus-based and corpus-driven analyses of language variation and use. In *the Oxford handbook of linguistic analysis*. Oxford University Press.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286.

Biber, D., & Conrad S. (2009). *Register, Genre and Style*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson, P. Rayson, & T. McEnery (Eds.), *Corpus linguistics by the lune: A festschrift for Geoffrey Leech*. Frankfurt/Main: Peter Lang.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371-405.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Bird, S. (2006). NLTK: The natural language toolkit. *Proceedings of the COLING/ACL on Interactive Presentation Sessions,* 69-72.

Bloch, J. (2003). Creating materials for teaching evaluation in academic writing: Using letters to the editor in L2 composition courses. *English for Specific Purposes, 22*(4), 347-364.

Bondi, M. (1999). *English across genres: Language variation in the discourse of economics.* Modena: Edizioni Il Fiorino.

Boote, D. N., & Beile, P. (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.

Boye, K., & Harder, P. (2009). Evidentiality: Linguistic categories and grammaticalization. *Functions of Language*, *16*(1), 9-43.

Bowden, D. (1999). *The mythology of voice*: Heinemann.

Buckingham, L. (2008). Development of English Academic Writing Competence by Turkish Scholars. *International Journal of Doctoral Studies Volume*, (*3)*, 1–18.

Buckingham, L. J. (2014). Building a career in English: Users of English as an additional language in academia in the Arabian Gulf. *TESOL Quarterly: a journal for teachers of English to speakers of other languages and of standard English as a second dialect, 48* (1), 6-33.

Brezina, V. (2016) Collocation networks. In Baker, P., & Egbert, J. (Eds.), *Triangulating Methodological Approaches in Corpus Linguistic Research* (pp.90-107). New York: Routledge.

Brezina, V., Mcenery, T., & Wattam, S. (2015). John Benjamins Publishing Company. *International Journal of Corpus Linguistics*, *20*(2), 139–173.

Bruce, I. (2016). Constructing critical stance in University essays in English literature and sociology. *English for Specific Purposes*, *42*, 13–25.

Bruce, I. (2014). Expressing criticality in the literature review in research article introductions in applied linguistics and psychology. *English for Specific Purposes*, *36*(1), 85–96.

Bruce, I. J. (2003). *Cognitive genre prototype modelling and its implications for the teaching of academic writing to learners of English as a second language*. University of Waikato.

Bybee, J. L., & Fleischman, S. (1995). *Modality in grammar and discourse* (Vol. 32): John Benjamins Publishing.

Chafe, W. L., & Nichols, J. (1986). *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: 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.

Charles, M. (2006a). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics, 27*(3), 492-518.

Charles, M. (2006b). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*(3), 310-331.

Charles, M. (2009). Stance, interaction and the rhetorical patterns of restrictive adverbs: Discourse roles of only, just, simply and merely. *Academic Writing: At the Interface of Corpus and Discourse*, 152-169.

Charles, M., Pecorari, D., & Hunston, S. (2009). *Academic writing electronic resource]: At the interface of corpus and discourse / edited by maggie Charles, Diane Pecora Ri and susan hunston*. London; New York: London; New York: Continuum c2009.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29*(1), 72-89.

Conrad, S. M. (1996). *Academic discourse in two disciplines: Professional writing and student development in biology and history*. Northern Arizona University. Unpublished Doctoral Dissertation,

Conrad, S.M., & Biber, D. (2000). Adverbial marking of stance in speech and writing. *Evaluation in Text: Authorial Stance and the Construction of Discourse*, 56-73.

Conrad, S., & Mauranen, A. (2003). The corpus of English as lingua franca in academic settings. *TESOL Quarterly, 37*(3), 513-527.

Conroy, C. A., Dailey, A. L., & Shelley-Tolbert, C. A. (2000). The move to agriscience and its impact on teacher education in agriculture. *Journal of Agricultural Education*, *41*(4), 51-61.

Cortes, V. (2001). The issue of corpus size in the study of lexical bundles. *Paper presented at the third north american symposium on corpus linguistics and language teaching*, Boston.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397–423. https://doi.org/10.1016/j.esp.2003.12.001

Cortes, V. (2007). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education, 17*(4), 391-406.

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, *3*(1), 43–57.

Cortes, V. & Hardy, J. (2013). Analyzing the semantic prosody and semantic preference of lexical bundles. In D. Belcher and G. Nelson (Eds.), *Critical and corpus-based approaches to intercultural rhetoric* (pp, 180-200). Ann Arbor, Ml: University of Michigan Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Crismore, A. (1989) *Talking with readers: Metadiscourse as rhetorical act*. New York: Peter Lang Publishers.

Crismore, A., & Farnsworth, R. (1990). Metadiscourse in popular and professional science discourse. InW. Nash (Ed.), *The writing scholar: Studies in academic discourse*. Newbury Park, CA: Sage

Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics*, *34*(3), 369–388.

Dahl, T. (2004). Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *Journal of Pragmatics*, *36*(10), 1807-1825.

Davies, A. (2000). Review. Other Floors, Other Voices: A Textography of a Small University Building. JM Swales. *Applied Linguistics, 21*(4), 585-590.

Degaetano-Ortlieb, S, (2015). Evaluative meaning and cohesion: the structuring function of evaluative meaning in scientific writing. *Discours,* 16, 3–29.

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.

Dong, J. & Jiang, F. (2018). John Flowerdew and Tracey Costley (eds): Discipline-Specific Writing Theory into Practice, Applied Linguistics, amy010, *Applied Linguistics*, *2018* (1):1-5.

Dong, J. & Jiang, F. Construing evaluation through patterns: register-specific variations. *Australian Journal of Linguistics*. (In press) 2019??

Donohue, J. P. (2006). How to support a one-handed economist: The role of modalisation in economic forecasting. *English for Specific Purposes, 25*(2), 200-216.

Dressen, D. (2003). Geologists' implicit persuasive strategies and the construction of evaluative evidence. *Journal of English for Academic Purposes, 2*(4), 273-290.

Dressen-Hammouda, D. (2008). From novice to disciplinary expert: Disciplinary identity and genre mastery. *English for Specific Purposes, 27*(2), 233-252.

