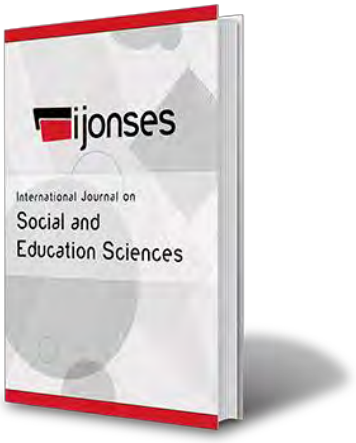

Design and Implementation of the Development of a Corpus-based Loose-leaf Textbook for Higher Vocational English Learners: Using the *Cross-border E-Commerce Operations English* as an Example

Jing Liu
Chengdu Polytechnic, China

www.ijonses.net

To cite this article:

Liu, J. (2023). Design and implementation of the development of a corpus-based loose-leaf textbook for higher vocational English learners: Using the cross-border e-commerce operations English as an example. *International Journal on Social and Education Sciences (IJonSES), 5*(1), 132-152. https://doi.org/10.46328/ijonses.535

International Journal on Social and Education Sciences
2023, Vol. 5, No. 1, 132-152 https://doi.org/10.46328/ijonses.535

Design and Implementation of the Development of a Corpus-based Loose-leaf Textbook for Higher Vocational English Learners: Using the *Cross-border E-Commerce Operations English* as an Example

Jing Liu

Article Info

Article History

Received:
12 May 2022
Accepted:
28 November 2022

Keywords

Corpus-based
Loose-leaf textbook
Specialized textbook corpus
Online learning platform

Abstract

Current higher vocational English textbooks have limitations in providing timely work-related information, representing the nature of vocational education, preserving authenticity of text materials, and using technology to enhance textbook interactivity. To address the limitations of current higher vocational English textbooks, this study adopts corpus-based approaches to developing a loose-leaf textbook for higher vocational English, a new form of textbook that delivers timely, authentic, job-oriented content enhanced by digitalization. Upon an on-the-job investigation, this study builds CECLOTED, a corpus congruent with the typical job tasks for cross-border E-commerce operations. Based on corpus data analysis upon metrics of frequency, keyness, dispersion, collocation and colligation, this study strikes a balance between authentic text feature and linguistic complexity appropriate for higher vocational English learners. This study also develops an online learning platform coherent with the printed textbook with B/S architecture so as to present information in a ubiquitous, dynamic and interactive manner. Compared with the conventional textbook, the loose-leaf textbook which comprises learning texts, vocabulary lists, learning tasks and an online learning platform is more conducive to enhancing language proficiency and job-related skills.

Introduction

The world today is undergoing major changes unseen in a century. New scientific and technological revolution and industrial transformation are driving the rapid development of new technologies, new industries and new business models, giving birth to new occupations, and in turn spurring changes in learning contents and thus in textbooks. In December 2019, the Ministry of Education of the People's Republic of China took the lead in issuing the *Administrative Measures for Vocational School Textbooks* (Ministry of Education of the People's Republic of China, 2019a), advocating developing new forms of textbook such as loose-leaf textbook or work manual-style textbook, and the focus of this study is on the former type.

The loose-leaf textbook is deemed as a new form both for its outer and inner characteristics. Externally, the constituent pages are not glued in such textbook, but rather are hole-punched and held together, most usually, by

a ring binder so that pages can be removed, added or put back undamaged. While the outer character traits of the loose-leaf textbook can be explicitly noted, the study on the essence of it is still in the exploratory stage. Quite a few scholars (Cai et al., 2021; Huang et al., 2021; Li, 2020; Wang et al., 2021; Wu et al., 2022) have tried to explicate this new form of textbook, and they proposed, among other qualities, that typical job tasks for certain positions in enterprises should serve as the backbone of a loose-leaf textbook and the entire textbook should be oriented toward career skills and emphasis be put on reflecting the complete work process of a real job, and hence granting distinguishing features to set loose-leaf textbooks from conventional discipline-based ones.

Status Quo and Limitations

To answer the call of national textbook requirement, reform has been undertaken on textbooks for higher vocational English learners. Despite ongoing efforts (Liu & Zhou, 2022; Peng, 2020; Wang & Li, 2018), there still exist limitations:

- The first on the list is a lack of timely job-related information. As the frequency of industrial renewal accelerates and the cycle of technological iteration shortens, the work situation of actual jobs in enterprises may change all the time, while the traditional long-cycle, single-form, paper-based textbooks cannot respond in time, resulting in the lagging behind of learning content and thus making it difficult to meet the needs of cultivating innovative talents.
- Second, the vocational nature is underrepresented. Despite benevolent attempts in terms of content reconstruction and layout design, existing learning materials for higher vocational English learners fail to distinguish themselves from their academically-oriented counterparts in that what they offer is more about disciplinary knowledge than practical communicative skills necessary for job performance or business needs, and hence leading to a disconnect between education and industry.
- Third, quantitative analysis of language data is insufficient to ensure the pedagogical feasibility of authentic textbook materials. Among the existing textbooks for higher vocational English learners, barely any was developed on the premise of a quantitative examination of authentic language data and the selecting and sequencing of language items depend to a great extent on the materials developers' intuitions about language use. Moreover, as Carter (1998, p.52) noted, it is not uncommon to see mediation on authentic data in the textbook "to achieve clarity, tidiness and organization for purposes of learning", practice as such runs the risk of being subjective and hence compromises the reality of target language use.
- Last but not least, textbook interactivity has not been given due weight to motivate learners. The way how existing higher vocational English textbooks present information is mostly static, rigid and unilateral, failing to meet individual learning needs and hence is not in a position to adapt to the paradigm shift from a teacher-centered learning environment to a student-centered one.

Therefore, the focus of this study was put on the practical application of corpus linguistic techniques to address the limitations in current textbooks for higher vocational English learners. Based on the outer and inner characters of the loose-leaf textbook, this study explored the path to developing a corpus-based textbook of such type for higher vocational English learners, by following which the *Cross-border E-Commerce Operations English* has

been developed so that the learning materials could respond to the dynamic changes at work and reflect authentic language use, the organization of learning modules could reflect the complete work process, and the compilation of learning texts could be in line with learners' cognitive process. Furthermore, an online learning platform has been developed by merging corpus-based and web-based technologies to generate new possibilities for user interactivity. In sum, this study aimed to provide new ideas and solutions for the development of loose-leaf textbooks for vocational education.

Material and Methods

It's nothing new to use corpus-based approaches to developing foreign language materials. However, a literature review on all hitherto existing corpus-based English teaching and learning resources yielded the conclusion that researchers at home and abroad have been devoting much attention to incorporating corpus technologies into developing dictionaries and grammar books, such examples include *Collins COBUILD English Dictionary for Advanced Learners*, *MacMillan Dictionary for Advanced Learners of English*, *Cobuild English Grammar*, *Cambridge Grammar of English*, *Collins Natural Grammar,* etc. Meanwhile, researches on corpus-based textbook development were also productive, *Touchstone*, *Collins Cobuild English Course*, *Innovations*, *Headway*, *Cutting Edge* and *New Era Applied College English* are all representative cases, yet they are intended for English for General Purpose rather than English for Specific Purpose. Recently, Xu (2022) has used corpus-based approaches to extracting the vocabulary for vocational purposes. However, so far, no existing vocational English textbooks, let alone loose-leaf textbooks in this regard, have benefited from corpus methods, which fact brings novelty to our research. Hu and Li (2016) built the Contemporary English Corpus for Textbooks and analyzed its use in the compilation of English textbooks. Li (2021) developed a platform to promote the application of corpus linguistic techniques in language teaching and linguistic studies.

According to the findings of their research, the advantages of using corpus linguistic methods to develop higher vocational English textbooks can be summarized into four aspects:

(1) corpora, if designed as monitor ones, can track the diachronic change of language data linked to the technological revolution and industrial upgrading;
(2) corpora data can be sequenced, grouped or categorized against certain criteria, allowing the feasibility of modularizing textbook materials based on work process;
(3) quantitative analysis of corpora data are indicative of language learnability and hence can help select text data appropriate for pedagogical purposes;
(4) corpus linguistic methods can be incorporated with web-based technology to allow digital enhancement of printed textbooks.

In a word, corpus-based approaches can be a viable solution to the limitation of existing textbooks for higher vocational English.

