UDK 811.111-13:371.3
Pregledni rad

Dr. sc. Edina Rizvić-Eminović
Azra Hadžić, BA

USING CORPORA IN ENGLISH LANGUAGE TEACHING – A BRIEF OVERVIEW AND SAMPLE CORPUS-BASED ACTIVITIES

Abstract

Corpus linguistics has seen an expansion in scope over the past several decades. It has also found its way into language classrooms. Corpus-based approaches have gradually become a common practice among English teachers. Some research studies have established that corpus-based approaches are extremely beneficial to both teacher and learner. Aiming to illustrate how corpora can be used in English language teaching, this paper offers brief explanations of what corpora is, elaborating on the most relevant advantages of such an application in language classroom in terms of the teacher's role and learners' autonomy, motivation, self-reflection. Finally, the paper offers sample corpus-based activities (with concordances from COCA and BNC) that teachers can use printed out as hand-outs in the classroom.

Key words: *corpus linguistics, corpus-based activities, English language teaching, autonomy, motivation*

Introduction

"A corpus is a large, principled collection of naturally occurring texts (written or spoken) stored automatically" (Reppen, 2010: 23), where naturally occurring text denotes the language from the real-life situations such as "friends chatting, meetings, letters, class assignments, and books rather than … made-up language" (ibid.). According to Baker (1995: 223), a corpus is "primarily a collection of text held in a machine-readable form and capable of being analysed automatically or semi-automatically in a variety of ways". Most often, corpus data is stored on computer systems which make it possible to use a software interface to access the data. Corpus linguistics, therefore, is the study of language through corpora. The term corpus linguistics was initially introduced by Geoffrey Leech. He identifies corpus linguistics as "a recognised branch of linguistics, just as (for example) sociolinguistics, psycholinguistics…" (Leech, 1992: 105). However, corpus linguistics cannot be compared to the mentioned linguistic disciplines since it is rather "a methodological basis for pursuing linguistic research" (ibid.). Such linguistic research "has paved the way for the emergence of modern corpus linguistics" (Yvonne, 2009: 1).

It was not until the late eighties and early nineties that the corpus linguistics found its way to classrooms and language teaching and the interest in corpus-based approaches rapidly developed in recent years (Flowerdew, 2009). Nevertheless, many were skeptical in accepting the corpus-based approaches to language learning classrooms. This resulted in a number of studies aimed at investigating the attitudes of the educational world towards the corpus-based teaching materials, such as the one by Friginal et al. (2020). Such studies contributed to the

"understanding of the value and benefits that language corpora can bring to language pedagogy" (Cheng, 2010: 320). Additionally, they also investigated whether "corpora can provide valid models for learners" (ibid.). Garcia (2012) maintains that corpora can be used to complement diverse language reception activities. Similarly, corpus-based approaches have been extremely beneficial to both language teaching and learning informally and formally, as stated by Walsh (2010). He argues that "one of the main advantages of a corpus is that teachers have a large resource from which to select texts to use with a group of learners" (ibid, 2010: 337).

The purpose of this paper is to give a brief overview of the benefits of using corpus-based activities in English language classrooms and to present a couple of corpus-based activities designed as a supplement for practicing the use of phrasal verbs *put something on, put something up* and *set something up*, and the idiomatic expressions *to be safe and sound, to be over, the black sheep of the family* taught in the course book *Challenges 4*.

Using corpora in ELT

Incorporating corpora or corpus-based materials into the language classroom requires a transition from a number of traditionally set beliefs to more innovative methods. The role of the teacher changes in the way that teachers are no longer the only source of knowledge about the target language. Teachers become facilitators who help the learners interpret and analyse the data they find. Teachers have a less central role than in traditional teaching (Gilquin et al., 2010: 366). To portray the change better, Hughes (2010: 405) states: "Often the individual teacher who becomes interested in moving beyond intuitions and traditional approaches… is often an IT-literate teacher comfortable with using new technology in front of a class, is often a keen champion of ITC in the language teaching classroom, and enjoys the challenging exploratory nature of students and teachers finding answers together."

