

Overwhelming, Time-consuming, User-unfriendly... and Now What? Teacher Training for the Successful Integration of Corpus Use into Language Teaching

Szilvia Szita
University of Strasbourg (<szita@unistra.fr>)

Abstract

This article presents a 30-hour training on state-of-the-art teaching techniques by the International Studies Center of the University of Pécs (Hungary). The training was attended by 48 in-service teachers, most of them teaching Hungarian to lower proficiency levels. I will first review previous research about teacher trainings on corpus use. Then, a detailed description of the four-step approach applied in this training as well as some activities will be provided. I will argue that effective teacher training should start with exploring specialized pedagogical corpora rather than large general ones and link innovative practices to well-known exercises. The participants' feedback reveals that, while their reactions to corpus use are positive, they express their insecurity when it comes to analyzing corpus data by themselves. They seem to prefer ready-made material and worksheets prepared in collaboration and supervised by experts with respect to creating material on their own.

Keywords

corpus-informed activities; corpus pedagogy; language learning; language teacher education; pedagogical corpora; teacher training

1. Introduction

This article presents a 30-hour long training on state-of-the-art teaching techniques offered by the International Studies Center of the University of Pécs (Hungary). The training took place in 2018 and 2019. 48 in-service teachers of Hungarian attended, most of them teaching learners at lower proficiency levels. One of the focal points of this course was corpus use

Online first, 30 novembre 2022

CC BY-NC-ND 4.0 | SFU – Studi Finno-Ugrici, n.s. 2 (2022), 1-28 | ISSN 1826-753X
<https://doi.org/10.6093/1826-753X/9852> | © Università di Napoli L'Orientale – UniorPress



for the classroom. First, I will briefly review the literature about teacher trainings on corpus use and corpus pedagogy. Then, I will give a detailed description of the activities proposed during our training. I will argue that effective trainings for teachers at lower proficiency levels should explore pedagogical corpora rather than large linguistic ones and should begin by presenting methods and activities that participants are familiar with before gradually introducing them to new ones. Proposed innovations seem to be accepted more easily when they are linked to existing practices and presented step by step. The last part of the publication is dedicated to the analysis of participants' feedback and considerations for future research.

2. Studies on teacher training and corpus-informed pedagogy

2.1 Findings on teacher trainings: benefits and problems

As teachers «are the main conduit between research and classroom» (Breyer 2011, 228), they should be the first to address when implementing new methods in language lessons. The ideal place for corpus training would therefore be pre-service language teacher education where future teachers have numerous opportunities to familiarize themselves with state-of-the-art approaches to language teaching and make corpus pedagogy an essential part of their toolkit (cf. Chambers 2019; Farr 2008). While benefits are clear, only a few institutions offer such pre-service courses and, as Qing Ma, Jinlan Tang, and Shanru Lin (2021, 3) point out, there are «only a handful of empirical studies focusing on teacher training».

These studies emphasize the participants' overall positive attitude (Breyer 2009; Çalışkan, Kuru Gönen 2018; Callies 2019; Ebrahimi, Faghih 2017; Farr 2008; Leńko-Szymańska 2014, 2017; Poole 2022; Zareva 2017) while also identifying potential difficulties. Key benefits include access to authentic data with different tools and from different entry points (Poole 2022) promoting enhanced awareness of particular linguistic items such as multi-word units or grammar phenomena (Çalışkan, Kuru Gönen 2018) and uncovering patterns not presented in grammar books autonomously (Farr 2008). Primary concerns are related to: (1) the time consuming-nature of creating corpus-based and/or corpus-informed activities, (2) lack of technical skills, and (3) lack of appropriate corpora for the language classroom (Poole 2022).

Farr's (2008) study shows that most pre-service teachers are willing to use corpora in their future work but also emphasizes their lack of self-confidence when interpreting corpus data as well as potential technical difficulties. Leńko-Szymańska (2017) confirms the perceived usefulness of corpus work by the participants of a course for future (non-native) English teachers, stressing however that a well-designed teacher training including both corpus literacy and corpus pedagogy as well as an extensive exposure to corpora in form of various tasks are needed to increase future teachers' willingness to integrate corpus use into their lessons. Zareva (2017, 70) highlights the importance of developing future teachers' corpus literacy skills during training programs and states that corpus work without such prior training can be perceived as «a complex and time-consuming endeavor for both teachers and students». Ebrahimi and Faghih (2017) also focus on pre-service teachers and suggest that courses on corpus linguistics should be included into the initial stages of degree programs in higher education and extended throughout the entire curriculum. They also emphasize the importance of user-friendly corpus analysis tools and skills for successful corpus exploitation. Çalışkan and Kuru Gönen (2018) focus on vocabulary teaching with corpora and identify use of technology and designing corpus-based materials as the two major challenges for (Turkish) EFL instructors. Finally, Callies (2019, 261) makes «a case for the inclusion of specifically designed content classes and modules on corpus linguistics in English language teacher education that include hands-on activities in which student teachers learn about the use of corpora from a researcher's, teacher's and learner's perspective».

Some of the studies also address the effect of ready-made materials on teachers' willingness to use corpora in their classrooms. Çalışkan and Kuru Gönen (2018) and Poole (2022) show that teachers appreciate when they are provided with ready-made activities and, in this case, they do not find corpus work particularly time-consuming. While this can be seen as a positive development, Poole raises the question of whether such a pragmatic approach does not challenge the very purpose of corpus exploration, namely that both teachers and learners should be trained to become autonomous corpus users over time, able to find answers to their language-related questions.

Addressing the afore-mentioned issues, Qing Ma, Jinlan Tang, and Shanru Lin (2021) propose a two-step training scheme focusing on (1) corpus literacy in the first step, then on (2) corpus-based language pedagogy in the second. While the effectiveness of their approach has been attested by the participants' feedback, the authors claim that a consistent corpus-informed pedagogy is still needed.

2.2 Elements of a corpus-informed pedagogy

Several issues related to a corpus-informed pedagogy have been discussed in the literature, the three most important ones being the following: (1) the need for pedagogical corpora and the criteria of their construction, (2) the accessibility of texts in the corpus *and* in their entirety, and (3) a consistent methodology to explore corpora in the language lessons. The next paragraphs will give an overview of each of these points.

