

# 6 Introducing corpus linguistic tools to EFL undergraduates and trainee teachers

Mária Adorján<sup>1</sup>

---

## Abstract

Many language teachers use Information and Communications Technology (ICT) in their classrooms to create tasks, quizzes, or polls with general online learning platforms. Few teachers have experience, however, of incorporating online corpus tools in their teaching or assessment practices. This paper will explore how autonomous learning can be fostered by gradually introducing freely available lexical databases, online collocation dictionaries, pronunciation guides, concordancers, N-gram extractors, and other text analysis tools for vocabulary building, skills practice, or self-checking. Tasks used with English as a Foreign Language (EFL) undergraduates and teacher trainees on a Master's Teaching English as a Foreign Language (MA TEFL) course will be presented. I will also explain why having some familiarity with linguistics research can enable teachers to use these applications more meaningfully.

**Keywords:** corpus linguistic tools, EFL, vocabulary learning, text analysis, language teacher training.

---

1. Károli Gáspár University of the Reformed Church, Budapest, Hungary; adorjan.maria@kre.hu; <https://orcid.org/0000-0001-8870-3947>

**How to cite this chapter:** Adorján, M. (2020). Introducing corpus linguistic tools to EFL undergraduates and trainee teachers. In K. Borthwick & A. Plutino (Eds), *Education 4.0 revolution: transformative approaches to language teaching and learning, assessment and campus design* (pp. 45-51). Research-publishing.net. <https://doi.org/10.14705/rpnet.2020.42.1086>

## 1. Introduction

Teachers are increasingly aware of online tools that can support the teaching of EFL and are willing to incorporate a large number of them into their courses in higher education. These tools are used for many purposes, from supporting individual work to networking or sharing content. The use of online language reference tools based on corpus data is still in its infancy, though, both in English language teaching and in teacher training (Granath, 2009). At Karoli Gaspar University, EFL undergraduates and trainee teachers are offered a range of courses where corpus linguistic tools and techniques are used regularly, including study skills, advanced writing, business English, and a dedicated course for students on the MA TEFL module: Foreign Language Assessment Methods.

Advanced level language learners have specific needs that make them distinct from other learners. Learning languages at higher levels necessitates different learning strategies than learning at beginners' levels (Politzer & McGroarty, 1985). The main differences lie in vocabulary acquisition and production: having learnt most words within the core vocabulary range (the first 2,000 word-families), they have to acquire words within a lower frequency band (Nation, 2006). This requires a more conscious effort on the learners' part to actively look for opportunities to learn more words, or read much more in order to facilitate incidental learning (Schmitt, 2000). If students are informed about word frequency lists, or other applications based on corpus linguistics, they can easily optimise their learning processes. Another advanced learner problem is finding the right collocations. A concordancer can help to produce multi-word expressions used by native speakers or competent L2 users. Students also need to be aware of diverse discourse characteristics to sound natural or appropriate in various communicative environments. Samples of these, again, are difficult to find in coursebooks.

Corpus linguistics research can help in the above areas by raising awareness, speeding up the learning process, and promoting learner autonomy (Gavioli, 2009). In language teaching and learning, three main types of corpora are studied most often: authentic corpora to observe language use (Szudarski, 2017),

learner corpora to study interlanguage (Granger, 2002), and multilingual corpora for translation studies and contrastive analysis (Flowerdew, 2012). The aim of this paper is to present a variety of corpus linguistic tools based on these three types of corpora, which were incorporated into the syllabi of several university courses for EFL learners.

## 2. Method

The use of online tools was first incorporated into several language courses in 2018, as early as in the first semester. Table 1 lists some of these tools with their functions and their online location, however this paper will only discuss the first two in detail. In the first term, students were trained by demonstrating how each tool worked, explaining research findings which related to the tool, and by encouraging students to experiment with real data (their own texts, for instance); also, to compare their findings with research results. The use of corpus tools had two main aims: to facilitate the self-assessment of work, and to help students improve their writing and speaking skills. From the second year onwards, the main focus was analysing authentic original and translated texts and textbook language.