Despite the fact that corpora have been present in the world of linguistics and teaching for decades now, there are not many commercially available textbooks for English Language Teaching (ELT) based on corpus data. The reason behind this might lie in "the fact that corpus linguists have spent so much time debating methods and frameworks for analysis that has led to a lack of a convincing 'story' about corpus informed pedagogic grammars, and hence slow uptake by the publishing world and the pedagogic profession" (Hughes, 2010: 403). There are also other reasons for the gap between the ELT publishers and the teaching community on one side and the adoption of corpus data on the other side. Some believe that the gap exists due to "the powerful influence of traditional pedagogic grammars" or even "the lack of communication between the teacher training community in applied linguistics and the research community" (Hughes, 2010: 402). As Garcia (2012: 34) states, practising teachers usually have a syllabus to follow and a set of course books they have to use. Hence, there is little room for them to stray away from the instructions prescribed by the educational authorities. The resistance to using corpora and corpus-based materials in language classroom might also lie in the fact that corpus examples do not neatly fit into a given target 'level' while traditional grammar textbooks contain a table of contents with clearly defined topics to be covered (Hughes, 2010: 406). One of the positive sides

of using corpora in language teaching is that it "bring(s) about exciting pedagogical perspectives in a wide range of areas of English language teaching (ELT) pedagogy: material design, syllabus design, language testing, and classroom methodology" (Cheng, 2010: 319). Although it is a relatively new phenomenon in language teaching and learning, the use of corpora has "very clear pedagogical aims" (Cheng, 2010: 325), such as allowing the learner to identify patterns of errors or even allowing both learners and teachers "to identify and prioritise very specific and general patterns of use" (ibid.).

The use of corpora in language teaching in the classroom and outside of it has witnessed a rising trend over the last decade. "Notwithstanding the fact that the initial reception was rather 'frosty', "corpora have become a part of the pedagogical landscape'" (Vaughan, 2010: 476). Since both teachers and students are now 'confident members of the online world', teaching is now becoming more technology-led depending less on the course books. The aspect of corpus-based teaching materials found to be one of the most important and useful in English language teaching and learning equally is the challenge it provides to students and teachers as well, especially if they are non-native speakers. When a non-native English teacher embraces the challenge of using the corpus data in teaching materials, the teacher also becomes a learner (Hughes, 2010: 407). A high level of motivation, support and encouragement is needed to get to the point where a teacher can successfully learn and at the same time teach the students at the times when the majority of teachers still prefer following the published teaching materials. However, once a teacher has acquired the basic skills of searching a corpus, it will be an easy task to incorporate the data into the materials used in the classroom. As McCarten (2010: 425) points out: "there is a bright future for corpus-informed course materials. Bringing into the classroom the language of the world outside gives learners greater opportunities to increase their understanding of natural language and the choice to use it. And this is surely the ultimate aim of all of us who are engaged in language teaching." On the other hand, the students, too, benefit from the use of corpus-based materials. Such materials encourage exploratory, autonomous and creative learning about language forms (Hughes, 2010: 405). The role of learners is more active, reflective and autonomous (Chambers, 2010: 346) when data-driven learning is engaged in the language classroom. Consequently, the role of teachers changes – they truly become facilitators of language learning and the classroom truly becomes student-centred. A variety of corpora may be used to that end, among them the learner corpora as "a collection of students' spoken or written work" (Walsh, 2010: 333). Apart from making the learners feel as researchers and learners at the same time, exploring a corpus and doing corpus-based activities also helps them understand that in their language use native speakers, too, sometimes deviate from the established rules of grammar or tend to use certain expressions more than others. When learning a language, a non-native speakers' fluency is measured by considering features such as coherency, frequency of hesitation, flexibility, and automaticity. Having the opportunity to investigate a corpus, a learner actually has access to natural occurring language. In that way, a learner can observe that native speakers also make hesitations while speaking, their speech is not always coherent, their vocabulary range is not as wide as thought (Walsh, 2010: 336). This

realization helps the learner overcome many problems that are an integral part of the learning process and gain confidence in using the target language.

As Garcia points out (2012: 35), teachers can use corpora to move away from the traditional course book language and expose the learners to authentic language, challenge them to find answers by themselves since "corpora provide evidence for our intuitions about language" (O'Keeffe et al, 2007: 21). It does not necessarily mean that the language classroom has to be technologically equipped. If there is a computer lab available, that is an advantage indeed. If not, that–should–discourage language teachers. Teachers can prepare corpus-based materials in advance. It is usually enthusiasts who are promoting corpus application to language teaching, as proposed by Flowerdew (2009) and it is their aim to popularise such an application.

