



## **Corpora and Language Learning: Ways to Use Corpora in Classroom**

Alif Ilham Baiquni  
Universitas Negeri Yogyakarta  
alifilham.2020@student.uny.ac.id

### *Abstract*

Corpus-based studies usually involve the comparison of two corpora, in which translated texts are compared with either their source texts or with another corpus constructed according to similar design criteria, either in the same or in another language. The ability of using corpora for language teaching has been questioned from different perspectives. The skeptics have expressed reservations about the ability of corpora to capture language use in providing a model for teaching. However, due to corpus linguistics is based on the theory that language varies according to context, its potential for finding out new facts about language is infinite. This includes designing them for teaching and learning in the classroom which can be applied indirectly and directly. In this paper, the writer proposes ways of using corpora in the classroom by considering how best corpora can aid language learning and teaching and how to deal with the theoretical and practical implications of the body of work dealing with corpora in the classroom.

**Keywords:** corpora, language learning, classroom.

### **Introduction**

In applying the genre-based approach to the teaching English as a foreign language in Indonesia as recommended in the curriculum, learners are to be equipped with adequate knowledge on the field and adequate understanding of the text structure and grammatical features of the target text as prerequisites for writing well the target text (Derewianka and Jones, 2016; Emilia and Hamied, 2015; Feez and Joice, 1998; Richards, 2006). The former prerequisite can be acquired through critically exploring the model texts through various activities such reading, viewing and listening to relevant. This can be an expansion and a deepening of reading comprehension. The latter can be gained through analyzing thoroughly the model texts in terms of text structure and grammatical features. This analytical activity might be very challenging for both the teacher and learners since the language features may be very detailed, involving the development of sensitivity to collocation phenomena and rhetorical consciousness. In answering such a challenge, the teacher can make use of corpus linguistic (CL) methods. Hyland (2008),

for example, stated that CL methods can help learners' understanding of the features of the target discourses.

However, in reflecting on the writer's own experiences in learning English through the text-based instruction the teachers of EFL has not made any use of CL methods. They tend to concentrate in involving students in superficial analysis and this seems inadequate to equip student with the second prerequisite. The reason might be that they have not gained any information on this field. This problem might be solved by providing information on the nature of corpora and proposing a way of using corpora in teaching EFL in this country.

In a given situation, it is known that one or the other method of English Language teaching (ELT) may be highly appropriate to a certain extent as far as the question of linguistic efficiency of the learners in the English language is concerned. Conventionally, in EFL classrooms worldwide, teachers use both implicit and explicit instruction. Teachers give students rules in explicit teaching to practice and make the students fully aware of what they learn (Talley and Hui-ling, 2014). While some methods have certain advantages in the development of listening and speaking skills of the learners, other methods have striking advantages in the areas of reading, writing, and communication, corpus-based had a hard time to be applied in the classroom. The utility of corpora for language teaching has been questioned from different perspectives. Widdowson, (1994) have expressed reservations about the ability of corpora to capture language use, while Dellar in Boulton (2013) argues that L1 corpora can intimidate learners or disempower teachers.

Nevertheless, using corpora in Language Learning is not impossible. The native-speaker corpora is useful in providing a model for teaching (Prodromou, 2003). Moreover, corpora and corpus evidence have not only been used in linguistic research but also in the teaching and learning of languages, probably a use that the compilers of corpora may not have foreseen (Johansson, 2007).

In universities, corpus-based research and teaching have been carried out predominantly. Therefore, many individuals in other educational settings including translation may think that corpora are not relevant to their teaching situation, nor that the knowledge, skills and technology required to integrate corpora into their teaching are beyond them. Johansson (2009) even makes the useful point that corpora should not be used in language teaching just because we now have this wonderful tool and would like to apply it in language teaching as well. This type of argument has been levelled at pedagogical applications of information and communications technology (ICT) as in general it seems axiomatic that practical developments should be theory and pedagogy driven rather than technology driven (Chambers 2005). However, there have been articles on how teachers with minimal computer resources can make use of corpora.

These corpus-based applications soon prepared the ground for the question of how to use corpus data and corpus-linguistic methods for the language teaching and learning purpose. Efforts to incorporate corpus into the actual language teaching seem missing and the awareness towards corpus and corpus tools is very low and, in some cases, there is a resistance toward corpora from learners and teachers. Many teachers of English are actually have been using the corpus-based language materials such as dictionaries and grammar books in their classrooms, however they did it without actually being aware of

this (Burnard and McEnery, 2007). This paper aims to show and propose how teachers may use corpora as teaching-learning material in the classroom. Conversely, the fact that corpus-based studies relevant to language learning concentrate on those issues of the utilization of corpora can offer insights may be misinterpreted as implying that corpora are the be all and end all of language teaching.

