

The *Corpus of Business English Textbooks*: A facilitative resource for digital Business English textbook compilation

Research Article

Yubin Qian, Congyue Wang*, Yu Hu

Department of Foreign Languages, University of Chinese Academy of Sciences, Beijing, 10004

Received 10 December 2025; accepted 08 March 2026

Abstract: In the digital era, textbooks are increasingly required to be both adaptable to the digital environment and sensitive to genre conventions, posing new challenges for material selection. Although several corpora document authentic business communication, few are textbook-based and designed to serve as benchmarks for evaluating textbook content. To address this gap, the present study introduces the *Corpus of Business English Textbooks* (CBET), designed specifically to support corpus-informed compilation of digital Business English textbooks. The CBET is built from 69 widely adopted textbooks in Chinese tertiary education, comprising 7 genres and 35 sub-genres and amounting to 786,183 tokens. Beyond offering a comprehensive record of textbook discourse, the CBET enables systematic comparison between textbook and workplace texts, thereby providing empirical benchmarks for assessing the pedagogical appropriateness of instructional materials. To demonstrate its application, we employed Coh-Metrix to compare textbook letters from CBET with workplace letters, answering two questions: 1. How do business letters in the textbook scenario differ from those in the workplace? 2. What do these differences imply about CBET and its potential contribution to the selection of business letter materials for textbooks? Results reveal that textbook letters differ significantly from workplace letters in syntactic complexity, lexical concreteness, and cohesion, suggesting that they are more informationally compact and intentionally structured to develop learners' command of abstract vocabulary and inference skills. These findings provide empirical insights for text selection in future digital Business English textbooks. Overall, this study highlights the CBET's dual value as both a research resource and a practical tool for guiding the selection and digitalization of Business English teaching materials, thereby advancing corpus-informed pedagogy in the context of digital education.

Keywords: *Business English • digital textbook • textbook corpus • workplace texts • text selection*

© Author(s)

1 Introduction

The expansion of globalization has made English a necessary tool in international business communication, creating a specialized variety known as Business English (Friginal & Udell 2020). The importance and extensive application of Business English worldwide have been well documented (Martins 2017; Nickerson 2005), and this significance has induced an equally significant boost in demand for Business English education and development of teaching materials (Bhatia & Bremner 2012; Ellis & Johnson 1994). More recently, the development and popularization of digital technologies have triggered a surge in new learning resources such as digital textbooks (Wan Sulaiman & Mustafa 2020), which have been proven to be highly beneficial to students' learning (Junco & Clem 2015). Consequently, compiling digital Business English textbooks that reflect commonly occurring genre features has become a crucial undertaking, one that necessitates the selection of instructional materials aligned with genre-specific characteristics. Currently, this selection process largely depends on the compilers' accumulated professional expertise and teaching experience, tacit knowledge that is embedded in a body of established, readily available materials. However, such tacit knowledge remains difficult for

* Corresponding author: Congyue Wang, E-mail: wangcongyue24@mailsucas.ac.cn

Open Access. © 2026 Author(s).

 This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

novices to internalize and apply effectively in text selection or evaluation, thereby affecting the overall accuracy and reliability of the process. Besides, the discursive features of genres in textbook contexts, tailored for pedagogical purposes, often differ from those in authentic workplace settings, which are primarily characterized by financial functionality and institutional intentions (Qian & Sun 2021). Identifying the nature and extent of these variances is essential for selecting materials that are truly appropriate for textbook compilation. In this context, developing generically diverse and representative corpora that serve both as research resources and practical tools in textbook evaluation and compilation has emerged as a key solution (Chang & Kuo 2011; Sheng & Li 2025). However, as the following section will demonstrate, there is currently a notable absence of such specialized resources, indicating a significant gap in the research area (Le Foll 2024). To address this gap, the present study introduces the *Corpus of Business English Textbooks* (CBET), a specialized, genre-based classified corpus compiled from widely used Business English textbooks.

The present study is presented in five sections. Section 2 is a survey of existing Business English corpora as well as the rationale for CBET's compilation. Section 3 specifies data collection, material processing, text renaming, and reorganization process. Section 4 presents a pilot study that employs Coh-Metrix to compare business letters drawn from textbooks with those from workplace contexts. The pilot study answers the following questions: 1. How do business letters in the textbook scenario differ from those in the workplace? 2. What do these differences imply about CBET and its potential contribution to the selection of business letter materials for textbooks? Section 5 discusses the findings of the pilot study, summarizes the overall contributions of the research, and elaborates on its pedagogical implications and limitations.

2 The position of CBET among other Business English corpora

The advent of corpus methods has enabled scholars to draw more generalizable conclusions in studies of language variation and use (Biber, Reppen & Friginal 2010), a development that has been indispensably facilitated by the compilation of general and specialized corpora. Business English is one of the areas that benefited from this surge, leading to the creation of many written and spoken Business English corpora.

Based on the text type they contain, existing Business English corpora generally fall into three categories: written, spoken, or mixed corpora. Mixed corpora contain both written and spoken data and currently represent the most common type. A well-known example is the *Nelson Business English Corpus* (BEC) built in 2000, one of the earliest in the field. It contains approximately 1 million running words, 56% of which are from written materials and 44% from spoken materials. Furthermore, both components are categorized into finer-grained genre classifications based on their function: those used to *conduct* business activities, e.g., annual reports, contracts, emails and letters, and those used to *discuss* business-related topics, e.g., books, newspapers, interviews (Nelson 2006). The materials were collected from both publicly available sources, such as newspapers, magazines, radio, and private channels, with particular attention to the spoken components.

Developed alongside the BEC was the *Published Material Corpus* (PMC), a 590,000-word corpus compiled from 33 widely used Business English textbooks. To ensure textual integrity, Nelson retained only complete materials in the PMC, excluding rubrics, incomplete exercises, and isolated wordlists (Nelson 2000). Because the PMC was designed primarily to contrast teaching materials with authentic Business English through lexical comparison with the BEC, it was not further subdivided beyond broad labels indicating whether each textbook focused on writing, speaking, or integrated skills. This limited level of categorization indicates that, unlike the BEC, the PMC is less suitable for applications requiring genre-specific or function-oriented material selection.

Another major mixed Business English corpus is the *Cambridge International Corpus* (CIC). Rather than being a single corpus, the CIC is essentially a collection of several sub-corpora, including the *Cambridge and Nottingham Spoken Business English Corpus* (CANBEC), the *Cambridge Corpus of*

Business English, and the *Cambridge Corpus of Financial English*, collectively amounting to over 200 million words. Despite its large scale and previous application in textbook compilation (Friginal & Udell 2020), the CIC does not contain a separate textbook sub-corpus that could directly inform material selection.

