

Exploring Pedagogical Approach in EFL Courses Based on Bilingual Parallel Corpus

Haozhi Sun^{1,a,*}, Jianzhe Qin^{2,†}, Yashi Li^{3,†}

¹*University of International College, Macao University of Science and Technology, 999078, Macau SAR, China*

²*Department of foreign language, Guangdong University of Technology, Guangzhou, 510006, China*

³*Foreign Language College, Minnan University of Science and Technology, Quanzhou, 362700, China*

a. 1909853fu011011@student.must.edu.mo

**correspondence author*

†Those authors contributed equally.

Abstract: This paper is to examine the feasibility of bilingual parallel corpus as a data-driven approach targeting to comprehensively enhance English competence of EFL (English as a foreign language) learners. In particular, corpus turns out to be functional of assisting EFL learners in acquiring high-frequency words and the collocation and colligation, indicating that corpus can be considered to be optional in pedagogy, as well. Bilingual parallel corpus, more importantly, provides learners with an ideal tool to focus on the translation techniques based on the function of comparing the same words or similar conceptions in corpora of two languages, providing authentic texts to expound translation theories, terminologies and techniques. In addition, corpus implies the use of improving learners' comprehension of original texts, further the reading proficiency. It also helps to extend the scope of reading based on the large quantity of authentic materials.

Keywords: Corpus Linguistics, Data-driven learning, Translation and language teaching, EFL, Vocabulary Teaching, Reading with Corpus.

1. Introduction

The word “corpus” (pl. corpora) is from Latin, which means “body”, and thereafter comes to denote “a collection of written or spoken texts”. Initially, the corpus method emerged as an empirical research method, which originated from large quantities of observation. After that, linguists focused on real materials of text, with the conviction that facts derived from instinct and introspection are inadequate [1]. This review paper will mainly focus on the exploration of how parallel corpus be exploited as a pedagogical approach in EFL courses.

Since Tim Johns proposed the conception of DDL in 1992, the integration of corpus linguistics (CL) into pedagogy has reached new heights [2]. According to Gilquin and Granger, incorporating corpus analysis tools, namely concordancers, into language pedagogy can effectively tap the utility

of language materials in many respects [3]. Learners can make sense of such language points like high-frequency words and its collocation not by intuition but KWIC (key word in context).

Given the particularity of aligned texts in both languages, bilingual parallel corpora are reasonably ideal for translation pedagogical purposes. Despite the large amount of time and energy to construct a parallel corpus, scholars like Wang have made great efforts to materialize it into pedagogy as assistance [4,5].

Moreover, a parallel corpus contains two sub-corpora which can be separated as required. The function of presenting various words of different types in concordances can be used in reading pedagogy, where learners will project greater language competence, both in processing information and grasping the theme. Reading materials as a relatively independent section can be used for creative pedagogical activity like jigsaw reading, which will be mentioned a little below.

2. CI as a Methodology for Education

As an emerging approach in EFL courses, there exists drawbacks to be honed. Braun concerned about the clash between currently existing general corpora and pedagogic requirement, particularly in terms of content, size, format, and annotation. To address these problems, a pedagogic processing is needed to “transfer corpus findings into pedagogically accepted materials and means [6]. For years “Pedagogic processing” of corpus is a profound reflection of many scholars on the application of corpus linguistics to language education, and that will be our focus.

2.1. The Development of Corpus and Pedagogy

The past few decades have witnessed an evolution of corpus linguistics, from its emergence as a tool combined with computers to applications in multiple fields. After the 1960s, it developed rapidly with the advent of electronic corpora like LOB and Brown; Since 1990s, corpora have advanced immensely in sizes and registers for use of diversity. Throughout decades of evolution, corpus can presently be defined as what McEnery & Hardie said: a collection of some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions [7]. Parallel corpus has been defined by multiple scholars, and in this paper, it adopts the definition given by McEnery & Hardie: “Source texts in one language plus translations into one or more other languages” [7]. As a descriptive and applied branch of linguistics, corpus linguistics (CL) has gained wide currency across disciplines and other cross-fields, and integrating corpus-based approaches into classroom as a pedagogy is a case in point. In China, Wang K.F as the spearhead has proposed several examples of leveraging corpus in vocabulary/translation teaching to endorse the utility of parallel corpus [4,5]. The integration of technological elements will add to diversity in EFL learning and provide learners new insight into language itself.

