

Corpus for Classrooms: Ideas for Material Design

Neveen Al Saeed & Salma Waly

Using corpora in language classrooms has proven to be an effective tool in teaching vocabulary, grammar and language use to learners of English as a second/foreign language. However, many EFL teachers find integrating corpus-based activities in their classrooms a challenging teaching practice. Moreover, some teachers avoid using corpora because they think that designing corpus-based activities is a challenging task. In this paper, the researchers will guide the readers on how corpora can help students understand how certain language forms are used by native speakers of English. They will give a brief review on studies discussing the benefits and the challenges of using corpora in language learning. The role of the teacher is also highlighted in this paper. The researchers will then go through some corpus-based materials and activities suitable for different language levels. Activities will aim at developing students' vocabulary, grammar, knowledge of idiomatic expressions and pragmatic constraints. All activities discussed in this paper could be used in any language classroom even those classrooms that are not equipped with computers.

Introduction

Technology and computers have become one of the main aspects of human life. No one can deny the fact that technology has a great effect on the way people lead their lives. However, according to Breyer (2008), teaching is one of the areas where technology has not had a strong impact. Corpus linguistics is one of the technology-based tools that could be very useful in teaching but still has not been widely used or tested. Nevertheless, in the last 30 years, the use of corpora in classrooms has started to develop (Varley, 2008).

A lot of research has been conducted to examine the effectiveness of using corpus linguistics as a teaching technique to highlight how native speakers of English use certain language forms, vocabulary items, and expressions. Maddalena (2001) explains that the use of authentic and real-life examples with L2 learners is more beneficial than examples that are made up by the teacher and do not simulate real life use of language. Another benefit of using corpora for language teaching is that it gives students the chance to actually work with language and analyze it.

In this paper, the two researchers aim at highlighting several ways through which corpora can be integrated in language classrooms to teach students of various levels. The first part of the paper is a brief review of literature that mainly aims at showing how other researchers went about using corpora in their language classrooms and how effective that was. The researchers also point out other elements that affect the use of corpus such as the role of the teacher and the challenges that face both students and teachers when using corpora. As for the second part of the paper, the researchers present a number of ideas for planning corpus-based activities. Those activities are tailored so that they meet the different language abilities of students in L2 classrooms. The researchers will demonstrate how vocabulary, grammar, idiomatic expressions and pragmatic constraints could be covered in EFL classrooms using corpus-based resources.

Literature review

Although corpora have started to be used in language classrooms, Varley (2008) argues that more research needs to be done to provide evidence that corpora are useful for language learning. According to the researcher, many studies have been conducted to explore the different types of corpus-based

activities. What is new in Varley's study is that she takes into consideration the students and the benefit they get from using corpora.

To fill this research gap, Varley, first, carried out a study on 19 EAL students to find out if corpora help develop students' language proficiency. She investigated how corpora improve students' vocabulary and grammar. In addition, she examined the attitudes of students towards using corpora in language learning. In other words, Varley explored whether they prefer to use corpora or traditional tools like dictionaries and grammar books. Lastly, the researcher pointed out the problems that might face students while using corpora.

During the 14 weeks of the study, participants attended 4 hours of instruction each week. Two hours were dedicated to introducing students to the theoretical background of corpora, research done on the use of corpora in classrooms, and other readings that raised the awareness of the students to the use of corpora. As for the other two hours, students spent them in computer labs looking at corpora, learning how to use them and how to search for certain language structures. At the beginning, they were given guided activities to search for a certain problematic language area. Towards the end of the study, however, participants were asked to submit a project where they were given the freedom to choose any language feature, search for it in the corpus and write a paper about the findings.

To collect data, Varley chose a qualitative design. She administered a questionnaire at the beginning of the study to find out whether they know what corpora are, and whether they like to learn languages through CALL or through the traditional tools. The researcher also asked the students to submit a project at the end of the study, where participants applied everything they learned about corpora. Attached to this project, each student was asked to fill out a log. In the log, students reflected on their experience of using a corpus, the strategies they used to analyze their findings and how corpora added to their knowledge about the language feature they were researching. These projects were then analyzed to highlight the benefit of each student. Last but not least, Varley chose 5 students and had an in-depth discussion with them. The discussion aimed at exploring students' attitudes more.

Findings revealed that most of the participants agreed that corpora are very beneficial for their language learning and they would depend on them for future reference. Varley found out that corpora helped students to learn about certain language uses that were not available in any of the traditional tools. In addition, corpora allowed students to examine these language features in context. Findings also suggested that even students who like to use dictionaries and grammar books as references benefited from corpora and used them as a 'complement' to these traditional tools. Another benefit was the fact that corpora helped students get a broader view of language (i.e. corpora could compare spoken and written languages).