Some Business English corpora focus singularly on the written form, such as the *Wolverhampton Corpus of Written Business English* (WBE), the more specific *Business Letter Corpus* (BLC), and the *Enron Corpus*. While the WBE does include a small proportion (0.7%) of transcribed speech (Fuertes-Olivera 2007), this spoken component is minimal and does not alter its overall characterization as a written corpus. The WBE is distinctive for its source materials, drawn from 23 business-related websites across genres such as product descriptions, annual reports, research papers, and government reports, totaling more than 10 million running words. By comparison, BLC is more specialized, containing 1 million running words from 11,610 business letters authored by British and American professionals (Zhang & Tong 2025). It was compiled in the form of an online concordancer, but maintenance of the site has now ceased. The *Enron Corpus* comprises approximately 500,000 email messages exchanged among 150 employees, primarily senior management and executives at the Enron Corporation. It was released to the public during a United States federal investigation into corporate fraud in the early 2000s, serving as a key dataset for examining workplace communication in a corporate context.

Although the number of spoken Business English corpora is considerably smaller than that of the written corpora (Friginal & Udell 2020), several notable examples have emerged. The *Vienna-Oxford International Corpus of English* (VOICE), for example, is a 1-million-word corpus of spoken English. It gathers data by transcribing naturally occurring, face-to-face business interactions, and classifies its contents into 10 categories according to the speech event types (Seidlhofer et al. 2021). Another prominent spoken Business English corpus is the CANBEC mentioned above as part of the CIC. The corpus also contains roughly 1 million words, built by collecting spoken data from a variety of companies located mainly in the UK. Different from the event-based categorization adopted by VOICE, the CANBEC's classification is based on conversational settings, which include but are not limited to internal meetings, external meetings, and office talk (McCarthy & Handford 2004).

From the above survey of Business English corpora, it is clear that existing corpora focus almost exclusively on written and spoken texts, with a notable absence of multimodal corpora. Furthermore, their data are mostly drawn from the workplace and used in real-world business communication, aiming to examine language use in practice. However, as noted in previous research, there are distinct differences not only between textbook and every-day genres (Kearsey & Turner 1999), but also between textbook and workplace genres (Bremner 2008). Therefore, while existing Business English corpora supply rich examples for textbook compilation and provide authentic context that could narrow the gap between classroom learning and real-world communication (Bhatia & Bremner 2012), they might also impede learning due to their disproportionate lexical, syntactic, and situational complexity. Furthermore, even corpora that comprise textbook materials, such as the PMC, often lack refined genre classification—an aspect widely regarded as crucial for facilitating English learning in non-Anglophone contexts (Yakhontova 2001). This shortage may undermine the consistency of material selection within individual genres, leading to variations in rhetorical structure across the final products, thereby discouraging students from grasping the conventions of each genre. Consequently, such corpora are less effective in supporting genre-specific material selection. It is also notable that most existing Business English corpora were compiled before or during the 2000s, which, considering the fast evolution of Business English, suggests that they may only partially reflect current language use. In sum, these factors may limit the suitability of existing corpora for selecting teaching materials that are better calibrated to the student's proficiency levels to achieve increased learning efficacy.

Motivated by these considerations, the present study introduces a corpus built from widely used Business English textbooks to facilitate material selection for future digital textbooks. The outcome of this effort is the CBET, the detailed description of which is presented in the next section.

3 Compilation of the CBET

3.1 Data source

The textbooks selected as the source of materials were all downloaded from the Super Star Library, one of the largest and most authoritative datasets of digitalized books in China. The dataset was queried with keywords such as ‘Business English’ to obtain an initial list of books, which was then manually inspected to exclude non-textbooks, textbooks that focus on spoken language, and textbooks that contain only Chinese examples. This process yielded a list of 69 textbooks, which form the source of materials used to compile the CBET. The selected textbooks were published by esteemed Chinese national publishers such as University of International Business and Economics Press and Fudan University Press, and they were widely recognized. For example, *Practical Business Writing in English* is endorsed by the Ministry of Commerce of China as one of the thirteenth Five-Year Plan Key Textbooks. This criterion ensures that the selected textbooks are peer-reviewed and their content is of high quality and representative, making them suitable for being used as prototypical examples of Business English in textbook settings. Further evidence of the textbooks’ suitability comes from their prefaces, which suggest that, though the examples were borrowed from sources such as workplace texts and other books about Business English, they were first adapted to take into account factors such as the students’ proficiency levels and the official teaching guidelines.

Two additional criteria were followed to retain completeness and facilitate genre-specific material selection. First, only complete texts were included in the corpus. Following Nelson’s (2000) principles, we excluded all the rubrics, cloze exercises, and wordlists, thus preserving the integrity of the materials. Moreover, a fine-grained genre classification was adopted to group the materials into sub-corpora. We employed an inductive approach to identify genres present in the materials. For the purpose of recording as much genre information as possible, a two-level genre classification scheme was adopted, assigning each text both a genre and a sub-genre label. This process resulted in a classification of 7 genres and 35 sub-genres, comprising 4,236 texts and amounting to 786,183 tokens (Table 1; for definitions of the sub-genres, see Appendix A).

Table 1. Genres and word counts of the CBET.

Genre	Sub-genres	Word Count
Business Letter	Trade Letter, Social Letter, Recruitment Letter, Complaint-Related Letter	291,587
Business Document	Contract and Agreement, Memorandum, Minute, Itinerary, Summary, Regulation	234,183
Business Report	Informational Report, Interpretive Report	50,615
Business Notice	Internal Notice, External Notice, Press Release, General Notice	25,623
Business Advertisement	Advertisement, Sales Letter, Introduction	56,210
Business Voucher	Bank Guarantee, Credit Inquiry, Certification, Letter of Credit, Debt	12,619
Miscellaneous	Questionnaire, Table and Chart Report, Financial News, Manual, Business Card, Conversation, Message, Curriculum Vitae, Speech, Congratulation Card, Invitation Card	115,346

The first genre is Business Letter. It comprises four sub-genres: Trade Letter, which includes negotiations of business details such as offers, counteroffers, and shipments; Social Letter, which encompasses socialization in the business context, such as invitations, offering condolences, and congratulations; Recruitment Letter, which pertains to employment processes such as interview notifications, applications, and dismissal notifications; and Complaint Related Letter, which covers complaints, including those made by a customer to a company and those made by one company to another, and their settlements. Altogether, texts in this genre amount to 291,587 tokens.