2.2. Key Characteristics of Parallel Corpus in Pedagogy

Not merely a parallel corpus, any given corpus for multiple applications should take several dimensions into consideration. The following characteristics are noteworthy:

Firstly, text in corpora should be machine-readable data. It is the large scale of the data used that explains the practicality of machine-readable text [7].

Secondly, all materials in corpora are authentic ones, or “used languages”. In other words, they are not intentionally fabricated with the purpose of teaching or examination, which makes them ideal materials for pedagogy. However, there may exist a potential rule of selection of materials for handout, in order to circumvent improper or even vulgar contents.

Thirdly, each corpus is sampled for a certain purpose as general or specialized corpus, in order to be representative and balanced [7]. Constructor of corpus should select an approach of sampling (e.g.

simple random sampling or stratified sampling.). Whichever the approach, the corpus is required to be representative of the language. Take BNC (British National Corpus) as an example, the corpus covers several genres (news, academic, literature, etc.), and set proportions for each section to keep it balanced.

One of the advantages of parallel corpus is that such a corpus contains two languages and can be separated into two subcorpora. They can function separately or collectively given different requirements to fulfill a certain direction. Recently (in 2021) a corpus named Yiyang English-Chinese Parallel Corpus was established with great efforts made by Beijing Foreign Studies University (Hereinafter referred to as Yiyang corpus) [8]. Unlike traditional parallel English-Chinese ones, which are slightly out-of-date and inadequate in scale, Yiyang corpus is characterized by its up-to-date and authentic materials of several genres, allowing the researchers to utilize flexibly.

2.3. Data-Driven Learning in Pedagogy Using CL

The awareness to integrate corpus with pedagogy was initially proposed by Tim Johns [2]. In 1991, who are convinced that learners are required to “discover” the language, while the teacher are merely the one to provide language materials. There exists a major conversion of teachers’ role from information transmitter to research director and collaborator with the aid of corpus [9-12]. It is defined as the informants of a specific type, and “give learners access to the facts of linguistic performance”, that is what he called Data-Driven Learning (DDL), a conception widely mentioned in EFL teaching materials. What makes it creative may lie in constructing a connection between “data” and “knowledge” [1]; that is, to gather scattered data and turn them into conclusion (or sometimes knowledge). Such an inductive approach “seem to be at the heart of what is generally accepted as DDL.

At first, learners can only be handed the outcome of concordance. After that, they are provided with access to concordancer, thus allowing learners to make spontaneously explorations in the KWIC (key word in context) form. At present, DDL has formed the backbone in pedagogy with CL.

But DDL approach in pedagogy seems to deepen the variation of how learners react to the lesson. On teacher’s part, the onus is on them to balance the proportion of DDL. Entirely applying DDL may not make a huge difference if it is mechanical and repetitive, and taking it as an auxiliary technique for clarifying specific language points may be more productive. On learners’ part, DDL entails much initiative of learners. As is noted by Tim Johns, “it would not work with students as unintelligent, unsophisticated, and poorly-motivated as theirs [2].”

3. Vocabulary Teaching

As an instrumental approach, Data-driven learning functions in a wide range when it comes to the application of language pedagogy. In effect, Breyer suggests that DDL, for its variability, is confined only by the learners’ imagination [11], which is to say, there are diverse choices available to choose from. However, it is pointed out by Gilquin and Granger that it is more suitable to be selective of the particular types of corpora to reach certain purposes of teaching [3]. It is considered that corpus, consisting a large quantity of linguistic data, performs little use if not utilized with other tools together for exploitation [3]. In this context, a new trend of concordance tools introduced to language teaching would greatly improve the efficiency of DDL [12].

Thus, based on Yiyang Parallel Bilingual Corpus as an ideal material for the comprehensive English competence improvement, with the aid of the concordancer, this paper discusses feasible application methods adapting to certain diverse aims of the enhancement of English learners’ competence in which teaching high-frequency words, collocation and semantic prosody are covered.

