

Biber's Corpus-based Multidimensional Analysis to Investigate Lexico-grammatical Patterns and the Underlying Communicative Functions

Woravit Kitjaroenpaiboon^{*a}, Benjamas Khamsakul^a, Samniang Fahkrajang^a, Sutan Mumdaeng^a, Prissana Fongsarun^a, Wilasinee Ploylermsaeng^a & Witchaya Tiyapongrapbhan^b

^a Faculty of Humanities and Social Sciences, Suan Dusit University, Bangkok, 10300, Thailand

^b Research and Development Institute, Suan Dusit University, Bangkok, 10300, Thailand

Article info

Article history:

Received: 1 July 2022

Revised: 10 August 2022

Accepted: 29 August 2022

Keywords:

Douglas Biber, Multidimensional analysis, Lexico-grammatical Pattern, Corpus Linguistics

Abstract

Douglas Biber is a veteran in corpus linguistics which is a field of linguistics studying natural written or spoken languages to better understand how the language is used and now is attracting numerous linguists' interests. One salient contribution of corpus linguistics for present linguistic studies is that it helps reveal lexico-grammatical patterns (or grammatical collocations) naturally used in the language together with their underlying communicative functions. Biber is said to be a pioneer in this field of study and to have employed a statistical approach (Multidimensional Analysis: MA) to investigate lexico-grammatical collocations in his corpus of written and spoken discourses. This paper thus elaborates to what extent Biber's Corpus-based Multidimensional Analysis has contributed to studies in the field. After reviewing literatures related to the topic, the writers of this paper found that numerous scholars have agreed that Biber's Corpus-based Multidimensional Analysis is relatively good for studying lexico-grammatical patterns and their communicative functions and can further be applied for other studies particularly in the corpus linguistics field.

Introduction

Numerous corpus scholars (e.g. Atkinson, 1999; Baoya, 2015; Biber, 1995, 2002; Biber & Finegan, 1994; Connor & Upton, 2013; Friginal, 2009; Getkham, 2010; Gozdz-Roszkowski, 2011; Kanoksilapatham, 2003; Kitjaroenpaiboon et.al., 2021; Kitjaroenpaiboon & Getkham, 2016a; 2016b) have mutually agreed that multidimensional analysis or MA assists in providing a more comprehensive linguistic description of texts and text varieties since it can help statistically reveal

lexico-grammatical collocations with the underlying communicative functions in a corpus. This paper therefore aims at reviewing previous research studies having exploited Biber's corpus-based multidimensional analytical approach to study lexico-grammatical patterns as well as underlying communicative functions in different corpus.

Corpus linguistics, a field of linguistics studying collected samples of natural written or spoken discourses in order to better understand how the language is used is

* Corresponding Author
e-mail: woravit_kit@dusit.ac.th

now gaining popularity among several linguists (Fuchs & Schoenberg, 2020). This has been confirmed from an increasing number of research studies in the field. One key contribution of corpus linguistics for linguistic research studies is that it helps linguists reveal lexico-grammatical patterns (or grammatical collocations) naturally used in the language together with their underlying communicative functions. Douglas Biber is said to be a pioneer with high reputation in the field. He is known for being well versed at employing a multidimensional analysis (MA), a quantitative statistical approach, to investigate lexico-grammatical collocations with their underlying communicative functions in his corpus of written and spoken discourses.

A multidimensional analysis is a statistically corpus-based quantitative approach often used to analyze linguistic variation among registers or genres (Biber, 1995). Unlike ordinary grammatical analyses with focus particularly on single lexico-grammatical features, the MA takes a multivariate statistical approach (known as Factor Analysis) to study sets of lexico-grammatical features in texts. Methodologically, the MA comprises many phases such as developing a range of lexico-grammatical features for the analysis, computing frequencies of the lexico-grammatical features in a corpus by a concordance program, normalizing them to allow comparison between texts of different sizes, and performing a factor analysis to identify 'dimensions' (Arunmanakul, 2012).

After performing a multidimensional analysis, the analyzed lexico-grammatical features cluster in dimensions. The dimension usually contains positive loadings and negative loadings features. The positive loading features are complimentary to the negative loading features, whereas the higher the frequency of the occurrence of lexico-grammatical features in one set, the lower the regularity of the features appear in the other set and vice versa. Each lexico-grammatical feature has its own loading indicating density of co-occurrence on each dimension. The more the feature's loading value, the more likely the feature is sharing the same function with its co-occurring features (Ravat et.al., 2007).

Subsequently, the dimension is correlative to the shared communicative functions of its composing lexico-grammatical features (Biber, 1995). For example, as observed in Biber's (1995) study of a 48-spoken-and-written-text-corpus, 6 dimensions were identified. Also, past tense verbs, 3rd person pronouns, perfect aspect verbs, public verbs, synthetic negation, and

present participial clauses simultaneously provide functions of 'Narrative versus Non-Narrative Concerns' Notwithstanding, a dimension score is calculated by summing up the frequencies of lexico-grammatical features, each of which has notable loading on a pattern. This average dimension score facilitates a comparison of relations among texts and text varieties with respect to each pattern (Ravat et.al., 2007).