The view that corpora created for linguistic explorations do not lend themselves easily for classroom use seems to be widely accepted today (e.g., André 2020; Braun 2007, 2010; Breyer 2009, 2011; Farr 2008; Frankenberg-Garcia 2012, 2014; Poole 2022). Frankenberg-Garcia (2012, 476) notes that «the majority of corpus resources are neither pedagogically oriented nor user friendly» and Poole (2022, 16) observes the «lack of ready-made, pedagogically-focused corpus materials». Whereas consulting large authentic corpora can be challenging for these levels, providing classrooms with pedagogical corpora that «take the learner's perspective» (Braun 2010, 82) can be a good starting point for corpus exploration (Aston 2001; André 2017, 2019; Braun 2007, 2010; Chambers 2019; Charles 2014; Kennedy, Miceli 2010, 2017). These 'curated' corpora can ensure that the linguistic complexity and the situational context of the utterances are in line with both the learners' needs and their level of proficiency. As Friginal (2018, 24) observes, «[t]he advantage of creating written and spoken corpora specifically intended for teaching purposes is that the corpora can be designed with a clear purpose».

It is clear that when building pedagogical corpora, one cannot strictly follow the principles identified for the creation of corpora for linguistic research. While authenticity, statistical representativeness, well-balanced-

ness are of paramount importance for linguistic corpora (cf. Brezina 2018; Stefanowitsch 2020), the key component of a good pedagogical corpus is that it contains *relevant and accessible linguistic input* (Braun 2010). Compromising on authenticity by amending authentic texts seems to be necessary as understanding fully authentic language is likely to go beyond learners' proficiency at the A1-B1 levels.

How to realize such corpora? Most researchers suggest that pedagogical corpora should be created based on the teaching material (e.g., Aston 2001; Braun 2010; Charles 2014; Flowerdew 2009; Hunston 2002; Meunier, Gouverneur 2009; Meunier, Reppen 2015; Timmis 2015; Tyne 2012; Widdowson 2003; Willis 2003). Timmis (2015, 3), quoting Willis, suggests for example that a corpus for teaching purposes can be «made up of the *texts already used by the learners in class, which is then exploited for the study of particular language features*. The advantage of such corpora [...] is that learners will already be familiar with the co-text, i. e. the textual environment of the target feature, as they will previously have studied the whole text in class» (emphasis added). The pedagogical corpus should be representative of the learner's needs and of manageable size, containing material that is relevant and linguistically accessible to the learner (Aston 2001). Furthermore, when the teaching material is corpus-informed and carefully designed, it can preserve a significant degree of linguistic authenticity even at lower proficiency levels (Szita 2014)¹.

Another issue addressed in the literature is the accessibility of texts. Most authors stress that for learners to optimally benefit from the corpus, they need to have access to both the *complete individual texts* and the *collection through analysis tools* (e.g., Braun 2007, 2010; Chambers 2019; Charles 2014; Flowerdew 2009; Kennedy, Miceli 2010, 2017; Tyne 2012). In this manner, learners can familiarize themselves with the text first, then work on its content, vocabulary, intonation, pronunciation, and other aspects. The use of the pedagogical corpus can ensure condensed exposure to relevant linguistic items in a second phase.

The last point concerns corpus pedagogy, i.e., the integration of corpus work and insights gained from corpus linguistic research into approaches

¹ In this chapter, we do not discuss suggestions about learners compiling their own corpora as they do not seem to be particularly relevant for lower proficiency levels.

to language teaching. The affordances and limitations of direct and indirect applications of corpora for language teaching have both been extensively discussed in the literature (e.g., Boulton 2010; Chambers 2010; Johns 1991a, 1991b; O’Keeffe, McCarthy, Carter 2007; Römer 2011). In the case of direct applications, learners are encouraged to explore the corpus on their own to find answers to their language-related questions. Indirect applications draw on corpus findings to create teaching materials and activities for the classroom. There are a number of excellent examples for both applications. Activities for hands-on applications are presented, among many others, by Tribble and Jones (1990), Johns (1991a, 1991b), Boulton (2010), Boulton and Thomas (2012) and Crosthwaite (2019). Indirect applications of corpus data have brought about corpus-informed teaching materials and grammars. Good examples of these are the *English Grammar Today* by Carter *et al.* (2016), Conrad and Biber’s *Real Grammar* (2009) or the *Longman Student Grammar of Spoken and Written English* by Biber *et al.* (2002). Corpus-informed textbooks include the *Touchstone* (McCarthy *et al.* 2005-2011) and the *Viewpoint* (McCarthy *et al.* 2012-2013) course book series. Activities for various proficiency levels in Hunston and Francis’s *Pattern Grammar* (2000) are available on the publisher’s website. Resource books with corpus-based activities such as Poole (2018) and Friginal (2018) can also enrich the teacher’s toolkit. However, both of these resources need to be adapted for lower proficiency levels and for languages other than English.

While the effectiveness of corpus-based activities seems to be confirmed (e.g., Boulton 2010; Boulton, Cobb 2017; Chambers 2019; Cobb, Boulton 2015; Tyne 2012), it is uncertain how far they have been integrated into the actual teaching practice of the average practitioner. As mentioned earlier, some studies suggest that well-designed, ready-made materials could incite teachers to adopt corpus use more readily (Çalışkan, Kuru Gönen 2018; Poole 2022). More pedagogical corpora and a consistent corpus pedagogy are also likely to contribute to the broader use of corpus use (cf. Braun 2007, 2010; Chambers 2019; Ebrahimi, Faghieh 2017).