Dressen-Hammouda, D. (2014). Measuring the voice of disciplinarity in scientific writing: A longitudinal exploration of experienced writers in geology. *English for Specific Purposes, 34*, 14-25.

Du Bois, J. W. (2007). The stance triangle. *Stancetaking in discourse: Subjectivity, evaluation, interaction*, 139-182.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1), 61-74.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, jsp004, 917-928.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 28*(3), 157-169.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, *38*(2), 165-193.

Ellis, N. C. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology: Section A, 49*(1), 234-250.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 375-396.

Fairclough, N. (1992). *Discourse and social change* (Vol. 73): Polity press Cambridge.

Feak, C. B., & Swales, J. M. (2011). *Creating contexts: Writing introductions across genres*. Ann Arbor, MI: University of Michigan Press.

Firth, J. R. (1935). Phonological features of some Indian languages. *Reprinted in Firth (1957)*, 47-53.

Flowerdew, J. (1993). Concordancing as a tool in course design. *System, 21*(2), 231-244.

Flowerdew, J. (2000). Discourse community, legitimate peripheral participation, and the nonnative-English-speaking scholar. *TESOL Quarterly*, 127-150.

Fortanet-Gómez, I. (2004). Verbal stance in spoken academic discourse. In G. Del Lungo Camiciotti & E. Tognini Bonelli (Eds.), *Academic Discourse–New Insights into Evaluation*. Bern: Peter Lang.

Francis, G. (1994). Labelling Discourse: An Aspect of Nominal-group Lexical Cohesion. *Advances in Written Text Analysis*. M. Coulthard (ed). London: Routledge. 83-101.

Francis, G., Hunston, S., & Manning, E. (1996). Collins COBUILD Grammar Patterns 1: *Verbs*. London: HarperCollins.

Francis, G., Hunston, S., & Manning, E. (1998). Collins COBUILD Grammar Patterns 2: *Nouns and Adjectives*. London: HarperCollins.

Freddi, M. (2005). Arguing linguistics: Corpus investigation of one functional variety of academic discourse. *Journal of English for Academic Purposes, 4*(1), 5-26.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, *67*, 155–179.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus‐based analysis. *TESOL Quarterly, 41*(2), 339-359.

Gardner, S., & Nesi, H. (2013). A classification of genre families in university student writing. *Applied Linguistics*, *34*(1), 25–52.

Giannoni, D. S. (2010). Mapping academic values in the disciplines: *A corpus-based approach* (Vol. 124): Peter Lang.

Gillaerts, P., & Van de Velde, F. (2010). Interactional metadiscourse in research article abstracts. *Journal of English for Academic Purposes, 9*(2), 128-139.

Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes, 19*(2), 115-135.

Gledhill, C. (2011). The 'lexicogrammar'approach to analysing phraseology and collocation in ESP texts. *ASp. La Revue Du GERAS,* (59), 5-23.

Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics, 13*(4), 403-437.

Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes, 4*(3), 257-277.

Halliday, M. A. K. (1978). *Language as social semiotic*. Arnold: London.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education, 5*(2), 93-116.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: London: E. Arnold 1994.

Halliday, M. A. K. (2003). On the "architecture" of human language. *On Language and Linguistics, 3*, 1-29.

Halliday, M. A., & Hasan, R. (1989). Language, context, and text: *Aspects of language in a social-semiotic perspective*. Victoria: Deakin University Press.

Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, *8*(2), 183-207.

Hasan, R. (2009). *Wanted: a theory for integrated sociolinguistics*. London: Equinox.

Hasan, R. (2014). Towards a paradigmatic description of context: systems, metafunctions, and semantics. *Functional Linguistics*, *1*(1), 1-54.

Hatzitheodorou, A-M. (2008). *Comprehension and Production of Written Discourse in a University EFL Context*. Thessaloniki: Zita Publications.

Hewings, M., & Hewings, A. (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. (2004). Textual colligation: a special kind of lexical priming. *Language and Computers*, *49*, 171-194.
Hoey, M. (2005). *Lexical priming: A new theory of words and language*: Psychology Press.
Hoey, M., & O’Donnell, M. B. (2008). Lexicography, grammar, and textual position. *International Journal of Lexicography*, *21*(3), 293–309.
Hoey, M., & O’Donnell, M. B. (2015). Examining associations between lexis and textual position in hard news stories, or according to a study by.... In Groom, N., Charles, M., & John, S. (Eds.), *Corpora, Grammar and Discourse. In Honour of Susan Hunston.* Amsterdam: John Benjamins.
Hood, S. (2006). The persuasive power of prosodies: Radiating values in academic writing. *Journal of English for Academic Purposes, 5*(1), 37-49.
Hood, S. (2010). *Appraising research: Evaluation in academic writing.* Palgrave Macmillan.
Hu, G., & Cao, F. (2011). Hedging and boosting in abstracts of applied linguistics articles: A comparative study of English-and Chinese-medium journals. *Journal of Pragmatics*, *43*(11), 2795-2809.
Hu, G., & Cao, F. (2015). Disciplinary and paradigmatic influences on interactional metadiscourse in research articles. *English for Specific Purposes*, *39*, 12–25.
Hübler, A. (1983) *Understatements and Hedges in English*. Amsterdam: John Benjamins.
Hunston, S. (1993). Evaluation and ideology in scientific writing. In Mohsen Ghadessy (Eds) *Register Analysis: Theory and Practice*, 57-74.
Hunston, S. (1994). Evaluation and organization in a sample of written academic discourse. *Advances in Written Text Analysis*, 191-218.
Hunston, S. (2002). *Corpora in applied linguistics*: Cambridge University Press.
Hunston, S. (2004). Counting the uncountable: Problems of identifying evaluation in a text and in a corpus. *Corpora and Discourse, 9*, 157-188.
Hunston, S. (2007a). Semantic prosody revisited. *International Journal of Corpus Linguistics*, 2, 249–268.
Hunston, S. (2007b). Using a corpus to investigate stance quantitatively and qualitatively. In Englebretson R. (Eds): *Stancetaking in discourse: Subjectivity, evaluation, interaction.* Amsterdam: John Benjamins.
Hunston, S. (2008). Starting with the small words Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics, 13*(3), 271-295.
Hunston, S. (2011). *Corpus approaches to evaluation: Phraseology and evaluative language.* New York: Routledge.
Hunston, S. (2013). [Review of the book *Corpus Linguistics: Method, Theory and Practice, by* McEnery, T. & A. Hardie]. *International Journal of Corpus Linguistics, 18(2), 290-294.*
Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
Hunston, S., Francis, G., & Johnson, C. (2000). Pattern Grammar: *A Corpus-Driven Approach to the Lexical Grammar of English*, *27*(2), 318–320.
Hunston, S., & Sinclair, J. (2000). *A local grammar of evaluation*. In Hunston, S., & Thompson, G. (eds.), 74–101.
Hunston, S., & Thompson, G. (2000). *Evaluation in Text: Authorial Stance and the Construction of Discourse: Authorial Stance and the Construction of Discourse*. Oxford University Press.
Huffman, W. E., & Evenson, R. E. (2008). Science for agriculture: *A long-term perspective*. John Wiley & Sons.
Hyland, K. (1996). Writing without conviction: hedging in scientific research articles. *Applied Linguistics*, *17*(4), 433–454.
Hyland, K. (1998a). *Hedging in scientific research articles* (Vol. 54). John Benjamins Publishing.
Hyland, K. (1998b). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, *30*(4), 437-455.
Hyland, K. (1999a). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, *20*(3), 341-367.
Hyland, K. (1999b). Disciplinary discourses: Writer stance in research articles. In C. Candlin & K. Hyland. (Eds.), *Writing: Texts, processes and practices* (pp. 99-121). London: Longman.
Hyland, K. (1999c). Talking to students: Metadiscourse in Introductory Coursebooks. *English for Specific Purposes*, *18*(1), 3-26.
Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, *20*(3), 207-226.