Review of Literatures

A thorough search of the relevant studies revealed that many prior studies applied Biber's corpus based multidimensional analysis to investigate lexico-grammatical features with their underlying communicative functions in different corpus (e.g. Baoya, 2015; Biber, 1995, 2002; 2004; Biber et.al., 2002; 2006; Biber & Finegan, 1994; Connor & Upton, 2013; Friginal, 2009; Getkham, 2010; Gozdz-Roszkowski, 2011; Guinovart, 2000; Kanoksilapatham; 2003; Kitjaroenpaiboon et.al., 2021; Kitjaroenpaiboon & Getkham, 2016a; 2016b Kanoksilapatham, 2003; Louwerse et.al., 2004). Prior studies are reviewed as follows.

In 1995, Douglas Biber (1995) tried employing a statistically multidimensional analysis approach to investigate lexico-grammatical features in his corpus originating from LOB (Lancaster-Oslo-Bergen) corpus and the LLC (London-Lund Corpus) corpus, representing several registers (e.g. radio broadcasts, press reportage, fiction, academic prose, letters, conversations, public speeches, interviews, etc.). In the study, four dimensions with five underlying communicative functions were revealed and shown below.

1) 'Narrative and Non-narrative discourse' which comprises *past tense verbs, 3rd person pronouns, perfect aspect verbs, public verbs, synthetic negation, and present participial clauses*.

2) 'Involved and Informational Production' which consists of *that-deletions, contractions, present tense verbs, 2nd person pronouns, 'do' as pro-verb, analytic negation, demonstrative pronouns, general emphatics, 1st person pronouns, 'it', 'be' as main verb, causative subordination, discourse particles, general hedges, amplifiers, sentence relatives, 'Wh'-questions, possibility modals, non-phrasal coordination, 'Wh'-clauses, final prepositions, nouns, word length, prepositions, type-token ratio, attributive adjectives, and indefinite pronouns*.

3) 'Situation-dependent and Elaborated Reference' which contains *time adverbials, place*

adverbials, adverbs, 'Wh' relative clauses in object positions, pied piping constructions, 'Wh' relative clauses in subject positions, phrasal coordination, and nominalizations.

4) 'Overt Expression of Persuasion' which composes of *infinitives, prediction modals, persuasive verbs, conditional subordination, necessity modals, split auxiliaries, and possibility modals.*

In 2000, Xavier Gomez Guinovart (Guinovart, 2000) applied Biber's corpus-based multidimensional analysis to identify the stylistic variables influencing the nature of spoken and written-to-be-spoken English texts, compiling the corpus from the British National Corpus (BNC). The multidimensional analysis led to the distinction of three dimensions with three underlying communicative functions, consisting of:

1) 'Notional Richness versus Dynamic Deictic Reference' in which *possibility modals, relatives and interrogatives, adjectives modified by adverbs, adjectives/nouns ratio, amplifiers, modal auxiliary, conditional adverbial subordinators, causative adverbial coordinators, infinitive forms, predictive modals, adverbs, necessity modal, type token ratio, place adverbials, and proper nouns* co-occur.

2) 'Explicitness versus Concision' in which *hapaxlegomena, relatives and interrogatives (infinitive forms), coordinating conjunctions, discourse particles, prepositions nouns ratio, negative constructions, adverbs, interjections, indefinite pronouns, adverbs prepositions ratio, personal pronouns, coleman-liau index, word length, attributive adjectives, nouns, prepositions plus nouns, prepositions plus phrases, adjectives, proper nouns, nominal pre-modifiers, determiners, past participles, and genitive markers* inhere.

3) 'Favored versus Disfavored' in which *sentence length in words, sentence length in characters, automated readability index, reflexive pronouns, and type token ratio* consist.

In 2002, Douglas Biber with Susan Conrad, Randi Reppen, Pat Byrd, and Marie Helt (Biber et.al., 2002) together explored the T2K-SWAL corpus (TOEFL 2000 Spoken and Written Academic Language) by using Biber's corpus-based multidimensional analysis to investigate lexico-grammatical features, finding the same five dimensions and five underlying communicative functions as Biber's 1995 and presented as follows.

1) 'Involved versus Informational Production' which comprises *private verbs, 'that' deletion, contractions,*

present tense verbs, second person pronouns, 'do' as pro-verb, analytic negation, demonstrative pronouns, general emphatics, first person pronouns, pronoun 'it', 'be' as main verb, causative subordination, discourse particles, indefinite pronouns, general hedges, amplifiers, sentence relatives, 'Wh' questions, possibility modals, nonphrasal coordination, 'Wh' clauses, final prepositions, nouns, word length, repositions, type/token ratio, and attributive adjectives.

2) 'Narrative versus Non-Narrative Discourse' which composes of *past tense verbs, third person pronouns, perfect aspect verbs, public verbs, synthetic negation, and present participial clauses.*

3) 'Non-impersonal Style' in which *conjuncts, agentless passives, past participial adverbial clauses, 'by' passives, past participial postnominal clauses, and other adverbial subordinators* cooccur.