The second genre is Business Document, which includes vital document types such as Contract and Agreement, Memorandum, Minute, Itinerary, Summary, and Regulation. Altogether, texts in this genre amount to 234,183 tokens.

The third genre is Business Report, which is divided into Informational and Interpretive Report sub-genres. Informational Report, such as *compliance reports*, delivers information as-is, leaving interpretation to the recipient; in contrast, Interpretive Report, such as *analyst reports*, offers a narrative that analyzes, explains, and evaluates corporate performance indicators, strategic decisions, and future prospects, aiming to clarify the story (Qian 2026). Altogether, texts in this genre amount to 50,615 tokens.

The fourth genre is Business Notice. Its four sub-genres are Internal Notice, which is intended for the company's own employees; External Notice, which targets current and potential customers; Press Release, which is to be published on papers and media; and General Notice, which does not circulate within or between companies, but is usually issued by and target specific individuals, such as engagement announcements and notices for exhibitions and seminars. Altogether, texts in this genre amount to 25,623 tokens.

The fifth genre is Business Advertisement. It contains three sub-genres: Advertisement, which explicitly promotes merchandises or services; Introduction, which is an overview of a company that indirectly prompts its prestige and its products' craftsmanship; and Sales Letter, which is promotional information taking the form of a letter. Altogether, texts in this genre amount to 56,210 tokens.

The sixth genre is Business Voucher, which includes certifications such as Letter of Guarantee, Credit Inquiry, and IOU. Altogether, texts in this genre amount to 12,619 tokens.

The final category consists of miscellaneous genres, defined as materials that do not function as primary instruments for conducting business transactions or performing core workplace tasks, and therefore cannot be assigned to any of the preceding six genres. Instead, they provide supplementary or contextual information related to business settings. Typical examples include questionnaires used for data collection, financial news reporting market conditions, instructional manuals, informal conversational transcripts, and curricula vitae. Collectively, texts in this category amount to 115,346 tokens.

3.2 Material processing and duplicate removal

Before being admitted into the corpus, the materials underwent two principal procedures to guarantee their eligibility, which are material processing and duplicate removal.

The first step involved OCR, cleaning, and renaming to clean and standardize the materials. After being converted from PDF to TXT format, the outputs were manually modified, removing instructional contents, cloze tests, and rubrics. The materials were then further cleaned to remove typos, punctuation errors, and redundant lines and spaces. Finally, each cleaned example text was saved as a separate TXT file and renamed according to the classification scheme. Each filename is composed of three parts, for example, in the filename 'BL_3_1', 'BL' indicates its genre (Business Letter), '3' denotes sub-genre (Recruitment Letter), and '1' is its sequence number within the sub-genre. The renaming process not only enabled ready identification of the file's genre but also facilitated automatic collection of metadata, such as genre information, token count, and publication details, via a Python script designed to analyze file path, thereby speeding up processing.

The second procedure was duplicate removal, during which repeated texts were identified and removed. There are primarily two reasons that could account for the presence of duplicates. The primary reason is that several source books are different editions of the same book, with usually only minor revisions or additions in later editions, resulting in substantial content overlap. Another reason is textbook compilers' preference for certain canonical examples, which recur across different textbooks. Duplicates occurred pervasively in the corpus materials, and removing them was more beneficial than retaining them, as supported by two reasons: massive repetitions may lead to inaccurate representation of the Business English textbook genres, undermining CBET's support for genre-specific textbook material selection; duplicates may inflate the frequency and keyness of words, syntactic structures, and rhetorical moves, thereby skewing corpus analysis at the lexical, syntactic, and discourse level.

To remove the duplicates, we adopted the Jaccard Coefficient, a metric for similarity evaluation, which is defined as the size of the intersection of two sets A and B divided by the size of their union (Jaccard 1912):

$$\text{Jaccard Coefficient} = (A \cap B) / (A \cup B)$$

In the context of the present study, the Jaccard Coefficient is the proportion of space-delimited strings (a token or a token with a punctuation) that two texts share, ranging from 0 to 1. The greater the Jaccard Coefficient is, the more words the two texts will have in common, and the more similar they will likely be. However, this index has several limitations. An inherent limitation of this index is its insensitivity to word order and frequency. Consequently, two texts with similar word composition but have different word order and frequency may be erroneously recognized as duplicates. Characteristics of the texts also pose challenges. Two texts may be similar or identical in content, but one has letterhead (i.e., sender information such as name, address, and affiliation) and the other does not, causing the Jaccard Coefficient to be diluted and the two texts to be recognized as non-duplicates. Furthermore, no commonly accepted benchmark exists for the Jaccard Coefficient, and the threshold varies across different scenarios. Despite its limitations, the Jaccard Coefficient remains an economical and sufficient tool for duplicate removal in the present study. To apply the Jaccard Coefficient, we first established a benchmark for duplicate identification through an exploratory procedure. Candidate thresholds were tested at 0.1 decrements (0.9 down to 0.1). For each pair of adjacent test values (e.g., 0.9-0.8, 0.8-0.7), text pairs whose Jaccard Coefficient fell within that interval were recorded and manually inspected to confirm duplication status. The results revealed a clear pattern: for intervals bounded by higher thresholds (e.g., 0.9-0.8), duplicate pairs outnumbered non-duplicates; as the thresholds decreased, duplicates diminished and non-duplicates increased. The appropriate benchmark was set at the upper bound of the first interval where nearly all pairs were non-duplicates, which was identified as 0.6 in this study. Accordingly, all text pairs with a Jaccard Coefficient greater than 0.6 were classified as duplicates, and one text from each pair was removed from the corpus. The duplication removal caused a considerable reduction of the corpus's scale from 5,239 to 4,236 files, and from 955,259 to the 786,183 tokens, which now form the CBET.

4 Pilot study

According to Swales (1990), genres are not static templates but dynamic constructs that may exhibit different linguistic features when realized in different contexts of use. Drawing on business letters as an example, a Coh-Metrix-based comparative analysis was conducted to confirm whether the same genre exhibits differences when situated in textbook and workplace scenarios, and to demonstrate CBET's value for textbook material selection. The pilot study addresses the following questions:

1. How do business letters in the textbook scenario differ from those in the workplace?
2. What do these differences imply about CBET and its potential contribution to the selection of business letter materials for textbooks?