3.1. High-Frequency Words Teaching

It is considered that high frequency words are of importance for EFL students to learn to reach proficiency [13]. Leech states that word frequency should be considered as a prior part in Language learning, arguing that the frequency of occurrence of words in a certain degree indicates the language using habits of the speakers [14]. Hence, Chen illustrates that based on language usage habits that native speakers have, it is reasonable to acquire the high-frequency vocabulary [15], so that the EFL learners could efficiently grasp the general idea of the native language users in communication.

It is also suggested that synonymy, as a challenging area in English Language teaching and learning for their meaning and usage varied to a great extent [16], is another part that would be confusing to learners, resulting the lack of authenticity in learners' output. Liu suggested that learners could improve their dictation accuracy in language use comparing the frequency of synonyms [17]. In the application of Yiyao Parallel Corpus, learners are then encouraged to concordance synonyms according to the frequency in the corpus by sorting.

Table 1: Frequency of 5 synonyms in Yiyao Corpus.

Words	achieve	gain	obtain	acquire	fetch
Frequency	29	23	10	6	5
Percentage	39.73%	31.51%	13.70%	8.22%	6.85%

In Table 1., for native speakers, tendency of discriminately using 5 synonyms with the same Chinese meaning “获取” is shown. In the data base of Yiyao Corpus consisting of authentic materials, native speakers use “achieve” mostly among the others, while “fetch” rarely occur in the corpus with the frequency of 5.

3.2. Collocation Teaching

For EFL learners, it is a general case of making mistakes both in oral and written forms regardless of their language proficiency in relation with vocabulary [18]. It is considered that collocation error is one of the most universal ones across language learners [19]. Koosha and Jafarpour find that one of causes of unsuitable words collocation presented in language output is due to the lack of knowledge in terms of collocation [12]. Further, Hill states that the phenomenon is in the light of their “limited collocation competence” [20]. Neglecting the acquisition of collocation, learners in practical learning put more emphasis on the separated semantic characteristics of each word [21]. Hence, it is the approach of Data-Driven Learning to cultivate the learners' collocation competence, by which concordance could play an importantly helpful part.

At the learner-led end [22] where learners are encouraged to actively explore the knowledge of language freely, it is recommended to utilize the “cluster” function of the concordance tool to find the authentic collocation of the selected word. By acquiring the lexical and grammar items collocating with the selected item, learners would find it much clear to learn different collocation forms from the presented list. Classified as one of the ten collocational patterns, collocation of prepositions has been a major problem among the others and suggested to be essential to produce language like natives [23]. Taking the example shown below of the word “agree” (see Table 2.), the prepositions that collocate with it is actually diverse, however, not random, that is to say, when acquiring the collocation knowledge, learners have to pay special attention to the practical use of the items.

Table 2: Frequency of different patterns collocation.

No.	Cluster	Rank	Freq	Range
1	agree with	1	9	8
2	agree that	2	8	6
3	agree to	3	7	7
4	agree on	4	4	3
5	agree or	5	2	1
6	agree among	6	1	1
7	agree mark	6	1	1
8	agree when	6	1	1

Additionally, the list ranked by frequency in the corpus also provides the opportunities for learners to acquire the (un)common collocating items, observing the authentic language using preference of the native-speakers. As shown in the frequency list, “with”, as the top preposition collocating with “agree” suggests that, to a certain extent, is more commonly used. Therefore, based on the corpus, this concordance example would be the beneficial assets for the improvement of “verb +adjective pattern” knowledge.

At the teacher-led end [22], the abundant authentic materials of Corpus make it an ideal sample for educators to choose. To ensure the learners’ sufficient mastering the key point of collocation by learner-led activities, teachers are recommended to make cloze tests by using the Key-Word-In-Context (KWIC). Exporting the concordance lines from the concordance tools, the cloze test could be easily handmade by deleting the key word while keeping the rest of the sentences. By applying the collocation knowledge learnt in the actual test, learners will consolidate their understanding further, and eventually, reaching the aim of data-driven learning.