This paper is therefore aimed at exploring the advantages of CL methods, the nature of corpora, and possible uses of corpora in teaching EFL. To reach this aim this paper will cover the points of teaching EFL for mastering English communicative competence, the nature of corpora and its relation to communicative competence, making use of corpora in EFL teaching through the text-based instruction, and conclusions.

### **Communicative Competence to be realized in Texts**

The concept of communicative competence has been shifted many years due to its being formulated as well as reconstructed it into different perspectives by a number of scholars. In addition, it is created based on a series of experiments conducted by the scholars with various propositions and also coverage. Celce Murcia et.al (1995) formulated the final model of communicative competence into six namely linguistic, strategic, sociolinguistic/socio-cultural, formulaic, interactional and discourse competences.

Regarding the aims of Indonesian curriculum, it is closely in line with how to use English as a means of communication as well as instruction medium in the context of teaching and learning. This is due to English has been a strategic way of teachers as well as students in terms of communicating and transferring their knowledge to others. In addition, it deals with how the teachers, as well as the students, competence generate the outcomes of the communication. It is related to the communicative competence which must be possessed by both the teachers and the students respectively. In other words, English has become the basic principles in terms of the implementation of English pedagogy.

There are numbers of suggestion regarding the principles for the design and implementation of language courses using the model of communicative competence that aim at giving learners the knowledge and skills they need to be linguistically and culturally competent in a second or foreign language (Celce Murcia 2007). In fact, the students' competence does not represent the communicative competence which used for their academic purposes as well as communication in general.

### **Producing Texts Through the Text-Based Instruction**

The Text-based instruction is an approach which aims at developing learners who can engage meaningfully with different texts in different contexts. This approach is developed from Halliday's Functional Linguistics (Cahyono and Widiati, 2011). Mohlabi-Tlaka (2016) defines text-based instruction as the teaching of language built on the exposure to, and the handling and manipulation of different text types. It is an approach which helps learners be able to handle a variety of texts to help them use language effectively. Text-based instruction helps learners to comprehend not only language but also other non-linguistic texts.

Text-based teaching and learning pursues to develop communicative competence by mastering different types of texts in which, have different purposes and are organized in different ways with different language features. An example of this, is the note to a friend will use informal language while the note to a principal may employ a polite and formal language. It shows that to communicate means to be able to use different kind of spoken and written texts in the specific contexts of their use.

Text-based approach allows the teachers to use texts which must be tailored to class communities. The teaching objectives must correspond to the function and purpose of the speech community in which learners derive interest from learning what happens in their communities. This allows them to relate what occurs in their community to the classroom situation. Mikan (2011) asserts that the aim of teaching is to immerse students in experiences with texts in order to reflect the richness of learners' everyday experience of texts. This means, teachers have to select and sequence an appropriate texts for planned and direct instruction for a class to learn the texts for them to participate actively in the class community and the world at large.

To implement this approach, especially in building text, teachers need to go through some steps started with building the context. The focus at this step is on the contents of the text according to (Mkandawire, 2018). The teacher introduces the text by generating backgrounds so that learners would be able communicate in a language that seem convenient for better self-expression and sharing ideas and eventually, write the text using their own expression. The next step is modelling. In this step, the teacher showing the learners what to do, demonstrating and practicing the steps, as preparation for dependent and independent work. The teacher should aid learners build up their understanding about the purpose of the text and its general language features so that learners can gain enough background knowledge (Abbaszadeb, 2013). This stage helps learners to investigate the structural pattern and language features, comparing the model with other examples of the same text type.

Following this, the next step is the joint construction of the text. The focus is the illustration of the process of creating a text (Dirgeyasa, 2016). The teacher and the learners decide on a topic. They both write the text together using the mind map developed in the first stage. The learners must do something more practical in writing. Afterwards in the Independent construction of the Text, the learners choose a text that is related to the one that is discussed in the first and second stages. Since they have enough scaffolding on the topic, learners can now write independently or in pairs. They may consult their teacher to make a first draft. They will use knowledge from previous stages to present texts in listening, reading comprehension, speaking, oral presentations and writing tasks. Finally, learners will have to reflect on the texts learnt and compare the use of text types across different fields. Learners can also role play what would happen if the same text is used by people in different roles and relationships.