4.1 Previous studies on business letters as a genre

Business letters form a genre that uses highly formalized language to realize communication of intents between writer and reader (Jenkins & Hinds 1987). This genre is chosen to demonstrate the application of the CBET in genre-specific textbook material selection for both its prevalence and its importance in Business English teaching. Although many more modern means of communication are now available, business letters, be they printed or electronic, remain an essential genre in business settings (Wu, Zhu & Yang 2022). They persist through nearly every stage in business activity, such as inquiry, quotation, offer, and transport arrangement, and are indispensable for building a successful career (Taylor & Gartside 2012), which makes the appearance of the business letters genre in teaching materials nearly compulsory. Therefore, business letters constitute a genre that is representative, accessible, and possesses critical

pedagogical value, making it appropriate for comparison between textbook and non-textbook Business English genres. Business letters have attracted considerable scholarly attention in the field of linguistics, both for their genre-specific lexico-grammatical features and their rhetorical structures.

On the one hand, research has focused on the lexico-grammatical features of business letters. Vergaro (2004) analyzed two corpora of English and Italian promotional business letters, discovering that the Italian letters prefer to use mood and modality to realize negative politeness, while their English counterparts associate mood and modality with positive politeness. Absence of metadiscourse elements and preference for reader-oriented perspectives in both corpora were also noted. Extending her inquiry from politeness to concession in business letters, Vergaro (2008b; 2008a) analyzed propositional and procedural concessive constructions in business letters, concluding that only procedural constructions were used pragmatically, and that preference for the two types of constructions depends more on text sub-genre than culture influence. Poole (2017) investigated evaluative adjectives in letters to shareholders and found them to be overwhelmingly positive. Huang & Rose (2018) studied metadiscourse markers in Chinese and English CEO letters and discovered that their presence was significantly more frequent in the English letters, showing reliance on affective and credibility appeal.

On the other hand, researchers have also attempted to summarize the general rhetorical structures of business letters. Bhatia (1993) constructed a seven-move model for promotional letters that includes 1) Establish Credentials; 2) Introducing the Offer; 3) Offering Incentives; 4) Enclosing Documents; 5) Soliciting Response; 6) Using Pressure Tactics; and 7) Ending Politely. Bhatia further reported similarity between his model and a model of job application letter, which only changes move 2 to Introducing the Candidate. Based on Bhatia's model, Vergaro (2004) developed a slightly different model for both the English and the Italian promotional letters. She found that, in both the English and the Italian corpora, the moves are consistent except for the absence of the Subject move in the English letters and different sequence of move presentation. Based on 117 authentic business letters of negotiations, Dos Santos (2002) developed a 4-move model, which is composed of 1) Establishing the Negation Chain; 2) Providing Information or answers; 3) Requesting Information/Action/Favors; 4) Ending. Vergaro (2005) discovered that, sharing a similar universal structure that comprises 1) Subject; 2) Opening Salutation; 3) Information; 4) End Politely; and 5) Closing Salutation, the English and Italian FYI letters demonstrate organizational differences, with the Information move usually being more elaborate and preceded by other explanatory moves in the English letters. Townley and Jones (2016) analyzed the generic structures of emails and covering letters during the negotiation of a legal contract, and identified discursive hybridity in these two genres between elements of a specialized technical discourse and those of a less technical, but more interpersonal and polite professional discourse.

It can be concluded from the above summary that current research on business letters focuses on authentic materials instead of textbook materials and on specific instead of overall lexico-grammatical features. Though scholars like Bhatia (1991), Bremner (2008), and Bhatia & Bremner (2012) have discussed teaching material selection and improvement based on genre analysis, they also unanimously treat workplace materials as the preferred source, stressing the importance of incorporating more realistic context for students to acquire the intertextual and interdiscursive competence real workplace requires. However, as pointed out by Widdowson (2010) and Gilmore (2007), workplace materials have their limits, since comprehensiveness of the materials does not necessarily mean that they are calibrated to the students' proficiency levels. In other words, workplace materials might be too difficult for their target users.

Therefore, the pilot study intends to compare textbook business letters and workplace business letters, and by doing so, explore the genre-specific linguistic features manifested by such differences, as well as their implications for the selection of business letter materials for textbooks.

4.2 Corpora for comparison

To contrast business letters in the textbook and workplace context, the present study employed the Coh-Metrix Tool to compare the Easability Component Scores of the CBET's Business Letter sub-corpus and that of the *UIBE Business English Corpus*. The *UIBE Business English Corpus* was originally created for a program led by Professor Wang Lifei at the University of International Business and

Economics (UIBE) to investigate nominalization in business discourse. For this purpose, approximately 200 million words of materials were collected, mostly from workplace settings, forming 14 sub-corpora corresponding to genres such as Economic News, Business Laws, Business Letters, and Business Reports. A two-million-word sample of this corpus was then obtained via stratified random sampling and made available on the BFSU CQPWeb, which the present study utilizes for comparison with the CBET (BFSU CQPweb). Therefore, by the *UIBE Business English Corpus*, the present study refers to the two-million-word sample corpus instead of the original. Similarly to the CBET's Business Letter sub-corpus, the Business Letter sub-corpus of the *UIBE Business English Corpus* includes various sub-genres of business letters and business emails, with the addition of memoranda, which falls within the scope of the Business Document genre in the CBET. Considering the vast number of texts in both corpora, the processing limit of the Coh-Metrix software, and the exclusive online accessibility of the *UIBE Business English Corpus*, a random sample was taken from each corpus for subsequent analysis. Specifically, 30 textbook letters were sampled from the CBET, which amount to 3,659 running words; and 30 workplace letters were sampled from the *UIBE Business English Corpus*, which amount to 3,231 running words. To maintain comparability, the random sampling excludes memoranda in the *UIBE Business English Corpus*. Otherwise, letters from the two corpora are largely comparable.