Students assessed their work in two main areas: receptive vocabulary knowledge and text production. The first tool they used was Vocabulary Profiler ([lex tutor.ca](http://lex tutor.ca)), which was used to analyse essays to find the ratio of academic words. Next, they compared their results with authentic text characteristics provided in the Typical Profiles section. The Vocabulary Size Test revealed the differences among students in their passive word knowledge. Doing the test stimulated a discussion on language learning styles and strategies, and discussing the results provided an opportunity to become familiar with some research into optimal vocabulary size for various purposes (Nation, 2006).

The process of using the tools to check individual texts went through these phases: drafting the text, familiarisation with the tool (e.g. what vocabulary levels exist, how to check words with the tool), analysing the text with the

tool (e.g. types of vocabulary used, the ratio of K1, K2, and academic words), reflecting on, analysing, and peer-discussing the results, then finally, rewriting the text and submitting it.

Table 1. The functions of corpus tools with the website

Tool name	Function	Link to website
Lextutor Vocabulary Profiler	textual analysis for vocabulary levels (K1, K2, ... AWL)	<a href="https://www.lexutor.ca/vp/eng/">https://www.lexutor.ca/vp/eng/</a>
Vocabulary Size Test	testing receptive vocabulary knowledge, multiple choice format	<a href="https://my.vocabularysize.com/">https://my.vocabularysize.com/</a> or <a href="https://www.lexutor.ca/tests/">https://www.lexutor.ca/tests/</a>
Cambridge English Dictionary	monolingual word search; checking CEFR <sup>2</sup> levels of word senses	<a href="https://dictionary.cambridge.org/dictionary/english/">https://dictionary.cambridge.org/dictionary/english/</a>
Vocabulary Kitchen	CEFR level visualiser for texts	<a href="http://vocabkitchen.com/profiler/cefr">http://vocabkitchen.com/profiler/cefr</a>
Coh-Metrix	cohesion and text readability analysis	<a href="http://tea.cohmetrix.com/">http://tea.cohmetrix.com/</a>
Oxford Collocation Dictionary (free version)	how words are used together in a sentence	<a href="http://www.ozdic.com/">http://www.ozdic.com/</a>
Sketch Engine for Language Learners	concordance, exploring immediate context of words	<a href="https://skell.sketchengine.eu/#home?lang=en">https://skell.sketchengine.eu/#home?lang=en</a>
Google Books N-Gram Viewer	historical use of words or expressions	<a href="https://books.google.com/ngrams">https://books.google.com/ngrams</a>
Wordnet Editor	finding and visualising words and synonym sets; differentiating word senses	<a href="http://wordventure.eti.pg.gda.pl/wne/wne.html">http://wordventure.eti.pg.gda.pl/wne/wne.html</a>
Multidimensional Analyser	six features characteristic of various genres (Biber, 1992)	<a href="https://sites.google.com/site/multidimensionaltagger/home">https://sites.google.com/site/multidimensionaltagger/home</a>

### 3. Results and discussion

Introducing corpus linguistic tools into various EFL courses resulted in an overall positive outcome, based on end-of-term student feedback<sup>3</sup>. During the

2. Common European Framework of Reference for languages

3. End-of-term course evaluation forms, unpublished raw data.

sessions, students were provided with numerous practice opportunities so that they knew the rationale behind the tools' use in teaching, and felt comfortable experimenting with them at home. With such scaffolding provided (Hubbard, 2013), corpus linguistic tools have become an essential part of the syllabus without an explicit focus on applied linguistics research. The students, according to their feedback, gained empirical evidence about the quality of their writing and were able to assess themselves, thus creating an atmosphere of involvement, interaction, individualisation, and independence (terminology from Dudley-Evans & St John, 1998, p. 200).

The heavily technology-enhanced L2 environment meant, however, that there were some anxiety issues at the start of the course that had to be overcome. Even though all students are surrounded daily by ICT, individual help was necessary for those who were new to monitoring language, focussing on syntax, or understanding grammatical terminology. Another important consideration is that for the successful use of corpus linguistic tools in such courses, the language teacher needs to be literate in corpus linguistic research so that results can be connected to existing corpus findings and language acquisition theories.