### **The nature of corpora and its possible contribution to text-based instruction**

Retrieved from RMS Iberia (2020) a corpus or text corpus is a language resource consisting of a large and structured set of texts. In linguistic field, these corpora are used to do statistical analysis and hypothesis testing, checking occurrences or validating linguistic rules within a specific language territory. Corpus can be monolingual corpus or

multilingual corpus. Each contain texts in a single language or text data in multiple languages respectively. The texts of a corpus is to contain are selected and stored in electronic format. Written texts, if they are not already in electronic form must be scanned, and spoken texts must be recorded and transcribed. The utility and flexibility of a corpus can be increased by adding line of coding for a computer to read and recognize.

Nowadays usually electronically stored and processed. In order to make the corpora more useful for doing linguistic research, they are often subjected to a process known as annotation. An example of annotating a corpus is part-of-speech tagging, or POS-tagging, in which information about each word's part of speech (Yoon and Hirvela, 2004). Tags or labels are attached to the words, phrases, sentences, paragraphs, sections, or to entire texts in the corpus. Information related to non-linguistic properties of the texts is referred to as mark-up that used to give information about the source of the text.

Corpora are the main knowledge base in corpus linguistics. One of notable areas of application include Language technology, natural language processing, computational linguistics. The analysis and processing of various types of corpora are also the subject work in computational linguistics, speech recognition and machine translation. Corpora and frequency lists derived from them are useful for language teaching. In line with this, Yoon and Hirvela also stated that corpora can be considered as a type of foreign language writing aid as the contextualized grammatical knowledge acquired by non-native language users through exposure to authentic texts. Due to this characteristic, CL method will allows learners to grasp the manner of sentence formation in the target language, enabling effective writing.

Atar and Erdem (2019) has listed the advantage of CL method. Firstly CL provides data for language studies by supplying huge amounts of examples about a specific structure and teachers can look up how a structure is used by looking at these occurrences. This way, corpus act as a decent tool for gathering data about language studies which provides huge amounts of data.

Secondly, in corpus based language studies, comments and evaluations about language use is not regarded as reliable evidence as they are totally subjective and it is difficult to test them empirically. In this way, empirical data is at the core of CL and unlike qualitative methods, it abstains from subjective comments or ideas in analysis. The data obtained from corpora is statistically analyzed and concrete evidence is used. Therefore CL depends on objective and empirical data and the data obtained from CL is scientific and it is open to objective verification.

The third advantage is that CL provides complete information about language structures. Due to CL use statistical software, frequencies and statistical information about certain structures in a study can be easily obtained. For example, when a language research about the use of quantifiers is undertaken, CL can supply frequencies of use and researchers can deduce some rules about these structures. In short, CL does not focus on single contexts, rather it gives an overall picture of a phenomenon. This enable the researchers to have a general understanding about the nature of a certain phenomenon. In this regard, CL is really successful at large scale analysis.

The final advantage of CL is that corpora include essential information about the data it offers. Most of the modern corpora include information about the socio-economic status, educational background, gender or place of birth of speakers. Moreover, corpora can provide data according to subject, field or genre. For instance, a researcher studying bilinguals can obtain data about bilingual speakers by looking at their background. Another example is that a sociolinguist studying sociolinguistic factors in language can obtain data by looking at the socio-economic status of the speakers in a corpus. This is another significant advantage of CL because researchers can get relevant data thanks to the information included about the speakers or subjects in a corpus.

In English language teaching, there is plenty of indirect corpus use, via dictionaries and course books. However, the writer argues that using direct corpus by teachers and even students are very possible if the dictionary or the course book does not provide a certain information needed. For instance, when the teacher want to find out what the word 'sacrilegious' means, they could start from a corpus but it will be slow and takes up their time. Hence, look it up in a dictionary is a sensible choice. However, if the teacher already know 'sacrilegious' and want to use it, but are not sure what verb to use it with, then the leading learners' dictionaries give little help. Utilizing Data Driven Learning (DDL), in which the writer will explain in few section, promptly shows that teacher could resume, restart, reopen, stall, drag on and even collapse (each item can be clicked), to see examples of the collocation in use. It can show them all the examples of a word in context. This collection of examples is called a concordance. By running their eye over the concordance, teachers can find all the meanings of the word, and phrases it is in.

This way, teachers can generate contents of the text faster than using dictionary or course book. Since the context are already shown, the teacher can also introduces the text using the backgrounds generated. This provide learners a language that seem convenient for better self-expression and sharing ideas. They would be able communicate and eventually, write the text using their own expression. This echoes Abbaszadeb's argument that the teacher should aid learners build up their understanding about the purpose of the text and its general language features so that learners can gain enough background knowledge. This condition opens another consideration. That is, student motivation. If a class is currently engaged with ocean, it would be nice for them to look at the English of 'ocean' and have an oceanic corpus. A corpus website such as The Sketch Engine (<https://www.sketchengine.co.uk>) has an instant corpus tool, where text on a topic is gathered from the web in a few minutes by a teacher or, as a class exercise, by the students which they could use the data for a mini research project.