Three-stage Design

The three-stage design scheme of a corpus-based loose-leaf textbook for higher vocational English learners is

shown in Figure 1. The three stages are the investigation of actual job tasks, the construction of specialized textbook corpus, and the transformation of corpus data into the printed textbook and the digital one, a detailed explanation of each stage is listed as follows:

(1) As the starting point, investigation into typical job tasks specified for a certain job position in enterprises needs to be conducted through a myriad of methods, including interviews, observation, questionnaires and document analysis. Then, the work-related information entailed in typical job tasks should be pedagogically processed to generate learning scenarios and learning tasks for instructional use (Cai et al., 2021).

(2) At the construction stage, material developers should construct a textbook corpus comprising language data about the learning tasks developed from the previous stage, preferably in cooperation with IT specialists.

(3) The transformation stage is pivotal in the development process, during which corpus data are calculated, analyzed, compared, and filtered with the aid of corpus-based techniques to flesh out the notion of the loose-leaf textbook in various forms including learning texts, vocabulary lists, learning tasks and an online learning platform.

In actual practice, those three stages are not entirely congruent with this linear sequence and the recursive process is not uncommon to ensure that the work process knowledge and language skill content in the textbook reach equilibrium.

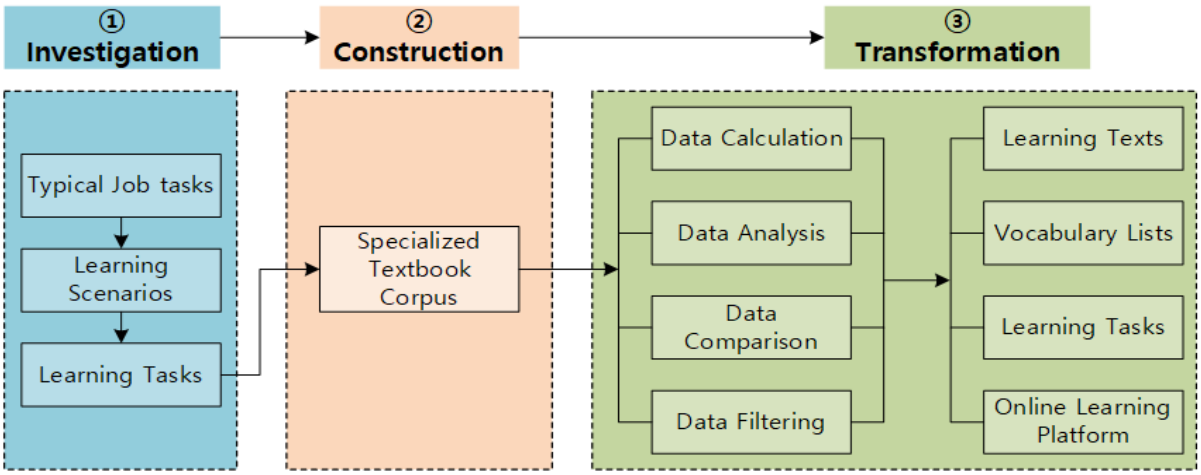

Figure 1. Three Stages of Developing a Corpus-based Loose-leaf Textbook for Higher Vocational English Learners

Implementation of Three-stage Design

The Research Center for Application of Big Data on Higher Vocational English Education affiliated with Chengdu Polytechnic, where the author works, developed a corpus-based loose-leaf textbook titled *Cross-border E-commerce Operations English* by following the above-mentioned procedure, and the concrete steps will be detailed in the following section.

Investigation

The loose-leaf textbook characterizes typical job tasks and work process knowledge as the main content and foregrounds vocational education through doing work-related tasks (Cai et al., 2021). Therefore, we went into cross-border e-commerce companies that have established school-enterprise cooperation with our college to investigate the know-how of cross-border e-commerce operations which constitutes a potential job common to the cross-border e-commerce major groups in our college. Upon investigation, we clarified the entire work process, typical job tasks and vocational competencies as prerequisite to getting the job done. And then, we grouped correlated vocational competencies into modules which served as the indicators for designing learning tasks contextualized in real work situations. Figure 2 shows the typical job tasks for cross-border e-commerce operations.

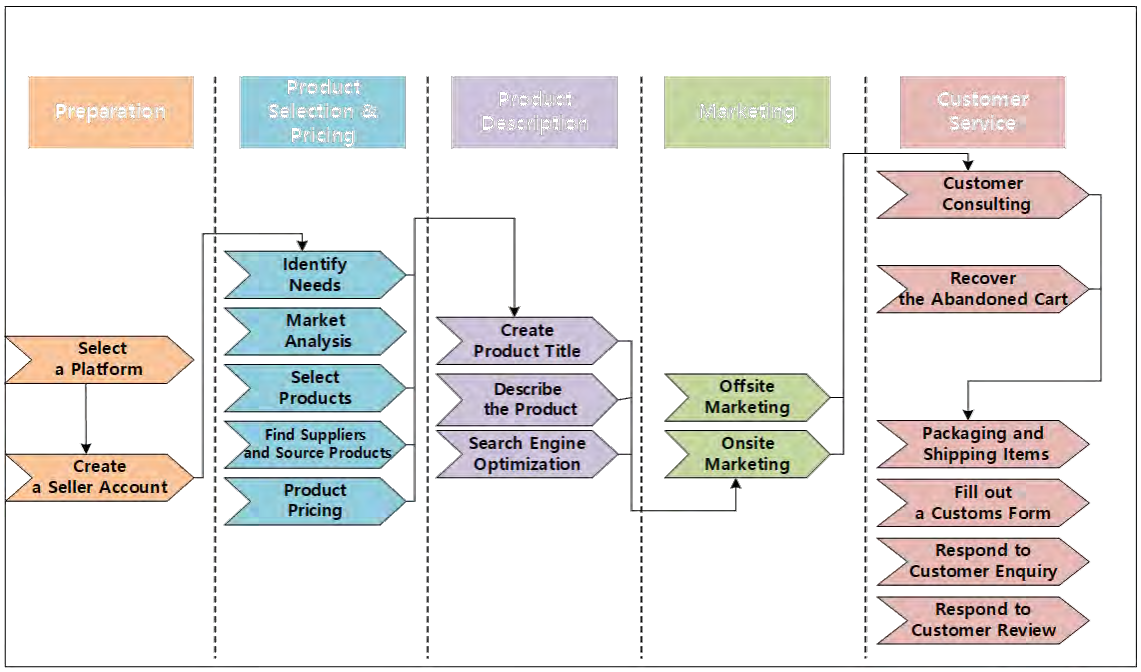

Figure 2. Typical Job Tasks for Cross-Border E-commerce Operations

Based on the analysis of typical job tasks pertaining to cross-border e-commerce operations, seven learning tasks were designed, each of which comprised several subtasks. As shown in Figure 3, those seven learning tasks are placed on the vertical axis and each of them forms a discrete module parallel to all the rest, while what's horizontally located within each module are subtasks arranged in a serial order indicative of the usual sequence to get a job done. Echoing the systematization of the work process, loose-leaf textbooks are modularized to answer the needs of enterprises as modularity in textbook design gives rise to functionally partitioned learning units that can be independently created, permutated or combined with other units and thus provides flexibility in knowledge representation and skills training. In addition, it should be noted that we maintained collaboration with cross-border e-commerce companies for constant observation on the changes in the industry. Should the changes become scalable to such a degree that the findings of previous analyses on typical job tasks need modification or replacement, we will start a new round of work-related investigation to ensure that textbook content precisely matches actual job requests.