Sample corpus-based activities

A proposed by Bennett (2010), designing corpus-based activities involves several steps. Those activities can be aimed at conducting mini research with students, in which they themselves explore a corpus and engage in qualitative or quantitative analysis, depending on their level of English proficiency, or they can be used as supplement activities, as is the case in this paper. Due to the lack of infrastructure and equipment in our context, supplement activities based on a corpus are more suitable for our classrooms. Such activities are created by teachers themselves and printed out as hand-outs to be used in the classroom and accompany a unit from a textbook. For the purpose of creating corpus-based activities presented in this paper, the concordances from the Corpus of Contemporary American English (COCA) and also the British National Corpus (BNC) were used. The target CEFR level is pre-intermediate, students' profile is 14 to 15 years old. The activities were designed to complement the existing activities proposed in the course book. The course book is *Challenges 4 Students' book* published by Pearson Education (2012). Each activity comprised instructions asking the students to observe a pattern or reach a conclusion (which adds to the exploratory nature of the activity), a sequence of examples with the target lexeme or a lexical bundle retrieved from the corpus (the purpose being to expose the students to the real life examples of use), and one or more gap-fill exercises (to provide the students with an opportunity to practice the use of a word or a lexical bundle). It is of paramount importance to plan the appropriate amount of time for such activities, otherwise they will be counterproductive. This paper presents two sample activities, one designed to practice the use of phrasal verbs *put something on, put something up* and *set something up*, and the other to practice the use of idiomatic expressions *to be safe and sound, to be over, the black sheep of the family*.

Sample corpus-based activity 1

The *Challenges 4* course book covers a number of phrasal verbs that pre-intermediate students should be familiar with after having completed the 9th grade. The corpus-based activity presented in this paper is a complement to Activity 4, *Word Builder* (p. 29) and is focused on three phrasal verbs that might be hard for students to master due to their similarity either in

meaning or in structure. Those are: *put something on, put something up* and *set something up.* The tasks are prepared based on the actual concordances from the COCA and the BNC. The main aim of the activity is to help the students master the meanings and usage of the phrasal verbs under consideration. Before the activity, the students should be reminded of the structure of phrasal words, the fact that they are one of the key features of the informal language, and that their meaning cannot be conveyed to Bosnian/Croatian/Serbian word for word. The corpus-based activities in general can be assigned either in class when the unit is being taught, during the revision class or as homework. The activity is designed as follows:

A Read the sentences below and underline the verbs *put* and *set.* Pay attention to the words around them and tell if they have something in common?

1. They just put on a Halloween dance at school as a fundraiser.
2. This is the guy who dominated the games, who put on a show.
3. He asked me to take a seat and listen to some music he would put on.

4. They set up special clubs which concentrate on mountain rescue.
5. When two or more people set up a business together, they form a partnership.
6. They set up a disaster fund for the widows and orphans.

7. They could not get the goal through. *Town* put up the barricades and blocked everything.
8. And judging from his own reactions, he knew that it was impossible. But she put up a barrier around her and allowed no one to pass.

B Match the verbs with their meaning according to the sentences above.

1. put up	a) organise
2. put on	b) begin a project
3. set up	c) build

C Rewrite the sentences under A as in the example below.

1. They just **put** *it* **on** at school as a fundraiser.

D Fill in the gaps with the phrasal verbs

put up put on set up

1. They have _____ a Christmas show that runs until the end of January.
2. Somebody came along and wanted to _____ farm buildings and a house.
3. The Federal theatre Project _____ play, musical and variety shows across the continent.
4. This year a travelling circus _____ its tent and offered the public a horse-riding show.
5. The charity's goal is to help unemployed youngsters _____ their own business and find jobs.
6. There are plans to _____ a fund in her name to assist wildlife film-makers.

The students should be encouraged to draw as many conclusions as they can about the form, meaning, pattern of use, even translation in the native language. Thus, Task A is designed to help the students recognise that the verbs *put* and *set* are in each and every example sentence followed by a preposition and that such a structure actually makes them phrasal verbs or multi-part verbs. Task B aims to help the students understand the meaning of the phrasal verbs. Task C provides an opportunity for the students to practice the use of such verbs when their object is a pronoun but in sentences retrieved from the Corpus, that is, from the real life language. Task D is somewhat more challenging. Since the students were previously taught on phrasal verbs and their grammatical features, they know that there are certain phrasal verbs that can be separated and there are some that are inseparable in nature. This task encourages the students to be active, exploratory and creative in the learning process and to undertake the role of a researcher who tests and examines issues. It also provides them with a chance to test themselves in terms of how much they have learned to that point and whether they need to practice more.