### **Corpora and Language Learning**

In recent era, as the needs of communication continue to grow, the ability to sounds like a 'native speaker' especially in English will, in fact, enable us to acquire advanced knowledge from various fields of science and technology controlled and contributed by the world. English will empower those who master it to deliver real and useful information about their life, living, history, tradition, culture, wealth, heritage and treasures to the external world in the terms and mediums which the external world understands.

It will help the society to construct a better impression of their knowledge, command, and mastery in the fields of arts and humanities, science, and technology. Moreover, it will make them linguistically competent and confident in the act of exchanging all kinds of knowledge and information globally. In classroom context the use of corpora will enable the teacher to make decisions about syllabus design, listing of frequent words or expressions, concordance lines, and tagged texts/searches, provide examples for teachable moments, and using the language input for activities. The impact of corpora on the study of language has been likened to a 'corpus revolution' (Atkins and Rundell, 2009). While the analysis of learner corpus data has indirectly impacted on references and materials, it now also increasingly features directly in applications for the classroom. Thus, learner corpus data represents the connection between indirect and direct corpus use for language learning and the indirect contributions corpora have made to teaching are the periphery of corpus-aided language teaching.

### **Utilization of Corpora in Language Learning**

Corpora can be accessed widely nowadays using technological advances such as computers and the internet. It is now possible to access numbers of articles, books and journals on the topic published every year essential resources of linguistic description of language, especially language corpora, which are also playing an increasingly relevant role in language learning and teaching. Corpora have already become a crucial tool for the analysis of languages, and a dynamic relationship has thrived between corpora and language teaching. However, as the writer has stated in the previous point, the utilization of corpora for language teaching has been questioned from different perspectives. Pedrosa et. al (2013) has given the example of expressed reservations about the ability of corpora to capture language use (e.g., Widdowson, 1991), or the usefulness of native-speaker (L1) corpora in providing a model for teaching (e.g., Prodromou, 1997), some going so far as to argue that L1 corpora can intimidate learners (Gabbrielli, 1998) and some argue that the use of corpora disempower teachers (Dellar, in Boulton, 2013).

On the contrary, the fact that corpus-based studies relevant to language learning concentrate on those issues into which the use of corpora can offer insights may be misinterpreted as implying that corpora are the end of all language teaching. The aim of this paper, therefore, is to define the corpora place within language teaching as a whole.

### **Utilizing Corpora in Learning Process**

Eventhough the first corpora were relatively small it still considerably large because it consisted millions of words. As corpus is a large database of language, it should be bear in mind, that even now, there exist corpora consisting of hundreds of millions of words (Gavioli and Aston, 2001). Such corpora included, but not limited to the British National Corpus which consisted 100 million words and the Bank of English (COBUILD) at Birmingham University, UK with over 500 million words.

Corpus techniques have created new knowledge about the behavior of lexis, grammar, semantics, pragmatics, and textual features. Due to corpus linguistics is based on the theory that language varies according to context, its potential for finding out new facts about language is infinite. If this theoretical insight is practical to pedagogy, then the case for the use of corpora in teaching can becomes very powerful. This is due to the fact that

neither dictionary nor grammar is able to fully describe the language, the educationist. The materials designed for classroom practitioner or the learners themselves may play an important role in identifying regularities in the language which could not to be found in such texts. Another proven benefit of corpus-based approaches to the study of language is that analysis is based on empirical. The native-speaker intuition about language is often wrong, yet, it has many uses (Aston, 2001). These uses included in traditional descriptions not occur with any frequency in large general corpora and, on the other hand, many uses which occur in corpus data are not recognized in the descriptions. This means that teachers and learners have been being given inaccurate and incomplete descriptions of the language.

### **Adopting Corpora in Language Education**

There are not many methodological transfer for adapting CL research area to the applied field of language learning and teaching, and thus learners were presented with the same tools of corpora and analytical tasks that used by professional linguists. It is not surprising that the attention to the pedagogic transfer which is essential in the adaptation of methods and corpora in classroom-based research were still lacking. Even though the research confirms the usefulness of language corpora in the classroom, it should be said that most of the subjects taking part in such innovative experiences are adult university language learners with a wide selection of analytical skills at their disposal.