3. The present study: Participants, context, and corpora

3.1 Context and participants

The training presented in this chapter was held at the University of Pécs (Hungary) twice, first in 2018, then in 2019. From the 48 participants (two groups of 24), 42 were native Hungarian speakers from Hungary and 8 bilinguals, speaking Hungarian and another language, from neighboring countries (Croatia, Serbia, Romania, Slovakia, Ukraine and Austria). The participants indicated in the pre-training questionnaire that they did not receive any training on corpus use during their pre-service years, nor since, and were not familiar with corpora. All participants were teaching lower proficiency level learners (A1, A2 and B1 levels of the Common European Framework of Reference for Languages) and enrolled to the course to find out more about state-of-the-art teaching methods. The majority of them (31 participants) had been teaching for at least five years, 4 participants had just started their carrier and 13 had been teaching between one and five years. Most of them reported that they were using the *MagyarOK* corpus-informed textbooks and supplementary online teaching materials (Szita, Pelcz 2013-2019), but none of them had used the pedagogical corpus that accompanies the teaching material before the training. The two instructors, Katalin Pelcz and Szilvia Szita, are authors of the *MagyarOK* teaching materials, practicing language teachers, linguists and teacher trainers with a strong interest in the applicability of corpus linguistic findings to language teaching.

The first day (six full hours) was dedicated to the presentation of and practice with the afore-mentioned pedagogical corpus and, to a lesser extent, to the work with two large general corpora. The training followed a four-step approach, each step including phases of reflection on language use and various corpus activities (see Section 4). In this way, teachers gradually familiarised themselves with corpora as well as with some corpus analysis tools and relevant findings from corpus linguistics. They could also see how corpus use could make their lives easier. On the other training days, corpora were used whenever they offered the best solution to a specific question. During the entire training, corpus work was integrated into broader issues such as (1) teaching vocabulary (collocations, synonyms, words with multiple meanings), (2) working with (semi-) authentic texts

to produce natural-sounding utterances in the language classroom, (3) improving students' written and oral production and (4) teaching grammar (especially word order and the two conjugations, two particularities of the Hungarian language that are considered to be particularly difficult to learners). It is important to note that we were cautious not to challenge beliefs about grammar rules right from the outset and we only occasionally invited teachers to verify grammar rules based on corpus data. While such exercises are undoubtedly meaningful in the training context as they raise teachers' awareness of possible deviations between grammar descriptions and real language use (Breyer 2011; Farr 2008; Heather, Helt 2012), this choice was motivated by previous training experiences and by the fact that «canonical» teaching methodologies for Hungarian are still strongly grammar-oriented. Losing sight of this was likely to provoke some resistance on the part of the participants. Instead, we focused on the interconnection between lexis and grammar, and on selected language-related questions.

3.2 Software, tools, and corpora

3.2.1 Software and tools

A major difference to the courses cited in Section 2.1 was that this training focused on teaching with specific, corpus-informed teaching materials and a particular corpus instead of presenting an overview of existing software for corpus building and analysis, and related activities.

As many studies address the demotivating nature of technical difficulties (e.g., Farr 2008; Leńko-Szymańska 2017), we decided against presenting software that allows teachers to compile their own corpus. It seemed unlikely that participants without any previous knowledge about corpus use would wish to build their own corpora after just one introductory training. Furthermore, most open-source software does not work perfectly for Hungarian².

Summarizing state-of-the art research, Poole (2022, 10) points out that «[c]orpus study is often viewed as appropriate only for more advanced learners for it is perceived as potentially too difficult and demanding for learners of

² It is due to the large number of different accents on vowels on the one hand, and to the morphologically complex, agglutinating nature of the language on the other hand.

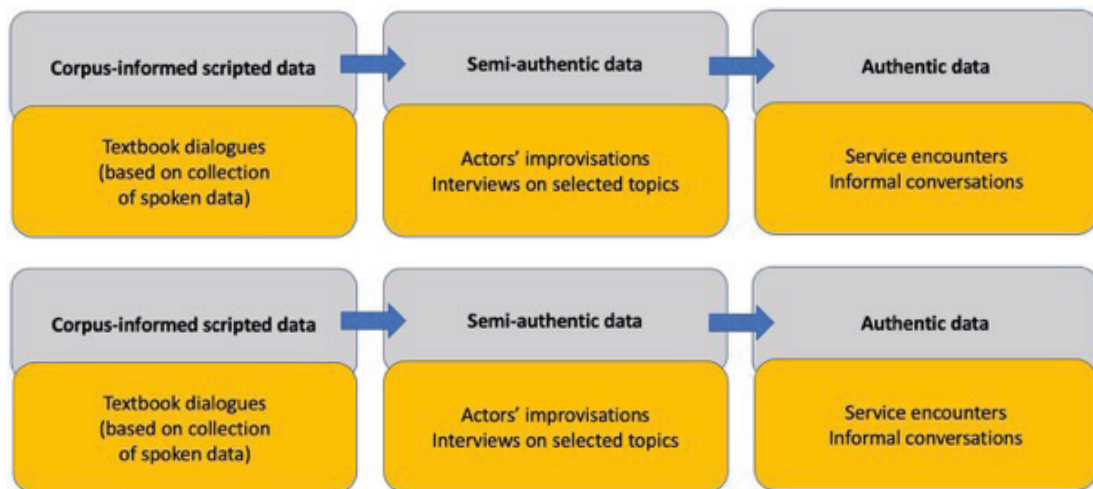
lower proficiency levels». It is true that large corpora include a number of occurrences that are rather difficult to understand due to their complexity, idiomaticity and/or the lack of the broader textual environment and situational context. Moreover, many utterances contain inappropriate content such as problematic views and opinions, making them unsuitable as authentic examples of language use for language teaching purposes. The emphasis was therefore placed on the exploration of the *MagyarOK* open pedagogical corpus hosted on Sketch Engine (Szita 2020). The interface of Sketch Engine is well-structured, user-friendly and easy to work with. From a range of available corpus analysis tools only three, the Concordancer, Word Sketch and Wordlist, were presented to the participants in order to simplify their first encounter with such tools. Participants were also presented the *huTenTen21* large general corpus on Sketch Engine³ and the Hungarian National Corpus (Magyar Nemzeti Szövegtár) which is less intuitive to use than Sketch Engine but freely accessible. These large corpora were consulted to identify frequently used collocations and to answer questions to which traditional grammars and dictionaries do not offer satisfying responses (see Section 4.4).