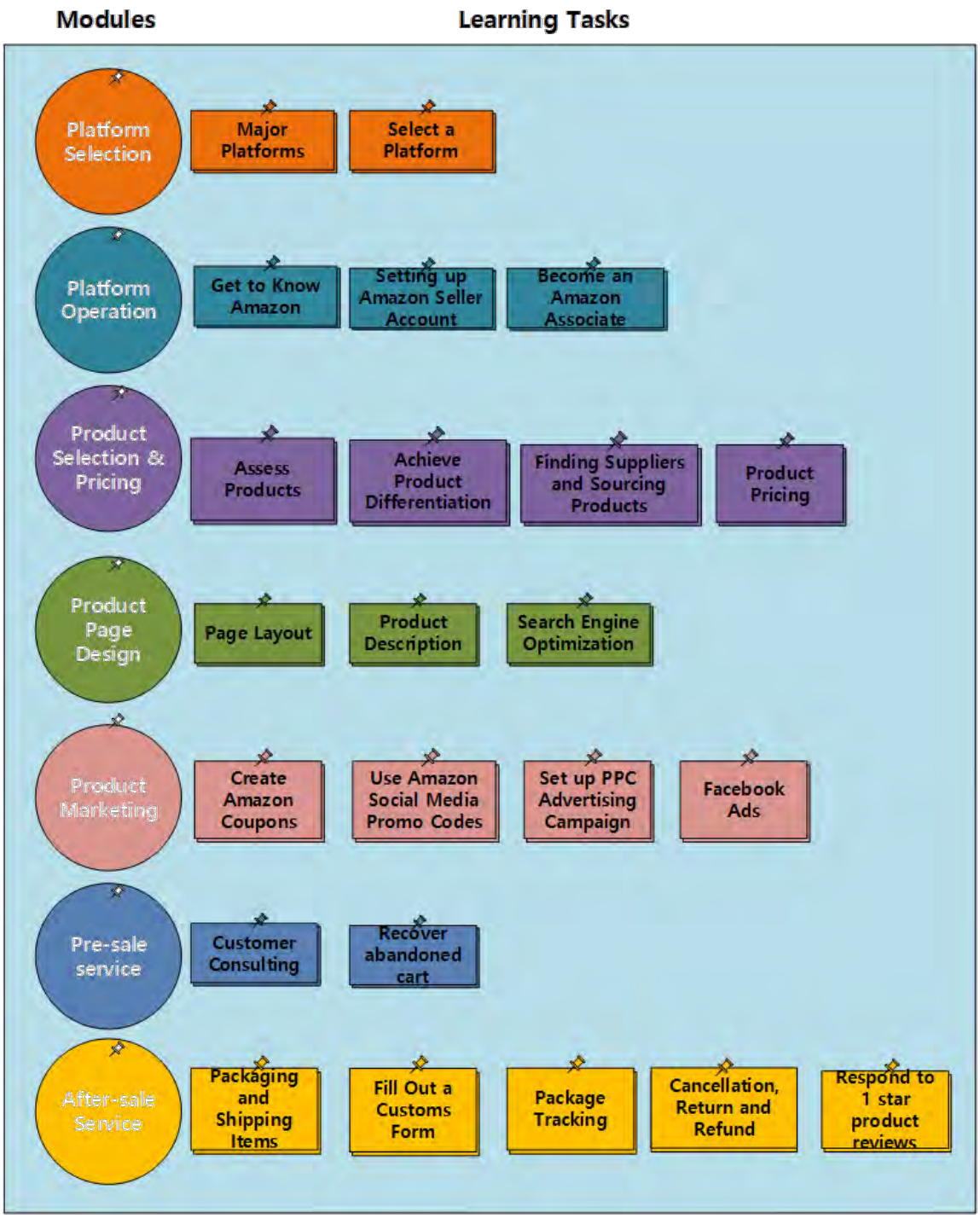

Figure 3. Modules of Learning Tasks and Subtasks in the *Cross-Border E-commerce Operations English*

Construction

No specialized corpus limited to the domain of cross-border e-commerce operations has hitherto been made readily available, not to mention a loose-leaf textbook developed based on it, so we worked closely with an IT company to build a textbook corpus called Cross-border E-commerce Corpus for Loose-leaf Textbook Development (henceforth CECLOTED). The creation of CECLOTED consisted of three steps: collection, preparation, and processing (see Figure 4).

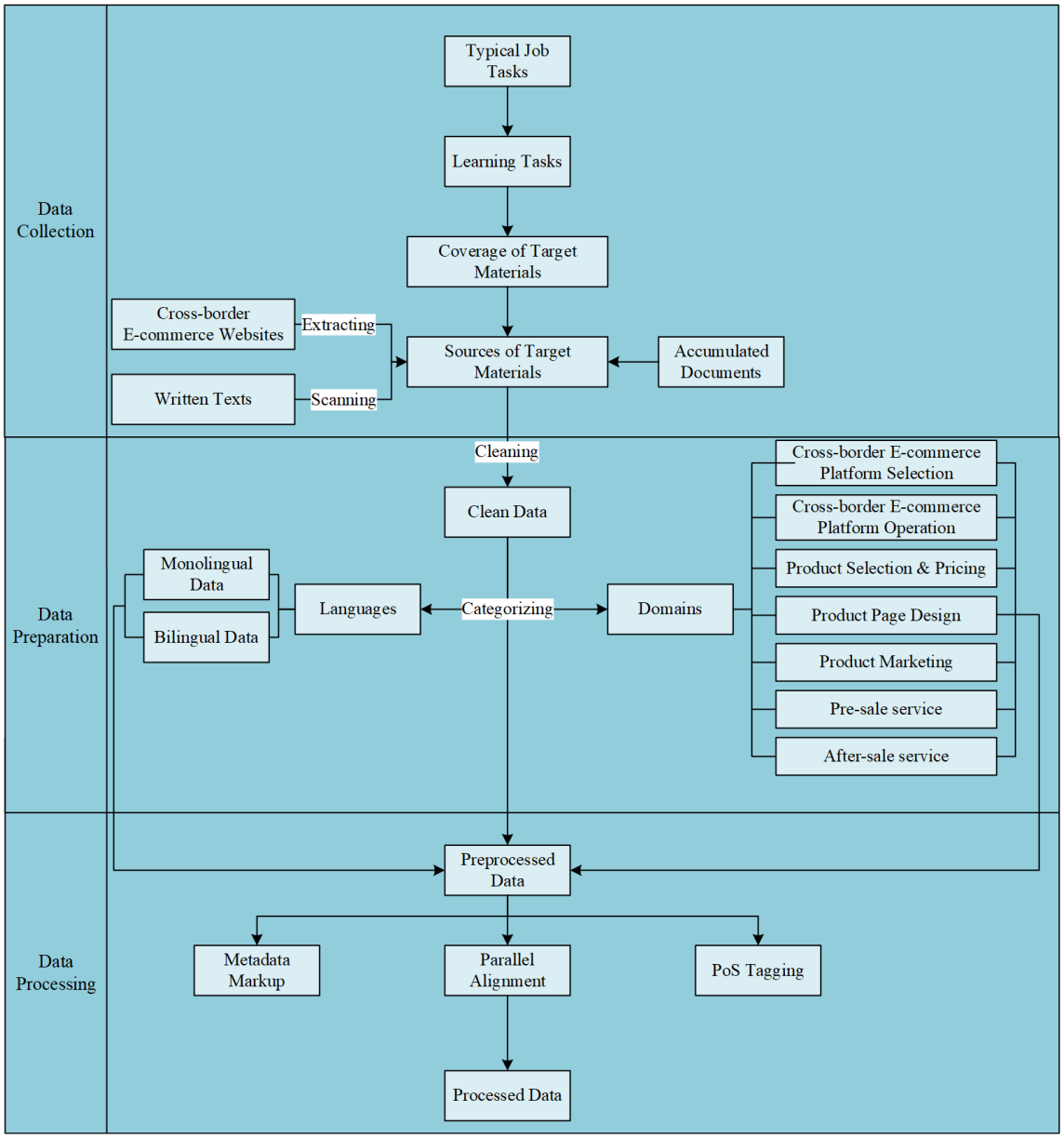

Figure 4. The Creation Process of CECLOTED

(1) The collection step drew on the results of work-related analysis from the prior stage to define the coverage and sources of target materials. CECLOTED covered subject fields related to those seven learning tasks, namely, cross-border e-commerce platform selection, cross-border e-commerce platform operation, product selection & pricing, product page design, product marketing, pre-sale service, and after-sale service. By doing so, CECLOTED painted a miniature of work scenarios, and a corpus as such is more conducive to enhancing practical language ability and professionalism. CECLOTED data were mainly sourced in three ways: first, scanning written texts from paper publications such as specialized literature, corporate case studies, technical manuals and converting the image of texts into machine-readable text formats through OCR software; second, extracting publicly available data from major cross-border e-commerce websites; third, sifting through relevant documents accumulated by the IT company from previous corpus-building projects. Also noteworthy was that we collected both monolingual and bilingual data, the latter of which was aimed, among other things, at pooling resources for designing

translation exercises in the later stage.

(2) In the preparation step, we pre-processed the corpus by cleaning and categorizing the data. In the case of CECLOTED, cleaning referred to the proofreading work to standardize the texts and to remove irrelevant digits and characters so that what's left was a high-quality dataset ready to be further processed by corpus tools. Categorizing meant dividing CECLOTED into subcorpora using two criteria, i.e. language and domain. Against the criterion of language, CECLOTED was initially divided into two corpora, one of the monolingual texts and one of the bilingual ones. Those two corpora were then further split into seven domains specific to above-mentioned learning tasks, giving rise to seven monolingual subcorpora and seven bilingual ones respectively.