4.3 Coh-Metrix analysis

The tool Coh-Metrix 3.0 was employed to derive the Easability Component Scores, indices that are indicative of the lexical, syntactic, and discourse level features of a genre. Another index, Coh-Metrix L2 Readability, was also used for comparison, as the CBET is composed of materials collected from Business English textbooks for English learners, and is designed to serve as a reference for material selection of such textbooks. Based on the scores calculated by Coh-Metrix, independent sample t-tests were conducted to compare differences between textbook and workplace letters. Coh-Metrix is a tool primarily used for measuring text readability, an essential reference for genre characterizations (Graesser, McNamara & Kulikowich 2011; McNamara et al. 2014). Unlike traditional text difficulty measurements, Coh-Metrix's indices are proven to be well-aligned not only with surface level information such as word and sentence frequencies but also with semantic and discourse knowledge, which are more indicative of genre categories (Graesser & McNamara 2011). Specifically, based on analysis of the TASA (*Touchstone Applied Science Associates*) corpus, Graesser and his colleagues (2011) summarized eight Easability Component Scores that corresponded to the multilevel text comprehension framework proposed by Graesser & McNamara (2011): genre, manifested through Narrativity, which measures whether a text is likely to be familiar to the reader and therefore narrative; situation model, manifested in Deep Cohesion, which measures the extent to which causal and intentional connectives are used, Connectivity, which measures the degree to which a text uses connectives to express logical relationship, Temporality, which reflects the degree of temporality cues such as tense and aspects usage in text, and Verbal Cohesion, which measures the degree of verb repetition across text; textbase, which manifests in Referential Cohesion, measuring the extent to which words and ideas overlap across text; syntax, which manifests in Syntactic Simplicity, measuring whether sentences in a text tend to be simple or not; words, which manifests in Word Concreteness, measuring the degree to which words in a text are concrete (Graesser, McNamara & Kulikowich 2011; McNamara et al. 2014). The Easability Component Scores also correspond to multiple levels of genre: vocabulary (Word Concreteness), syntax (Syntactic Simplicity), discourse and rhetorical structure (Narrativity, Deep Cohesion, Connectivity, Referential Cohesion, Verbal Cohesion, Temporality), making them ideal references for comparing different genres. The Coh-Metrix L2 Readability index is designed to predict readability of texts for second language readers. It is calculated through three indices, which are content word overlap, sentence syntactic similarity, and word frequency (McNamara et al. 2014). Crossley and colleagues (2011) found that the Coh-Metrix L2 Readability index outperforms traditional readability indices such as Flesch-Kincaid Grade Level and Flesch Reading Ease in classifying text level for second language learners. Supplementing the Easability Component Scores with this index helps to interpret the way genre-specific features influence second language learners' overall perception of these two sub-genres of business letters. After obtaining the scores from Coh-Metrix

3.0, independent samples *t*-tests were conducted with SPSS 26 to investigate on which indices textbook and workplace letters exhibited significant differences. Based on these differences, this study further analyzed differences in linguistics features between the two genres, and indications of these differences on CBET's potential value to future Business English textbook material selection.

4.4 Results

The comparison shows that textbook letters are significantly different from workplace letters across several dimensions. As shown in Table 2, textbook letters show significantly higher scores on Syntactic Simplicity ($p < 0.001$), but lower scores on Word Concreteness ($p < 0.05$), Referential Cohesion ($p < 0.001$), and Verbal Cohesion ($p < 0.001$), indicating distinctive linguistic characteristics of textbook letters that set them apart from their workplace counterparts.

Table 2. Independent sample *t*-test on the Easability Component Scores.

Easability Component Scores	Mean		t	sig.
	Textbook Letters	Workplace Letters		
Narrativity	0.177	-0.046	1.168	0.247
Syntactic Simplicity	-0.046	-0.720	3.388	0.001
Word Concreteness	-0.353	0.353	-2.639	0.011
Referential Cohesion	-0.805	1.805	-10.096	0.001
Deep Cohesion	0.317	0.536	-0.683	0.497
Verb Cohesion	-0.784	1.357	-7.039	0.001
Connectivity	-0.865	-0.399	-1.466	0.148
Temporality	-0.159	-0.478	0.847	0.400
Coh-Matrix L2 Readability	11.393	11.889	-0.283	0.778

The textbook letters are syntactically simpler than their workplace counterparts ($p < 0.001$). Syntactic Simplicity measures whether the sentences in a text tend to be elaborate or not. The more syntactically elaborate a text is, the longer and more complex its sentences will tend to be, and the lower it will score on Syntactic Simplicity (McNamara et al. 2014). Receiving a higher score on this index indicates that textbook business letters potentially adopt simpler syntactic structures, which would make them easier to understand. Six measures contribute significantly to the Syntactic Simplicity index, one of which is 'left embeddedness', or 'words before main verb of main clause' (Graesser, McNamara & Kulikowich 2011: 229; McNamara et al. 2014: 250). For example, example (1) and (2) each have three main clauses, with example (1) having 6, 2, and 2 words respectively before the main verbs, and example (2) having 5, 1, and 28 words before the main verbs. Therefore, example (1) has an average left embeddedness degree of 3.33, while that of example (2) is 14.67. Calculations of the present study may be different from the Coh-Matrix due to the parsing mechanism, but this distinct comparison serves to demonstrate that the textbook letters tend to have fewer words before main clauses, which leaves little room for complicated adverbial clauses (as in example (2)) and poses less challenge to readers' working memory. Therefore, the overall lower score of left embeddedness contributes to the elevated Syntactic Simplicity of textbook letters.

- (1) **In compliance with your request**, I have reviewed the proposal for updating our Brand Image. **I totally** agree with the recommendation. **Your proposal** has been approved by the Board of Directors. (BL_1_1314)
- (2) **As a cash customer you** will be advised of all special sales, and **we** know that you will find our prices and services competitive enough to allow us to continue serving you. **If you have any questions about this decision, or if I may be of any help in any way with regard to your dealings with our company, please** contact me at the above office. (BL00583)

The second Easability Component on which textbook and workplace letters differ is Word Concreteness. Word Concreteness measures the extent to which the words in a text are concrete. It is measured using the *MRC (Medical Research Council) Psycholinguistic Database*, which collects human ratings on several thousand words' properties such as Age of Acquisition, Familiarity, and Concreteness (Coltheart 1981; McNamara et al. 2014). Features that have a significant impact on this index include meaningfulness, concreteness, and imagery (Graesser, McNamara & Kulikowich 2011). In example (3), nouns and noun phrases such as *inquiry*, *flexibility*, *discount rate* and *settlement of accounts* are all abstract concepts, which have no connection to tangible objects, suggesting lower degree of concreteness; whereas in example (4), *wood paper*, *board papers*, *cards* and *parchment* correspond to common items in daily life, indicating a higher degree of word concreteness. A lower score on this index suggests that textbook letters are inclined to use more abstract vocabulary, which may place greater demands on working memory, increasing processing difficulty and formality.