Apart from certain collocation patterns of prepositions, language learning requires in-depth understanding of words on the semantic level, that is to say, it is often the case that some words habitually collocating with other words indicate the semantic prosody of words including positive, negative and neutral connotation [24]. For instance, Sinclair suggests that the phrase “set in” tends to co-occur with the words with unpleasant connotation such as “infection”, “rot” and “decay” displaying a negative prosody [25]. Therefore, based on Yiyin Parallel Corpus, learners could use the authentic material and concordance the words with positive, negative and neutral prosody to acquire the words semantically by identifying the words accompanying with the “KWIC”.

Table 3 illustrates sentences with the concordance word “terrorist” which possesses negative prosody. It is found that in the corpus “terrorist” accompanies the words such as “attack”, “shot dead”, “bombings” and “assassination” with negative meanings, implying that “terrorist” with negative prosody often appears with the accompany of negative words.

Table 3. The negative prosody of “terrorist”.

Interesting Fact: After the 9/11	terrorist	attacks on America,
Occasional	terrorist	attacks take place
A presidential assassination. A	terrorist	attack on the scale of 9/11.
Indonesian special forces shot dead an Islamic	terrorist	linked to the fatal July bombings of two hotels
the Taliban, he will be arrested and tried as a	terrorist	by a military tribunal.
Was a	terrorist	around the corner? Was there a firearm being pointed

4. Translation Teaching with Corpus

As a cross-cultural behavior, translation process is intrinsically rich in complexity for its function of transformation in languages as well as their cultures. Thus, translation teaching stresses the transformation of cultural connotation but preservation of the original meaning, so translators are required to have a good command of source and target languages [9]. In courses oriented to translation, large accumulation pertaining to materials of a wide range and techniques are frequently highlighted. Parallel corpora are exactly an ideal tool for application in translation research (Baker in 1993 initially conducted) and translation pedagogy [9].

4.1. Techniques & Terminologies of Translation Teaching with Parallel Corpora

Applying parallel corpus to providing pedagogical aid is a corpus-based approach, which can provide documentary evidence as reference and assistance for translation [5]. A parallel corpus is strictly aligned and translated by experts, which ensures the utility and value of teaching and researching.

Based on concordancer, here gives an example of how the Chinese word “提供” (means provide, offer sth.) is translated in English. The materials are extracted from news, editorials and comments sections. All of them are processed and retrieved by BSFU Paraconc 1.2.1 which allows the users to retrieve words of both Chinese and English.

From Figure 1, noticeably the meanings between “provide” and “提供” are the closest(53%). Second to “provide” are cases where the English context do not contain such conception (16%). Other frequently verbs are offer (11%), support (6%), deliver (3%), power (3%), and the only adjective is “available” (3%).

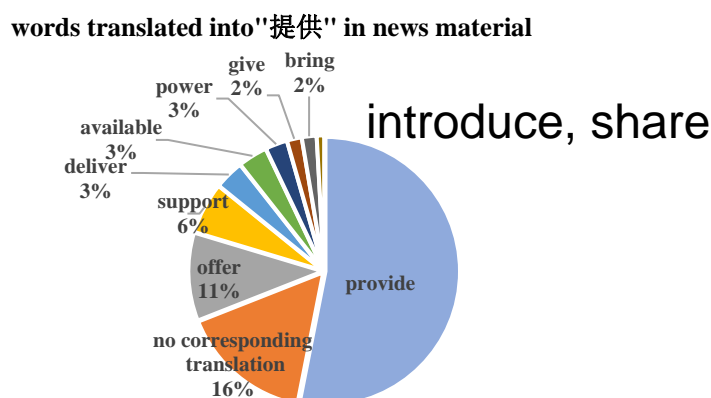


Figure 1: “提供” translated percentages (Photo credited: original).

However, Wang. argues that these abstract data provided to learners directly make no sense. He highlighted promoting the comprehension of examples behind the data [5], from which to recognize the laws of language from a holistic perspective. Thus, analysis on concrete cases is requisite.

In “provide” cases, most of them are “provide services, power for sth. / sb.”, while in some cases the provision action does not reflect in English text, the translator is required to add the word: The FDA declined to approve the product for this use, requesting additional information. 去年十月，美国食品药品监督管理局拒绝了 Geodon 的该用途，要求辉瑞公司提供更多的信息。

In this context, the object of request is Pfizer(omitted), a corporation. But when translated into Chinese, a gap between Pfizer and information must be bridged so as to conform to the habit of writing in Chinese. Thus, the conception “provision” must be reflected in Chinese texts to get readers across. Strategy like this is called “amplification”. While sometimes a Chinese text containing “提供” does not have the according word in English version; strategy like this is called “omission” [26].