4) 'Overt Expression of Persuasion' in which *infinitives, prediction modals, suasive verbs, conditional subordination, necessity modals, split auxiliaries, and possibility modals* consist.

5) 'Situation-dependent versus Elaborated Reference' which *time adverbials, place adverbials, adverbs, 'Wh' relative clauses on object positions, pied piping constructions, 'Wh' relative clauses on subject positions, phrasal coordination, and nominalizations* inhere.

In 2003, Similarly though, Bussaba Kanoksilapatham (Kanoksilapatham, 2003) applied Biber's corpus-based multidimensional analysis to investigate research articles from international journals in biochemistry and found seven lexico-grammatical dimensions with seven underlying communicative functions and presented as follows.

1) 'Evaluative Stance' which comprises *extraposed 'it', 'that' clause controlled by adjectives, predicative, adjectives, and 'to' clause controlled by adjectives.*

2) 'Expression of Purpose' which contains *'to' infinitives, whether/if, 'to' clause controlled by verbs, first person pronouns, 'to' clause controlled by adjectives, prepositions, and type / token ratio.*

3) 'Framing Claims' which composes of *demonstratives, quantifiers, and 'that' clause controlled by verbs.*

4) 'Conceptual versus Specific References' in which *word length, attributive adjectives, nouns, numerals, and technical jargon* are found.

5) 'Attribute Knowledge and Current Findings' where *present tense verbs, references, type / token ratio,*

common nouns, past tense verbs, pointers, and prepositions co-occur.

6) 'Concrete Action versus Abstract Notion' in which *passives, coordinating conjunctions, nominalization, past tense verbs, definite articles, prepositions, and modals* inhere.

7) 'Expression of Contradiction' which composes of *concession, pointers, 'not' negation, and adverbs*.

In 2004, Max Louwerse, Phillippe McCarthy, Danielle McNamara, and Arthur Graesse (Louwerse et.al., 2004) attempted to prove whether Biber's corpus-based multidimensional analysis is effective. These four researchers thus similarly applied the same corpus and methods as Biber (1995) had done, revealing six dimensions as well as six underlying communicative functions. All six dimensions are similar to Biber (1995), Therefore, confirming that a Biber's corpus-based multidimensional analysis is reliable, useful and effective for investigating language in the corpus and presented as follows.

1) 'Informational and Declarative' where *positive temporal connectives, polysemy (adjectives), meaningfulness, LSA paragraph to paragraph, familiarity, LSA sentence to sentence, negative temporal connectives, paragraph length, argument overlap, LSA sentence to paragraph, LSA paragraph to text, ratio of causal particles to causal verbs, LSA paragraph to paragraph, type-token ratio for nouns, LSA paragraph to text, imageability, concreteness, LSA sentence to sentence, LSA sentence to sentence, concreteness, negative causal connectives, frequency, (verbs), causal particles, average syllables per word, positive causal connectives, and age of acquisition* are found.

2) 'Speech and Writing' where *frequency, concreteness, imageability, meaningfulness, polysemy, flesch reading ease, ambiguous quantification, pronoun density, higher level constituents per word, abstract nouns, hypernym, polysemy, LSA sentence to sentence, ratio of causal particles to causal verbs, LSA paragraph to paragraph, paragraph to text, vague adverbs, type/token ratio for nouns, concreteness, argument overlap, average paragraph length, age of acquisition, average syllables per word, mean number of modifiers per noun-phrase, stem overlap, and flesch kincaid grade level* co-occur in.

3) 'Topic Consistency and Topic Variation' where *frequency conditionals, frequency negations, causal verbs, positive additive connectives, polysemy, LSA paragraph to paragraph, positive causal connectives,*

LSA sentence to text, LSA paragraph to paragraph, LSA paragraph to text, type-token ratio, and noun density are found.

4) 'Factual and Situational' which comprises *clarification connectives, causal particles, negative causal connectives, noun overlap, ratio of causal particles to causal verbs, vague adjectives, negative additive connectives, positive causal connectives, ambiguous quantification, argument overlap, vague verbs, vague nouns, polysemy, imageability, causal verbs, and mean hypernym of verbs* inhere in.

5) 'Narrative and Non-narrative' which comprises *ambiguous temporal relation, vague nouns, positive connectives, temporal connectives, LSA sentence to text, LSA paragraph to text, and LSA sentence to sentence*.

6) 'Elaborative and Constrained' in which *type-token ratio, negative additive connectives, vague adjectives, vague verbs, positive additive connectives, LSA paragraph to text, LSA paragraph to paragraph, and LSA sentence to text* consist.

In 2006, Douglas Biber, Mark Davies, James Jones, and Nicole Tracy-Ventura (Biber et.al., 2006) investigated lexico-grammatical features in a Spanish language corpus, revealing six dimensions with six underlying communicative functions and are shown below.