3.2.2 The *MagyarOK* open pedagogical corpus

The *MagyarOK* open pedagogical corpus offers several subcorpora for the same proficiency level, with progressively increasing complexity. The first and most accessible collection for each level comprises corpus-informed edited language, i.e., adapted authentic and natural-sounding texts. The second subcorpus contains semi-authentic data and is organized by topic. It includes non-scripted improvisations by actors, interviews and texts by (teacher and non-teacher) native speakers on selected topics. The third subcorpus offers fully authentic input. This three-step approach ensures that the linguistic content of the corpora is relevant and accessible to the learners right from the beginning but does not remain limited to the texts presented in the course books. The corpus can be searched by book chapter, type of data and/or level. The following table shows an overview of the different subcorpora.

³ This corpus was freely accessible in the educational context until March 2022, thus, during the time of the training.

Spoken corpus



Written corpus

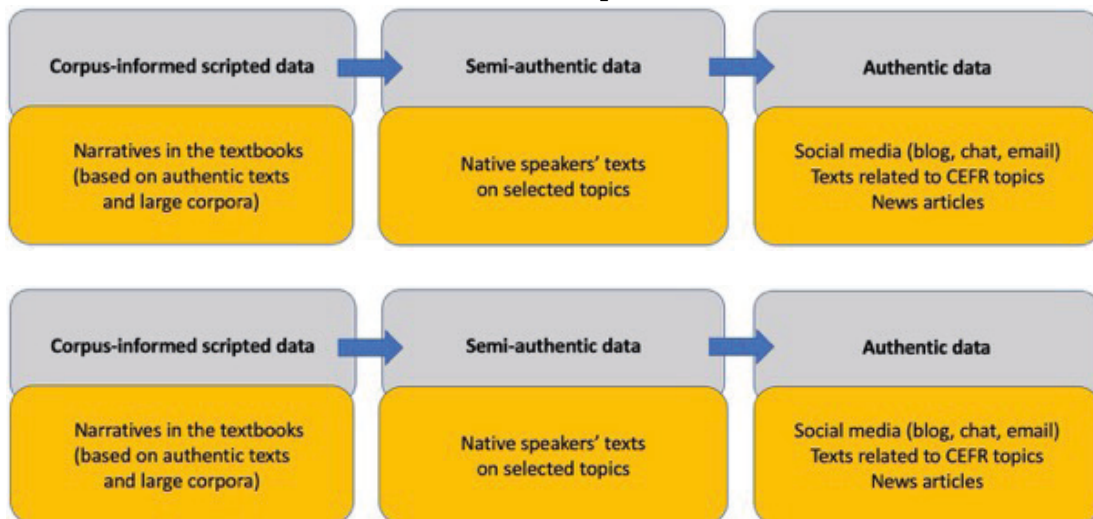


Table 1 – Subcorpora of the *MagyarOK open pedagogical corpus* (A1-B2 levels).

The written collection contains about 710,000 tokens at present and the spoken collection (A1 to B1 level) counts about 262,200 tokens (more than 38 hours of recorded material, approximately 115 token/minute) at present. Both collections are continuously being expanded. This open pedagogical corpus is relatively large, which has two major advantages: (1) repetitions and variations across the texts become noticeable and (2) collocation lists with frequently occurring linguistic items are fairly reliable from a statistical point of view⁴.

4. Activities

During the training, participants familiarized themselves with (1) tools allowing them to prepare activities and worksheets for their lessons, (2) hands-on activities for the language learning classroom, (3) ways of using the corpus to improve the quality of the learners' linguistic products, and (4) finding answers to specific linguistic questions.

A clear advantage of a training for in-service teachers was that most participants had an extensive teaching experience and they were aware of learners' needs at lower proficiency levels. They could therefore evaluate the usefulness of the activities as well as identify potential issues.

The next section presents the four stages of corpus exploration. This program was carried out on the very first day which was dedicated to corpus work. On other training days, corpora and corpus-based activities were used whenever they seemed appropriate, as mentioned previously. This allowed participants to gain more practice with corpora, identify useful applications and create exercises in small groups. When designing the training, we relied on recommendations and existing activities such as Johns (1991a, 1991b), Tribble and Jones (1990), Sinclair (2003, 2004a) and Poole (2018), adapted for Hungarian and complemented with new ones.

⁴ Search for collocations with frequent items in this corpus produces very similar results to large authentic corpora. The results concerning less frequently occurring items cannot, however, be considered as statistically reliable.

4.1 Awareness-raising activities (Step 1)

The training started with awareness-raising activities in which participants were invited to reflect on real language use and observe its characteristics. Activities included identification of multi-word units in authentic texts and noticing that these represented the core of many utterances. Participants also listened to genuine everyday-life interactions and observed conversational strategies (turn-taking, asking back, hesitation, self-correction, etc.) as well as frequently used vocabulary items. As language learners are expected to participate in simple but authentic everyday-life interactions right from the outset, the teachers understood the importance of these tasks immediately. Follow-up activities carried out with corpus tools allowed to raise teachers' awareness of the interconnection of vocabulary and grammar and to introduce them to the concept of linguistic patterns. For this, they studied concordance lines and collocations with selected language items. Finally, practitioners adopted the learners' position and explored short texts as well as concordance lines in a foreign language of their choice⁵.

These activities prepared the ground for the next step: they made teachers aware of situations in which corpora can be of valuable help in their daily work. For example, they made practitioners realize that authentic spoken interactions contain fairly simple vocabulary, thus, the difficulty is not necessarily related to linguistic complexity but rather to the fact that simple but highly frequent words may have multiple meanings (e.g., participants found that the simple word *jó* whose basic meaning is 'good' displayed at least 18 different meanings in the conversations observed). Such findings raised their interest to explore common words in depth and challenged some of their beliefs about what teaching basic vocabulary at lower proficiency levels implies. These activities also raised the participants' awareness for the need of models for the students' linguistic production, i.e., the necessity of proposing at least a few examples of native speakers' likely utterances in a given situation. Corpora (predominantly the pedagogical corpus) were

⁵ For English, we used Sinclair's examples «budge» (Sinclair 1998) and «naked eye» (Sinclair 2004b), and Stubbs's example of «cause» (2009) for the analysis of multi-word units and their grammatical, lexical, and semantic patterns.

perceived as potential sources of such models. Last but not least, these activities drew the teachers' attention to the importance of noticing activities in the language lessons (cf. Schmidt 1990).