Besides, key terminologies as foreignization and domestication feature the beginning of translation teaching, which were originally proposed by Laurence Venuti in his work *The Translator's Invisibility* in 1995, who argues that domestication is an ethnocentric reduction of the foreign text to language cultural values, bringing the author back home, while foreignization is an ethno deviant pressure on those values to register the linguistic and cultural difference of the foreign text, sending the reader abroad [27]. One fairly typical sort of foreignization is transliteration, transforming a word from one alphabet of language to another, so that people can pronounce words in foreign languages. Here is an example:

“我们接到一个电话.....

发现她在自己公寓外面的街上，没有受伤但她有些歇斯底里。”

"We've had a call...

We found her on the street outside her apartment. No one's hurt but she's a bit hysterical."

“歇斯底里” is a Chinese slang imitating the pronunciation of the English word “hysterical”, but the two have the identical meaning: unable to control one’s feelings or behavior because one is extremely frightened, angry, excited, etc. (from Cambridge Dictionary)

But the word “hysterical” can be translated as its original meaning. That’s domestication:

她解不开儿童椅上复杂的绑带，于是孩子哭得愈发厉害。

She could not figure out the car seat's elaborate buckle, so the child's crying became hysterical.

“哭得愈发厉害” can be literally understood as “cry even more” in English. It preserves the meaning of “hysterical” and intelligible to the readers.

In E-C translation, it’s typical of the translator to use a kind of set phrase in Chinese (usually composed of four characters) which is highly consistent in connotation, and sometimes use allusion. That’s notably domestication, like this:

他们都不承认，出轨的丈夫都会这样，但只要谈到这个话题，他们的脸上就会挂起尴尬的嬉笑，目光就会变得呆滞，似乎是在回忆曾经的巫山云雨。

They all denied it, like cheating husbands do, but, when the subject came up, little shit-eating grins would appear on their faces and their eyes would glaze over as if they were remembering the wild time they'd had. (p.s. “cheating” here means unfaithful.)

“巫山云雨” is a set phrase from myth in ancient Chinese, literally meaning “goddess in Mount Wu brings cloud to rain”. But now it represents the love affair of a couple tinged with mystery and romance. The application of it implies that their love was once beautiful but fragile. By using such phrase, both sentences are highly aligned in meaning. All of these cases abound in corpus especially in terms of novels.

(All the examples are from novel sections of Yiyan corpus)

4.2. Back Translation Approach

The word “back translation” was originally proposed in Newmark’s work *A Textbook of Translation* in 1988. In *Dictionary of Translation studies*, it is defined as a process in which a text which has been translated into a given language is re-translated into source language [28]. As a common practice in China’s translation pedagogy, the integration of corpus linguistics may strengthen the effectiveness of acquisition of translation.

Huang showed that there are 3 main uses of back translation, one of which is use for translation quality assessment and translation pedagogy [28]. Learners can spot inappropriate words and sentence structures they have used when they are compared with the teacher’s version. The amounts of discrepancies signify the extent to which translator may misunderstand the original texts.

In Huang’s view, translation teaching has to evolve from outcome-oriented to process-oriented [28]. Merely giving model answers to each sentences seems to be ineffectual, for learners can only know how to translate without knowing why to translate in that way (that is, translation techniques mentioned above). In the back translation process, the wording and phrasing, and cultural comprehension are improved. More importantly, learners with the awareness to process back translation will foster an autonomy for translation learning. As Zanettin argues, corpus-based translation activities can enhance learning autonomy of the students and the use of corpus resources affects the role of the teacher in such a context [10].

5. Reading with Corpus

This paper investigates a model of corpus-assisted reading teaching, i.e., presenting various key words in discourse from the bottom up through a corpus and assigning reading tasks to guide learners to read the text, get the gist of theme and develop their language competence and thinking patterns. Through observation, classification, and evaluation, learners are taught to comprehend the text's thematic significance, subtopic connotations, chapter organization, and even writing style, therefore enhancing reading proficiency [29].