According to Romer (2011), there are two ways of applying corpora in language teaching: indirect and direct. In indirect application, the selection of language features, the order of which to teach them in, and how to teach them has been influenced by research based on large corpus collections has impacted their usage on pedagogical textbooks and syllabus design. In direct application, the application is more focused into teacher and learner (Bernardini, 2002). Instead of having to rely on the researcher as mediator and provider of corpus-based materials, language learners and teachers get their hands on corpora and concordance tools themselves and find out about language patterning and the behavior of words and phrases in an autonomous way.

#### ***Indirect Application***

New understandings gained from corpus analysis using computer or technology driven has had a significant impact on the discussion of what to teach. On syllabus design, corpora have been used extensively by a number of scholars who also used corpus data directly to look critically at existing teaching English as a foreign language or TEFL. These syllabuses and teaching materials provide more accurate descriptions of language use, in which. Hunston (2012) suggesting that the experience of using corpora should lead to rather different views of syllabus design. The type of syllabus in which discussed extensively by the particular researcher, is a lexical syllabus. According to Sinclair and Renouf in Moudraia (2001), a lexical syllabus would focus on the most common word forms in a language, the central patterns of usage, and the combinations which they usually form.

While the term may occasionally be misinterpreted to indicate a syllabus consisting solely of vocabulary items, a lexical syllabus actually covers all aspects of language, differing



from a conventional syllabus only in that the central concept of organization is lexis (Hunston, 2002).

In addition, corpora may provide frequency data, which may further change what is taught. For example, on the basis of a comparison of the frequencies of modal verbs, future time expressions and conditional clauses in native English corpora and their grading in textbooks used widely. Mindt in Kupske (2019) concludes that one problem with non-corpus-based syllabuses is that the order in which those items are taught in syllabuses are very often does not correspond to what one might expect from corpus data of spoken and written English. In his article, Kupske argues that teaching syllabuses should be based on empirical evidence rather than tradition and intuition, with frequency of usage as a guide to priority for teaching. While frequency is certainly not the only determinant of what to teach and in what order, it can indeed help to make learning more effective. For example, in the book authored by McCarthy, McCarten and Sandiford's innovative Touchstone book series (2006), which is based on the Cambridge International Corpus, aims to present the vocabulary, grammar and functions learners may encounter most often in real life.

The writer found an article by Kaszubski and Wojnowska (2003) which presented a corpus-driven computer program called TestBuilder, for building sentence-based ELT exercises. The program is able to process raw corpora of plain texts using another linked computer program that assigns the part-of-speech category to each word in the corpus automatically in real time. The annotated data is used in turn as input for test material selection. The writer found this to be a good resource since corpora have recently been used by major providers of test services for a number of purposes such as an archive of examination scripts; to develop test materials; to optimize test procedures; to improve the quality of test marking; to validate tests; and to standardize tests.

### ***Direct Application***

Parallel corpora is a collection of texts in one language that is aligned with translations into or from one or more other languages. It plays important role mostly in translation studies or comparable studies of language. The lines featuring both the target language and translations from another language are displayed when concordance lines are produced. However, they have only seen limited use in language learning contexts. There are two general beliefs about corpora that limit their use in language learning contexts (McErney and Xiao, 2007). The first belief is to assume that barriers such as lack of electricity or access to the internet are not issues. The first belief is to assume that barriers such as lack of electricity or access to the internet are not issues. To be able to use the said corpora, an access to electricity and internet connection is needed in order for the machine to works. The second believe is that even if the language were of the appropriate level, the tools for manipulating and searching corpora are not learner-friendly. Most of professional tools available are designed with researchers in mind. They have complex interface that are not designed for learner. Moreover, parallel corpora face some issues such as a lack of awareness and a lack of accessible tools for searching parallel corpora.

Although the classroom use of corpora and Data-Driven Learning (DDL) have seen a significant amount of research in recent years, little of the discussion has focused on

parallel corpora. Communication alone cannot lead to language learning and that attention to form a grammar is an essential component of language learning these days (Pincas, 2000; Ellis, 2005). The CL community will need to gain further insight into the role of different types of corpora in different types of learning contexts if we want to guarantee a widespread use of CL methods in language teaching and learning.

The interest of applied linguists in CL methods and their introduction in language education has been pioneered and carried out by many researchers such as Johns (1986) and Tribble (2008). However it is later discovered that CL method can be used to reveal language facts (Sinclair, 2004). In this paradigm, CL is an instrument in exploring language entities such as collocations, colligation patterns or semantic, not to mention more recent and elaborated accounts such as linear unit grammar (Sinclair and Mauranen, 2006) or lexical priming theory (Hoey, 2005).