Sample activity 2

As with phrasal verbs, there is a number of idiomatic expressions that need to be mastered at all levels of language learning. The corpus-based activity for vocabulary learning presented in this paper is a complement to Activity 4, *Word Builder* (p. 89). The course book unit covers five idiomatic expressions, however, the sample corpus-based activity is concentrated on three. The main aim of the activity is to help the students master the meanings and usage of the idiomatic expressions under consideration. Its structure is similar to sample activity 1.

A Read the sentences below and underline the following expressions *the black sheep of the family, to be safe and sound, to be over.* What do these expressions have in common?

1. My prayer for you is that when you get home, you can come home safe and sound.
2. She was the black sheep of the family. We're never really that close to her.
3. Charles and Warren arrived safe and sound in the Republic of Texas.
4. 'The show appears to be over for tonight,' he said.
5. He is considered the black sheep of the family. He has been in and out of trouble for years.
6. The war will not be over until the U.S. troops come home.
7. Don't worry about Quinn. I'm sure Axel will bring him back safe and sound.
8. Yeah, I was the black sheep in the family because I was a lawyer.
9. Summer days will soon be over. Soon the autumn starts.

B Match the expressions with their meaning according to the sentences above.

1. the black sheep of the family	a) unhurt
2. to be safe and sound	b) finished
3. to be over	c) somebody who is different from his/her family

C Complete the sentences with the correct idiomatic expression.

1. They are back by 2 o'clock. Their work on the trip appears _____, although they'll spend another 24 hours on Barbados.
2. We made it back, _____, and cuddled up on the couch to eat crackers and cheese.
3. Once we graduated from high school I became _____ and my sister was the responsible one, loved and praised by the family.
4. If I don't do this, my baseball career could literally _____ tomorrow.
5. Every family has a _____ and if you think you are perfect, wait, you will get your turn.
6. The worst was over. The worst had _____ or else he didn't think he could keep breathing.
7. Our only goal should be to return Lady Rosaline _____.
8. One of the daughters was overweight, kind of _____, not as into riding as her family.

Idiomatic expression is in essence a characteristic feature of native speech and learners have hard time learning and using the expression in an adequate way. Sometimes, teachers need to provide additional exercises to bring the meaning of a certain idiomatic expression closer to the students. The aim of the Task A is to encourage the students to recognise the idiomatic expressions in the sentences retrieved from the COCA corpus and get the idea of how the expressions are incorporated in a sentence. Since idiomatic expressions are known to have non-literal meaning and their meaning has to be figured out from the context, Task B asks the students match the expression with its meaning according to the contexts provided in Task A. Task C is designed to help the student be more confident in using idiomatic expressions in a variety of contexts or in their writing.

Conclusion

The activities presented in this paper illustrate that students do not necessarily need to be in a computer lab to access the corpus data. Corpus-based activities can be prepared by teachers as hand-outs to complement specific units covered in class. This means that no time has to be invested in teaching and training the students on how to use the corpus. Teachers do not need to inform the students that a corpus has been used to prepare the activities. If a corpus-based

activity is carefully designed, the pattern of a single activity can be used again when needed, to address the areas of difficulty, or to further develop the language issues that are not covered in sufficient detail (Garcia, 2012).

Using corpus-based materials in language classrooms boosts incidental learning and promotes learner autonomy. “Because learners will be engaged in looking up solutions to problems in the forefront of their minds, learning is bound to take place.” (Garcia, 2010: 41). Consequently, students may feel more motivated to learn English.

If the activity has proven to be useful and successful and proven to yield positive results in terms of students’ learning, it is up to the teacher to promote the use of corpus-based materials among the language teaching staff. The type of the tasks in the sample activities presented in this paper show that corpus-based activities can be used in a language classroom without disrupting the usual classroom routine. Their formulation resembles the one the students are accustomed to in their course books; however, examples used are from the real life language retrieved from the COCA corpus.