On the other hand, corpora have their disadvantages as well as their benefits. One disadvantage is that some corpus software is not available for students. That is why the researcher recommended that teachers introduce their students to free corpora online. Another obstacle facing students is that when searching any corpus, students are confused because of the large amounts of search results. Therefore, Varley suggested that teachers need to simplify these findings for those students with limited experience using corpora.

Research has proven that corpora could help improve students' language proficiency (Varley, 2008). However, because corpora is a relatively new trend in teaching, one can still find teachers who are skeptical with regards to corpora and whether they could be used to teach challenging language features such as modals and idioms. Therefore, Neff et al. (2001) and Simpson & Mendis (2003) do not

focus on using corpora for planning activities and tasks in their studies. They, instead, highlight the usefulness of using corpus analysis to get ideas about the problems that non-native speakers of English encounter while writing or studying in English.

Neff et al. (2001) investigated the differences in the use of modal verbs between Spanish learners of English and native speakers of English. They examined a corpus of 400,000 words that consisted of argumentative essays written by both groups. The researchers aimed at looking at how native and nonnative speakers use modal verbs like *can*, *could*, *may*, *might* and *must*. They also focused on the reasons why L2 learners over-use or under-use those modal verbs. Analysis of the argumentative essays led to a number of results. Spanish learners seemed to use *can* and *must* more than any other modal verbs. *Can* and *must* tended to occur with the pronoun *we* more than any other pronoun. Students seemed to under-use *might* and *may* and over-use *can*. One reason for such a case was that *can* was one of the first modals that those Spanish learners were taught. Another reason was that non-native students assumed that *can* could be used anywhere the Spanish modal *poder* could be used. The use of *we must*, on the other hand, reflected the way writer stance was expressed in Spanish.

The results of the study imply that many of the problems that English language learners have when acquiring English modal verbs can be attributed to discourse differences between their L1 and English. Such problems are challenging since teachers should find ways to overcome language transfer errors.

Simpson and Mendis (2003), on the other hand, investigated the usefulness of corpora in acquiring idioms. They wanted to explore whether idioms appeared in corpora or not. They examined a 1.7 million word corpus that included academic spoken English. They found that idioms are found in corpora and that corpora could be used to design teaching materials. Researchers also argued that corpora provide students with information about the pragmatic use of idioms. Moreover, corpora provide students with authentic contexts, which means that students learn language in context and not in isolation.

Although the previously reviewed studies agreed that corpora are effective in language teaching, other studies highlighted that there are certain challenges that face teachers when using corpora. Teachers should keep these challenges in mind because they could affect the effectiveness of corpora. For example, Maddalena (2001) attempted to use a well-established corpus in order to teach his students the differences between seemingly synonymous words in English. The researcher was asked by one of his Japanese students to explain the difference between *ruin* and *destroy*. He realized that he could not fully explain the difference to her because, rather than explaining the difference in meaning, he wanted to show how the two words are used to communicate ideas.

The researcher with the help of another EFL teacher started the experiment with six year three Japanese high school students who needed to pass a respected standardized test in Japan that qualifies students for good social, educational and professional positions. The researcher decided to use corpus concordance lines as a way to expose non-native speakers of English to how native speakers contextualize and use the problematic words. The choice of which corpus to use was not very easy as the researcher was looking for a corpus of both manageable size and enough instances with the target words. The researcher also wanted to use a corpus that was not very technical so that it would be interesting for the students. The Lancaster-Oslo-Bergen or the LOB was finally chosen to use with the students. The experiment consisted of five stages. During the first stage students were involved in a simple exercise using a dictionary. Then, they were asked to examine concordance lines and see how the words are used. After that, they were given a cloze exercise. At stage four, they were asked to produce sentences using these words. Finally, they had to formulate rules of using the target words.

The results of the study showed that using concordance lines was not very useful for the students for a number of reasons. First, students were not used to such a teaching methodologies in which they are given something to look at and then to report on. Second, the main goal for the students was to pass the exam. Knowledge of word meanings was more important for them than how to use words. This article shows that the role of the instructor or the researcher is crucial when it comes to using corpus-based activities. In some conditions, teachers can use concordance lines to extract real life instances and use them with the students rather than ask students to examine pages of sentences. If this is to be done, it should be done with more proficient and advanced learners who are interested in how language works.