4.2 Exploring the pedagogical corpus (Step 2)

From this step onwards, participants carried out familiar tasks with corpus tools before moving on to new activities. In Step 2, the instructors first showed in what ways the *MagyarOK* textbooks and supplementary materials were corpus-informed. For example, they explained how data for the books was collected and adapted, how core vocabulary was identified and multi-word units were selected. They also pointed out that the books contained activities in which sentences were presented similar to concordance lines to help learners to notice patterns.

In the next phase, the open pedagogical corpus was presented as the first corpus to explore at lower proficiency levels. The first activities simply showed teachers how they could use the corpus as a collection of examples, download examples and create various exercises based on them. These included, among others, gap-filling tasks (Johns 1991b) with follow-up activities such as changing one item per sentence, reacting to the information, etc. Afterwards, several activities to review vocabulary were presented; for example, demonstrations were given on how learners could review the most frequent adjectives, verbs or nouns of a given chapter with the help of the Wordlist tool and the Concordancer and learn their most frequent collocates using Word Sketch.

By this time, participants were, at least to some degree, familiar with three basic tools (Concordancer, Word Sketch and Wordlist) and could use them on their own. In the next phase, they were invited to think about how they can explore the open pedagogical corpus together with their learners. As the starting point, participants formed small groups and collected questions frequently asked by their learners. Then, they decided in small groups which questions could be meaningfully explored in the corpus. They observed, among others, (1) the various meanings of frequently occurring words such as *jó* 'good', 'fine', 'alright' and *jár* 'go', 'go regularly', 'circulate', (2) the textual environment of topic-related keywords, (3) the use of

selected suffixes and prefixes, (4) word order tendencies (the position of parts of speech depending on the context), and (5) the use of words without an exact equivalent in other languages such as *pedig* that could mean 'and', 'or' or 'although'. Finally, the exploration process and findings were presented to the entire group.

As homework for the training week, participants were invited to create activities, worksheets and collections of examples for their lessons that were shared on the last training day and, at a later stage, between the two training groups. In this manner, participants departed with some ready-to-use collocation lists, examples and activities to implement into their lessons. As Section 5.3 will show, this was one of the most appreciated aspects of the training.

4.3. Improving the quality of learners' linguistic products (Step 3)

The next step focused on a different application of the open pedagogical corpus with the primary purpose of improving the quality of learners' linguistic products. For this, Subcorpus 2 of the written corpus, containing topic-related texts written by native speakers, was used. This corpus was presented as *a collection of model utterances* provided by expert language users that learners could choose from when writing their own texts (see also Kennedy, Miceli 2010; Szita, Pelcz, forthcoming for detailed methodology). These texts reviewed and expanded the vocabulary covered in a selected textbook chapter and integrated it in a natural-sounding text. Learners can read the texts in their entirety to mark multi-word units that they wish to «recycle» in their narratives and use corpus tools to identify frequently occurring items. We showed examples of students' texts written with and without models, and invited participants to compare them for naturalness. Participants agreed that recycling lexical items (high number of correctly used multi-word units, cohesive devices, etc.) from expert users' utterances improved the naturalness of learners' texts significantly. Reactions to this finding were, however, mixed. Some teachers saw «fitting ready-made phrases into their texts» as a serious limitation on learners' creative language use or even plagiarism. One of them said that she felt «as if someone had already done the work» for her students. It required some time to convince her that it was actually the native

speakers' community that «has done the work» for them by establishing their preferred multi-word units.

Furthermore, participants expressed their worries that students would just copy and paste some sentences and phrases without actually learning them. These are indeed valid concerns and participants were invited to discuss how to reduce potential risks.

4.4 Large corpora to identify collocations and answer linguistic questions (Step 4)

In the last step, participants explored two larger corpora (*huTenTen12* on Sketch Engine and the Hungarian National Corpus). This step served two purposes: (1) teachers were shown how corpora could help them identify useful multi-word units and (2) how they could find answers to questions that dictionaries and grammar books fail to address or treat only partially.

While the knowledge of multi-word units (collocations) is essential for natural-sounding language (e.g., O'Keeffe, McCarthy, Carter 2007; Pawley, Syder 1983; Sinclair 1991), their use is far from simple in a foreign language (Nesselhauf 2005; Nation 2013). Such units are therefore particularly suitable to demonstrate the utility of large general corpora. This step began with an analysis of learners' written and spoken products; participants were asked to identify mistakes in these texts and make a list of them by frequency. They realized that many of these errors were related to difficulties using appropriate multi-word units such as incorrect modifier preceding a noun, verb or adjective, errors in noun-verb collocations, incorrect adjective in front of a noun, etc. Then, participants were invited to discuss possible ways of working on multi-word units in their lessons. Since they all agreed that explicit teaching of multi-word units was a meaningful activity for the classroom, they were invited to use the Word Sketch tool on the *huTenTen12* corpus. As homework, collocation lists were prepared by the participants for ten frequent nouns, adjectives, and verbs from the pedagogical corpus, with adapted examples. The results were also shared with the entire group as well as participants across the two training groups.

Participants were also invited to encourage their students to explore large corpora with the collocation generator as, once they learned how to

use it, they could look for frequently occurring collocations on their own. Most teachers, however, did not think that this task would be appropriate for lower proficiency levels due to the complexity of the examples and, when examples were not consulted, to difficulties related to the learnability of some multi-word units without a broader context.

In the second phase of Step 4, the two large corpora were used to answer linguistic questions to which dictionaries and grammar only gave partial answers. First, participants explored the use of the synonyms *tűnik* and *látszik* (~ 'seem' and 'appear'), then that of the two conjugations (a particularity of the Hungarian language)⁶. These tasks aimed at demonstrating that lexis and grammar were not separable and participants were invited to identify both vocabulary and grammar patterns related to both phenomena. The analyses were carried out with the help of the Concordancer and the respective collocation generator of the two large corpora.

