Building and Applying an English Corpus for Animated Special Purposes

Yingying Wang

School of International Exchange, Jilin Animation Institute, Changchun, 130032 Jilin, China

Keywords: Animation english, Special purpose english, Corpora

Abstract: The research on the integration of specialized English and corpora is a natural development trend. The study of building a corpus of English for animation special purposes is of positive significance to meet the needs of the globalized animation industry chain and the needs of teaching English for animation. The construction of the corpora of English for special purposes in animation can follow the three principles of professionalism, authenticity and scale, and be implemented through three steps: corpora collection, corpora sampling and corpora assignment. The constructed corpora can be widely used in the compilation of animation high-frequency vocabulary handbooks and vocabulary research.

1. Introduction

English for Special Purposes (ESP) refers to English that is closely related to a specific discipline or profession. It has distinctive stylistic features and a clear linguistic purpose. Moreover, it is characterized by clear objectives, strong relevance, and high practical value. Animation English is a subcategory of special purpose English, which has a very important role in animation production, product development, industry communication, and professional teaching. The combination of corpora and special-purpose English is a development trend. Small special-purpose corpora are important for the research and application of special-purpose English. Since the current corpora research is mostly focused on arts, engineering, medicine, agriculture and other directions, less attention is paid to art subjects, especially in animation English corpora, which cannot meet the needs of animation globalization industry chain and animation English teaching needs. Therefore, the construction of animation ESP corpora has high academic value and application value. In terms of academic value, it is conducive to promoting the interdisciplinary development of corpora and ESP, is an extension of corpora theory in the field of ESP, and provides a more complete theoretical framework for the study of ESP. In terms of application value, it has a guiding role in the application of English for animation in the actual working environment, promotes the research and teaching application of ESP in animation-related majors, and provides a new way to improve the foreign language ability of international composite animation talents.

2. Status of Research on Corpora of ESP

The research and development of ESP corpora comes into being with the rise of corpora research. Quantitative data has been used to describe language teaching since 1920, and it is particularly important in special English teaching. Focusing on actual linguistic phenomena is an important part of ESP research. Barber, for example, used a corpus of 23,000 words, which contained three types of text: engineering, chemistry, and astronomy [1]. Barber's research could respond to the contemporary controversy over whether progressive tense should be taught in English for Science and Technology [2]. With the rise of corpora linguistics, it is also a natural trend for corpora to be used in ESP teaching. Dudley-Evans & St. Johns pointed out that the research on the combination of corpora and specialized English has great potential [4]. Hunston believed that “what to teach” is more important than “how to teach” in the field of specialized use [5]. In the teaching of academic English writing, it is necessary to guide learners to be aware of the distinctive characteristics of academic genres. In the past decade, studies on ESP corpora have become more systematic and perfect. Laura Gaviloli explored the influence of corpora-based learning on SLA in ESP teaching,

  DOI: 10.25236/icited.2022.010

and explored the methods and possibilities of corpora-based students' autonomous learning [6]. In addition, the construction and research and development of medium and small corpora in specific disciplines such as business English and medical English are also in full swing [7]. In China, the research on ESP corpora mainly focuses on academic English, lexical block research and dictionary compilation. Jiang Feng published *Corpora and Academic English Research*, discussing the research philosophy and paradigm of academic English corpora [8]. Hu Chunyu focused on the large number of terms, idioms, fixed or semi-fixed collocations and multi-word sequences in business English [9]. Xu Jiajin made a pioneering exploration on the compilation of his own medical English dictionary based on the corpora of international journal articles of medical English [10].

The construction and development of specialized purpose corpora has opened up a new way for specialized language research and is the organic combination of language and technology. In the context of corpora construction in full swing, animation English corpora construction is rarely involved. There are still many aspects to be explored and developed.

3. The Construction of the English Corpus for Animated Special Purposes

Due to the influence of the object of use, scope of application, resource characteristics, research content and other comprehensive factors, the principles and plans of the construction of the English corpora for animated special purposes have certain particularities.