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. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, *13*(2), 133-151.

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 2005.

Hyland, K. (2006). Disciplinary differences: Language variation in academic discourses. In K Hyland & M. Bondi (Eds.), *Academic Discourse Across Disciplines* (pp. 17–48). Bern: Peter Lang AG.

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing, 16*(3), 148-164.

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

Hyland, K. (2008b). Persuasion, Interaction and the Construction of Knowledge: Representing Self and others in Research Writing. *International Journal of English Studies*, *8*(2), 1–23.

Hyland, K. (2009). Academic Discourse: *English In A Global Context*. London: Continuum.

Hyland, K. (2011). Disciplines and Discourses: Social Interactions in the Construction of Knowledge. In D. Starke-Meyerring, N.Artemeva, M.Horne & Yousoubova, L.(Eds.), *Writing in knowledge societies. Perspectives on writing* (pp. 193-214). Fort Collins: The WAC Clearinghouse.

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, *32*, 150–169.

Hyland, K. (2013). Disciplinary identities: *Individuality and community in academic discourse*, ken hyland. cambridge university press (cambridge applied linguistics, ed. carol A. chapelle and susan hunston), cambridge (2012).

Hyland, K., & Hamp-Lyons, L. (2002). EAP: Issues and directions. *Journal of English for Academic Purposes*, *1*, 1–12.

Hyland, K., & Jiang, K. F. (2017). "We Believe That … ": Changes in an Academic Stance Marker. *Australian Journal of Linguistics*, (December 2017), 1–23.

Hyland, K., & Jiang, K. F. (2018). "In this paper we suggest": Changing patterns of disciplinary metadiscourse. *English for Specific Purposes*, *51*, 18–30.

Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, *6*(2), 183-205.

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes, 24*(2), 123-139.

Hyland, K., & Tse, P. (2009). Academic lexis and disciplinary practice: corpus evidence for specificity. *International Journal of English Studies*. Retrieved from http://dialnet.unirioja.es/servlet/articulo? codigo=3104270%5 Cnhttp://www.eric.ed.gov/PDFS/EJ878445.pdf

Ifantidou, E. (2005). The semantics and pragmatics of metadiscourse. *Journal of Pragmatics, 37*(9), 1325-1353.

Ivanic, R. (1998). Issues of identity in academic writing. *Writing and Identity: The Discoursal Construction of Identity in Academic Writing*, 75–106.

Jespersen, O. (1924). *The philosophy of grammar*. London: George Allen & Unwin.

Jiang, F. K. (2015). Nominal stance construction in L1 and L2 students' writing. *Journal of English for Academic Purposes*, 20, 90-102.

Jiang, K. F., & Hyland, K. (2016). Nouns and academic interactions: A neglected feature of metadiscourse. *Applied Linguistics*, 37 (4), 1-25.

Jiang, F. K., & Hyland, K. (2017). Metadiscursive nouns: Interaction and cohesion in abstract moves. *English for Specific Purposes*, *46*, 1-14.

Jiang, K. F. (2017). Stance and voice in academic writing. *International Journal of Corpus Linguistics*, *22*(1), 85–106.

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

Johns, A. M., & Swales, J. M. (2002). Literacy and disciplinary practices: Opening and closing perspectives. *Journal of English for Academic Purposes, 1*(1), 13-28.

Johnston, J. E., Berry, K. J., & Mielke Jr, P. W. (2006). Mesuares of effect size for Chi-squared and likelihood-ration goodness-of-fit tests. *Perceptual and Motor Skills*, *103*(2), 412-414.

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences. *Formulaic Sequences*, 269-292.

Jurafsky, D. & Martin, J. H. (2000). Speech and Language Processing: *An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Englewood Cliffs, NJ: Prentice-Hall.

Kaltenböck, G., Mihatsch, W., & Schneider, S. (2010). *New approaches to hedging* (Vol. 9): BRILL.

Khedri, M., Heng, C. S., & Ebrahimi, S. F. (2013). An exploration of interactive metadiscourse markers in academic research article abstracts in two disciplines. *Discourse Studies*, *15*(3), 319-331.