This was the only step in which teachers were asked to analyze complex data on their own. This was also the only step that entirely failed its purpose as most participants doubted whether they would ever be able to interpret the data correctly. While this is in accordance with Farr's findings (2008) that teachers find corpus analysis a challenging task and do not always feel confident when it comes to interpreting the results, the participants' reactions were rather unexpected since, despite their doubts, they *did manage* to interpret corpus data correctly. They were, above all, confused when their findings challenged widely accepted rules presented in grammar books and, although they carried out similar research on most training days, their perception about their analytical skills did not change until the end of the training.

⁶ Hungarian uses the so-called indefinite conjugation when the sentence does not contain a direct object, or the direct object is indefinite (such as direct objects preceded by the indefinite article). The definite conjugation is used with definite direct objects (such as direct objects preceded by the definite article).

5. Participants' feedback

5.1 Questions

Participants were invited to discuss their experiences on the last training day as well as to complete a feedback sheet after the training. Out of 48 teachers, 44 completed the questionnaire that contained the following three questions related to corpus use:

- (1) On a scale of 1 to 5, how useful do you consider (pedagogical and large linguistic) corpora for language teaching (1=not useful at all, 2=rarely useful, 3=useful from time to time, 4=generally useful, 5=very useful)?
- (2) What do you think are the key benefits of (pedagogical and large linguistic) corpora for teaching lower proficiency levels?
- (3) Would you be interested in participating in other training sessions related to corpus use in the classroom? (yes/no) If yes, what would you like to explore in more detail?

We provided some answers to question 2 that participants could tick and/or complete with other perceived benefits. They could also comment upon every question and explain their point of view more precisely. We encouraged them to write down all their thoughts (or send us a voice message if they found it easier) so that we could improve the training contents and the design of the open corpus if needed. Many participants gave very detailed answers that will be presented in the next paragraphs.

5.2 Key benefits of corpora for language teaching

24 participants considered corpora to be very useful for their teaching, 12 participants found corpora useful, and 8 participants thought that corpora could be useful from time to time. Availability of the complete texts in the datasets was also seen as a valuable asset. According to the participants, it ensured that «the broader textual environment of the utterances can always be consulted» (Participant 12) and that «any text could be used as linguistic input, for example for discussions» (Participant 34). Answers to Question 2 about the key benefits of pedagogical corpora, are presented in the table below (multiple answers were possible).

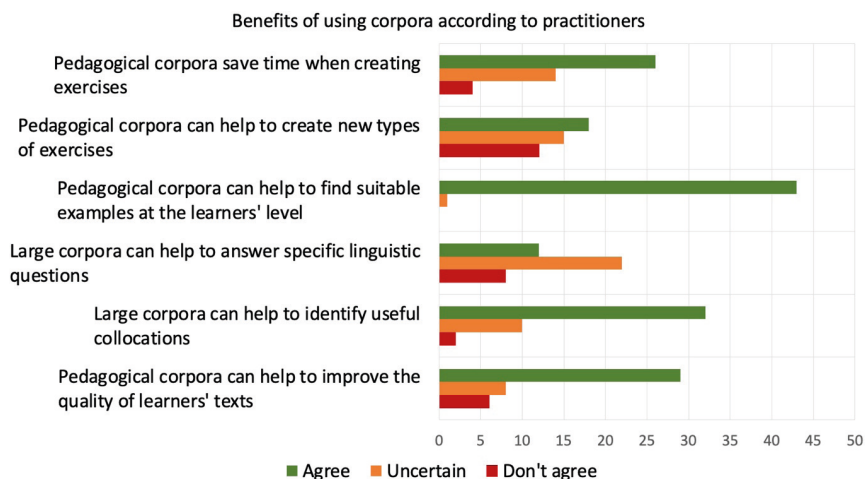


Table 2 – Key benefits of using corpora according to participants.

Additional comments about the key benefits included: «I can use the texts without much preparation beforehand» (Participant 15, similar thoughts were expressed by five other participants), «It's good that you (the authors of the open corpus) thought of providing texts that are in accordance with the level descriptions of the CEFR» (Participant 22, similar thoughts were expressed by three other participants). Including textbook contents and supplementary materials in the open corpus was appreciated by many participants: «It's great that the coursebooks are in the corpus» (Participant 37) and «It's really good that the starting point for this corpus are the coursebooks» (Participant 28, similar thoughts were expressed by eleven other participants). Other comments mentioned the usefulness of the examples: «My students can read many texts about a given topic and explore them with corpus tools» (Participant 18, similar thoughts were expressed by two other participants), «Nice to have so many examples that my students can understand» (Participant 42), «the pedagogical corpus can provide examples for various questions at my learners' level» (Participant 27). Participants also raised some issues related to the transparency of the corpus: «The corpus could be annotated for topics so that students can find

all relevant texts about a topic easily» (Participant 37, similar thoughts were expressed by four other participants), «There are some questions that the pedagogical corpus cannot provide answers for because it contains too few or no examples. The problem is that you never know beforehand if this is going to be the case» (Participant 45).

5.3 Comments about corpora

While the work with the *MagyarOK* open pedagogical corpus was generally appreciated, the two large corpora as sources of linguistic input generated some distrust. Practitioners stated almost unanimously (39 out of 44) that they would be reluctant to illustrate linguistic phenomena with example sentences from large general corpora because most utterances were «too difficult or too weird» (Participant 31) for learners to understand in their original form. Adapting them was considered to be «a time-consuming activity of little benefit to the learner. I could invent natural-sounding examples much faster» (Participant 42, similar thoughts expressed by eight other participants). These answers alluded to the time-consuming nature of finding appropriate examples for lower proficiency levels in a large general corpus. Other answers included concerns regarding complex vocabulary difficult to recontextualize and possible mistakes: «The sentences in these corpora are far too complex. My students could not understand it.» «Sometimes even I cannot figure out the meaning of the utterances. How could I expect from my learners that they will?» (Participant 19, similar thoughts expressed by three other participants) «I get irritated by the masses of linguistic errors in this corpus. I do not want my learners to see such sentences, at least not at beginner level» (Participant 43), «I think that I would need a lot of time to prepare activities based on these data and I would still show my students selected (and perhaps even adapted) sentences only. What would then be the benefit compared with the open corpus or examples that I create for them on the spot?» (Participant 35).