1) 'Oral and Literate' discourse' which contains *indicative mood, causal subordinate clauses, time adverbs, first person pronouns, copula SER, demonstrative pronouns, specific single-word conjuncts. first person pro-drop, mental verbs, place adverbs, existential haber, que verb complement clauses, tag questions, present tense, future ira, perfect aspect, communication verbs, third person pronouns, progressive aspect, yes-no questions, que relative clauses, manner adverbs, augmentatives. quantifiers. CU verb complement clauses, pre-modifying demonstratives, conditional subordinate clauses, nu listed, desire verbs, general single-word conjuncts. verbs of facilitation, simple occurrence verbs, singular nouns, postmodifying adjectives, definite articles, prepositions, plural nouns, simple NPs (without determiners, etc.), derived nouns, type token ratio. postnominal past participles, pre-modifying attributive adjectives, long words, other adjectives, and se passives.*

2) 'Spoken 'Irrealis' Discourse' which comprises *subjunctive verbs, que relative clauses (subjunctive), que verb complement clauses (subjunctive), verb+infinitive, conditional verbs, obligation verbs, future tense, infinitives without preceding verb or article, que verb*

complement clauses (indicative), verbs of facilitation, progressive aspect, conditionals in dependent clauses, and que noun complement clauses.

3) 'Narrative Discourse' which *clitics, imperfect tense, possessives, third person pronouns, se (not passive or reflexive), preterit tense, aspectual verbs, se (reflexive), se (emocion), infinitives without preceding verb or article, verb+infinitive, derived notins, and postmodifying adjectives* are in.

4) 'Addressee Focused Interaction' which *third person pro-drop, tu, exclamatives, Ct/questions, simple NPs (without determiners, etc.), yes-no questions, diminutives, que relative clauses (indicative), and other -merUe adverbs* co-occur in.

5) 'Informational Reports of Past Events' in which *proper nouns, preterit tense, long words, prepositions, pre-modifying attributive adjectives, present tense, predicative adjectives, and verb+infinitive* inhere.

6) 'Formal' Written Style' which includes *relative clauses, and other clauses.*

In 2013, Jocelyne Daems, Dirk Speelman, and Tom Ruetten (Daems, Speelman, & Ruetten, 2013) studied language in weblogs and identified with four patterns with four underlying communicative functions. Daems, Speelman, and Ruetten reported that their revealed patterns are quite identical to Biber's (1995, 2004, 2006) and Biber et al.'s (2002) and argued that Biber's patterns proved predominantly universal, not only with the 'entire' spectrum of registers within a language, but also when inspecting a specific register. The revealed dimensions are as follows.

1) 'Narration and Instruction' where *third person pronouns, past tense verbs, possessive pronouns, adverbs, particles, word length, and proper noun* occur in.

2) 'Formal and casual' in which *subordinating prepositions and conjunctions, determiners, past participles, 'Wh'-determiners, and adjectives* consist.

3) 'Diary and Background Story' which is compounded from first person pronouns, personal pronouns, and nouns.

4) 'Reflection and Report' which comprises of *second person pronouns, third person singular, present tense verbs, modals, base form verbs, 'Wh'-adverbs, and 'Wh'-pronouns.*

Again in 2013, Allar Connor and Thomas Upton (Connor & Upton, 2013) investigated lexico-grammatical features in three hundred and sixteen mail letters taken from a hundred and eight organizations. Their results displayed four dimensions and four underlying

communicative functions and are shown as follows.

1) 'Narrative and Non-Narrative Concerns' which composes of *past tense verbs, third person pronouns, perfect aspect verbs, public verbs, synthetic negation, present participle clauses, present tense verbs, attributive adjectives, past participles, whiz deletion, and word length.*

2) 'Involved and Informational Production' in which *private verbs, 'that' deletion, contractions, present tense verbs, second person pronouns, 'do' as pro-verb, analytic negation, demonstrative, pronouns, general emphatics, first person pronouns, pronoun 'it', 'be' as main verbs, causative, subordination, discourse particles, indefinite pronouns, general hedges, amplifiers, sentence relatives, 'Wh' questions, possibility modals, non-phrasal coordination, 'Wh' clause, final prepositions, nouns, word length, prepositions, type token ratio, and attributive adjectives* inhere.

3) 'Overt Expression of Persuasion Reference' in which *'Wh' relative clauses on object position, pied piping constructions, 'Wh' relative clauses on subject position, phrasal coordination, nominalizations, time adverbials, place adverbials, and adverbs* co-occur.

4) 'Explicit and Situation Dependent' where *'to' infinitives, prediction modals, suasive verbs, conditional subordination, necessity modals, and split auxiliaries* are found.

In 2009, Eric Friginal (Friginal, 2009) employed Biber's corpus-based multidimensional analysis to investigate spoken discourse from outsourced call centers. Three dimensions were discovered. All revealed dimensions mark characteristics distinguishing the functional attribute of speakers' discourse. The dimensions are as follows.