Based on the findings above, CL is a pedagogically-relevant tool for the use of authentic corpus data to increase learners' sensitivity both on collocation related phenomena and language learning strategies. It shows that it is certainly possible to use language corpora and CL methodology in the language classroom especially for teaching English texts. This possibilities scenario has been researched by Pérez-Parades (2010) who takes the research tools and methodology of the CL research paradigm straightaway to the language classroom.

The possibilities of scenario that has been outlined by Pérez-Parades has proved a fruitful and innovative instrument to enrich the learning experiences of language students, who are empowered with new skills and input opportunities. However, the full integration of these new skills and input into language education presents important challenges. In this context of discussion, the writer would like to refer to mainstream language education as language teaching which addresses non-university language learners, the largest population of foreign language students in our modern societies. In doing so, the writer seek to expose one area where, the utilization of language corpora still lacking, especially the pedagogic corpora.

The influence of CL research methods and classroom applications in which CL method can be exploited in the classroom has been hinted by Pérez-Paredes & Alcaraz (2009). However, language corpora in the classroom need to consider the appropriateness of the tools that have been used in CL research. This appropriation is confirmed by the fact that classroom activities involving language corpora have been characterized by the use of research-oriented corpora, the reliance in concordance as a mediation tool between the corpus and the learner, accompanied by the absence of an effort to make software user interfaces more user-friendly for general users and, the very high cognitive demand which is put on learners when prompted to reading concordance lines, which involves the interpretation of centrality. If possible, the tagging of a corpus should be instructed by the teachers who will use it as a resource. This will result in a full integration of the different elements that are part of the curriculum can be achieved.

### ***The Proposed Way of Using Corpora in Classroom***

As corpora deals with language patterning and the behavior of words and phrases, it can show us the word frequency. For example, what are the top 500 words in this

language? How often does a certain word occur? It can also show us word clusters of common short expressions such as “can you give me...” or even as simple as “I’m sorry”. The writer proposes ways of using corpora in the classroom for making decisions about syllabus design, language testing, lists of frequent words or expression, concordance lines, and language input for activities.

These activities will give the learners exposure to non-textbook language pattern, access to a much larger language sample than classes can normally provide and answer questions about everyday usages like the questions of “do people really use/say this?”; “which vocabulary is commonly use?”; “is this word widely use in speech/writing?”

Language Testing can be achieved by using free applications available such as Google Form (<https://docs.google.com/forms/u/0/>) combined with Flubaroo (<http://www.flubaroo.com>). These are best suited for teachers who need to quickly launch testing and calculate results, and teachers who conduct surveys. The output in a Google Form can be used to create simple tests without installing any tools and create tests in collaboration with colleagues. Another free tool that teachers could utilize is QuizStar (<http://quizstar.4teachers.org/index.jsp>). The application offers to create a simple online classroom with login and password. It can be used to quickly track students’ results without a learning management system. With the output of an online quiz, the application best suited for creating simple drill-and-practice exercises and vocabulary quizzes.

In the making of syllabus, instead of relying on the intuitions about what grammar should be taught, research from corpora can tell us what grammatical structures and what vocabulary are most common in the language we want the learners to learn. We can also provide learners with online sources to find examples of each type of usage via DDL or Data Driven Learning which might help learners to evaluate frequency of usage types and pragmatic applications. There are some general corpora consisting of millions of words such as British National Corpus (BNC), COCA (The Corpus of Contemporary AmericanEnglish), Brown (2005), and ANC (American National Corpus) and such as Lextutor for studying vocabulary. The benefits of applying corpus linguistics by using these corpora is that it can provide an accurate description of language, and its new potentials for language structure and use, and have many applications in language learning. This could contributes to rendering learning a foreign language more effective since learners will be faced with real language.

Pérez-Parades (2010) have argued that utilizing CL within the framework of the feasibility scenario should be easy to install and distribute, easy to access and should integrate as many modes, such as text, audio, and video as required by the learning situation. This made possible by SACODEYL (<http://www.um.es/sacodeyl>), which adopts a small pedagogic corpus approach. Each of the seven corpora, namely English, French, German, Italian, Lithuanian, Rumanian and Spanish had video-recorded interviews of about 10 minutes each. To ensure thematic comparability, a common set of interview questions was used, covering a wide range of topics including personal information, home and family, present and past living routines, hobbies and interests, holidays, school and basically, any kinds of open discussion topics.