Breyer (2008) supported Maddalena's view that the teacher has a crucial role in using corpus in classrooms. She believed that in order for the students to benefit from corpora, teachers should be properly trained on how to use corpora first. According to Breyer, the teacher plays a very important role in using corpora to teach languages. However, this role has not been researched enough. Therefore, in her study, Breyer first investigates the role of the teacher in the use of corpora, and the challenges that face the teachers. She concludes that teachers have to be trained to use corpora in their classrooms. The researcher then reports on a case study that she did to support her view that teacher training is essential.

One of the reasons why teachers find using corpora challenging is that in order to use corpora in teaching, they have to know how to use it very well. Moreover, teachers have to develop materials and activities using corpora because there are no 'ready-made' ones. They also need to make the activities suitable for their students' proficiency level and culture. In addition, teachers need to learn how to shift their role in the class to be monitors of the learning process. As a result of all the above mentioned challenges, Breyer came to the conclusion that teachers need to be trained on how to use corpora in their classrooms. Therefore, she carried out a study on 18 student teachers in Germany. The researcher designed a training course where she introduced them to corpora, how to use them, ideas for using them in teaching, and how to develop corpus-based materials.

Teachers were asked to write a teaching diary, fill in a questionnaire, and create materials and activities. According to findings, teachers need to understand the nature of corpora, and how to manage this learner-centered activity. This could be done by training teachers at the beginning of their career. Teachers also pointed out that it was very challenging to design activities. Therefore, the researcher reached the conclusion that some ready-made tasks are needed. This will be discussed in the second part of the paper. Like Simpson and Mendis (2003), Breyer highlighted that corpora provide both teachers and students the language in context. This will help students discuss language features more deeply in class. Beyer also argued that after the training course, the teachers not only became familiar with using corpora, they became more aware of the language they teach and how to teach it.

Unlike the previous studies, Fuentes and Rokowski (2002) had a different context for using corpora to teach English. The researcher proposed a model for integrating corpus-based exercised in ESP classrooms. Although this is just a model of syllabus-design, it is good to include it in this review because it gives a number of suggestions for tasks that could be used in ESP classes. He mainly focused the Business and Information Technology which integrates ESP skills with EAP or English for Academic Purposes skills. The use of corpus in this class basically revolves around the idea that students need to build some kind of language competence through active task completion so that they use such knowledge to fulfill the requirements of the tasks of the course. The tasks of the course include a number of writing tasks and presentations. For such tasks, students need to build a repertoire of technical terms and possible collocations. A small corpus with a limited number of instances is recommended to be used as it will be focused on the academic fields that the students will be learning.

Therefore, the researcher thinks that the corpus of Business and Information Technology or the BIT could be used. Some of the activities that could be used to help students become aware of certain collocations is concordance *analysis*. Students can explore a number of words that are frequent in the field of IT. Then, they can explore the collocations of such words. Another idea is to have students compare instances in the BIT corpus with instances in other corpora that are either specialized ones like the IST (Information Science and Technology Corpus) or general like the BNC (the British National Corpus). Students can then see how such words are used technically and generally by native speakers of English. By doing so, students become aware of lexical chunks that are useful when it comes to completing course tasks. After students are given the chance to explore the different corpora, worksheets can be designed by the teacher to guide them. They could be given worksheets which they have to complete by finding words that collocate with others. They could also be given a number of chunks and then asked to match them with text types or writing genres.

The researcher gives ideas of good corpus-based activities that could be used with high intermediate and advanced students. Like Breyer (2008), this study proves that it is the role of the teacher to align the level of activities with the ability level of the EFL students he/she is teaching.

To sum up, it becomes clear that corpora have a very big potential in language classrooms. It has been proven that corpus is beneficial but there are certain challenges that any teacher needs to keep in mind before using corpora. Both teachers and students need to be properly trained on how to use corpora. It is the role of the teacher to simplify corpus findings at the beginning until students get a better understanding of how to screen their corpus findings and analyze them. Otherwise, the process will become very confusing (Varley, 2008). One of the challenges that have been discussed in the previous review is the fact that designing corpus-based activities is difficult. In addition, it is very beneficial that teachers have ideas about these activities. Therefore, the second part of this paper will focus on how to use corpora to teach different language features.

Corpus-based Activities and Ideas for Material Design

Elementary students: Collocations: make/do

Verb pairing game

The verb pairing game can be done after the teacher highlights some collocations to the students. After that, students will be divided into two groups. Group 1 will have some corpus-derived sentences without the verbs *make* and *do*. Group 2 will have the verbs *make* or *do*. Students in both groups will stand in lines facing each other. The ones carrying verbs will take turns trying to pair their verbs with the sentences the other group has.