(3) Next came the processing step which involved metadata markup, PoS tagging and alignment of bilingual texts. For easier data retrieval, we used a set of descriptors to mark up such beyond-text properties as the original source, publisher, publication date and author, as well as in-text structural references such as title, heading, paragraph and passage. Also, we used the PoS tagging technique to label each word with its part of speech, which made it possible to conduct syntactic analysis on corpus texts. Moreover, we employed a combination of automated methods and manual revision to have the bilingual subcorpora of CECLOTED (henceforth CECLOTED_PL) aligned with sentence and paragraph delimiters.

Transformation

Once built, CECLOTED served as a treasure box from which materials can be selected and compiled to transform into a textbook. Compilation held a crucial position in the transformation stage. Therefore, we weighed the merits and demerits of two broad schools of thought about how far corpus linguistics can and should influence the content of language teaching, namely, the "modeling" approach and the "corpus-driven" approach. As Carter et al. noted (2011, p. 94), " the differences and distinctions between corpus-informed and corpus-driven materials are useful here, with the category of corpus-driven suggesting a full adherence to the evidence of the corpus and the former corpus-informed category suggesting that some modification, manipulation and careful choice on the part of the materials writer should be preferred", corpus-informed materials presupposes modeling data on authentic patterns. By contrast, corpus-driven approaches prioritize authenticity and presume that changes made on corpus data of any kind, be it rewriting, simplification, modification, edition or revision would compromise authenticity (He, 2010). The pedagogic concerns behind the inclination towards the "modeling" approach, amongst other things, are "range, learnability and usefulness" (Ranalli, 2003, p. 8), all of which can be addressed owing to the advancement of corpus linguistic techniques and supports from the school-enterprise cooperation established for this project. First, the range factor was taken into consideration at the stage of corpus construction, we built CECLOTED as a corpus specific to the domain of cross-border e-commerce operations. Second, as for learnability of textbook materials, we resorted to readability tools and newly-issued *English Curriculum Standards for Higher Vocational Education 2021 Edition* (Ministry of Education of the People's Republic of China, 2021) as the criterion to ensure that the selected materials would neither pose an insurmountable barrier to learners nor extend beyond or fall below national curriculum standards. Third, regarding the usefulness of textbook materials, we benefited from multidisciplinary collaboration in a team made up of linguistic researchers, higher vocational English teachers, corporate specialists and IT professionals, and hence triangulation of perspectives can enhance

the validity and reliability of research findings.

Results

With *Module Four: Product Page Design* as a a case example, below will be a detailed recount of the statistical and linguistic findings during the process of transforming CECLOTED data into learning texts, vocabulary lists, learning tasks and an online learning platform, all of which are constituent parts of the *Cross-border E-Commerce Operations English*. *Module Four: Product Page Design* relied as source texts on the identically-named sub-corpus of CECLOTED (henceforth CECLOTED_PPD), data of which were extracted from product detail pages on Amazon and we ensured that those data bore no identification of and hence no potential harm to any individual users (Ainscough et al., 2018; Dawson, 2014; Eysenbach & Till, 2001). In order to ensure the representativeness and balance of CECLOTED_PPD, our team referred to the *State of the Amazon Seller* reports from 2020 to 2022(Jungle Scout, 2020; Jungle Scout, 2021; Jungle Scout, 2022), thus narrowing the coverage of data collection to the top 10 most popular Amazon product categories over the latest three years, namely, Home & Kitchen, Beauty & Personal Care, Toys & Games, Health & Household & Baby Care, Sports & Outdoors, Kitchen & Dining, Office Products, Garden & Outdoor, Tools & Home Improvement and Pet Supplies. As shown in Table 1, the size of the collection was limited to around 150,000 tokens for each of the above categories so that CECLOTED_PPD was balanced and representative of trending products in the current cross-border e-commerce market, and the size of CECLOTED_PPD is 1,306,983 in tokens and 62,640 in types.

Table 1. Statistics of Each Category of Corpus Data in CECLOTED_PPD in terms of Type-token Ratio (TTR), Standardized Type-Token Ratio (STTR), Moving Average Type-token Ratio (MATTR)

Categories	Types	Tokens	TTR	STTR	MATTR
Beauty & Personal Care	11070	132702	0.08341999	0.73411167	0.7341141
Garden & Outdoor	13850	143068	0.096807115	0.7580649	0.7596851
Health & Household & Baby Care	10514	98994	0.10620846	0.7501828	0.7491133
Home & Kitchen	13477	134883	0.09991623	0.7602598	0.7608536
Kitchen & Dining	12467	136553	0.9129789	0.7559156	0.7563213
Office Products	11828	132960	0.08895908	0.76007575	0.7593586
Pet Supplies	11768	115882	0.101604186	0.74108684	0.7403655
Sports & Outdoors	12608	142765	0.08831296	0.7542994	0.75464255
Tools & Home Improvement	13742	132190	0.101913914	0.7615131	0.76141924
Toys & Games	13214	137046	0.096420184	0.7404154	0.7398923

Learning Texts

Learning texts of a loose-leaf textbook should provide, in a systematic, comprehensive and detailed manner, theoretical knowledge encompassing the complete work process, so that learners, in the meantime of improving reading comprehension, can gain an overview of the whats, the whys and the hows of job tasks, e.g., “tools, methods, principles, procedures, skill requirements and operational guidelines for getting the job done” (Cai et

al., 2021, p. 91). With the assistance of corpus linguistic techniques, we drew on the synergy of combining quantitative and qualitative methods to measure the readability of texts and the Wordlist of the *English Curriculum Standards for Higher Vocational Education 2021 Edition* (henceforth WECSHVE 2021) (Ministry of Education of the People's Republic of China, 2021) was used as the benchmark for comparative analysis. With the aid of the metadata variables encoded within each corpus file, we started by sifting for relevant texts from CECLOTED_PPD and then pinpointed the ones that qualified as potential learning texts. Once chosen, those textbook materials candidates would undergo readability measurement by corpus tools such as AntWord Profiler 2.0.1. (Anthony, 2022b).

To be more exact, we divided the WECSHVE 2021 (Ministry of Education of the People's Republic of China, 2021) into three base lists as suggested by its developers (Ma et al., 2022), namely, WECSHVE_1st_2100 (2100 words that should have been mastered prior to college), WECSHVE_2nd_500 (500 words for fundamental-level higher vocational English learners), WECSHVE_3rd_400 (400 words for advanced-level higher vocational English learners), with which as the reference, we used AntWord Profiler 2.0.1 (Anthony, 2022b) to measure the percentage of words beyond the range of the WECSHVE in the preliminarily selected texts and filtered out the ones whose proportion of Off-list words exceeded 2%, which was a threshold of reading comprehensibility proposed by Laufer and Ravenhorst-Kalovski (Laufer & Ravenhorst-Kalovski, 2010). Then we conducted further analysis on the remaining texts in terms of vocabulary dispersion across the three base lists and Off-list respectively, thus ensuring that learning texts were sequenced by increasing difficulty. In this way, we can guarantee the authenticity and integrity of the learning texts while seeing to it to the greatest extent that the order of compilation is in line with learners' cognitive process. Figure 5 presents the dispersion of words in the four learning texts in *Module Four*: *Product Page Design* across the above-mentioned three base lists and the Off-list.