- (3) This letter is in response to your **inquiry** regarding our **flexibility** in the **discount rate** we offer for early **settlement of accounts**. (BL_1_934)
- (4) We manufacture from ground **wood paper** which is often use for mimeographing, **board papers** for special **cards**, **parchment**, **desk pads** and stationery with **envelops** for executives. (BL00678)

Textbook letters also score significantly lower on Referential Cohesion. Referential Cohesion, also called 'coreference', is an index that reflects overlap in words or semantic references between units in the text (McNamara et al. 2014). Thirteen factors bear significant influence on this index, including noun overlap between sentences, content word overlap between and across sentences, and LSA (Latent Semantic Analysis) overlap between and across sentences. For instance, in example (5), there is overlap neither in specific word nor in meaning between the two adjacent sentences. The text moves abruptly from the source of the company's information to request of merchandise information, with no clear connection. However, in example (6), nouns such as *our firm*, *this organization*, and *name of the company* all refer to the company which the employee is about to leave, and the sentences are unified under the same topic of the employee's leaving. Together, they demonstrate a higher level of noun overlaps and LSA overlaps as compared to example (5), boosting the paragraph's referential cohesion. Some have attempted to explain referential cohesion with priming, according to which if the previous sentence mentions a word that also appears in the following sentence, its recognition would be accelerated (Dell, McKoon & Ratcliff 1983). The textbook letters' lower level of referential cohesion therefore indicates fewer connections between sentences, often resulting in abrupt topic shifts, as in example (5), offering fewer points for establishing connections between sentences, which makes comprehension more difficult.

- (5) We avail ourselves to the information about your esteemed company from the Internet. One of our customers needs above 3,000 kgs of walnut meat, and please kindly offer me the price, the lead time, payment terms, and details of packing. (BL_1_1246)
- (6) The fact that you will be leaving **our firm** shortly has been brought to my attention. We are very sorry to lose you inasmuch as your work has always been most satisfactory and we were hoping that you would remain with the (**name of firm**) for many years. I understand that you are leaving for personal reasons that have nothing to do with **this organization**. (BL00801)

The final feature in which textbook and workplace letters demonstrated a significant difference is Verbal Cohesion. Verb Cohesion is the repeated use of verbs across the entire passage. This feature is calculated using Latent Semantic Analysis (LSA) and WordNet. Verbs are usually associated with actions, so their repeated usage provides a clue about the succession of events, fostering the construction of coherent event structures. Factors that were found to have significant influence on this index include polysemy, verb overlap between adjacent sentences, and LSA verb overlap between

adjacent sentences. In example (7), the verbs in adjacent sentences *thank*, *guarantee*, and *inform* have little semantic overlap, whereas in example (8), *submit*, *brought*, *quote* all refer to the initial action of submitting the proposal, facilitating comprehension and increasing readability. The textbook letters' lower score on this index suggests a tendency towards presenting events in a more disjointed manner, which in turn indicates a faster narrative pace and informationally dense style, though at the expense of decreased readability.

- (7) Thank you very much for your order of August 10 for five hundred units of electronic lighters. All these items are in stock, and we can guarantee shipment to San Francisco well before September 30. As requested, we will inform you of the date of dispatch immediately upon completing shipment. (BL_1_1182)
- (8) Our proposal to (proposed service) was **submitted** to you some time ago. Since we haven't heard from you, we wanted to be sure that it was **brought** to your attention. We are enclosing a copy of our original proposal for your review. As you will note, the price **quoted** is good for 30 days following the submission of our bid. (BL00967)

5 Discussion and conclusion

The present study introduces the *Corpus of Business English Textbooks* (CBET), a facilitative resource for material selection in future digital Business English textbooks. By selecting complete texts from representative Business English textbooks and employing a genre-based classification, the CBET complements existing Business English corpora in that it replaces workplace materials with textbook materials, dated materials with more up-to-date materials, and coarse-grained classification with fine-grained classification. Its benefit for genre-specific textbook material selection is also demonstrated with a pilot study that compares business letters in the textbook and workplace contexts.

The pilot study reveals that, while sharing some similarities and about the same level of readability for second language learners, textbook letters are different from workplace letters on multiple aspects. Specifically, of the eight Easability Component Scores, textbook letters score lower on Word Concreteness, Referential Cohesion, and Verbal Cohesion, but higher on Syntactic Simplicity, while no significant differences were identified on Narrativity, Deep Cohesion, Temporality, and Connectivity. These significant indices show some genre-specific linguistic features of textbook letters as opposed to workplace ones: lexically, a high level of word abstractness is usually associated with academic language that is often intricate, formal, and imbued with discipline-specific words (Zhang et al. 2022). A lower score on this index suggests that textbook letters generally employ a more formal vocabulary, carrying more words that are exclusive to the business setting; at the syntactic level, higher syntactic simplicity indicates that textbook letters generally apply sentences with shorter length and less complicated structures (McNamara et al. 2014); at the discourse level, both Referential Cohesion and Verbal Cohesion manifest connections between sentences. Lower scores on these two indices signal that there are less cohesive devices in textbook letters, which may result in the readers having to make inferences to achieve comprehension (Medimorec et al. 2015). The disjuncture between ideas is also evidence of rapid information progression, with less repetition of known knowledge, causing a condense rhetorical structure. Given that textbook and workplace letters show no significant difference on the Coh-Matrix L2 Readability score, the differences observed on Syntactic Simplicity, Word Concreteness, Referential Cohesion, and Verbal Cohesion align with previous research which reports a similar tradeoff among linguistic features reflected by the Easability Component Scores (Medimorec et al. 2015; Pyburn & Pazicni 2014; McNamara et al. 2014). In the present case, the textbook business letters appear to prioritize formality and generic conciseness through greater word abstractness and lower cohesion, achieved at the expense of syntactic simplicity.

The present study shows that, though they share many similarities, textbook and workplace genres are essentially different. The differences may be caused by miscellaneous factors such as official guidelines, teaching objectives, and competence levels of the students. These factors necessitate compilers' adaptations of genre-specific linguistic features in textbook materials at the lexical, syntactic,

semantic, and discourse level. The linguistic features ultimately manifest themselves in statistical scores that can be obtained from the CBET, and can be used as reference points to compare with scores obtained from candidate materials in future digital textbook material selection, thus enabling prediction of their suitability.

The present study fills the current gap of a specialized business textbooks corpus, and its potential in facilitating the selection of materials for creating future digital Business English textbooks is demonstrated by the pilot study. Textbook material selection and development have long relied on “creative intuition” (Tomlinson 2023: 99), characterized by intuition and spontaneity. Though efforts have been undertaken to strengthen coherence and consistency in material selection, with scholars and educators proposing guiding principles and taking measures such as trial use, peer feedback, and reviewing (Tomlinson 2012), there is still need for corpus-informed quantitative measurements that could minimize the uncertainty and inconsistency in this process. The CBET could provide such measurements, which reflect the complex considerations in textbook material selection pertaining to word abstractness, syntactic complexity, cohesion, and coherence. Accordingly, it could support the identification of and improvement on such difficult points in Business English teaching as lexical and semantic patterns (Qian, Boas & Zheng 2025). In addition, the CBET provides basic language resources that can be used to identify the differences and connections of the same genre in pedagogical and workplace settings in the future. For example, it could be used to examine whether business reports for teaching purposes have similar lexical distributions and move structures as those for business communication (Qian 2020). Clarifying this point will contribute to the construction of resources specifically for teaching, which can be used in teaching activities, instructional design, and textbook development.