While large corpora as sources of examples were, thus, generally rejected, practitioners expressed a keen interest in the Word Sketch tool that allowed them to identify frequent multi-word units. The enthusiasm for this tool was due to the fact that the degree of difficulty to understand

two- to three-word units was estimated to be lower than the comprehension of longer extracts (sentences) from the corpus and the usefulness of these multi-word items was estimated to be high.

5.4 Other perceived benefits

Participants emphasized that a great advantage of the datasets and tools was that they allowed them to create more effective exercises (gap-filling, sentence transformation, using the written texts and multimedia material for reading and listening comprehension, etc.) while significantly reducing preparation time. Interestingly, whereas most participants agreed that using the *MagyarOK* open pedagogical corpus could save time when writing exercises and they felt that they could link corpus tools to their existing practice, many of them also admitted that they found data analysis too challenging. The principal reason for this was perceived lack of corpus analytic skills: «I don't think I could identify the right patterns to practice» (Participant 17, similar thoughts were expressed by seven other participants), «First, I need to become more familiar with the tools» (Participant 38, similar thoughts were expressed by five other participants). Six participants mentioned lack of time for preparing not only corpus-informed exercises but any kind of activities due to an already over-charged curriculum. The answers are in line with the teachers' self-perception who said during the training that they saw themselves principally as mediators who provided students with effective teaching materials and activities that were prepared by «experts» such as linguists, material writers and teacher trainers. Questioning their own competences can, thus, explain their reluctance to create materials in an «unsupervised» fashion, i.e., with the help of corpora but without an «expert» looking over their shoulders.

35 out of 44 participants reported that one of the most appreciated activity during the training was to create worksheets and multiword-unit lists together with colleagues. The fact that the results of this collaboration were validated by the instructors gave participants a feeling of security that «nothing can go wrong» (Participant 39). They also expressed the wish that the instructors provide them with ready-made materials: «I would be happy to use more materials based on corpora.» (Participant 32, similar thoughts

were expressed by six other participants). These reactions confirm that more training is needed to increase teachers' confidence in their abilities of creating materials of high quality on their own.

5.5 Further training on corpus use

The last question whether participants would follow another training on corpus use was met with enthusiasm. All teachers but one (43 out of 44) answered it with «yes». Participant 44 indicated that she felt sufficiently prepared for using corpora in her classroom occasionally. 23 participants expressed their wish to improve their skills to interpret corpus data correctly and fast. As one of the participants put it: «I'm never sure what exactly I see» (Participant 12) and 26 participants were interested in enriching their repertoire of corpus-based and corpus-informed activities.

It is interesting to note that major difficulties identified by other researchers (see Section 1.1) such as the time-consuming nature of many corpus-related tasks (building corpora, create exercises and activities), the lack of appropriate corpora and the lack of user-friendly, easy-to-use interfaces were not explicitly mentioned by our participants. The explanation probably lies in the fact that, contrary to other trainings, participants were not introduced to free software to build their own corpus, nor were they invited to work with large general corpora extensively. Concerns about the nature and linguistic accessibility of the data were also not voiced, presumably for the same reason, but participants questioned the usability of large general corpora with lower proficiency levels. As pedagogical datasets have serious limitations in terms of statistical reliability, it is, however, important in my view that such introductory training be complemented by follow-up courses on the use of large corpora.

6. Conclusions

The 30-hour course presented in this chapter was held for in-service teachers of Hungarian as a foreign or second language. It intended, above all, to provide teachers of lower proficiency levels with state-of-the-art pedagogical tools and methods as well as some relevant findings from

linguistic research. Parts of the training dedicated to corpus use aimed to show practitioners how pedagogical corpora can be used (1) to optimize lesson preparation by creating more effective activities (in less time) for the classroom; (2) to facilitate vocabulary review and consolidation by increasing the number of learners' encounters with relevant lexical items and (3) to help learners to improve the linguistic quality of their texts. Finally, large general corpora were used (4) to identify frequently occurring multi-word units (collocations) and (5) to answer selected linguistic questions. Practitioners, thus, learned how corpus work could be integrated into their existing toolkit and enrich their repertoire in the interest of effective language teaching. Throughout the training, they could explore the broad spectrum of corpus applications from multiple angles.

The overall positive feedback of participants shows that in-service teachers appreciate training on corpus use. The fact that the training focused on exploring a carefully designed pedagogical corpus for lower proficiency levels, could account for participants not mentioning the usual problems identified in previous research, such as the time-consuming nature of corpus work, the user-unfriendliness of corpus software and the lack of appropriate resources. Most participants, however, doubted whether they could ever be able to interpret large corpus data correctly and questioned their ability of performing systematic and reliable analyses on their own. Their reaction suggests that, in accordance with Leńko-Szymańska's findings (2017), regular guided corpus work is needed in teacher education (ideally both in pre-service education programs and in continuous in-service trainings) to ensure that teachers receive continuous practice and sufficient exposure to corpus data. This can allow them to build up expertise as well as confidence in their own analytic skills.

It is also clear from the participants' feedback that a one-week training on state-of-the-art methodology, even if it includes a fair amount of corpus work is far from enough. Follow-up trainings could present more activities for discovery learning, detailed analyses of large corpus data and tasks for increased learner autonomy. Providing a list of activities for lower proficiency levels that teachers can easily implement with the help of a pedagogical corpus, could also be a way to incite them to integrate more such exercises into their lessons. Developing materials in small groups,

together with other participants could also be an asset of future teacher trainings. Finally, close collaboration between linguists and practitioners could contribute to building more on-target corpora for the classroom and to creating a consistent, corpus-informed language pedagogy.