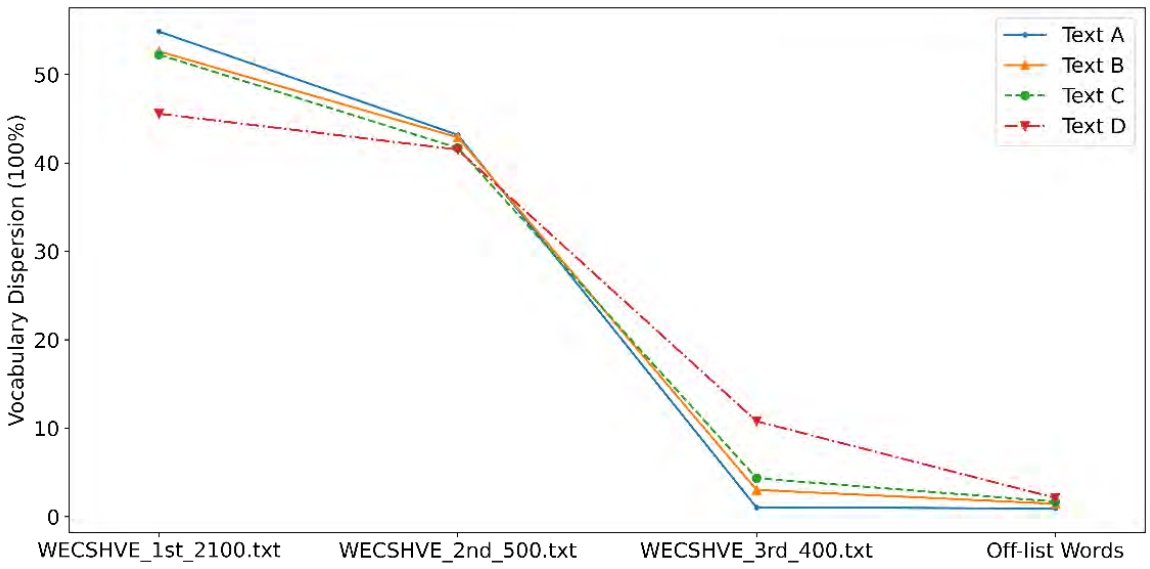

Figure 5. Vocabulary Dispersion of the Learning Texts in *Module Four* across the WECSHVE

As can be seen from Figure 5, these four texts (from Text A to Text D) are placed in descending order in terms of lower-level vocabulary (WECSHVE_1st_2100 and WECSHVE_2nd_500) whereas the trend reverses when it

comes to higher-level vocabulary (WECSHVE_3rd_400 and Off-list words), such is the statistical evidence for compiling those four texts (from Text A to Text D) in an order reflective of increasing lexical complexity.

Vocabulary Lists

Under the banner of vocational English, cross-border e-commerce operations English features domain-specific vocabulary which poses a greater challenge to learners than that of grammar or discourse. As Xu (2022) noted, vocational vocabulary or terminology is the fundamental obstacle to learning vocational English. As such, we used corpus tools to automatically extract the vocational vocabulary and terminology per module since each module is characterized by distinctive workplace scenarios. Also cite *Module Four: Product Page Design* as an example, we applied automatic identification of Keywords to the learning texts. With Crown and CLOB as reference corpora (Xu & Liang, 2013) and the learning texts of *Module Four* as the study corpus, we pinpointed the Keywords specific to the scenario of cross-border e-commerce product description using Antconc 4.1.2 (Anthony, 2022a), thus generating a list of 140 keywords ranked by keyness value and frequency. Further confirmation was undertaken by our multidisciplinary team of corpus linguistic researchers, higher vocational English teachers and corporate specialists to sift out function words, proper nouns and general vocabulary, and then the remaining words were lemmatized, thus reducing the vocabulary list to 57 words. Table 1 provides the list of Keywords, their frequencies and keyness values in terms of the 4-term Log-Likelihood keyness measure.

Table 2. The List of 57 Keywords of *Module Four*, Their Frequencies and Keyness Values in terms of 4-term Log-Likelihood Keyness Measure

keyword	**Frequencies**	**Keyness (LL4)**
SEO	27	297.96
vitamin	16	129.68
item	21	119.76
engines	10	64.49
pants	10	61.52
overalls	6	60.47
website	11	59.05
vitality	7	54.28
couture	5	49.77
template	6	48.31
prêt-à-porter	4	44.13
haute	5	42.83
keywords	5	40.06
blanket	7	39.32
bib	4	39.14
copy	8	35.3
boots	8	34.76

bling	4	34.6
purchase	8	33.25
beater	3	33.1
du jour	3	33.1
preppy	3	33.1
squalene	4	33.2
vitamix	5	33.3
intent	6	31.17
striped	4	28.92
skater	3	28.61
reviews	6	27.03
exclusivity	3	26.38
brand	8	25.78
boost	6	25.43
tags	4	25.17
faux pas	3	24.8
oatmeal	3	23.57
accessories	3	22.55
alliteration	2	22.07
dropper	2	22.07
glitterati	2	22.07
PDP	2	22.07
purees	2	22.07
squalene	2	22.07
testimonials	2	22.07
trucker	2	22.07
UPC	2	22.07
visibility	4	21.98
headline	4	21.06
optimize	3	20.94
immortality	3	20.94
recipes	4	20.78
blog	5	19.81
chic	3	19.67
evoke	3	19.13
sneakers	3	19.13
enticed	2	18.25
lingo	2	18.25
passe	2	18.25

Learning Tasks

To increase learner autonomy, we developed a series of learning tasks, the sum of which formed the backbone of the textbook. Each learning task consists of five sections titled “Learning Context”, “Learning Objectives”, “Task Description”, “Task Allocation”, “Task Preparation” and “Task Implementation” respectively, and the intent of a task-centered organization as such was to help learners reap the benefits inherent in learning through doing. Among the five sections, “Task Implementation” claimed the largest space. This section served the dual roles of task operation instruction and vocational English training, and the latter was fulfilled by incorporating the findings of corpus linguistic analysis, thus giving rise to exercises on word formation, collocation, grammar and translation. First, word formation exercises were designed on the basis of the Vocabulary List (see Table 2) compiled in the earlier stage.

By combing through the Vocabulary List, we identified words featuring a great variation in derivative forms, which served as base forms to do searches using the wildcard asterisk (*) in CECLOTED_PPD. By doing so, we tracked all derivatives of base forms in CECLOTED_PPD. Then, we probed the context of their usage through the keyword in context (henceforth KWIC) view to pinpoint example sentences semantically and structurally appropriate for designing word formation exercises. A case example is “optimize” tracked as a base form following the procedure mentioned above, and then we used optim* as the search term with Antconc 4.1.2 (Anthony, 2022a), and found that the word family of “optimize” in CECLOTED_PPD included “optimize”, “optimized”, “optimal”, “optimally”, “optimum”, “optimization” and upon further KWIC review, we chose as example sentences characteristic of the work situation of describing product details for cross-border e-commerce operations, and hence designed word formation exercises on their basis.

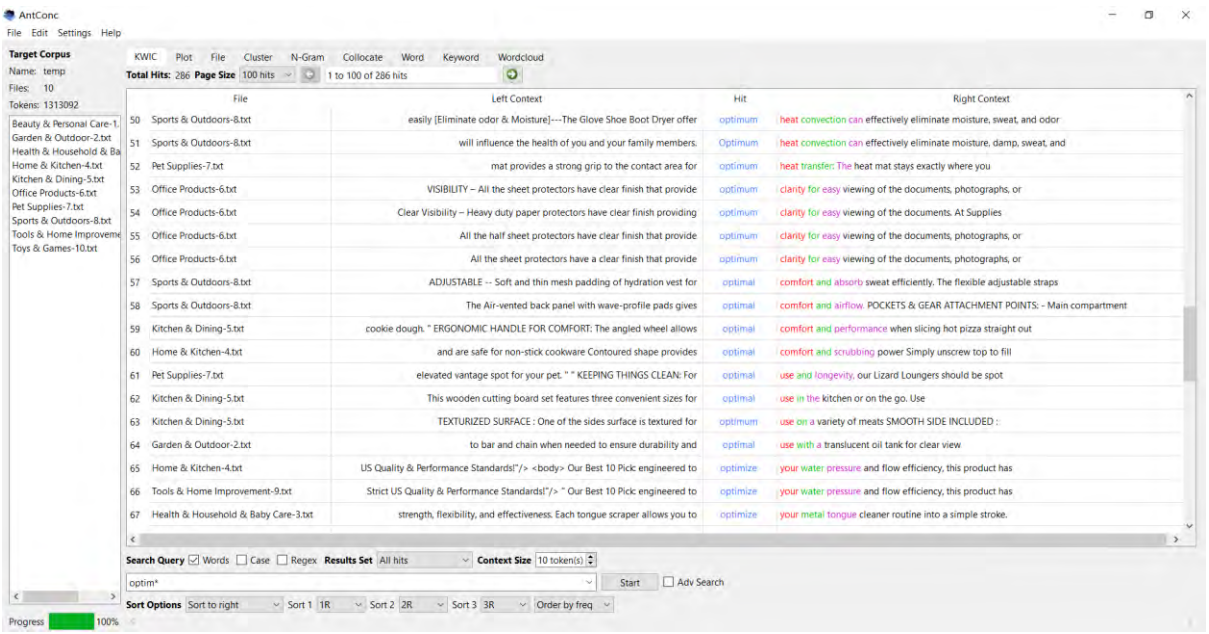

Figure 6. A Screenshot of KWIC Concordances for the Term optim* Generated Using the Antconc 4.1.2

Second, the design of collocation exercises was realized by using the collocation technique of corpus analysis

software. We used corpus-based measures to provide quantitative evidence to identify collocates. Operationally, for word A in the Vocabulary List of *Module Four*, a collocation investigation was conducted with corpus tools, rendering word B as its collocate and the strength of association between word A and word B in CECLOTED_PPD. Given the assumption that collocations are linked to fluent and natural production associated with native speakers of the language (Erman et al., 2016; Howarth, 1998; Hyltenstam et al., 2016; Paquot & Granger, 2012; Sinclair, 1991;), we would be in a position to hypothesize that the higher the value of collocational strength, the greater the necessity to master the collocation. Also with "optimize" as an example, we used the collocation measure embedded in #LancsBox 6.0 (Brezina et al., 2021) to identify the collocates of it in CECLOTED_PPD.