1) 'Addressee-Focused, Polite, and Elaborated Information versus Involved and Simplified Narrative' which compounds of *past tense verbs, third person pronouns, non-factual /communication verbs + that-clause, communicative verbs, 'that' deletions, and present tense verbs.*

2) 'Planned, Procedural Talk' which consists of *pronoun 'it', first person pronouns, past tense verbs, 'that' deletion, private verbs, 'Wh' clauses, perfect aspect verbs, verb 'do', second person pronouns, word length, please, nouns, possibility modals, nominalizations, length of turns, thanks, and ma'am/ sir.*

3) 'Managed information flow' in which *word count, length of turns, type / token ratio, second person pronouns, next/then, word length, time adverbial,*

prepositions, please, present tense verbs, nominalizations, and because/so inhere.

In 2010, Kanyarat Getkham (Getkham, 2010) studied co-occurring patterns of lexico-grammatical features in applied linguistic research article, unveiling six dimensions with six underlying communicative functions which are as follows.

1) 'Expression of Purposes' where *suasive verbs, 'that' clause controlled by adjectives, public verbs, 'that' clause controlled by verbs, 'that/which' relatives, predicative adjectives, 'that' clause controlled by nouns, and extraposed 'it'* are found.

2) 'Established Knowledge and Expression of Ownership' where *'to' infinitives, coordination, concessive connectors, whether/if, perfect, references, 'that' clause controlled by verbs, extraposed 'it', passives, and pointers* co-occur.

3) 'Expression of Generality' which composes of *private verbs, hedges, other connectors, public verbs, and cause connectors.*

4) 'Evaluative Stance' which comprises *word length, attributive adjectives, nouns, synthetic negations, and participial modifiers.*

5) 'Conceptual Complexity' which contains *present tense verbs, extraposed 'it', 'that' clause controlled by nouns, first person pronouns, place adverbials, result connectors, and past tense verbs.*

6) 'Framing Claims' in which *passives, coordinating conjunctions, nominalization, whether/if, references, prepositions, type/token ratio, first person pronouns, analytic negation, amplifiers, and pointer* inhere.

In 2015, Zhang Baoya (Baoya, 2015) utilized Biber's (1995) corpus-based multidimensional analysis to investigate his Educational Research Article Corpus (ERC), finding seven dimensions and seven underlying communicative functions as presented as follows.

1) 'Current Information versus Procedural Concerns' which comprises *word length, present tense, nominalization & gerunds, attributive adjectives, phrasal coordination, and past tense.*

2) 'Evaluative Stance versus Past Actions and States' which consists of *'be' as main verbs, predicative adjectives, pronoun 'it', present tense, pragmatic expressions, complements, and past tense.*

3) 'Logical Probability versus Integrated Information' which composes of *dispreferred forms, modals, passives, adverbs, pronoun 'it', to-infinitives, nouns, and prepositional phrases.*

4) 'Commentary' which contains *complements,*

public verbs, suasive verbs, private verbs, and pragmatic expressions.

5) 'Personal Engagement versus Modified Information' which composes of *personal pronouns, to-infinitives, present tense, proverb 'do', relative clauses, nouns, and participial clauses.*

6) 'Unsatisfactory Status Quo Versus Research Conduct' which includes *present tense, existential 'there', adverbial clauses, present tense aspect, negations, adverbs, and past tense.*

7) 'References to Present Research versus Information Relevant to the Past' which contains *demonstratives, prepositional phrases, independent clause coordination, past tense, and type/token ratio.*

In 2016, Woravit Kitjaroenpaiboon and Kanyarat Getkham (Kitjaroenpaiboon & Getkham, 2016a) presented the results of a multidimensional analysis investigating patterns of linguistic features in nursing research articles from the top 5 international leading journals. The study revealed, four lexico-grammatical patterns with four underlying communicative functions and are presented as follows.

1) 'Evaluative Stance Focused' where *predicative adjective, verb to 'be', adverbs, analytic negations, emphatics, 'that' verb complement, prepositional phrases, public verbs, adverbial subordinators, and present tense* are found.

2) 'Established Knowledge Focused' which comprises *present tense, average word length, attributive adjectives, present perfect aspect, demonstrative adjectives, nominalizations, split auxiliary, and phrasal co-ordination.*

3) 'Claim Focused' where *possibility modals, first person pronoun, conjuncts, pronoun 'it', 'that' verb complement, that relative clauses on subject position, sentence relative, 'that' deletion, causative adverbial subordinator, demonstrative pronouns, private verbs, predictive modals, and pipe-piping relative clause* are found.

4) 'Intention Focused' where *'to' infinitive, time adverbial, suasive verbs, third person pronouns, gerunds, and 'Wh' relative clauses on subject positions* co-occur.

Later in 2016, Woravit Kitjaroenpaiboon and Kanyarat Getkham (Kitjaroenpaiboon & Getkham, 2016b) together again applied Biber's corpus-based multidimensional analysis to investigate stylistic patterns and communicative functions in language teaching research articles. The findings were that language teaching research articles contained six stylistic patterns

and six underlying communicative functions as shown below.

1) 'Persuasion Focused' which *present perfect aspect verbs, split auxiliary, type/token ratio, possibility modals, adverbs, 'to' infinitive, and 'that' relative clauses on subject position* are found.

2) 'Evaluative Stance Focused' where *predicative adjectives, verb to 'be', analytic negations, hedges, emphatics, demonstrative pronoun, and predictive modal* together cooccur.