Kline, R. B. (2004). Beyond significance testing: *Reforming data analysis methods in behavioral research.* Washington, DC: American Pyscological Assocaition.

Koutsantoni, D. (2004). Attitude, certainty and allusions to common knowledge in scientific research articles. *Journal of English for Academic Purposes, 3*(2), 163-182.

Koutsantoni, D. (2007). Developing academic literacies: *understanding disciplinary communities' culture and rhetoric* (Vol. 4): Peter Lang.

Lancaster, Z. (2011). Interpersonal stance in L1 and L2 students' argumentative writing in Economics: Implications for faculty development in WAC/WID programs. *Across the Disciplines*, 8(4). Retrieved from http://wac.colostate.edu/atd/ell/lancaster.cfm.

Lancaster, Z. (2016). Expressing stance in undergraduate writing: Discipline-specific and general qualities. *Journal of English for Academic Purposes*, *23*, 16-30.

Larsson, T. (2016). 'The introductory it pattern: Variability explored in learner and expert writing'. *Journal of English for Academic Purposes 22*: 64–79.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York and London: Routledge.

Leech, G. (1992). Corpora and theories of linguistic performance. Directions in corpus linguistics. *Proceedings of Nobel Symposium* 82, ed. by Jan Svartvik, 105–122. Berlin, New York: Mouton de Gruyter.

Leech, G., P. Rayson, and A. Wilson. (2001). *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. Longman.

Lemke, J. L. (1995). Textual Politics: *Discourse and Social Dynamics*. London: Taylor & Francis.

Lemke, J. L. (1998). Resources for attitudinal meaning: Evaluative orientations in text semantics. *Functions of Language*, *5*, 33–56.

Lewin, B. A. (2005). Hedging: An exploratory study of authors' and readers' identification of ‘toning down’ in scientific texts. *Journal of English for Academic Purposes*, *4*(2), 163-178.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta‐analysis. *Language Learning*, *60*(2), 309-365.

Lillis, T., & Curry, M. J. (2013). Academic writing in a global context: *The politics and practices of publishing in English*. Routledge.

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.

Lim, J. M.-H. (2012). How do writers establish research niches? A genre-based investigation into management researchers' rhetorical steps and linguistic mechanisms. *Journal of English for Academic Purposes*, *11*(3), 229-245.

Lin, P. M. S. (2009). Recent developments in research on phraseology in foreign language teaching and learning. *International Journal of Corpus Linguistics*, *14*, 255–269.

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes, 31*(1), 25-35.

Liu, M. S. M., & Lim, J. M. H. (2014). How do writers evaluate their own empirical research? A genre-based inquiry into economics journal papers. *Procedia-Social and Behavioral Sciences*, 134, 13-22.

Liu, Y., & Buckingham, L. (2018). Journal of English for Academic Purposes The schematic structure of discussion sections in applied linguistics and the distribution of metadiscourse markers. *Journal of English for Academic Purposes*, *34*, 97–109.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in SLA: A collection of empirical studies* (pp. 408–452). New York: Oxford University Press.

Mahlberg, M. (2007). Clusters, key clusters and local textual functions in Dickens. *Corpora*, *2*(1), 1-31.

Mahlberg, M. (2009). Local textual functions of move in newspaper story patterns. In U. Römer & R. Schulze (Eds.), *Exploring the lexis-grammar interface* (pp. 265–287). John Benjamins.

Mahlberg, M. & O'Donnell, M. B. (2008). "A fresh view of the structure of hard news stories". In S. Neumann & E. Steiner (Eds.), *Online Proceedings of the 19th European Systemic Functional Linguistics Conference and Workshop*, Saarbrücken, 23–25 July 2007.

Magnet, A., & Carnet, D. (2006). Letters to the editor: Still vigorous after all these years?: A presentation of the discursive and linguistic features of the genre. *English for Specific Purposes, 25*(2), 173-199.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Massachusetts: MIT press.

Marco, M. J. L. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for Specific Purposes, 19*(1), 63-86.

Markkanen, R., & Schröder, H. (1997). Hedging and discourse: *Approaches to the analysis of a pragmatic phenomenon in academic texts*. Walter de Gruyter.

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

Martin, W. M. (2006). Theories of judgment: *Psychology, logic, phenomenology / wayne M. martin*. Cambridge, UK; New York: Cambridge, UK; New York: Cambridge University Press 2006.

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, *28*(3), 183-198.

Martínez, R., & Schmitt, N. (2012). A phrasal expression list. *Applied Linguistics*, *33*(3): 299-320.

Matsuda, P. K. (2001). Voice in Japanese written discourse: Implications for second language writing. *Journal of Second Language Writing, 10*(1), 35-53.

Matsuda, P. K., & Tardy, C. M. (2007). Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes, 26*(2), 235-249.

Mauranen, A. (1993). Contrastive ESP rhetoric: metatext in Finnish–English economics texts. *English for Specific Purposes*, 3–22.

Mauranen, A. (2004). Where next? A summary of the round table discussion. *Academic discourse: New insights into evaluation. Bern: Peter Lang*, 203-215.

Mauranen, A., & Bondi, M. (2003). Evaluative language use in academic discourse. *Journal of English for Academic Purposes*, *2*(4), 269–271.

Mauranen, A., Hynninen, N., & Ranta, E. (2010). English as an academic lingua franca: The ELFA project. *English for Specific Purposes, 29*(3), 183-190.

McEnery, T. (2006). Swearing in English: *Bad Language, Purity and Power from 1586 to the Present*. Abington, UK: Routledge.

Macmurray, J. (1935) *Reason and Emotion*, Princeton University Press.

Moreno, A. I. (2004). Retrospective labelling in premise-conclusion metatext: An English-Spanish contrastive study of research articles on business and economics. *Journal of English for Academic Purposes*, *3*(4), 321–339.

Morton, J. (2009). Genre and disciplinary competence: A case study of contextualisation in an academic speech genre. *English for Specific Purposes, 28*(4), 217-229.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes, 25*(2), 235-256.

Muñoz, V. L. (2015). The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *English for Specific Purposes*, *39*, 26-44.