Figure 7 presents in graphical terms the words that repeatedly co-occur with "optimize", thus they were deemed as the collocates of it. A KWIC view uncovered a series of co-occurrences of "optimize" with its above-mentioned collocates within a specified window span of 0 word to the left and 5 words to the right of the node (0L-5R), for example: "optimize storage space patented bit-bar design", "optimize bioavailability", "optimize lean muscle growth", "optimize workouts 24/7 heart rate", "optimize the next 7 days", "optimize the performance of your air purifier", etc. These are formulaic sequences, and as noted by Wood (2002, p. 13), "given our abundant knowledge about the role of formulaic language in acquisition and production, it appears high time that we began to teach formulas and facilitate their acquisition more directly in the classroom. By doing so, we could open the door to improvements in how learners acquire second languages." Therefore, we developed collocation exercises focused on the above-mentioned formulaic language units common in the context of cross-border e-commerce.

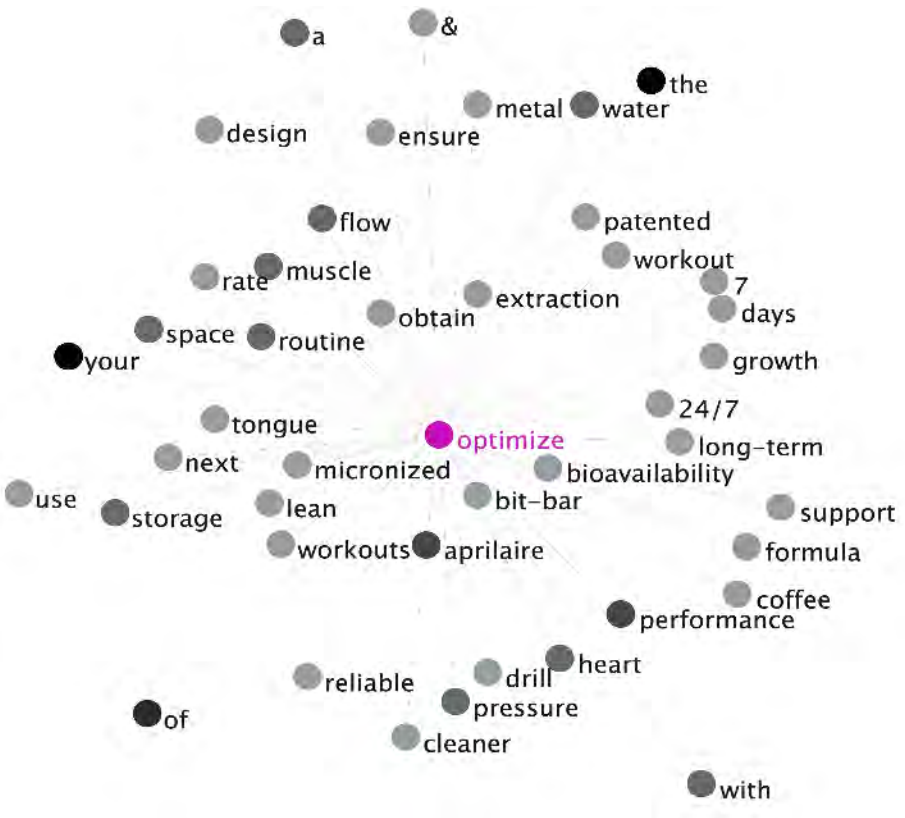

Figure 7. Graphical Display of the Collocates of *Optimize* in CECLOTED_PPD

Third, exercises were generated by drawing insights from the results of a colligation analysis on PoS-tagged data in CECLOTED_PPD. According to Firth (1968), colligation refers to the syntagmatic attraction between grammatical categories while collocation, for him, was the syntagmatic attraction between lexical items. As with collocation, awareness of colligation should accelerate the process that Hoey (2010) described as priming so as to shorten the length of time required to prime any grammatical structure. For the sake of coherence with the previous explication, "optimize" will be used again for illustration. We used #LancsBox 6.0 (Brezina et al., 2021) to find and visualize grammatical categories that co-occurred with "optimize" in CECLOTED_PPD.

As shown in Figure 8, there exists the strongest association between "optimize" and possessive pronoun (PP$) as the latter is placed closest to the former. As far as colligation frequency is concerned, common nouns (NN) claim the top spot as indicated by the intensity of the color of NN. Additionally, The positioning information in Figure 8 reflects the exact position of grammatical categories co-occurring with "optimize" in CECLOTED_PPD, i.e., coordinating conjunction (CC), possessive pronoun (PP$), comparative adjective (JJR), past tense verb of lexical verb (VVD), article and determiner (DT), common noun (NN) appear to the right of "optimize", by contrast, VVN (past particle of lexical verb), TO (to), JJS (superlative adjective) appear to the left of "optimize" while cardinal number (CD) appear sometimes left and sometimes right (as indicated by the middle position in Figure 8). All the above-mentioned colligational information concerning association strength, frequency and position serve as indicators for designing exercises to raise learners' awareness of grammatical patterns identified in the co-text of "optimize".

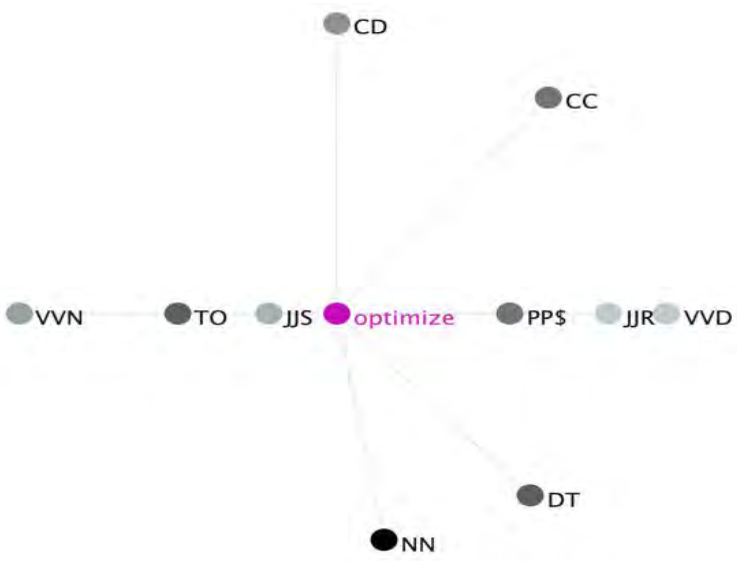

Figure 8. Colligation Graph Visualizing Grammatical Categories Co-occurring with "Optimize" in CECLOTED_PPD

Last but not least, we used CECLOTED_PL as source texts to design translation exercises, which were aimed to improve higher vocational English learners' command of lexico-grammatical patterns. Along the lines of using"optimize" as an instance, we will exemplify the hows of designing "optimize"-related translation exercises. Using "optimize" as the search term with bilingual concordancer such as BFSU Paraconc1.2.1 (Xu et al., 2012) and ticking its "lemmatize" function, we retrieved English-Chinese bilingual concordance lines containing

"optimize" and its inflected forms in CECLOTED_PL to design translation exercises, through which learners can infer the lexico-grammatical patterns of "optimize" and its inflections through translation practice.