5.1. Date-driven Based on Corpus in English Teaching

Unlike traditional paper-based reading, DDL, proposed by Tim Johns, used in reading corpus is a unique methodology in response to the changes in the age of big data, as well as aids in the development of learners' reading skills by highlighting important information or providing language frequency information [30].

DDL can provide learners with exposure to real-life contexts that can aid lexical and pattern awareness and can drive a type of cognitive process intensification through contextualizing [31]. The application of DDL to the teaching of reading in English, adopting the reading discourse as its point of departure using Unit1 of the compulsory second book of English (2019) on the PEP (People's Education Press) as an example, questions and solutions were picked to develop a lesson plan [30]. The teaching process is shown in Figure 2.

In the survey of the effectiveness of corpus-assisted reading, the vast majority of respondents said the corpus-assisted deep reading approach was useful for uncovering the content and logical progression of texts, as well as focusing on the form of language and understanding it more deeply in relation to the expression of meaning [29].

The corpus-assisted reading teaching allows learners to infer the primary concept of a text using keywords, which helps them acquire reading skills, and cultivates critical thinking, including observation, comparison, categorization, evaluation, and inference. It is crucial to achieve date-driven based on corpus in English teaching with the purpose of accomplishing contextual reproduction and increasing learners' interest in reading, as well as mastering the meaning of the high-frequency words

in the corpus. Additionally, teachers can explain contextual collocations and grammatical information of the corresponding vocabulary [32]. It is a way to shift from a monotonous approach to teaching and learning English, which means to increase learners' interest in reading and assist them to imitate writing styles from articles they prefer.

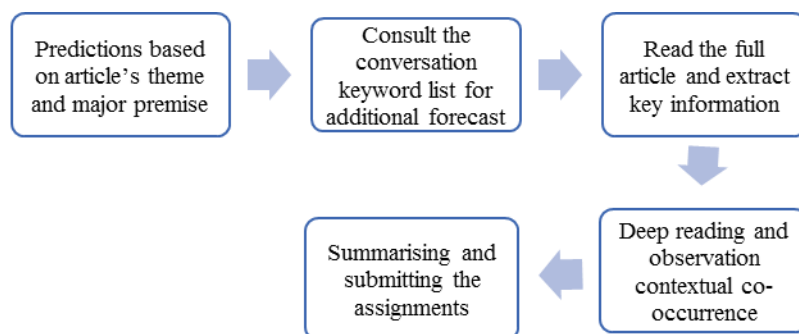


Figure 2: The teaching process.

5.2. Extensive Reading

Palmer H. E., a British linguist, introduced the concept of extensive reading in 1917. Extensive reading means reading in quantity and in order to gain a general understanding of what is read. It is intended to develop learners' good reading habits, buildup knowledge of vocabulary and structure, and encourage a liking for reading [33]. It is worth emphasizing that Richard Day and Julian Bamford's ten principles for an extensive reading approach to teaching reading address the nature of extensive reading as well as the conditions and methods required for its effectiveness. It mentioned the top ten principles for teaching extensive reading, including the reading material is easy, wide-ranging reading material must be offered, learners choose what they want to read, learners read as much as possible, reading for enjoyment, information, and comprehension, teachers orient and guide their students, etc. [34].

Through extensive reading, teachers change their identity, reading with learners and exchanging ideas with each other rather than assigning reading homework alone. In the course of implementing instruction, teachers must pay closer attention to learners' aptitude and motivation to select open reading tasks. They can help learners build corresponding corpora to improve their learning interest and reading experience and achieve the purpose of interest teaching.

5.3. Jigsaw Reading

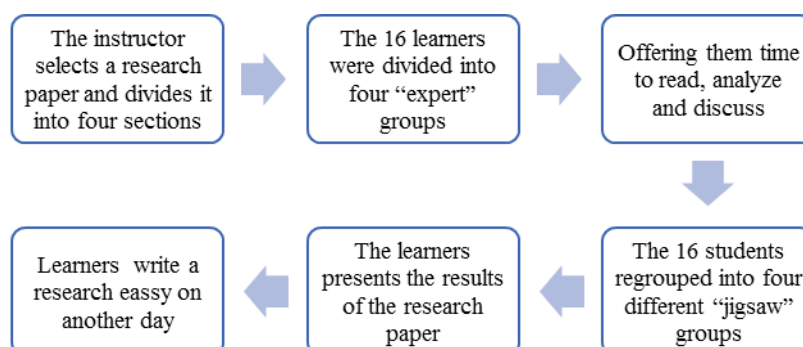


Figure 3: The jigsaw reading process.