3.1 The Principal of Building the Corpus

The development of network information technology and computer technology has facilitated and technically guaranteed the construction of an English corpora for animated special purposes, making it possible for small teams to carry out corpora development. In the process of building the corpora, the author follows the following principles:

3.1.1 Specialty

ESP corpora is different from general corpora, which is used by groups in specific professions or fields. Therefore, it needs to be targeted in corpora collection. In the collection process, apart from language researchers, it also needs the guidance and cooperation of professionals. Based on this, the establishment of English corpora for animated special purposes should be closely combined with the development needs of the animation industry and the learning needs of students majoring in animation. With the joint participation of animation professionals, the corpora should be selected according to professional background and professional needs, and the selected corpora should be representative and authoritative.

3.1.2 Authenticity

The principle of authenticity is followed in the collection of original materials from the corpora. “Analysis of authentic forms of language use in natural texts” using the corpora means “using a collection of natural texts collected in a certain volume and according to certain principles as a basis for analysis”: Extensive use of computers “to analyze the linguistic data, including automatic and interactive techniques”: “The analysis relies on both quantitative and qualitative analysis techniques”. These features guarantee the scope and reliability of corpora-based analysis, which is not possible with other methods [3]. Therefore, the construction of the animation English corpora should be based on the real corpora, collecting English materials such as authoritative publications, monographs, textbooks, news and information of the animation industry, with a certain proportion of each type of language materials respectively, with the theme close to the profession and covering more representative professional vocabulary.

3.1.3 Scale

Scale is the basis for ensuring the reference of corpora resources, which means that the collected corpora should reach a certain scale. “In terms of scale, as long as conditions permit, corpora should

have the largest scale, the larger the better" [11]. At present, the size of large general English corpora can reach hundreds of millions of words. Medium-sized ones are around 500,000-1 million. Small corpora are mostly special-purpose English corpora, usually between 50,000-100,000 words, which can meet the phonological research. In view of the overall resource situation of the animation industry and the actual needs of research, it is suitable to build a small corpus with a total corpus of about 100,000 words, and the definition of this word count meets the basic requirements of a small corpus.

3.2 The Plan of Building the Corpus

3.2.1 Linguistic Data Collection

We collected authoritative publications, monographs, textbooks, news and information materials of the animation industry, including authoritative works such as *Original Animation Basic Course - Animator's Survival Manual*, *Animator's Workbook*, *The Laws of Animation Performance: Bringing Your Character to Life*, *Animation Character Design*, and also news and information from international mainstream animation websites and scientific papers from international mainstream animation websites, with a total library capacity of over 100,000 words. For web resources, we used offline browsing and downloading tools (such as webdup, HTTrack, etc.) to download the text of the whole website to personal computers, and for paper text, we use scanning input, graphic conversion, manual input, etc. to input into computers.

3.2.2 Sampling of Linguistic Data

A combination of stratified random sampling and equally spaced sampling was used for corpus sampling. A total of five categories were formed, and the total number of texts in each category varied due to genre differences and volume differences, but the number of words drawn from the texts was the same, all 20,000. This method of evenly distributing the sample is more economical and efficient, and it is also easier to make comparisons between corpus categories. In addition, in order to satisfy both the periodicity characteristics and the authority of the corpus, the selected linguistic data of books and textbooks were within the last 30 years, animated journals and academic papers within the last 10 years, and news within the last 5 years (see Table 1). *Original Animation Basic Course - Animator's Survival Manual*, *Animator's Workbook*, *The Laws of Animation Performance: Bringing Your Character to Life*, *Animation Character Design*

Table 1 Sample Details Of the Linguistic Data

Category	Total number of texts	Number of using the words	Source
Literature	3	20,000	*Animator's Workbook: Motion Breakdown*, *Animator's Workbook: Animation Film and Video Production*, *The Art of Animated Special Effects*
Teaching material	3	20,000	*Original Animation Basic Course - Animator's Survival Manual*, *Animation Character Design*, *The Laws of Animation Performance: Bringing Your Character to Life*
Publications	10	20,000	Various authoritative English animation publications such as*Stopmotion Magazine*, *Animation Magazine*
News	40	20,000	Authoritative animation website various English information
Professional papers	10	20,000	English academic papers in Animation
Total Text	66	Total number of using the words	100,000

3.2.3 Tagging of the Linguistic Data

The corpus was lexically tagged using the corpus tagging software TreeTagger, which has an accuracy rate of 96%-97% and requires manual proofreading after automatic tagging. The lexical assignment of the corpus is equivalent to "value-added", which is convenient for researchers to search, and facilitates the investigation of the usage and recurrence of animation terms, the study of semantic associations, lexical features, and the linguistic habits of the discourse community.