The present study has several limitations. First, the CBET primarily supports text-based material selection and offers limited insights into contextual factors, such as official guidelines, teaching objectives, and students’ proficiency levels. Furthermore, as the corpus is built from textbooks designed for Chinese learners, its linguistic features may be influenced by both pedagogical considerations and cultural factors. Additionally, the CBET remains a moderate-sized corpus – some sub-genres contain fewer than 10,000 words – and would benefit from further expansion. These limitations suggest several avenues for future work. To begin with, future studies may explore material selection methods that integrate the aforementioned contextual factors through ethnographic study on textbook compilation processes. Second, the CBET may be enriched with materials from textbooks representing other cultural backgrounds, enabling cross-cultural comparison and identification of universally shared principles in textbook material selection. Finally, although peripheral to the present study, adopting more accurate duplicate removal methods may reduce the number of erroneously removed non-duplicate texts and enable fuller use of the materials.

References

- BFSU CQPweb. Business English Corpus. <http://114.251.154.212/cqp/business/index.php?thisQ=corpusMetadata&uT=y>. (9 September, 2025).
- Bhatia, Vijay K. 1991. A genre-based approach to ESP materials. *World Englishes* 10 (2). 153–166.
- Bhatia, Vijay K. 1993. *Analysing genre: Language use in professional settings*. New York: Routledge.
- Bhatia, Vijay K. & Stephen Bremner. 2012. English for business communication. *Language Teaching* 45 (4). 410–445.
- Biber, Douglas, Randi Reppen & Eric Friginal. 2010. Research in corpus linguistics. In Robert B. Kaplan (ed.), *The Oxford handbook of corpus linguistics*, 548–568. 2nd edn. New York: Oxford University Press.
- Bremner, Stephen. 2008. Intertextuality and business communication textbooks: Why students need more textual support. *English for Specific Purposes* 27 (3). 306–321.
- Chang, Ching-Fen & Chih-Hua Kuo. 2011. A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes* 30 (3). 222–234.
- Coltheart, Max. 1981. The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology* 33A (4). 497–505.

- Crossley, Scott A., David B. Allen & Danielle S. McNamara. 2011. Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language* 23 (1). 86–101.
- Dell, Gary S., Gail McKoon & Roger Ratcliff. 1983. The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior* 22 (1). 121–132.
- Ellis, Mark & Christine Johnson. 1994. *Teaching Business English*. Oxford: Oxford University Press.
- Friginal, Eric & Rachele Udell. 2020. Corpus analysis of Business English. In Carol A. Chapelle (ed.), *The concise encyclopedia of applied linguistics*, 303–313. Chichester, UK: John Wiley & Sons Inc.
- Fuertes-Olivera, Pedro A. 2007. A corpus-based view of lexical gender in written Business English. *English for Specific Purposes* 26 (2). 219–234.
- Gilmore, Alex. 2007. Authentic materials and authenticity in foreign language learning. *Language Teaching* 40 (2). 97–118.
- Graesser, Arthur C. & Danielle S. McNamara. 2011. Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science* 3 (2). 371–398.
- Graesser, Arthur C., Danielle S. McNamara & Jonna M. Kulikowich. 2011. Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher* 40 (5). 223–234.
- Huang, Ying & Kate Rose. 2018. You, our shareholders: Metadiscourse in CEO letters from Chinese and Western banks. *Text & Talk* 38 (2). 167–190.
- Jaccard, Paul. 1912. The distribution of the flora in the Alpine zone. *The New Phytologist* 11(2). 37–50.
- Jenkins, Susan & John Hinds. 1987. Business letter writing: English, French, and Japanese. *TESOL Quarterly* 21 (2). 327–349.
- Junco, Reynol & Candrianna Clem. 2015. Predicting course outcomes with digital textbook usage data. *The Internet and Higher Education* 27. 54–63.
- Kearsey, John & Sheila Turner. 1999. Evaluating textbooks: The role of genre analysis. *Research in Science & Technological Education* 17 (1). 35–43.
- Le Foll, Elen. 2024. *Textbook English: A multi-dimensional approach*. Amsterdam: John Benjamins.
- Martins, Hélder. 2017. Perspectives on Business English as a lingua franca in business communication. *Teacher Education and Curriculum Studies* 2(5). 61–67.
- McCarthy, Michael & Michael Handford. 2004. Invisible to us: A preliminary corpus-based study of spoken Business English. In Ulla Connor & Thomas A. Upton (eds.), *Discourse in the professions: Perspectives from corpus linguistics*, 167–201. Amsterdam: John Benjamins.
- McNamara, Danielle S., Arthur C. Graesser, Philip M. McCarthy & Zhiqiang Cai. 2014. *Automated evaluation of text and discourse with Coh-Metrix*. New York: Cambridge University Press.
- Medimorec, Srdan, Philip I. Pavlik, Andrew Olney, Arthur C. Graesser & Evan F. Risko. 2015. The language of instruction: Compensating for challenge in lectures. *Journal of Educational Psychology* 107 (4). 971–990.
- Nelson, Mike. 2000. *Corpus-based study of the lexis of Business English and Business English teaching materials*. Manchester: University of Manchester dissertation.
- Nelson, Mike. 2006. Semantic associations in Business English: A corpus-based analysis. *English for Specific Purposes* 25 (2). 217–234.
- Nickerson, Catherine. 2005. English as a lingua franca in international business contexts. *English for Specific Purposes* 24 (4). 367–380.
- Poole, Robert. 2017. “New opportunities” and “Strong performance”: Evaluative adjectives in letters to shareholders and potential for pedagogically-downsized specialized corpora. *English for Specific Purposes* 47. 40–51.
- Pyburn, Daniel T. & Samuel Pazicni. 2014. Applying the multilevel framework of discourse comprehension to evaluate the text characteristics of general chemistry textbooks. *Journal of Chemical Education* 91 (6). 778–783.
- Qian, Yubin. 2020. A critical genre analysis of MD&A discourse in corporate annual reports. *Discourse & Communication* 14 (4). 424–437.
- Qian, Yubin. 2026. Corporate reports and genre analysis. In Hilary Nesi & Petar Milin (eds.), *International encyclopedia of language and linguistics*. 3rd edn. Amsterdam: Elsevier.
- Qian, Yubin, Hans C. Boas & Qun Zheng. 2025. Frame-based semantic patterns in business discourse: A case study. *English for Specific Purposes* 80. 126–139.