Nathan, P. (2013). Academic writing in the business school: The genre of the business case report. *Journal of English for Academic Purposes, 12*(1), 57-68.

Nation, I. S. P. (2001). *Learning vocabulary in another language*: Ernst Klett Sprachen.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.

Nesi, H., & Gardner, S. (2012). Genres across the Disciplines: *Student writing in higher education*. Cambridge: Cambridge University Press.

Nuyts, J. (2001). *E*pistemic modality, language, and conceptualization: *A cognitive-pragmatic perspective* (Vol. 5): John Benjamins Publishing.

Oakes, M.P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

O'Donnell, M. B., Scott, M., Mahlberg, M., & Hoey, M. (2012). Exploring text-initial words, clusters and concgrams in a newspaper corpus. *Corpus Linguistics and Linguistic Theory*, *8*(1), 73–101.

Phillips, M. (1989). *Lexical Structure of Text*. University of Birmingham: English Language Research.

Paltridge, B. (2004). The exegesis as a genre: An ethnographic examination. *Analysing Academic Writing: Contextualized Frameworks*, 84-103.

Paltridge, B. (2007). Beyond the text: A textography of Chinese College English writing. *University of Sydney Papers in TESOL, 2*(2), 149-165.

Paltridge, B. (2008). Textographies and the researching and teaching of writing. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*, *15*(15), 9-24.

Paltridge, B. (2012). *Discourse analysis: An introduction*: Bloomsbury Publishing.

Partington, A. (2004). "Utterly content in each other's company": Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, *9*(1), 131–156.

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: theory and practice in corpus-assisted discourse studies (CADS)* (Vol. 55): John Benjamins Publishing.

Pascual, M., & Unger, L. (2010). Appraisal in the research genres: An analysis of grant proposals by Argentinean researchers. *Revista Signos, 43*(73), 261-280.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication, 191*-225.

Peacock, M. (2012). High-frequency collocations of nouns in research articles across eight disciplines. *Ibérica*, *23*(23), 29-46.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series*, *50* (302): 157–175.

Plonsky, L., & Oswald, F. L. (2014). How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning*, *64*(4), 878–912.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics, 13*, 519-549.

Rayson, P., & Garside, R. (2000). *Comparing corpora using frequency profiling*. Paper presented at the Proceedings of the Workshop on Comparing Corpora, 1-6.

Read, J., Hope, D., & Carroll, J. (2007), "Annotating expressions of appraisal in English", *Proceedings of the Linguistic Annotation Workshop*, ACL 2007, pp. 93-100.

Recski, L. (2005). Interpersonal engagement in academic spoken discourse: A functional account of dissertation de fenses. *English for Specific Purposes*, *24*(1), 5-23.

Rescher, N. (2005). *Epistemic logic: A survey of the logic of knowledge*. Pittsburgh: University of Pittsburgh Press.

Riloff, E., & Wiebe, J. (2003, July). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing* (pp. 105-112). Association for Computational Linguistics.

Rongen Breivega, K., Dahl, T., & Fløttum, K. (2002). Traces of self and others in research articles. A comparative pilot study of English, French and Norwegian research articles in medicine, economics and linguistics. *International Journal of Applied Linguistics*, *12*(2), 218-239.

Römer, U. (2008). Identification impossible? A corpus approach to realisations of evaluative meaning in academic writing. *Functions of Language, 15*(1), 115-130.

Römer, U. (2011). Observations on the phraseology of academic writing: Local patterns - local meanings? *The Phraseological View of Language: A Tribute to John Sinclair*.

Römer, U., & O'Donnell, M. B. (2010, May). *Positional variation of n-grams and phrase-frames in a new corpus of proficient student writing*. Paper presented at the ICAME 31 Conference, Giessen, Germany.

Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes*, *21*(1), 1–17.

Samraj, B. (2005). An exploration of genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes,* 24, 141–156.

Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes, 7*(1), 55-67.

Schelle, C & Field, E. (2015, March 5). *Agricultural studies enrolments on the rise at Australian universities*. Retrieved from http://www.weeklytimesnow.com.au/news/national/agricultural-studies-enrolments-on-the-rise-at-australian-universities/news-story/858cbc11d4322733085ecd89b80e5ac6.

Scott, M. R. (1996). *WordSmith Tools*. Oxford: Oxford University Press.

Scott, M. (1997). PC Analysis of key words—and key key words. *System,* 25, 233–245.

Sharpe, M. E. (1996). Adam Smith: *Right and Wrong* [Review of *Adam Smith in His Time and Ours*]. *Challenge*, *39*(5), 62–64. Retrieved from http://www.jstor.org/stable/ 40721710.

Shaw, P. (2003). Evaluation and promotion across languages. *Journal of English for Academic Purposes*, *2*(4), 343–357.

Shaw, P. (2004). How do we recognise implicit evaluation in academic book reviews. *Academic discourse: new insights into evaluation. Bern: Peter Lang*, 121-140.

Silver, M. (2003). The stance of stance: A critical look at ways stance is expressed and modeled in academic discourse. *Journal of English for Academic Purposes, 2*(4), 359-374.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.

Sinclair, J. (1991). *Corpus, concordance, collocation* (Vol. 1): Oxford University Press Oxford.

Sinclair, J. (1995). *Collins COBUILD English dictionary. Dictionary*. London: HarperCollins.

Sinclair, S. (2003). Computer-assisted reading: reconceiving text analysis. *Literary and Linguistic Computing*, *18*(2), 175-184.

Sinclair, J. (2004). *Trust the Text: Language, Corpus and Discourse.* London: Routledge.

Sinclair, J., & Coulthard, R. (1975). *Towards an Analysis of Discourse: the English Used by Teachers and Pupils*. Oxford: Oxford University Press.

Sinclair, J., Jones, S., & Daley, R. (2004). *English Collocation Studies: The OSTI Report*. London: Continuum.

Smith, A. (1759). (1976): The Theory of Moral Sentiments. *Indianapoli: Liberty Classics*.

Song, Y., & Yan, H. (2013). Empirical study on graduate ESP Curriculm system construction of agricultural colleges and universities. *Anhui Agricultural Sciences*. *41*(20), 8782-8784.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214-225.