Pedagogical mediator in the language learning process of young Europeans, exploiting web multimedia resources to deliver learning experiences based on data driven,

constructivist approaches to language acquisition. SACODEYL is simultaneously different things. It acts as a collection of corpora and a web based system for the assisted compilation and open distribution of pedagogic corpora which can be created by users other than the SACODEYL hosts. SACODEYL has developed tools which are open source and freely available to the language learning and teaching community, including a multimodal transcription tool, an annotation tool, a multimodal search tool as well as integration with online repositories of meta information. Each of its corpus consists of orthographical interview transcripts in XML format. Each transcript is structured on the basis of short thematic sections and annotated with regard to pedagogically relevant characteristics, such as topic, grammatical and lexical properties, discourse markers and CEF level.

## **Conclusion**

The utility of corpora for language teaching has been questioned from different perspectives. The skeptics have expressed reservations about the ability of corpora to capture language use or the usefulness of native-speaker corpora in providing a model for teaching. However, due to corpus linguistics is based on the theory that language varies according to context, its potential for finding out new facts about language is infinite. This includes designing them for classroom practitioners or the learners themselves. It can be applied indirectly and directly. In indirect application, corpora can be used for syllabus and textbook design, while in direct application, corpora can be used for DDL.

The writer propose the ways of using corpora in classroom in the classroom for making decisions about syllabus design, language testing, lists of frequent words or expression, concordance lines, and language input for activities as these activities will give the learners exposure to non-textbook language pattern, access to a much larger language sample than classes can normally provide. In the writer's opinion, teachers can put a good use of free applications such as Google Form combined with Flubaroo, or QuizStar for creating a language test. The writer also proposed to provide learners with online sources to find examples of each type of usage via DDL or Data Driven Learning which might help learners to evaluate frequency of usage types and pragmatic applications by utilizing some general corpora such as British National Corpus (BNC), COCA (The Corpus of Contemporary AmericanEnglish), Brown, and ANC (American National Corpus), and Lextutor.

The DDL could also utilized by teachers to teach English texts. This made possible by SACODEYL which adopts a small pedagogic corpus approach including English. Since each transcript is structured on the basis of short thematic sections and annotated with regard to pedagogically relevant characteristics, the teachers can utilize this to help them build a text with easy installation and distribution. It is also easy to access and should integrate as many modes including audio and video.

Utilizing various corpora, teachers can generate contents of the text faster than using dictionary or course book. Moreover, since the context are already shown in those corpora, the teacher can introduce the text using the backgrounds generated. This will provide learners a language that convenient for better self-expression and sharing ideas in which students would be able communicate and eventually, write the text using their own expression.

## References

- Abbaszadeh, Z. (2013). Genre-based approach and second language syllabus design. *Procedia-Social and Behavioral Sciences*, 84, 1879-1884.
- Boulton, A., & Tyne, H. (2013). Corpus linguistics and data-driven learning: a critical overview. *Bulletin suisse de Linguistique appliquée*, 97, 97-118.
- McEnery, A. M., & Xiao, R. Z. (2007). Parallel and comparable corpora: What are they up to? In G. James, & G. Anderman (Eds.), *Incorporating Corpora: Translation and the Linguist* (Translating Europe). Multilingual Matters.
- Atar, C., & Erdem, C. (2019). The advantages and disadvantages of corpus linguistics and conversation analysis in second language studies. Proceedings of IX Scientific and Practical Internet Conference of Young Scientists and Students.
- Atkins, B. T. S., & Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford University Press.
- Bernardini, Silvia. (2002). Exploring new directions for discovery learning. In Bernhard Kettemann and Georg Marko(eds.), teaching and learning by doing corpus analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19–24 July, 2000. Rodopi, 165–182.
- Brown, M. (2015). *Using Parallel Corpora for Language Learning*. Chiba.
- Cahyono, B. Y., & Widiati, U. (2011). *The teaching of English as a foreign language in Indonesia*. State University of Malang Press.
- Casas-Pedrosa, A. V., Domínguez, J. F., & Sintes, A. A. (2013). Introduction: the use of corpora for language teaching and learning. *Research in Corpus Linguistics*, 1-5.
- Celce-Murcia M, D. Z. (1995). A pedagogical framework for communicative competence: A Pedagogically motivated model with content specifications. *Applied Linguistics* , 5-35.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied linguistics*, 6(2), 5-35. DOI: 10.5070/L462005216
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language learning & technology*, 9(2), 111-125. DOI: 10125/44022
- Dirgeyasa, I. W. (2016). Genre-based approach: What and how to teach and to learn writing. *English Language Teaching*, 9(9), 45-51. DOI: :10.5539/elt.v9n9p45
- Ellis, R. (2005). Principles of instructed language learning. *The Asian EFL Journal*, 7 (3), 209-224. DOI:10.1016/j.system.2004.12.006
- Emilia, E., & Hamied, F. A. (2015). Systemic functional linguistic genre pedagogy (SFL GP) in a tertiary EFL writing context in Indonesia. *TEFLIN journal*, 26(2), 155-182. DOI:10.15639/teflinjournal.v26i2/155-182