Apart from the traditional paradigm of pedagogy about reading, the jigsaw reading methodology, often known as the "piecemeal approach," was established in the 1970s by the American social psychologist Elliot Aronson and his pupils. It divides reading tasks, learner groups, and individual tasks [35]. From the standpoint of educational ecology, the jigsaw is an ecological model of teaching and learning [36]. It lays greater emphasis on the multilateral interaction between learners and teachers, as well as between learners and reading resources, with teachers playing the roles of guide, collaborator, and facilitator.

When jigsaw reading was used in the classroom, there was a lesson plan in Figure 3 [37].

This study discovered that jigsaw reading is an engaging and motivating method for transforming learners' passive reading activities into active involvement, thereby meeting the teaching criteria of "less teaching, more learning [37]."

6. Conclusion

The bilingual parallel corpus offers a new way of thinking and a new strategy for EFL courses. This paper has presented an innovative approach to language teaching and learning, the "pedagogical approach in EFL courses based on the bilingual parallel corpus." We identified the effectiveness of corpus linguistics and divided it into several sections including vocabulary, high-frequency words, collocation, semantic prosody, translation, etc. The corpus provides a better understanding of the authentic language using the preference of native speakers. Also, we have combined corpus with reading instruction and applied it to the classroom with better results. The DDL facilitates learners to memorize vocabulary with greater interest and efficiency, and teachers are able to explain language through contextualization, which could be applied to study how to strike a balance between teachers and learners, and promote mutual learning and teaching. This approach has implications for the teaching of English in Chinese universities. In the future, in terms of both viability and operability, the creation, exploitation and popularization of a bilingual parallel corpus will be encouraging. Through the parallel bilingual corpus of word diversification, learners can get a comprehensive improvement in reading, writing and translation in the process of learning English. Additionally, it will also achieve the effect of injecting new vitality into Chinese English education. Pedagogical approach in EFL courses based on bilingual parallel corpus still has great potential for development and yet much more work remains to be done, which will be a great innovation and breakthrough in English education.

Acknowledgements

All authors contributed equally to this work and should be considered co-first authors.

References

- [1] He, & Peng. (2011). *A Survey of English Corpus Studies: Retrospect, Status Quo and Prospect*. *Foreign Language Education* (01), 6-10+15. doi:10.16362/j.cnki.cn 61-1023/h.2011.01.017.
- [2] Johns, T. (1991). *Should you be persuaded: Two samples of data-driven learning materials* (Vol. 4, pp. 1-16).
- [3] Gilquin, G., & Granger, S. (2010). *How can data-driven learning be used in language teaching*. *The Routledge handbook of corpus linguistics*, 359-370.
- [4] Wang, K. (2004). *The use of parallel corpora in translator training*. *Computer-assisted Foreign Language Education*, 6, 27-32.
- [5] Wang, K., & Qin, H. (2015). *The use of parallel corpora in the teaching of translation*. *Foreign Language Teaching and Research*, 47(5), 763-772.
- [6] Braun, S. (2005). *From pedagogically relevant corpora to authentic language learning contents*. *ReCALL*, 17(1), 47-64.
- [7] McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