#### 4. Conclusions

The use of corpus linguistic tools in a variety of courses at the university had not been considered previously: it emerged rather as a response to EFL learners' needs. The rationale behind introducing these tools was to raise language awareness and enable learner autonomy. The students' direct exposure to corpus linguistic tools occurred without offering an introductory course into the field. An active search to incorporate additional tools followed this phase, when it became clear that the students were motivated and involved in the explorations.

An important conclusion to be drawn from the end-of-term assessment of the courses was that scaffolding was essential twice: at the students' first attempts

with the tools and when they wanted to interpret the results. This latter was essential particularly in the case of tools which only provided numerical data as results. These percentages and ratios lent themselves to discussion, and the teacher could direct the group to further readings about the features of the language analysed.

I hope that sharing these ideas will provide the language teaching community with potential corpus linguistic activities, and the widespread use of these tools will have a significant impact on traditional higher education language courses.

## 5. Acknowledgements

I would like to thank my students who let me use their work for textual analysis.

## References

- Biber, D. (1992). On the complexity of discourse complexity: a multidimensional analysis. *Discourse Processes*, 15(2), 133-163. <https://doi.org/10.1080/01638539209544806>
- Dudley-Evans, T., & St John, M. J. (1998). *Developments in ESP. A multidisciplinary approach*. Cambridge University Press.
- Flowerdew, L. (2012). Definitions, purposes and applications of corpora. In L. Flowerdew (Ed.), *Corpora and language education* (pp. 3-35). Palgrave Macmillan. [https://doi.org/10.1057/9780230355569\\_1](https://doi.org/10.1057/9780230355569_1)
- Gavioli, L. (2009). Corpus analysis and the achievement of learner autonomy in interaction. In L. Lombardo (Ed.), *Using corpora to learn about language and discourse* (pp. 39-69). Peter Lang.
- Granath, S. (2009). Who benefits from learning how to use corpora? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 47-65). John Benjamins Publishing.
- Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds), *Computer learner corpora, second language acquisition, and foreign language teaching* (vol. 6, pp. 3-33). John Benjamins Publishing.

- Hubbard, P. (2013). Making a case for learner training in technology enhanced language learning environments. *Calico Journal*, 30(2), 163-178. <https://doi.org/10.11139/cj.30.2.163-178>
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82. <https://doi.org/10.3138/cmlr.63.1.59>
- Politzer, R. L., & McGroarty, M. (1985). An exploratory study of learning behaviors and their relationship to gains in linguistic and communicative competence. *Tesol Quarterly*, 19(1), 103-123. <https://doi.org/10.2307/3586774>
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
- Szudarski, P. (2017). *Corpus linguistics for vocabulary: a guide for research*. Routledge.



Published by Research-publishing.net, a not-for-profit association  
Contact: [info@research-publishing.net](mailto:info@research-publishing.net)

© 2020 by Editors (collective work)  
© 2020 by Authors (individual work)

**Education 4.0 revolution: transformative approaches to language teaching and learning, assessment and campus design**  
**Edited by Kate Borthwick and Alessia Plutino**

**Publication date:** 2020/08/10

**Rights:** the whole volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; **individual articles may have a different licence.** Under the CC BY-NC-ND licence, the volume is freely available online (<https://doi.org/10.14705/rpnet.2020.42.9782490057665>) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

**Disclaimer:** Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book is believed to be true and accurate on the date of its going to press, neither the editorial team nor the publisher can accept any legal responsibility for any errors or omissions. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

**Trademark notice:** product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

**Copyrighted material:** every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net  
Cover illustration: [iStock.com/SiberianArt](https://www.istock.com/SiberianArt)  
Cover layout by © 2020 Raphaël Savina ([raphael@savina.net](mailto:raphael@savina.net))

ISBN13: 978-2-490057-66-5 (Ebook, PDF, colour)  
ISBN13: 978-2-490057-67-2 (Ebook, EPUB, colour)  
ISBN13: 978-2-490057-65-8 (Paperback - Print on demand, black and white)  
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.  
A cataloguing record for this book is available from the British Library.

**Legal deposit, France:** Bibliothèque Nationale de France - Dépôt légal: août 2020.