Stefanowitsch, A., & Gries, S.T. (2003). Collostructions. Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, *8*(2), 209–243.

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.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*: Blackwell Publishers.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. (2004). *Research genres: Explorations and applications*. Cambridge: Cambridge University Press.

Swales, J. M. (1998). Textography: Toward a contextualization of written academic discourse. *Research on Language and Social Interaction, 31*(1), 109-121.

Swales, J. M. (2002). Integrated and fragmented worlds: EAP materials and corpus linguistics. In J. Flowerdew (Eds), *Academic Discourse* (pp. 150-164). London: Longman.

Swales, J. M., & Feak, C. B. (2004). Academic writing for graduate students: *Essential tasks and skills* (Vol. 1). Ann Arbor, MI: University of Michigan Press.

Teubert, W. (2005). Evaluation and its discontents. *Strategies in Academic Discourse. Amsterdam: John Benjamins*, 185-204.

Thetela, P. (1997). Evaluated entities and parameters of value in academic research articles. *English for Specific Purposes, 16*(2), 101-118.

Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics*, *22*(1), 58-78.

Hunston, S., & Thompson, G. (Eds.). (2000). *Evaluation in text: Authorial stance and the construction of discourse: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.

Thompson, G., & Thetela, P. (1995). The sound of one hand clapping: The management of interaction in written discourse. *Text-Interdisciplinary Journal for the Study of Discourse*, *15*(1), 103-128.

Thompson, G., & Ye, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics, 12*(4), 365-382.

Thompson, P. (2000). Citation practices in PhD theses. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 91-101). Frankfurt: Peter Lang.

Thompson, P. (2014) Exploring Hoey's notion of textual colligation in a corpus of student writing. In Alcaraz-Sintes, A., & Valera-Hernndez, S. (Eds.). *Diachrony and Synchrony in English Corpus Linguistics* (pp.347-372). Peter Lang International Academic Publishers.

Thompson, P. & Tribble, C. (2001). Looking at citations: Using corpora in English for academic purposes. *Language Learning & Technology*, *5*(3), 91–105.

Thurston, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes, 17*(3), 267-280.

Trebits, A. (2009). Conjunctive cohesion in English language EU documents–A corpus-based analysis and its implications. *English for Specific Purposes, 28*(3), 199-210.

Tse, P., & Hyland, K. (2007). Is there an "academic vocabulary"?. *TESOL Quarterly*, *41*(2), 235-253.

Tucker, P. (2003). Evaluation in the art- historical research article. *Journal of English for Academic Purposes, 2*(4), 291-312.

Vande Kopple, W.J., 1985. Some exploratory discourse on metadiscourse. *College Compos. Commun*. 26, 82--93.

Vande Kopple, W. J., & Crismore, A. (1990). Readers' reactions to hedges in ascience textbook. *Linguistics and Education, 2*(4), 303-322.

Van Dijk, T. A. (1997). Discourse as interaction in society. *Discourse as Social Interaction, 2*, 1-37.

Varttala, T. (1999). Remarks on the Communicative Functions of Hedging in Popular Scientific and Specialist Research Articles on Medicine. *English for Specific Purposes*, *18*(2), 177–200.

West, M. (1953). *A general service list of English words, with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. London, New York: Longmans, Green.

Wierzbicka, A. (1994). Semantics and epistemology: The meaning of 'evidentials' in a cross-linguistic perspective. *Language Sciences*, *16*(1), 81-137.

Wharton, S. (2012). Epistemological and interpersonal stance in a data description task: Findings from a discipline-specific learner corpus. *English for Specific Purposes*, *31*(4), 261–270.

Williams, T. (2015, November 5). *Accounting and Economics Careers Will Be In-Demand in 2016*. Reterived from https://www.goodcall.com/news/top-10-hottest-careers-for-accounting-and-economics-degree-holders-02339.

Wray, A. (2002). *Formulaic language and the lexicon* (Vol. 5): Cambridge University Press Cambridge.

Wu, R. Y. F., & Chin, J. S. (2010). An investigation into the English language needs of banking and economics professionals in Taiwan. In *Proceedings of the 12th Academic Forum on English Language Testing in Asia, Taipei* (pp. 73-87).

Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, *27*(1), 103–129. https://doi.org/10.1093/applin/ami045

Xie, J. (2016). Direct or indirect? Critical or uncritical? Evaluation in Chinese English-major MA thesis literature reviews. *Journal of English for Academic Purposes*, *23*, 1-15.

Xu, J. (2009). Log-likelihood ratio calculator [Computer Software]. *Beijing: National Research Center for Foreign Language Education*, Beijing Foreign Studies University.

Yeo, J. Y., & Ting, S. H. (2014). Personal pronouns for student engagement in arts and science lecture introductions. *English for Specific Purposes*, *34*, 26-37.

Zhang, G. (2015). It is suggested that… or it is better to…? Forms and meanings of subject it-extraposition in academic and popular writing. *Journal of English for Academic Purposes*, 20, 1-13.

Zuck, J. G., & Zuck, L. V. (1986). Hedging in news writing. In A.-M. Cornu, J. Van Parjis, M. Delahaye & L. Baten (Eds.), Beads or bracelets? *How do we approach LSP*, Selected papers from the fifth European symposium on LSP (pp. 172-180). Oxford: Oxford University Press.