- [8] Xu, Xiu., & Xu, J. (2021). *The construction of Yiyen English-Chinese Parallel Corpus*, [J]. *Corpus Linguistics*, This resource can be found in <http://corpus.bfsu.edu.cn/info/1082/1693.htm>.
- [9] Liu, K. (2020). *Corpus-Assisted Translation Teaching*. Springer Singapore.
- [10] Zanettin, F. (2009). *Corpus-based translation activities for language learners*. *The Interpreter and Translator Trainer*, 3(2), 209-224.
- [11] Breyer, Y. (2006) "My Concordancer: tailor-made software for language teachers and learners," in S. Braun, K. Kohn and J. Mukherjee (eds) *Corpus Technology and Language Pedagogy*, Frankfurt am Main: Peter Lang, pp. 157-76.
- [12] Jaefarpour, A.A & Koosha, M. (2006). *Data-Driven Learning and Teaching Collocation of Preposition*. *Research on Foreign Language*. *Journal of Faculty of Letters and Humanity*, 49, 200.
- [13] Okamoto, M. (2015). *Is corpus word frequency a good yardstick for selecting words to teach? Threshold levels for vocabulary selection*. *System*, 51, 1-10.
- [14] Leech, G. (2001). *The role of frequency in ELT: New corpus evidence brings a re-appraisal*. *Foreign Language Teaching and Research*, 33(5), 328-339.
- [15] Chen Xiuchun, (2022). *Corpus-based Approach to College English Vocabulary Teaching* (07), 68-72.
- [16] Shahzadi, A., Asghar, A., & Javed, S. (2019). *Effectiveness of corpus in teaching English synonyms*. *Corporum: Journal of Corpus Linguistics-CJCL*, 2(1), 51-65.
- [17] Liu. Ping, (2016). *Data-Driven ESP Vocabulary Learning*, 219-225.
- [18] Bahns, J., and Eldaw, M. (1993). *Should we teach EFL students collocations?* *System*, 21, 101-114.
- [19] Faghih, E. and Sharafi, M. (2006). *The Impact of Collocation on Iranian EFL learners Interlanguage*. *Science and Research Quarterly*, 16, 58.
- [20] Hill, J. (1999). *Collocational competence*. *English Teaching Professional*, 11, 3-7.
- [21] Flowerdew, L. (1999). *A corpus based-analysis of referential and pragmatic errors in students' writing*. Hong Kong University of Science and Technology.
- [22] Mukherjee, J. (2006) "Corpus linguistics and language pedagogy: the state of the art – and beyond," in S. Braun, K. Kohn and J. Mukherjee (eds) *Corpus Technology and Language Pedagogy*, Frankfurt am Main: Peter Lang, pp. 5-24
- [23] Zarei, A. (2002). *Patterns of Iranian advanced learners' problems with English collocations*. Unpublished doctoral dissertation, University of Isfahan, Iran.
- [24] Stubbs, M. (1995). *Collocations and semantic profiles: On the cause of the trouble with quantitative studies*. *Functions of language*, 2(1), 23-55.
- [25] Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.
- [26] Molina, L., & Hurtado Albir, A. (2002). *Translation techniques revisited: A dynamic and functionalist approach*. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 47(4), 498-512.
- [27] Yang, W. (2010). *Brief study on domestication and foreignization in translation*. *Journal of Language Teaching and Research*, 1(1), 77-80.
- [28] Huang. (2020). *Influence of Back Translation on College English Translation Teaching*. *Journal of Longdong University* (03), 125-128.
- [29] Xu, M., & He, A. (2016). *Corpus Technology for Teaching English as a Foreign Language in Depth Reading* - Kns.cnki.net.
- [30] Wang, Z., & Xia, H. (2022). *The application of corpus data-driven model in English reading teaching* - Kns.cnki.net.
- [31] O'Keeffe, A., & Mark, G. (2022). *TEMPORARY REMOVAL: Principled pattern curation to guide data-driven learning design*. *Applied Corpus Linguistics*, 100028.
- [32] Li, Y. (2021). *On the Application of Corpus Data-driven Learning Model in College English Vocabulary Teaching*. Kns.cnki.net.
- [33] Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics*. Longman.
- [34] Day, R., & Bamford, J. (2002). *Top Ten Principles for Teaching Extensive Reading*.
- [35] Zhou, J. (2018). *Design of Junior High School English Teaching based on Jigsaw Reading Model*. Kns.cnki.net.
- [36] Wang, A. (2018). *The practice of Jigsaw model in "Advanced English" teaching*. Kns.cnki.net.
- [37] Esnawy, S. (2016). *EFL/EAP Reading and Research Essay Writing Using Jigsaw*. *Procedia - Social and Behavioral Sciences*, 232, 98–101.