3) 'Claim Focused' which composes of *suasive verbs, pronoun 'it', 'that' verb complements, private verbs, 'Wh' relative clause on subject position, 'that' deletion, and adverbial subordinator*.

4) 'Established Knowledge versus Past Action Focused' in which *present tense verbs, conjuncts, attributive adjectives, phrasal coordination, past participial WHIZ deletion relative, past tense verbs, and agentless passive* inhere.

5) 'Ownership Focused and Result Focused' which comprises *first person pronouns, public verbs, third person pronouns, and causative adverbial subordinator*.

6) 'Modified Information Focused' which compounds of *gerunds, present participial WHIZ deletion relative, average word length, present participial clauses, and demonstrative adjectives*.

Recently as of 2021, Woravit Kitjaroenpaiboon, Benjamas Khamsakul, Samniang Fahkrajang, Sutthasinee Kesprathum, and Prissana Fongsarun (Kitjaroenpaiboon et.al., 2021) analyzed co-occurring patterns of lexico-grammatical features and their underlying communicative functions in applied linguistic research abstracts corpus. Biber's corpus-based multidimensional analysis revealed four dimensions with

four different underlying communicative functions and are presented below.

1) 'Showing Significance and Clearly Specifying' where *present tense verb, verb to 'be', predicative adjectives, 'that' verb complement, suasive verbs, adverbs, prepositional phrases, and nouns* are found.

2) 'Providing Comments' which comprises *past tense verbs, type / token ratio, public verbs, present tense verbs, and nominalizations*.

3) 'Focusing on Actions' where *average word length, nominalizations, past tense verbs, agentless passives, and nouns* co-occur.

4) 'Evaluating Value' which *possibility modals, attributive adjectives, average word length, verb to 'be', and predicative adjectives* inhere in.

As can be seen, several corpus linguists applied the Biber's corpus-based multidimensional analysis to investigate co-occurring patterns of lexico-grammatical features in their corpora. The dimensions were seen to be named differently. The name given to the dimension is based on the researchers' own interpretation of mutual communicative functions of the co-occurring lexico-grammatical features. For example, in Woravit Kitjaroenpaiboon and Kanyarat Getkham's (2016a) dimension I, both interpreted that *predicative adjective, verb to 'be', adverbs, analytic negations, emphatics, 'that' verb complement, prepositional phrases, public verbs, adverbial subordinators, and present tense* cooccur to communicatively function as 'Evaluative Stance Focused' while in the other research of theirs within the same year (Kitjaroenpaiboon & Getkham, 2016b), they interpreted that *present perfect aspect verbs, split auxiliary, type/token ratio, possibility modals, adverbs, 'to' infinitive, and 'that' relative clauses on subject*

Table 1: Recapitulation of the Revealed Dimensions in the 10 Previous Research of Written and Spoken Corpus

Research	Corpus	Dimension
Biber (1995)	Spoken, fiction, magazines, newspapers, and academic language corpus (LOB corpus and LLC corpus)	1. Narrative and Non-narrative discourse 2. Involved and Informational Production 3. Situation-dependent and Elaborated Reference 4. Overt Expression of Persuasion
Guinovart (2000)	Spoken, fiction, magazines, newspapers, and academic language corpus (BNC corpus)	1. Notional Richness and Dynamic Deictic reference 2. Explicitness and Concision 3. Favored and Disfavored (sentential length)
Biber, Conrad, Reppen, Byrd, and Helt (2002)	TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL corpus)	1. Narration and Instruction 2. Formal and casual 3. Diary and Background Story 4. Reflection and Report
Biber (2004)	Korean and Somali language corpus	1. Information-focused versus Interactive discourse 2. Stance vs. Context-focused discourse 3. Narrative-focused discourse

Table 1: (Continue)