Appendix A list of journals used in the two corpora

Journal No	Agriculture Journals	No.*	Economics Journals	No.
1	Acta Oecologica	3	Accounting Horizons	2
2	Advances in Agronomy	2	Accounting Review	2
3	Advances in Ecological Research	1	Accounting, Auditing and Accountability Journal	2
4	Advances in Water Resources	4	Accounting, Organizations and Society	3
5	Agricultural and Forest Meteorology	1	American Economic Review	1
6	Agricultural Sciences in China	3	Annual Review of Financial Economics	2
7	Agricultural Systems	8	Applied economics	1
8	Agricultural Water Management	5	Asian Economic Journal	1
9	Agriculture and Human Values	4	Asian Economic Papers	3
10	Agriculture, Ecosystems and Environment	4	Asian Economic Policy Review	1
11	Analytica Chimica Acta	1	Asia-Pacific Journal of Accounting and Economics	1
12	Analytical and Bioanalytical Chemistry	4	Asia-Pacific Journal of Financial Studies	1
13	Animal Feed Science and Technology	1	Australian Accounting Review	2
14	Antonie van Leeuwenhoek, International Journal of General	3	British Journal of Management	2
15	Applied and Environmental Microbiology	4	Business Horizons	3
16	Applied Soil Ecology	2	China and World Economy	1
17	Arab Gulf Journal of Scientific Research	1	China Economic Review	7
18	Archives of Biological Sciences	3	Chinese Journal of Development Economics	1
19	Archives of Dermatological Research	4	Contemporary Accounting Research	2

20	Arid Land Research and Management	4	Corporate Governance	1
21	Basic and Applied Ecology	1	De Economist	2
22	Biological Control	4	Discrete Dynamics in Nature and Society	1
23	Biology and Fertility of Soils	4	Econometric Reviews	2
24	Bioresource Technology	6	Economic Journal	3
25	Biotropica	1	Economic Modelling	2
26	BMC Bioinformatics	4	Economic Theory	1
27	Calcolo	1	Economics Research Letter	1
28	Catena	1	Emerging Markets Review	3
29	Cell and Tissue Research	2	Empirical Economics	3
30	Chemical Research in Chinese Universities	5	European Accounting Review	3
31	Chemoecology	1	European economic review	1
32	Chemosphere	2	European Financial Management	3
33	Chinese Geographical Science	6	European Journal of Health Economics	2
34	Chromosoma	1	European View	1
35	Current genetics	1	Federal Reserve Bank of St. Louis Review	2
36	Ecological Entomology	1	Finance Research Letters	3
37	Ecological Informatics	1	Financial Analysts Journal	2
38	Ecological Monographs	2	Financial Management	2
39	Ecological Research	3	FinanzArchiv	3
40	Ecology Letters	3	Fiscal Studies	2
41	Entomologia Experimentalis et Applicata	3	Forbes	2
42	Environmental Pollution	2	Geneva Papers on Risk and Insurance Theory	2
43	Environmental Science and Policy	4	Geneva Papers on Risk and Insurance: Issues and Practice	1

44	Environmental Science and Pollution Research	2	Geneva Risk and Insurance Review	4
45	Epigenomics	1	Health Policy	2
46	Euphytica	2	Healthcare Financial Management	1
47	European Journal of Agronomy	1	Healthcare financial management: journal of the Healthcare Financial	2
48	European Journal of Plant Pathology	4	IMF Economic Review	3
49	European Journal of Soil Biology	2	Industrial Management & Data Systems	2
50	European Journal of Soil Science	1	Information Sciences	1
51	Fitoterapia	4	International Business Review	1
52	Food Chemistry	2	International Journal of Bank Marketing	3
53	Food Control	1	International Journal of Central Banking	2
54	Food Research International	3	International Journal of Educational Development	1
55	Forest Ecology and Management	7	International Journal of Health Care Finance and Economics	2
56	Forest Pathology	1	International Journal of Manpower	2
57	Functional Ecology	2	International Journal of Production Economics	3
58	Fungal Diversity	1	International Journal of Project Management	2
59	Fungal ecology	3	International Journal of Social Economics	1
60	Genetica	1	Investment Analysts Journal	3
61	Genomics	2	Journal of Accounting and Economics	4
62	Geoderma	4	Journal of Accounting and Public Policy	2
63	Horticulture Environment and Biotechnology	1	Journal of Accounting Research	1
64	Indian Journal of Pathology and Microbiology	3	Journal of Applied Accounting Research	1
65	Insect Biochemistry and Molecular Biology	4	Journal of Banking & Finance	17
66	International Food Research Journal	1	Journal of Business Ethics	4
67	International Insolvency Review	2	Journal of Business Research	1

68	International Journal of Hydrogen Energy	1	Journal of Business; Industrial Marketing	1
69	Irrigation Science	5	Journal of Comparative Economics	2
70	Journal of Agricultural and Food Chemistry	8	Journal of Computer Science & Technology	1
71	Journal of applied Ecology	1	Journal of corporate Finance	1
72	Journal of Applied Phycology	5	Journal of Credit Risk	2
73	Journal of Arid Environments	4	Journal of Development Economics	3
74	Journal of Bacteriology	4	Journal of Econometrics	1
75	Journal of Biomedicine and Biotechnology	1	Journal of Economic Entomology	1
76	Journal of Chemical Ecology	5	Journal of Economic Growth	3
77	Journal of Electrostatics	6	Journal of Economic Literature	1
78	Journal of Environmental Sciences	2	Journal of Economics and Finance	1
79	Journal of Experimental Botany	1	Journal of Financial and Quantitative Analysis	3
80	Journal of Food Engineering	9	Journal of Financial Economics	5
81	Journal of Invertebrate Pathology	2	Journal of Financial Intermediation	3
82	Journal of Medicinal Chemistry	1	Journal of Financial Management of Property and Construction	1
83	Journal of Mountain Science	1	Journal of Financial Markets	2
84	Journal of Plant Nutrition and Soil Science	4	Journal of Financial Services Research	4
85	Journal of Plant Pathology	1	Journal of Financial Stability	4
86	Journal of Plant Research	4	Journal of Futures Markets	5
87	Journal of the Science of Food and Agriculture	6	Journal of Housing Economics	1
88	Landscape Ecology	1	Journal of Industrial Economics	4
89	Microbial Ecology	3	Journal of International Financial Management and Accounting	4
90	Molecular Biology Reports	2	Journal of International Marketing	3
91	Molecular Plant Pathology	2	Journal of Islamic Accounting and Business Research	1