Considering that the purpose of this corpus is lexical research, semantic and syntactic assignments can be avoided for the time being.

4. Application of the English Corpus for Animated Special Purposes

Corpus-based research has penetrated into various fields of language research and language teaching. Corpus research organically integrates quantitative research with qualitative research, which effectively enhances the efficiency and scientificity of language research. The author focused on the development of a high-frequency vocabulary handbook for animated English and the study of animated English vocabulary to explore the specific applications of the English corpus for animated special purposes.

4.1 The Handbook of High Frequency Words in Animated English

4.1.1 High Frequency Word Extraction

In the 1980s, the concept of “vocabulary syllabus” was introduced in applied linguistics, which suggested that English learners should focus on the most commonly used word forms, their core usage, and the typical word combinations formed by these word forms [12]. Based on the English corpus of animated special purposes, the extraction of high-frequency vocabulary in animated English is more convenient, efficient and accurate. We used corpus retrieval and analysis tools (e.g. AntConc, WordSmith Tool, etc.) for text processing, the top 3,000 words in terms of word frequency were counted, the function words appearing in the corpus, such as crown words, prepositions, conjunctions, etc., were blocked out, and finally the top 1,500 representative words in animation professions were selected through software processing and manual screening to form a high-frequency word list.

4.1.2 The Translation of High Frequency Words

The 1,500 words were translated manually by consulting dictionaries, professional guidance and literature cross-referencing. In the process of translation, special attention should be paid to the difference in meaning between professional and general English, for example, the word “inbetween” means “in the middle” in general English, but in animation English it means In between” means “the picture in the middle, the frame in the middle.”

4.1.3 Example Sentences and Illustrations

Example sentences were added for specialized high-frequency words. These example sentences were made by a combination of corpus resource indexing and manual sentence construction. Each professional vocabulary was equipped with one or two example sentences. The selected example sentences must be representative, with at least 10 words per sentence. At the same time, in order to reflect the vividness and interest of the vocabulary handbook, illustrations were chosen for the representative vocabulary, and the pictures could fully reflect the meaning of the vocabulary or further explain the content of the example sentences.

The handbook of animated high-frequency words can be considered as an in-school teaching aid for English for special purposes after it is printed. It can also be contacted to publishers for public distribution. In terms of the form of the manual, it can be developed in the direction of “three-dimensionality”, creating both paper and electronic versions to enhance the experience of teachers and students through the cross-complementation of different media platforms and to meet the development needs of teaching materials in the new era. The development of the handbook of high-frequency words in animated English solves to a certain extent the problem of the lack of resources for teaching animated English, provides reference and support for teachers in teaching specialized vocabulary, and helps to improve the quality of vocabulary teaching in specialized English courses.

4.2 Study of English Vocabularies in Animation

4.2.1 Word Collocation

Word collocation refers to “the co-occurrence of two or more words within a short distance of each other in a text” (Sinclair, 1991). This definition is also suitable for the study of lexical collocation in the corpus. PowerConc software can be used to search the lexical collocations of specific animation terms to understand the collocation patterns and application contexts of specific terms. This content is easy to use, simple and intuitive, and can be used in teaching English for animation specializations and in students' self-learning.

4.2.2 Word Colligation

Word colligation is a summary of forms and word classes from grammatical forms, and is a more abstract form than the collocation of words. For example, if N+N represents a class concatenation, then animation production, CG design, animation movie, etc. are concrete examples of this class concatenation. colligation is important for lexicography and foreign language teaching. It can not only distinguish the meaning of words, but also distinguish the nature of words. In addition, colligation is also important for learning prepositional collocations, which can be useful for both general and specialized English learning.