Research	Corpus	Dimension
Louwerse, McCarthy, McNamara, and Graesser (2004)	Spoken, fiction, magazines, newspapers, and academic language corpus (LOB corpus)	<ol style="list-style-type: none"> 1. Informational and Declarative 2. Speech and Writing 3. Topic Consistency and Topic Variation 4. Factual and Situational 5. Narrative and Non-narrative 6. Elaborative and Constrained
Biber, Davies, Jones, and Tracy-Ventura (2006)	Spanish Spoken and Written Language corpus	<ol style="list-style-type: none"> 1. Oral and Literate Discourse 2. Spoken 'Irrealis' Discourse 3. Narrative Discourse 4. Addressee Focused Interaction 5. Informational Reports of Past Event 6. Formal Written Style
Friginal's (2009)	Outsourced call center corpus	<ol style="list-style-type: none"> 1. Addressee-Focused, Polite, and Elaborated Information vs. Involved and Simplified Narrative 2. Planned, Procedural Talk 3. Managed Information Flow
Connor and Upton (2013)	Mail letter corpus	<ol style="list-style-type: none"> 1. Narrative and Non-Narrative Concerns 2. Involved and Informational Production 3. Overt Expression of Persuasion 4. Explicit and Situation Dependent Reference
Daems, Speelman, and Ruetten (2013)	eblogs	<ol style="list-style-type: none"> 1. Narration and Instruction 2. Formal and casual 3. Diary and Background Story 4. Reflection and Report
Kanoksilapatham (2003)	Biochemistry research article corpus	<ol style="list-style-type: none"> 1. Conceptual and Specific Reference 2. Concrete Action and Abstract Notion 3. Evaluative Stance 4. Framing Claims 5. Attribute Knowledge and Current Findings 6. Concrete Action versus Abstract Notion 7. Expression of Contradiction
Getkham (2010)	Applied linguistic research article corpus	<ol style="list-style-type: none"> 1. Expression of Purposes 2. Established Knowledge/Expression of Ownership 3. Expression of Generality 4. Evaluative Stance 5. Conceptual Complexity 6. Framing Claims
Baoya (2015)	Educational research article corpus	<ol style="list-style-type: none"> 1. Current Information versus Procedural Concerns 2. Evaluative Stance versus Past Actions and States 3. Logical Probability versus Integrated Information 4. Commentary 5. Personal Engagement versus Modified Information 6. Unsatisfactory Status Quo versus Research Conduct 7. References to Present Research versus Information Relevant to the Past
Kitjaroenpaiboon and Getkham (2016a)	Nursing research article corpus	<ol style="list-style-type: none"> 1. Evaluative Stance Focused 2. Established Knowledge Focused 3. Claim Focused 4. Intention Focused
Kitjaroenpaiboon and Getkham (2016b)	Language teaching research article corpus	<ol style="list-style-type: none"> 1. Persuasion Focused 2. Evaluative Stance Focused 3. Claim Focused 4. Established Knowledge versus Past Action Focused 5. Ownership Focused and Result Focused 6. Modified Information Focused
Kitjaroenpaiboon et.al. (2021)	Applied Linguistics research abstract corpus	<ol style="list-style-type: none"> 1. Showing Significance and Clearly Specifying 2. Providing Comments 3. Focusing on Actions 4. Evaluating Value

position occur together in dimension I to signify 'Narrative-focused discourse'.

All mentioned above are a review of research studies exploiting Biber's corpus-based multidimensional analysis approach to analyze lexico-grammatical patterns as well as their underlying communicative functions. For ease of understanding, all aforementioned research studies together with their revealed lexico-grammatical patterns as well as their underlying communicative functions are chronologically presented in Table I.

Table I recapitulates the aforementioned reviewed research studies and shows that numerous corpus linguists applied Biber's corpus based multidimensional analysis as a key approach to reveal lexico-grammatical patterns and their communicative functions in their different corpus each of which contains different languages, styles, genres, and registers. Each study revealed lexico-grammatical patterns which are both different from and identical to others. This might be due to some lexico-grammatical patterns are found to be universally used in the languages. For instance, adjective and verb to 'be' is usually collocated to function as evaluation (e.g. It is difficult). However, when adverbs are added in between verb to 'be' and an adjective, the structure becomes showing attitude of the writers or speakers (e.g. it is extremely important). Also, when 'to' infinitive is put in a sentence, it implies an intention (e.g. I want to conduct a research study). Simply put, it is usual that a pattern with a communicative function is specifically found in a genre not in others while some is plausible to be found in languages, styles, registers, and genres and provide similar underlying communicative function.

Conclusion

Grounded on Biber's (1995, 2004) assumption that some lexico-grammatical features, co-occurring in one dimension, could have some shared communicative functions, Douglas Biber is the first linguist to apply a multidimensional analysis (MA) to his corpus studies. The MA takes a multivariate approach to study language in a corpus. Lexico-grammatical features are tagged and analyzed by a factor analysis to disclose some co-occurring patterns, which are then called dimensions. The dimension is then interpreted in relation to the shared communicative functions of its composing features. Numerous linguists have applied the Biber's corpus-based multidimensional analysis to analyze co-occurring patterns of lexico-grammatical features in their corpora and confirmed that Biber's corpus-based

multidimensional analysis approach is effective for analyzing lexico-grammatical patterns and their underlying communicative functions in corpora (Baoya, 2015; Biber, 1995, 2002; 2004; Biber et al., 2002; 2006; Biber & Finegan, 1994; Connor & Upton, 2013; Friginal, 2009; Getkham, 2010; Gozdz-Roszkowski, 2011; Guinovart, 2000; Kanoksilapatham, 2003; Kitjaroenpaiboon et al., 2021a, 2021b; Kitjaroenpaiboon & Getkham, 2015, 2016a; 2016b; Louwerse et al., 2004). In summary, the insight given by multidimensional analysis is the individuation of the space of variation for these texts and their role within the ecosystem of the English language. An interested linguist is now able to predict which lexico-grammatical features patterns to expect in texts and what their underlying communicative functions are. Such knowledge, for instance, can empower forensic linguists who are interested in base-rate knowledge of forensic lexico-grammatical patterns. Similarly, for more theoretical purposes, analyses as the one presented constitute another piece of the puzzle in search for a comprehensive descriptive and predictive framework of the registers of English and for the understanding of the nature of the linguistic features, their extra-linguistic predictive factors, and their history and evolution (Nini, 2019).