- Qian, Yubin & Ya Sun. 2021. The correlation between annual reports' narratives and business performance: A retrospective analysis. *Sage Open* 11 (3). 1–14.
- Santos, V. B. M. Pinto dos. 2002. Genre analysis of business letters of negotiation. *English for Specific Purposes* 21 (2). 167–199.
- Seidlhofer, Barbara, Marie-Luise Pitzl, Daniel Schopper, Angelika Breiteneder, Hans-Christian Breuer, Nora Dorn, Theresa Klimpfinger, et al. 2021. The Vienna-Oxford International Corpus of English. <https://voice.acdh.oew.ac.at/>. (18 May, 2025).
- Sheng, Dandan & Xin Li. 2025. A corpus-based assessment of vocabulary in interpreting textbooks. *Humanities and Social Sciences Communications* 12 (1). 1–9.
- Swales, John Malcolm. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Taylor, Shirley & L. Gartside. 2012. *Model business letters, e-mails and other business documents*. 7th edn. London: Prentice Hall.
- Tomlinson, Brian. 2012. Materials development for language learning and teaching. *Language Teaching* 45 (2). 143–179.
- Tomlinson, Brian (ed.). 2023. *Developing materials for language teaching*. 3rd edn. London: Bloomsbury Academic.
- Townley, Anthony & Alan Jones. 2016. The role of emails and covering letters in negotiating a legal contract: A case study from Turkey. *English for Specific Purposes* 44. 68–81.
- Vergaro, Carla. 2004. Discourse strategies of Italian and English sales promotion letters. *English for Specific Purposes* 23 (2). 181–207.
- Vergaro, Carla. 2005. “Dear Sirs, I hope you will find this information useful”: Discourse strategies in Italian and English “For Your Information” (FYI) letters. *Discourse Studies* 7(1). 109–135.
- Vergaro, Carla. 2008a. Concessive constructions in English business letter discourse. *Text & Talk* 28 (1). 97–118.
- Vergaro, Carla. 2008b. On the pragmatics of concessive constructions in Italian and English business letter discourse. *Multilingua* 27(3). 255–283.
- Wan Sulaiman, Wan Noor Azhar & Siti Ezaleila Mustafa. 2020. Usability elements in digital textbook development: A systematic review. *Publishing Research Quarterly* 36 (1). 74–101.
- Widdowson, H. G. 2010. *Defining issues in English language teaching*. Oxford: Oxford University Press.
- Wu, Huizhen, Huimin Zhu & Xiaohu Yang. 2022. The effects of providing peer feedback on learners' genre awareness in English as a foreign language business letter writing. *Frontiers in Psychology* 13. 1–16.
- Yakhontova, Tatyana. 2001. Textbooks, contexts, and learners. *English for Specific Purposes* 20. 397–415.
- Zhang, Haomin, Yuting Han, Xing Zhang & Liuran Cui. 2022. Frequency, dispersion and abstractness in the lexical sophistication analysis of a learner-based word bank: Dimensionality reduction and identification. *Journal of Quantitative Linguistics* 29 (2). 195–211.
- Zhang, Lei & Fuzhi Tong. 2025. ‘Could you send us your latest catalogue?’: A local grammar of requesting in English business letters. *English for Specific Purposes* 79. 43–55.

Appendix A: Composition and word counts of the CBET

Genre	Sub-genre	Definition*	Running Words
Business Letter	Trade Letter	Letters used for exchanging and negotiating details about trade, such as offer, counter offer, status inquiry, etc.	196,408
	Social Letter	Letters used for socialization within the business context, such as invitation letter, letter of condolences, etc.	31,147
	Recruitment Letter	Letters pertaining to processes in employment, such as application, offer, resignation, etc.	29,863
	Complaint Related Letter	Letters regarding making complaint and its settlement, such as complaint, claim, letter of apology, etc.	34,169
Business Document	Contracts and Agreement	"A legal binding agreement".	164,198
	Minute	"The official record of the proceedings of a meeting".	18,852
	Memorandum	"A usually brief communication written for interoffice circulation".	21,016
	Itinerary	"The route of a journey or tour or the proposed outline of one".	13,800
	Summary	Main ideas of reports.	6,128
	Regulation	Written rules of a company.	10,189
	Informational Report	Business reports that are written solely to convey critical information, without making judgement or recommendations.	8,241
Business Report	Interpretive Report	Business reports that provide not only facts but also conclusions or recommendations.	42,374
	Internal Notice	Notices released by and circulated within a company.	6,502
Business Notice	External Notice	Notices released by a company to the public or other companies regarding its operations.	6,767
	Press Release	"An official statement that gives information to newspapers, magazines, television news programs, and radio stations".	8,264
Business Advertisement	General Notice	Notices released by individuals, often on newspapers.	4,090
	Advertisement	"A paid notices that is published or broadcasted".	31,299
	Introduction	Overviews of a company or product, often with the purpose of advertising.	12,380
Business Voucher	Sales Letter	Letters sent by sellers to customers with the purpose of promoting products or services.	12,531
	Bank Guarantee	"An undertaking given by a bank to settle a debt should the debtor fail to do so".	1,019
	Credit Inquiry	A request for information regarding a company's credit.	661
	Certification	Proofs to qualifications, degree, etc.	4,845
	Letter of Credit	"A letter from one banker to another authorizing the payment of a specified sum to the person named in the letter on certain specified conditions".	4,998
	Debt	Debt covers vouchers such as IOU and receipt.	1,096
	Questionnaire	"A set of questions for obtaining statistically useful or personal information from individuals".	8,890
Miscellaneous	Table and Chart Report	Descriptions of the contents in table or chart.	6,225
	Financial News	News related to finance.	35,188
	Manual	"A book that is conveniently handled", distributed by companies to inform customers about their products and services.	8,180
	Business Card	"A small card bearing information (such as name and address) about a business or business representative".	356
	Conversation	Simulated dialogue under certain business setting.	21,632
	Message	Information to colleagues in the form of note.	5,041
	Curriculum Vitae	"A short account of one's career and qualifications prepared typically by an applicant for a position".	17,885
	Speech	"A usually public discourse", addressed at certain events such as welcome party, retirement party, etc.	7,136
	Congratulation Card	Congratulations in the form of card.	1,317
	Invitation Card	Invitations, its acceptances, and refusals issued in the form of card.	3,496

* The sources of the quoted definitions include the *Merriam-Webster Dictionary* (<https://www.merriam-webster.com/>) and *Oxford dictionary of finance and banking*. The unquoted definitions represent cases where a definition of the genre or sub-genre cannot be found in either source. In these instances, we provide our own summary based on reading of the corpus materials.