Online Learning Platform

Among a number of revolutionary changes that the loose-leaf textbook brought to conventional print textbooks, digital enhancement broke new ground for "participation, flexibility and personalization" (Yerushalmy et al., 2014). Therefore, we availed the advantage of electronic datasets in CECLOTED to develop a corpus-based online platform coherent with the print textbook version, thus creating new spaces of connectivity and interactivity for higher vocational English learners. The corpus-based online platform was based on a B/S architecture and it was divided into backend and frontend components. The backend includes two modules, one is for data management which enables customized corpus building to facilitate autonomous learning, and the other is for user management which grants permission to the inclusion of new data into the embedded corpus after quality insurance. The frontend is the user interface which includes four modules, i.e., "My Corpus", "My Document", "My Terminology", and "Cross-border E-commerce in China", each having different functionality. Figure 9 below illustrates the formation mechanism for the corpus-based online learning platform.

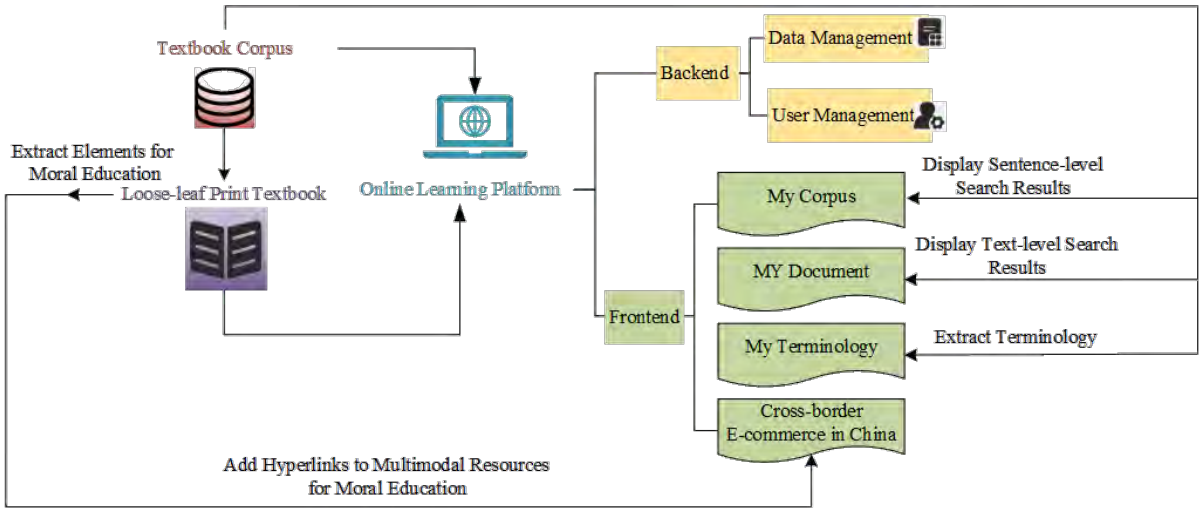

Figure 9. The Formation Mechanism for the Corpus-based Online Learning Platform

"My Corpus" displays the KWIC view through the use of CECLOTED. Then, "My Document" presents the context of the inputted search terms, i.e., texts such as specialized literature, corporate case studies and technical manuals in CECLOTED. "My Terminology" provides translation, definition and hyperlinks for quick and easy access to information related to the technical terms specific to cross-border e-commerce in CECLOTED. Furthermore, as the *Implementation Plan of National Vocational Education Reform* (Ministry of Education of the People's Republic of China, 2019b) unequivocally stated, moral education should be at the core of vocational education, so in conformity with national policy, we designed "Cross-border E-commerce in China" specifically to unleash the potentials of loose-leaf textbooks in instilling moral values and promoting cultural confidence. Albeit being an explicit part of national policy on vocation education, imparting moral values explicitly would run the risk of being too abrupt to be resonating with the learners. Therefore, we extracted from the printed version

of the *Cross-border E-Commerce Operations English* China-specific information before adding hyperlinks to the multimodal representation of such information to "Cross-border E-commerce in China", so that connectivity and coherence can be ensured between the printed version and digital version of the loose-leaf textbook in terms of moral teaching. Meanwhile, with a click of the mouse, learners can gain an awareness of the development of China's cross-border e-commerce, for example, the supportive policies such as the Belt and Road Initiative, China-Europe Railway Express and the "Going Global Strategy" as well as international situations such as the Covid-19 global pandemic, global inflation and digital divide, all of which are aimed at increasing higher vocational English learners' consciousness of the responsibility on their shoulders and the challenges awaiting to be hurdled by them as stakeholders.

Discussion

Upon the practice of developing the *Cross-border E-Commerce Operations English*, we derived inductively three key points worth noting for similar studies in the future.

Orientation towards Jobs

Being oriented towards typical job tasks is an intrinsic characteristic of a corpus-based loose-leaf textbook for higher vocational English learners and such is the essence that makes this type of textbook distinct from its disciplinary-based counterparts. Orientation towards job tasks entails a new type of corpus which embodies not only the authentic language use at work but also the complete work process to get a job done, therefore on-the-job investigation rather than introspective thoughts should be the premise for determining what materials the corpus will comprise. As such, loose-leaf textbook corpus features what enterprises need in terms of language, which distinguishes itself from traditional corpus revolving around disciplinary subjects. To be more specific, analysis of typical job tasks should be targeted at the same or similar job that certain major clusters are inclined towards (Wang et al., 2021, p. 104), and then the context of this very job should be converted pedagogically into learning situations, in accordance with which learning tasks can be developed and be ensured of authenticity and practicality. In other words, corpus-based loose-leaf textbooks should be expressive of both typical job tasks and authentic communicative context, thus linking EFL to vocational job requirements.

Maintenance of Authenticity

Authenticity plays a vital role in EFL learning. Liu & Wang (2021, p. 4) noted that once learners gain the ability and accustomization to acquire language in authentic communicative activities, learner autonomy can be substantially increased. As for higher vocational English, being authentic means that linguistic data arise from actual work activities and are reflective of the communicative context of real working process. Therefore, the loose-leaf textbook for vocational might as well position itself towards the right side of the corpus-informed versus corpus-driven continuum, maintaining to the greatest extent possible a full adherence to the evidence of the corpus. Nonetheless, a balance should be stricken between language authenticity and pedagogical feasibility so as not to pose undue cognitive difficulty (Carter et al., 2011, p. 89). A viable approach is examining data through corpus

linguistic lenses which include but are not limited to frequency, keyness, dispersion, collocation and colligation, by doing so, learning materials can be kept within the grasp of the learners while staying true to the original source to the maximum extent practicable.

Digital Enrichment

Wang et al. (2021, p. 58) remarked that digital enrichment should be conceived as an integral part of loose-leaf textbooks for vocational education. In the case of corpus-based loose-leaf textbooks for higher vocational English, a companion online learning platform is a viable form of digitization of conventional textbooks given the machine-readable feature of corpus data. The platform should be tightly connected to the print textbook yet much more than simply a digital version of it. Rather, it should provide an effective supplement to the print version by incorporating corpus-enabled functions such as KWIC view, file view, as well as terminology search, all of which are to facilitate autonomous learning beyond the classroom. Furthermore, it can be connected via hyperlinks to other sites so as to grant unlimited multimodal resources to help learners assimilate information related to the development of Cross-Border E-Commerce in their own country and raise their awareness of challenges and ensuing responsibility. In so doing, the loose-leaf textbook can fulfill the role of moral education in conjunction with the development of language proficiency and vocational skills.

Conclusion and Recommendation

To date, there has been no publicized endeavor in designing a loose-leaf textbook using corpora and corpus linguistic tools. This study targeted at the limitations of current higher vocational English textbooks and made attempts to overcome those constraints by incorporating corpus linguistic methods in textbook development. Upon the practice of compiling the *Cross-border E-Commerce Operations English*, this study proved the practicability of using corpus linguistic approaches to help EFL textbooks deliver timely, authentic, work-based content enhanced by digitalization. As has been noted in the opening part, loose-leaf textbook materials need to adjust to the ever-changing new technologies, new industries and new business models, and thus the development of such textbooks is an evolving process, providing a rich field for further exploration, be it an updated syllabus, a compatible teaching methodology or the possibility to automate the process to update the textbook corpus data echoing changes in the cross-border e-commerce industry.

Acknowledgments

This research is supported by the “2021 Research Project of Foreign Language Education Reform in Vocational Schools” and the grant reference is [WYJZW-2021-2006], granted by the Advisory Committee of Foreign Language Teaching in Vocational Education, Ministry of Education, P.R. China.

Notes

#LancsBox 6.0 use TreeTagger 3.0 for annotating texts with PoS information.