Acknowledgement

The writers of this article would like to thank Associate Professor Kanyarat Getkham who is a role model of being a researcher, who always provides academic inspiration, guide us whenever there are obstacles to overcome. The original version of this article was read and revised by complying her valuable comments. This article could not have been published without her insight and knowledge. Again, our sincere thanks are extended, Associate Professor Kanyarat Getkham (Ph.D.)

References

- Arunmanakul, W. (2012). *Corpus Linguistics: Principles and Uses*. Bangkok: Chulalongkorn University Press.
- Atkinson, D. (1999). *Scientific discourse in socio-historical context*. London: Laurence Erlbaum.
- Baoya, Z. (2015). *Moves and inter-move linguistic variation in education research articles*. (Unpublished doctoral dissertation). Suranaree University of Technology, Nakhon Ratchasima.
- Biber, D. & Finegan, E. (1994). *Intra-textual variation within medical research articles*. Amsterdam: Rodopi.
- Biber, D. (1995). *Variation across speech and writing*. Cambridge: Cambridge University Press.

- Biber, D. (2002). On the complexity of discourse complexity: A multidimensional analysis. *Discourse Processes*, 15(1), 133-163.
- Biber, D. (2004). *Modal use across registers and time*. Berlin: Mouton de Gruyter.
- Biber, D., Conrad, S., Reppen, R., Byrd, P. & Helt, M. (2002). Speaking and writing in the university: A multi-dimensional comparison. *TESOL Quarterly*, 36(1), 9-48.
- Biber, D., Davies, M., Jones, J. K. & Tracy-Ventura, N. (2006). Spoken and Written Register Variation in Spanish: A multi-dimensional analysis. *Corpora*, 1(1), 7-38.
- Connor, U. & Upton, T. A. (2013). *Language for Specific Purposes Section, The Encyclopedia of Applied Linguistics*. New York: John Wiley and Sons.
- Daems, J., Speelman, D., & Ruetten, T. (2013). Register Analysis in Blogs: Correlation between professional sector and functional dimensions. *Leuven Working Papers in Linguistics*, 2(1), 1-27
- Friginal, E. (2009). *The language of outsource call center: A corpus-based study of cross-cultural interaction*. Amsterdam: John Benjamin.
- Fuchs, I. & Schoenberg, M. (2020). *The Role of Corpus Linguistics in Focus on Grammar*. Retrieved July 22, 2021, from <https://longmanhomeusa.com/blog/tag/corpus-linguistics/>.
- Getkham, K. (2010). *A corpus-based study of applied linguistic research articles: A multidimensional analysis*. Bangkok: National Institute of Development Administration.
- Gozdz-Roszkowski, S. (2011). *Explorations across languages and corpora*. Bern: Peter Lang.
- Guinovart, J. (2000). A multidimensional corpus-based analysis of English spoken and written-to-be-spoken discourse. *Cuadernos de Filología Inglesa*, 9(1), 39-70.
- Kanoksilapatham, B. (2003). *A corpus-based investigation of scientific research articles: Linking move analysis with multidimensional analysis* (Unpublished doctoral dissertation). Georgetown University, Washington, D.C.
- Kitjaroenpaiboon, W. & Getkham, K. (2015). An Analysis of Interactional Meta-discourse Devices in Communication Arts Research Articles. *International Journal of Management and Applied Science*, 1(9): 125-131.
- Kitjaroenpaiboon, W. & Getkham, K. (2016a). Patterns of linguistic feature and their communicative functions in nursing research articles. *International Journal of Management and Applied Science*, 3(3), 98-103.
- Kitjaroenpaiboon, W. & Getkham, K. (2016b). Stylistic Patterns in Language Teaching Research Articles: A Multidimensional Analysis. *PASAA*, 52(1). 169-208.
- Kitjaroenpaiboon, W., Wongwisukul, S., Puksa, T., & Khamsakul, B. (2021a). A Multidisciplinary Corpus-based Comparative Analysis: Lexical Bundles in Language Teaching, Health Sciences, and Business Management Research Articles. *NIDA Journal of Language and Communication*, 26(40). 41-68.
- Kitjaroenpaiboon, W., Khamsakul, B., Fahkrajang, S., Kesprathum, S., & Fongsarun, P. (2021b). Rhetorical Move and Multidimensional Analyses of Applied Linguistics Research Articles Abstracts. *Journal of Language and Culture*, 40(2), 137-165.
- Louwerse, M., McCarthy, M., McNamara, D. & Graesser, A. (2004). *Variation in Language and Cohesion Across Written and Spoken Registers*. Mahwah: Erlbaum.
- Nini, A. (2019). *Multidimensional Analysis: Research Methods and Current Issues*. New York: Bloomsbury Academic.
- Ravat, F., Teste, O., Tournier, R., & Zurlfluh, G. (2007). *Conceptual Model for Multidimensional Analysis of Documents*. Proceeding presented at the 26th International Conference on Conceptual Modeling, Auckland, New Zealand, November 5-9, 2007, DOI:10.1007/978-3-540-75563-0_37