Computer cloze activity

To integrate the use of technology in our classes, computer cloze activities might be used to help students practice the uses of both verbs and to learn their collocations. The teacher might prepare a passage or a number of sentences on Microsoft Word with every *do* and *make* being deleted. Students should complete the passages and then print them out for the teacher. Another way is to use peer-evaluation first before answers are discussed in class.

Examples:

1. The little boy, Zohair, was sent to _____ coffee.

Proceedings of the 10th METU ELT Convention

2. I remember coming back from school and before you could _____ homework or go out to play there were always chores to do.
3. He is going to _____ decision based on the recommendations of his field commanders.
4. I just want to be her mommy and I want to _____ her hair.

Screen shot analysis

After introducing the idea of corpus in class, a data show could be used to illustrate how frequent nouns occur with certain verbs and not others.

Intermediate students: Modal verbs

Contextual Analysis

Students will be given a number of sentences obtained from any of the two corpora used in this analysis. They should try to analyze the context of each to see why one modal verb is used rather than the other. For example, students could be given the following sentence and asked why *have to* is used instead of *must*:

In New Mexico, for example, you ***have to*** renew your license every year after age 75.

Students will be expected to explain that *have to* is used because the obligation comes from outside. The context indicates that this is the law in New Mexico. This activity can be done in pairs or small groups.

Examples:

1. You **should** be out there doing something for somebody else instead of just feeling sorry for yourself---
2. You **must** be angry at him. Laura, what are you saying to people now?
3. I **have to** go to Pittsburgh on Wednesday. I will be admitted on Monday and Wednesday----
4. You **should** be getting the call for proposals soon.

Cloze activities

Students will be given a worksheet with some sentences from a certain corpus. Modal verbs will be deleted. They have to fill in the blanks with the modal verb they think is most appropriate to use. After they finish, they will compare their answers with a partner. The teacher will then display the correct answers to the whole class and they can all discuss why certain modal verbs are used.

You _____ take better care of your teeth!
The correct answer is *should* because it is a piece of advice.

I _____ take a break now. I am very tired.
The correct answer is *have got to* because the speaker wants to take immediate action.

Examples:

1. I _____ admit, I was shocked. I didn't expect it.
2. I think you _____ give me some credit for that.
3. I _____ confess, when we started this program I thought of discussing common ground at the --

Spot the error activity

Students will be given a worksheet that includes a number of sentences with wrong modal verbs. Students will work in pairs trying to spot the error and decide on the best modal verb to use in each sentence. Results can be either discussed as a whole class activity or in pairs.

If you are practicing without a license, you **should** pay a \$500 fine or serve 30 days.

The correct answer in this case is *have to* because it is an obligation that comes from outside.

Examples:

1. I **have got** to take a quick break. We will continue here right after this.
2. Susan **must** be in New York by now.

Multiple-choices

Students will be given sentences obtained from a certain corpus. For each sentence, 3 choices will be provided as explanations of the meaning of the sentence.

You **should** plan to go to Houston. It would be a lot of fun.

- a. The speaker and the listener are probably friends. ✓
- b. The speaker is the listener's boss. X
- c. The speaker and the listener barely know each other. X

Examples:

1. I **must** interfere because of course I'm, in the opinion of all our people, the one responsible for all of this.
 - a. The speaker is talking to friends.
 - b. The speaker seems to be in a formal meeting.
 - c. The speaker is giving advice.
2. Well, I **have got to** get to the dentist before he closes.
 - a. The speaker will go immediately.
 - b. The speaker is not in a hurry.
 - c. The speaker is planning to go sometime soon.

Express yourself

A number of sentences that express suggestion/obligation will be written on small folded pieces of paper. The teacher will then write a situation on the board. Students will randomly draw pieces of paper and read the sentences on them. Some sentences will be relevant to the situation on the board and some will not. This is a fun activity that exposes students to a large number of sentences using modal verbs. Students can negotiate the meaning of the sentences and decide which can be used and which cannot.

Proceedings of the 10th METU ELT Convention

Situation on the board: give your sick friend advice.

Sample sentences:

You have to attend the meeting at 7:00 (irrelevant) X

You should take a break. (Advice)✓

You must see a doctor. (Strong advice)✓

You should go ask the superintendent of the cemetery. (irrelevant/funny)X

Upper intermediate students: Phrasal verbs

Cloze activity

Students will be given a worksheet with some sentences from the corpus. Phrasal verbs will be deleted. They have to fill in the blanks with the family-related phrasal verb that is suitable to the text. After they finish, they will compare their answers with a partner. The teacher will then display the correct answers to the whole class and they can all discuss why certain phrasal verbs are used.

Example:

Fill in the gaps with the correct form of the phrasal verbs below.