- Feez, S., & Joyce, H. D. S. (1998). *Text-based syllabus design*. National Centre for English Language Teaching and Research, Macquarie University.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.
- Hunston, S. (2012). *Corpora in Applied Linguistics*. Cambridge University Press.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, 27(1), 4-21. DOI:10.1016/j.esp.2007.06.001
- Johansson, S. (2007). Using corpora: from learning to research. In *Corpora in the foreign language classroom* (pp. 17-28). Brill.
- Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. *Corpora and language teaching*, 33-44. DOI: 10.1075/scl.33.05joh
- Sinclair, J. (2004). *How to use corpora in language teaching*. John Benjamins.
- Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, 14(2), 151-162. DOI:10.1016/0346-251X(86)90004-7
- Jones, P., & Derewianka, B. (2016). *Teaching Language in Context*. Oxford University Press.
- Kupske, F. F. (2020). The impact of language attrition on language teaching: the dynamics of Kupske, F.F. (2019). The impact of language attrition on language teaching: the dynamics of linguistic knowledge retention and maintenance in multilingualism. *Ilha do Desterro: A Journal of English Language, Literatures in English and Cultural Studies*, 72, 311-330. DOI:10.5007/2175-8026.2019v72n3p311
- Gavioli, L., & Aston, G. (2001). Enriching reality: language corpora in language pedagogy. *ELT journal*, 55(3), 238-246. DOI:10.1093/ELT/55.3.238
- Bumard, L., & McEnery, T. (2000). *Rethinking language pedagogy from a corpus perspective*. Bern: Peter Lang.
- Mickan, P. (2011). *Text-based teaching: Theory and practice. Monograph on foreign language education*. Tokushima University.
- Mkandawire, S. B. (2018). Literacy Versus Language: Exploring their Similarities and Differences. *Journal of Lexicography and Terminology*, 2(1), 37-56.
- Mohlabi-Tlaka, N. (2016). *The contribution of a text-based approach to English education for communicative competence* (Doctoral dissertation, University of Pretoria).
- Kaszubski, P., & Wojnowska, A. (2003). Corpus-informed exercises for learners of English: the Test Builder program. *Research and Scholarship in Integration Processes: Poland-USA-EU*, 337-354.
- Pérez-Paredes, P. (2010). Corpus linguistics and language education in perspective: Appropriation and the possibilities scenario. *Corpus linguistics in language teaching*. Bern: Peter Lang, 53-73.

Pérez-Paredes, P., & Alcaraz-Calero, J. M. (2009). Developing annotation solutions for online data driven learning. *ReCALL*, 21(1), 55-75. DOI:10.1017/S0958344009000093

Pincas, A. (2000). New literacies and future educational culture. *ALT-J*, 8(2), 69-79. DOI: 10.1080/0968776000080208

*Preliminary Recommendations on Corpus Typology*. (n.d.). Retrieved from EAGLES: <http://www.ilc.cnr.it/EAGLES96/corpusstyp/corpusstyp.html>

Prodromou, L. (2003). In search of the successful user of English How a corpus of non-native speaker language could impact on EFL teaching. *Modern English Teacher*, 12(2), 5-15.

Richards, J. C. (2006). *Communicative Language Teaching Today*. Cambridge University Press.

Romer, U. (2011). Corpus Research Applications in Second Language Teaching. *Annual Review of Applied Linguistics*, 205-255. DOI: 10.1017/S0267190511000055

Sinclair, J., & Mauranen, A. (2006). *Linear Unit Grammar. Integrating speech and writing*. John Benjamins.

Talley, P. C., & Hui-Ling, T. (2014). Implicit and explicit teaching of English speaking in the EFL classroom. *International Journal of Humanities and Social Science*, 4(6), 38-45.

Tribble, C. (2008). From Corpus to Classroom: Language Use and Language Teaching. *ELT Journal*, 213-216. DOI: DOI:10.1093/elt/ccn005

Price, D. (2018, March 23). *Laziness does not exist*. Medium. <https://humanparts.medium.com/laziness-does-not-exist-3af27e312d01>

*What is corpora*. (2020, December 31). Retrieved from RMS Iberia: <https://rmsiberia.com/kari-wiki-vne/aa7652-what-is-corpora>

Widdowson, H. (1994). Aspects of Language Teaching. *International Journal of Lexicography*, 105-107.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of second language writing*, 13(4), 257-283. DOI:10.1016/j.jslw.2004.06.002