References

1. Leech, G. (1992) “Corpora and theories of linguistic performance” in J. Startvik (Ed.) *Directions in corpus linguistics,* Mouton de Gruyter, Berlin, 105-122
2. Reppen, R. (2010) *Using Corpora in the Language Classroom*, Cambridge University Press, UK
3. Baker, M. (1995). “Corpora in Translation Studies: An Overview and Some Suggestions for Future Research” in Target*,* Vol. 7(2), 223–243
4. Bennett, Gena R. (2010) *Using Corpora in the Language Classroom: Corpus Linguistics for Teachers,* Michigan ELT, USA
5. Breyer, Yvonne (2009) "Learning and teaching with corpora: reflections by student teachers" in Computer Assisted Language Learning Vol. 22 (2), 153-172
6. Flowerdew, J (2009) “Corpora in Language Teaching*”*in *The Handbook of Language Teaching,* eds. Long, Michael H. Doughty, Catherine J., Blackwell Publishing Ltd., USA
7. Friginal, Eric, Dye, Peter and Nolen, Mathew (2020), “Corpus-Based Approaches in Language Teaching: Outcomes, Observations, and Teacher Perspectives” in Boğaziçi University Journal of Education, Vol. 37 (1), (43-58)
8. Cheng, W. (2010) “What can a corpus tell us about language teaching” in *The Routledge Handbook of Corpus Linguistics*, ed. O’Keeffe, A. and McCarthy M., London and New York, 319-332
9. Frankenberg-Garcia, A. (2012) “Integrating corpora in everyday language teaching” in *Input, Process and Product Developments in Teaching and Language Corpora*, Masarykova Univerzita, Brno, 36-53
10. Walsh, S. (2010) “What features of spoken and written corpora can be exploited in creating language teaching materials and syllabus?” in *The Routledge Handbook of Corpus Linguistics*, ed. O’Keeffe, A. and McCarthy M., London and New York, 333-344

11. Gilquin, G. and Granger, S. (2010) “How can data-driven learning be used in language teaching” in The Routledge Handbook of Corpus Linguistics, ed. O’Keeffe, A. and McCarthy M., London and New York, 359-370
12. Hughes, R. (2010) “What a corpus can tell us about grammar teaching materials” in “The Routledge Handbook of Corpus Linguistics”, ed. O’Keeffe, A. and McCarthy M., London and New York, 401-412
13. Vaughan, E. (2010) “How can teachers use a corpus for their own research?” in *The Routledge Handbook of Corpus Linguistics*, ed. O’Keeffe, A. and McCarthy M., London and New York, 471-484
14. McCarten, J. (2010) “Corpus-informed course book design” in *The Routledge Handbook of Corpus Linguistics*, ed. O’Keeffe, A. and McCarthy M., London and New York, 413-427
15. Chambers, A. (2010) “What is data-driven learning?” in *The Routledge Handbook of Corpus Linguistics*, ed. O’Keeffe, A. and McCarthy M., London and New York, 345-358
16. Harris, M., Mower, D., Sikorzynska, A., Mešić, A. (2007) *Challenges 4, Students’ Book* Pearson Education Limited, UK

PRIMJENA KORPUSA U NASTAVI ENGLESKOG JEZIKA – KRATAK PREGLED SA PRIMJEROM PRAKTIČNIH AKTIVNOSTI

Sažetak

Korpusna lingvistika je u posljednjih nekoliko desetljeća sve prisutnija ne samo među lingvističkim istraživanjima nego i u učionicama u kojima se podučava engleski kao strani jezik. Korpusne aktivnosti postepeno postaju uobičajena praksa u nastavi engleskog jezika. Prethodna istraživanja su pokazala da je upotreba korpusnih aktivnosti izuzetno korisna i za nastavnike i za učenike. S ciljem ilustriranja načina upotrebe korpusa u nastavi engleskog jezika, ovaj rad ukratko objašnjava šta je korpus i nudi najvažnije prednosti korištenja korpusnih aktivnosti u nastavi s osvrtom na ulogu nastavnika i autonomiju, motivaciju, te samorefleksiju učenika. Konačno, rad nudi praktične primjere korpusnih aktivnosti (izvedenih iz COCA i BNC korpusa) koje nastavnici engleskog jezika mogu koristiti kao odštampane materijale u nastavi.

Ključne riječi: *korpusna lingvistika, korpusne aktivnosti, nastava engleskog jezika, autonomija, motivacija*