References

Ainscough, E., Smith, S. N., Greenwell, K., & Hoare, D. J. (2018). Findings and ethical considerations from a thematic analysis of threads within tinnitus online support groups. *American Journal of Audiology*, *27*(3S), 503–512. https://doi.org/10.1044/2018_aja-imia3-18-0013

Anthony, L. (2022a). AntConc (Version 4.1.2) [Computer Software]. Retrieved from http://www.laurenceanthony.net/software.

Anthony, L. (2022b). AntWordProfiler (Version 2.0.1) [Computer Software]. Retrieved from http://www.laurenceanthony.net/software.

Brezina, V., Weill-Tessier, P., & McEnery, A. (2021). #LancsBox (Version 6.0) [Computer Software]. Retrieved from http://corpora.lanc.ac.uk/lancsbox/download.php

Cai, Y., Wang, S., & Li, J. (2021). The essence and features of the loose-leaf textbook as a new form of textbook for vocational education and the key points in its development. *Chinese Vocational and Technical Education*, (11), 88–96.

Carter, R. (1998). Orders of reality: CANCODE, Communication, and culture. *ELT Journal*, *52*(1), 43–56. https://doi.org/10.1093/elt/52.1.43

Carter, R., Hughes, R., & McCarthy, M. (2011). Telling tails: grammar, the spoken language and materials development. . In B. Tomlinson (Ed.), *Materials Development in Language Teaching* (pp. 78–100). Cambridge University Press.

Dawson, P. (2014). Our anonymous online research participants are not always anonymous: Is this a problem? *British Journal of Educational Technology*, *45*(3), 428–437. https://doi.org/10.1111/bjet.12144

Erman, B., Lundell, F. F., & Lewis, M. (2016). Formulaic language in advanced second language acquisition and use. In K. Hyltenstam (Ed.), *Advanced Proficiency and Exceptional Ability in Second Languages* (pp. 111–148). Walter de Gruyter.

Eysenbach, G., & Till, J. E. (2001). Ethical issues in qualitative research on internet communities. *BMJ*, *323*(7321), 1103–1105. https://doi.org/10.1136/bmj.323.7321.1103

Firth, J. R. (1968). A new approach to grammar. In F. R. Palmer (Ed.), *Selected Papers of J. R. Firth, 1952–1959* (pp. 114–125). Indiana University Press.

He, L. (2010). Reflections on the Authenticity of Language Materials for Elementary Chinese Speakers-Study based on the Practical Chinese Textbook and the New Practical Chinese Textbook. In L. M. Liu (Ed.), *Beijing Forum of Foreign Languages and Literatures* (Vol. 3, pp. 711–723). Central Compilation & Translation Press.

Hoey, M. (2010). *Lexical priming: A new theory of words and language*. Routledge.

Howarth, P. (1998). Phraseology and Second language proficiency. *Applied Linguistics*, *19*(1), 24–44. https://doi.org/10.1093/applin/19.1.24

Hu, K. B., & Li, Y. (2016). Research on the creation and application of contemporary textbook corpus. *Foreign Language e-Learning*, (3), 34–39.

Huang, H., Yang, M. E., & Kuang, Q. X. (2021). The essence of the loose-leaf textbook as a new form of textbook for vocational education and the methodology for its development. *Education and Vocation*, (2), 99–103.

Hyltenstam, K., Erman, B., Forsberg, L. F., & Lewis, M. (2016). Formulaic language in advanced second language

acquisition and use. In *Advanced proficiency and exceptional ability in second languages* (pp. 111–148). De Gruyter Mouton.

Jungle Scout. (2020, February). *2020 state of the amazon seller report*. Retrieved October 1, 2022, from http://www.junglescout.com/wp-content/uploads/2020/02/State-of-the-Seller-Survey.pdf

Jungle Scout. (2021, February). *2021 state of the amazon seller report*. Retrieved October 1, 2022, from http://www.junglescout.com/wp-content/uploads/2021/02/State-of-the-Amazon-Seller-2021-Jungle-Scout.pdf

Jungle Scout. (2022, February). *2022 state of the amazon seller report*. Retrieved October 1, 2022, from http://www.junglescout.com/wp-content/uploads/2022/02/The-State-of-the-Amazon-Seller-2022-Jungle-Scout.pdf

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30.

Li, X. Q. (2021). The development of intelligent language teaching and research platform and its application in translation teaching. *Foreign Language e-Learning*, (3), 106–111.

Li, Z. (2020). New forms of textbooks for vocational education. *Vocational & Technical Education Forum*, (4), 21-26.

Liu, J., & Zhou, J. H. (2022). The theoretical framework of introducing traditional Chinese culture into textbooks for vocational English education. *Chinese Vocational and Technical Education*, (11), 76–81.

Liu, R., & Wang, S. (2021). Using authentic materials to develop TCSL textbooks: Theory and practice. *Language Teaching and Linguistic Studies*, (1), 1–11.

Ma, J. B., Ling, S. Y., & Zhou, R. J. (2022). The general wordlist for vocational college English: Development and application. *Foreign Language Education in China*, *5*(1), 37–42.

Ministry of Education of the People's Republic of China. (2019a, December 19). *Administrative Measures for Vocational School Textbooks*. Retrieved September 29, 2022, from http://www.moe.gov.cn/srcsite/A26/moe_714/202001/t20200107_414578.html

Ministry of Education of the People's Republic of China. (2019b, February 13). *Implementation Plan of National Vocational Education Reform*. Retrieved August 30, 2022, from http://www.gov.cn/zhengce/content/2019-02/13/content_5365341.htm

Ministry of Education of the People's Republic of China. (2021, April 1). *English Curriculum Standards for Higher Vocational Education 2021 Edition*. Retrieved August 23, 2022, from http://www.moe.gov.cn/srcsite/A07/moe_737/s3876_qt/202104/t20210409_525482.html

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149. https://doi.org/10.1017/s0267190512000098

Peng, S. S. (2020). The Situation and Reflection of the Development of Vocational English Textbooks in Vocational Colleges. *Theory and Practice of Education*, *40*(30), 30–32.

Ranalli, J. M. (2003). *Elt coursebooks in the age of Corpus Linguistics: Constraints and Possibilities*. Retrieved September 30, 2022, from https://www.scribd.com/document/576116471/ELT-coursebooks-in-the-age-of-corpus-linguistics-constraint-and-possibilities

Sinclair, J. (1991). *Corpus, concordance, Collocation*. Oxford University Press.

Wang, A. X., & Li, X. X. (2018). Research on the guiding principles of ESP textbook development for maritime

students. *Chinese Vocational and Technical Education*, (20), 84–91.

Wang, Y. S., Zhang, C. Y., & Yu, C. X. (2021). *Systematic design and development guidelines of loose-leaf textbooks, work manual-style textbooks and integrated media-featuring textbooks for vocational education*. Chemical Industry Press.

Wood, D. (2002). Formulaic language acquisition and production: Implications for teaching. *TESL Canada Journal*, *20*(1), 1–15. https://doi.org/10.18806/tesl.v20i1.935

Wu, Q. Q., Wang, X. W., Yan, Z. Y., & Wang, Y. J. (2022). Concept Clarification and Structure Selection of the Loose-leaf Textbooks of Vocational Education in the New Era. *Vocational and Technical Education*, *43*(4), 31–38.

Xu, J. J. (2022). Lexical syllabus design of higher vocational English and its instructional implications informed by the Principles of Four Usages. *Foreign Language Education in China*, *5*(1), 43–49.

Xu, J. J., & Liang, M. C. (2013). A tale of two C's: Comparing English varieties with Crown and CLOB (the 2009 Brown family corpora). *ICAME Journal*, *37*, 175–183.

Xu, J. J., Liang, M. C., & Jia, Y. L. (2012). BFSU ParaConc (Version 1.2.1) [Computer Software]. Retrieved from http://corpus.bfsu.edu.cn/TOOLS.htm.

Yerushalmy, M. (2014). Challenges to the authoritarian roles of textbooks. In K. Jones, C. Bokhove, G. Howson, & L. Fan (Eds.), *International conference on mathematics textbook research and development 2014 (ICMT-2014)* (pp. 13–20). University of Southampton, United Kingdom.

Author Information

Jing Liu

https://orcid.org/0000-0002-9465-2135

Chengdu Polytechnic

No.83, Tianyi Street, Hi-tech Zone District

Chengdu, Sichuan

China

Contact e-mail: *48087577@qq.com*