4.2.3 Vocabulary Tests

Vocabulary tests are widely used in practice as one of the basic tools for considering foreign language learning. Specialized vocabulary tests in ESP have always been an important part of the test to check students' mastery. However, the limitation of the range of topics makes it more difficult for teachers to give questions. It is difficult to ensure the correctness and typicality of some self-developed questions as well, which is particularly problematic in teaching animated English. The combination of corpus and vocabulary tests can solve this problem well. For example, using Test Builder software, select a built corpus of ESP in animation, input the animated English vocabulary that needs to be tested. The software can generate fill-in-the-blank, fill-in-the-blank according to initial letters and fill-in-the-blank and multiple-choice questions according to demand, which is simple and convenient. The correctness and professionalism can be well guaranteed, while reducing the time of a large number of teachers to create questions. The test questions can be saved as word documents, which are easy to print on paper and upload to the online teaching platform for batch import of test questions.

It can be seen that multiple applications of vocabulary research and teaching can be realized based on the English corpus of animated special purposes. Based on the corpus, in addition to word collocation, colligation, and vocabulary tests, research on word frequency, semantic tendencies, sentence patterns, and pragmatics can also be conducted, which will not be discussed further due to space limitations.

5. Conclusion

Compared to traditional language research and learning methods, linguistic data can be more intuitive and effective in verifying language patterns and language intuition, increasing the correctness and accuracy of language learners in language use and improving language user confidence. The combination of corpora and ESP has become an inevitable trend of the times. “Corpora can highlight recurring language features and facilitate learners' learning of special-purpose English. Small corpora are well suited for the teaching and research of ESP” [13]. The construction and development of an English corpus for animated special-purposes is still in its infancy and requires interdisciplinary cooperation between foreign language researchers and animation professionals to lay the foundation for a larger-scale dynamic corpus of animated English in the future. There are still many aspects to be explored in the application research based on the English corpora of animated special purposes. The purpose of this paper is to draw inspiration, and we look forward to working with colleagues to share the resources of the ESP corpora and to conduct more in-depth exploration in the application research of semantics, syntax and discourse.

Acknowledgment

Foundation project: “Critical Analysis and Research on Chinese Images in American Animated Films” of Jilin Animation Institute Social Science Foundation, Project No. KY21SY10; “Construction and Application of English Corpus for Special Purpose in Animation, Manga and Games”, Jilin Animation Institute Social Science Foundation.

References

[1] Barber C L. Some measurable characteristics of modern scientific prose [J], 1962, pp.3-14.

[2] Swales J M. ESP Comes of Age? -21 years after Some Measurable Characteristics of Modern Scientific Prose [J]. 1984, pp.14.

[3] Biber D, Conrad S, R.Reppen. Corpora linguistics [M]. Cambridge: Cambridge University Press,1998, pp.04.

[4] Dudley-Evans T, M. St. John.(1988).Developments in English for special purposes [M]. Cambridge: CUP, 1988, pp.103.

[5] Hunston S. Corporal in Spplied Linguistics [M]. Cambrideg: CUP, 2002, pp.211.

[6] Laura Gavioli. Exploring Corpora for ESP Learning [M]. Tsinghua University Press, 2016, pp.12.

[7] Sinelair J.M. Corpora, Concordance, Collocation [M]. 0xford: Oxford University Press, 1991, pp.70.

[8] Feng Jiang. Corpora and academic English research [M]. Shanghai: Foreign Language Teaching and Research Press, 2019, pp.89.

[9] Chunyu Hu, Lidan Chen, Anping He. A corpus-assisted study on teaching business English phrases [M]. Shanghai: Foreign Language Teaching and Research Press, 2019, pp.32.

[10] Jiajin Xu. A dictionary study of academic medical English from a genre phraseology perspective [J]. Foreign Languages and Their Teaching, 06, pp.52, 2017.

[11] Huizhong Yang. An introduction to corpus linguistics [M]. Shanghai: Foreign Language Education Press, 2002, pp.136.

[12] Zhengbin Feng, Feng Wang. A conceptualization and pedagogical application of a financial English news corpus [J].Foreign Language Education, 04, pp.57, 2016.

[13] Sha Liu. Teacher-student corpus building: English curriculum resource development for civil aviation - A case study of Exploring the Corpus of English for Special Purposes Learning [J]. Journal of Civil Aviation, 03, pp.40, 2019.