92	Mycologia	1	Journal of Mathematical Economics	1
93	Nature	4	Journal of Monetary Economics	2
94	Nature Biotechnology	4	Journal of Money, Credit and Banking	4
95	Nature Chemistry	2	Journal of Operational Risk	3
96	Nature Communications	2	Journal of Pension Economics and Finance	2
97	Naturwissenschaften	2	Journal of Policy Modeling	1
98	Neotropical Entomology	1	Journal of Portfolio Management	1
99	New Phytologist	4	Journal of Public Economics	2
100	Notulae Botanicae Horti Agrobotanici Cluj-Napoca	1	Journal of Real Estate Finance and Economics	2
101	Oecologia	1	Journal of Risk and Insurance	3
102	Oncogene	1	Journal of Risk and Uncertainty	2
103	Paddy Water Environ	1	Journal of Risk Model Validation	2
104	Pedosphere	2	Journal of Risk Research	2
105	Persoonia	1	Journal of the European Economic Association	1
106	Pest Management Science	2	Journal of World Business	3
107	Pesticide Biochemistry and Physiology	1	Management Accounting Research	3
108	Philosophical Transactions of the Royal Society B-	1	MANAGEMENT INTERNATIONAL REVIEW	2
109	Plant and soil	11	North American Journal of Economics and Finance	3
110	Plant Cell Reports	4	Open Economies Review	2
111	Plant Genome	1	Petroleum Economist	2
112	Plant Physiology	2	Petroleumscience	1
113	Plant Protection Science	3	Public Administration Review	1
114	Plant, Cell and Environment	1	Public Choice	1

115	PLoS Biology	2	Quality & Quantity	1
116	Plos One	1	Quantitative Marketing and Economics	2
117	Polymer Degradation and Stability	7	Review of Accounting Studies	2
118	Precision Agriculture	6	Review of Financial Studies	2
119	Progress in Natural Science	1	Review of Income and Wealth	1
120	Science China Life Sciences	3	Safety Science	1
121	Science in China, Series D: Earth Sciences	3	Scientific Research	6
122	Science of The Total Environment	1	Small Business Economics	4
123	Scientia Agricola	6	The Geneva Papers on Risk and Insurance Theory	1
124	Scientia Horticulturae	3	The Journal of Behavioral Economics	1
125	Soil and Tillage Research	7	The Journal of Finance	2
126	Soil Biology and Biochemistry	4	The Manchester School	1
127	Soil Use and Management	6	World Bank Economic Review	2
128	South African Journal of Botany	1	World Development	3
129	The Science of the Total Environment	1	Total	283
130	Theoretical and Applied Genetics	2		
131	Topics in Catalysis	5		
132	Tree Physiology	4		
133	Virus Research	1		
	Total	372		

* The number of research articles selected from the journal.

Appendix B The stance phrases with multiple functions

Note: * the phrases with multifunction.

1. account for*

(1) take up: to form a particular amount (judgement stance)

i.e., While it is possible that there are effects on the balance of fungal-to-bacterial balance within the decomposer community, these effects are too small to be discernible when grazing ***accounts for*** <10 % of the overall PLFA variation. (ARAC#022:3)

i.e., The first component ***accounts for*** more than 80% of variance in the five series; it is used as the financial liberalization index (FLI) in our paper. (FRAC#014:5)

(2) explain: to be the reason (analysis stance)

i.e., Different phenomena may ***account for*** this fact: (i) Evolutionary rates do not differ because generation times or recombination rates are similar among the interacting species. (ARAC#111:2)

In the second case, our estimates provide us with a gross measure of efficiency since firm- specific (i.e., measured as mean at the national industry level) and macro-economic factors are considered as determinants of inefficiency effects to *account for* country differences. (FRAC#076:1)

2. according to

(1) following or agreeing with (analysis stance)

i.e., The concentrated sample extract was eluted according to the EPA methods and re-concentrated in a Syncore Analyst and Supelco Visidry. (ARAC#39:2)

i.e., Thus *according to* the tenets of RDT a firm must respond to the external environment (deal with contingencies) by managing interorganizational relations in order to acquire and maintain resources (tangible and intangible) crucial to the firm's ability to compete (i.e., survive) within a given market. (FRAC#42:1)

*(*2) as stated or reported by someone (extra-reference stance)

i.e., According to Wolfram et al. (2010), a project's financial benefit is measured by its net present value (NPV), which is determined by discounting all arising cash flows to the start time of the project. (FRAC#50:1)

i.e., According to Magid et al. (1999), microorganisms lose some of their ability to degrade complex substrates during desiccation. They partly regained that activity upon rewetting, but not to the extent maintained by microorganisms in CWC conditions. (ARAC#125:1)

3. close to

(1) very near (analysis stance)

i.e., The initial surface tension values of *B. papyrifera and S. aucuparia* at time zero were very close to those obtained for the control drops of pure water. (ARAC#132:3)

i.e., We also estimated the specification of column (4) using the CCEP estimator, and the results are very ***close to*** each other. (FRAC#065:2)

(2) almost (hedges stance)

i.e., However, ESP increased significantly in BV-amended soils to reach values ***close to*** 15 when compared with V and BVV-amended soils. (ARAC#026:2)

i.e., By contrast, in several developing countries the elasticity estimate is not significantly positive, and R2 is ***close to*** zero. (FRAC#073:5)

4. contribute to*

(1) help to make it successful (intuition stance-attitude stance)

i.e., Indeed, government-owned banks ¨C like any other state-owned enterprise ¨C should address market failures and therefore ***contribute to*** economic development (e.g., by granting loans to socially valuable investment projects that do not receive private funding) (FRAC#046:1)

i.e., A knowledge of drought performance will therefore ***contribute to*** our understanding of the processes that determine tropical forest community composition, such as coexistence in highly diverse communities and gradients of species diversity. (ARAC#026:2)

(2) one of the causes of something (analysis)

i.e., The loss of organic matter ***contributes to*** the loss of soil structure, and makes the soil less resistant and resilient (Shepherd et al., 2001) (ARAC#125:3)

i.e., In an overinvestment situation, FRQ ***contributes to*** decreasing investment excess. We note that all coefficients are positive and significant, indicating that higher FRQ reduces the overinvestment problem (we confirm H1a), that is, it is a mechanism that help firms to decrease their investment and so move towards their optimal level. (FRAC#056:5)