1. The little girl said that when she _____, she wanted to be a flight attendant.
2. Many parents don't like to punish their children by hitting them; they prefer to _____ the children _____ instead.
3. He was _____ by his aunt and uncle because his parents died in a car accident.
4. I _____ very well with my sister - we spend a lot of time together.
5. Everyone says that she _____ her mother - they both have beautiful eyes and an aptitude for science.

bring up / grow up / take after / tell off / get on

Advanced students: Idiomatic Expressions:

Concordance Analysis

After training students on how to use online corpora, you can ask them to search for the idiomatic expressions and analyze the concordance lines to see how these idioms are contextualized and used by native speakers.

Following are some idiomatic expressions with the verb *break*.

1. **Break** a leg meaning to wish good luck
With luck, I will break a leg or something.
2. **Break** bread meaning to eat together
Joshua Morris had not invited him to break bread in Clungunford-----
3. **Break** the mould meaning to do something in different way
This seemed a refreshing attempt to break the mould of British politics.

4. **Break** a heart meaning to emotionally abandon somebody.
That would break his heart.
5. **Break** the bank meaning to cost a lot.
Splash out on a luxury meal that won't break the bank at Tesco.
6. **Break** the ice meaning to try to get through difficulties especially when talking to someone for the first time
To break the ice, blow the candle out.
7. **Break** free meaning to get away of tensions, fears, and difficulties.
Panic gave her the strength to break free.
8. **Break** new ground meaning to make something new.
It is difficult to break new ground in this field.

Pragmatics: Situation analysis

This activity could be done after students are introduced to the different uses and meanings of apologies. For this activity, students will be divided into groups. Each group will receive corpus-based dialogues on flash cards, and students will decide the meaning and the level of sincerity of the apology.

Examples of corpus-based dialogues:

Determine the meaning of the apology used in the following dialogues:

1. We apologize for the delay, which has been caused by the traffic jam.
2. I am sorry but you will have to leave.
3. I am sorry, the table is reserved.
4. Mr. Ahmed: Go and get me the papers, Matthew.
Matthew: I'm sorry, are you talking to me?
5. Manager: We regret the delay, naturally.
Lucy: Is that all you have to say?
6. George: Sorry, you failed your exam.
Barbra: You should be too.
7. Ed: Excuse me. I think that is my seat.
Ashley: I'm sorry, but I don't see a reserved sign.
8. I am sorry to hear about your accident.

A variation of this activity:

If your students are at the intermediate level, you could provide them with situations, and in their groups, they determine whether the speaker should apologize or say he/she is sorry.

Example:

9. You accidentally spill your friend's coffee.
10. You have to tell your friend that his team has lost the game.
11. Your boss called you to cancel your appointment.

Conclusion

Although using corpora in language teaching is challenging, it has a big potential in EFL classrooms. If teachers are trained on how to design suitable corpus-based tasks, they can help their students get exposed to a broader framework of how English is used for communication by native speakers. Certain points to bear in mind while designing corpus-based tasks include the ability level of the students, cultural and educational backgrounds and the age group of the students. It is also recommended to use online corpora as they are available for anybody at any given time. In class a specialized corpus is needed, EFL instructors and researchers need to make sure it is available at their institutions. The researchers' next plan is to examine some of the activities listed above using a more in-depth approach to highlight the benefits, challenges and drawbacks of each.

References:

- Breyer, Y. (2008). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22(2), 153-172.
- Fuentes, A. C. & Rokowski, P.E. (2002). Use of corpus in the business English classroom: From lesser to more specific. Paper presented at the *IVACS Conference*, June 2002: Ireland, UK.
- Maddalena, S.R. (2001). An investigation into how corpus analysis may be used in the second language classroom to solve some of the problems surrounding non-native speakers' understanding of seemingly synonymous words. Retrieved August 15, 2009 from ERIC database.
- Neff, J., Martinez, F. & Rica, J. P. (2001). A contrastive study of qualification devices in native and nonnative argumentative texts in English. Paper presented at the annual meeting of the *American Association of Applied Linguistics*, February, 2001: St. Louis, MO.
- Simpson, R. & Mendis D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, 27 (3), 419-441.
- Varley, S. (2008). I'll just look that up in the concordance: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22 (2), 133-152.

Neveen Al Saeed (vicky31@aucegypt.edu) is a teaching assistant at Ain Shams University. Her research interests include second language writing, curriculum design and online teaching and learning. She has presented in local and international conferences.

Salma Waly (swaly@aucegypt.edu) is an MA TEFL holder. She is an EFL instructor whose research interests include teaching and learning online, using technology and material design. She has presented in a number of local and international conferences.