Works cited

- André, Virginie. 2020. «Faire de la linguistique de corpus avec des apprenants de français langue étrangère» [Doing corpus linguistics with learners of French as a foreign language]. In *La didactisation du français vernaculaire*, edited by Pierre Larrivée and Florence Lefeuvre, 37-66. Caen: Presses Universitaires de Caen.
- 2019. «Des corpus oraux et multimodaux authentiques pour acquérir des compétences sociolangagières» [Authentic oral and multimedia corpora for the acquisition of sociolinguistic competences]. In *Variation, plurilinguisme et évaluation en français langue étrangère*, edited by Laurent Gajo et al., 209-223. Frankfurt am Main: Peter Lang.
- 2017. «Un corpus multimédia pour apprendre à interagir en situations universitaires en France» [A multimedia corpus for interactions in higher education]. *Actes du troisième colloque international de l'ATPF «Enseigner le français: s'engager et innover»* du 19 au 20 octobre 2017, 292-315. Bangkok: Association thaïlandaise des professeurs de français.
- Aston, Guy. 2001. *Learning with corpora*. Houston: Athelstan.
- Conrad, Susan, Geoffrey Leech, Douglas Biber. 2002. *Longman Student Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Boulton, Alex. «Data-driven learning: Taking the computer out of the equation». *Language Learning* vol. 60, n. 3 (2010): 534-572. DOI: <<https://doi.org/10.1111/j.1467-9922.2010.00566.x>>.
- . 2012. «Hands-on/hands-off: Alternative approaches to data-driven learning». In *Input, Process and Product. Developments in Teaching and Language Corpora*, edited by James Thomas and Alex Boulton, 153-169. Brno: Masaryk University Press.
- Boulton, Alex, Tom Cobb. «Corpus use in language learning: A meta-analysis». *Language Learning* vol. 67, n. 2 (2017): 348-393. DOI: <<https://doi.org/10.1111/lang.12224>>.
- Boulton, Alex, Henry Tyne. 2014. *Des documents authentiques aux corpus: démarches pour l'apprentissage des langues* [From authentic documents to corpora: Initiatives for language learnings]. Paris: Didier.

- Braun, Sabine. «Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora». *ReCALL* vol. 19, n. 3 (2007): 307-328. DOI: <<https://doi.org/10.1017/S0958344007000535>>.
- . 2010. «Getting past Groundhog day. Spoken multimedia corpora for student-centred corpus exploration». In *Corpus Linguistics in Language Teaching*, edited by Tony Harris and Maria Moleno Jaén, 75-98. Frankfurt am Main: Peter Lang.
- Breyer, Yvonne. «Learning and teaching with corpora: Reflections by student teachers». *Computer Assisted Language Learning* vol. 22, n. 2 (2009): 153-172. DOI: <<https://doi.org/10.1080/09588220902778328>>.
- . 2011. *Corpora in Language Teaching and Learning: Potential, Evaluation, Challenges*. Frankfurt am Main: Peter Lang.
- Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics. A Practical Guide*. Cambridge: Cambridge University Press.
- Çalışkan, Gamze, Safiye İpek Kuru Gönen. «Training teachers on corpus-based language pedagogy: Perceptions on vocabulary instruction». *Journal of Language and Linguistics Studies* vol. 14, n. 4 (2018): 190-210.
- Callies, Marcus. 2016. «Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English». In *Angewandte Linguistik in Schule und Hochschule. Neue Wege für Sprachunterricht und Ausbildung*, edited by Rolf Kreyer, Steffen Schaub, and Barbara Ann Güldenring, 391-415. Frankfurt am Main: Peter Lang.
- . 2019. «Integrating corpus literacy into language teacher education». In *Learner Corpora and Language Teaching*, edited by Sandra Götz and Joybrato Mukherjee, 245-263. Amsterdam: John Benjamins.
- Carter, Ronald et al. 2016. *English grammar today: An A-Z of spoken and written grammar*. Cambridge: Cambridge University Press.
- Chambers, Angela. «Towards the corpus revolution? Bridging the research-practice gap». *Language Teaching* vol. 52, n. 4 (2019): 460-475. DOI: <<https://doi.org/10.1017/S0261444819000089>>.
- Charles, Maggie. «Getting the corpus habit: EAP students' long-term use of personal corpora». *English for Specific Purposes* vol. 35, n. 1 (2014): 30-40. DOI: <<https://doi.org/10.1016/j.esp.2013.11.004>>.
- Cobb, Tom, Alex Boulton. 2015. «Classroom applications of corpus analysis». In *The Cambridge handbook of English Corpus Linguistics*, edited by Douglas Biber and Randi Reppen, 478-497. Cambridge: Cambridge University Press.

- Conrad, Susan, Douglas Biber. 2009. *Real Grammar. A Corpus-Based Approach to English*. New York: Pearson Education.
- Crosthwaite, Peter. 2019. *Data-Driven Learning for the Next Generation. Corpora and DDL for Pre-tertiary Learners*. London: Routledge.
- Ebrahimi, Alice, Esmali Faghieh. 2017. «Integrating corpus linguistics into online language teacher education programs». *ReCALL* vol. 1, n. 1 (2017): 1-16. DOI: <<https://doi.org/10.1075/scl.92.12cal>>.
- Farr, Fiona. «Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users». *Language Awareness* vol. 17, n. 1 (2008): 25-43. DOI: <<https://doi.org/10.2167/la414.0>>.
- Flowerdew, Lynne. «Applying corpus linguistics to pedagogy. A critical evaluation». *International Journal of Corpus Linguistics* vol. 14, n. 3 (2009): 393-417. DOI: <<https://doi.org/10.1075/ijcl.14.3.05flo>>.
- Frankenberg-Garcia, Ana. «Raising teachers' awareness of corpora». *Language Teaching* vol. 45, n. 4 (2012): 475-489. DOI: <<https://doi.org/10.1017/S0261444810000480>>.
- . «The use of corpus examples for language comprehension and production». *ReCALL* vol. 26, n. 2 (2014): 128-146. DOI: <<https://doi.org/10.1017/S0958344014000093>>.
- Friginal, Eric. 2018. *Corpus Linguistics for English Teachers. Tools, Online Resources, and Classroom Activities*. London: Routledge.
- Heather, Julian, Marie Helt. «Evaluating corpus literacy training for pre-service language teachers: six case studies». *Journal of Technology and Teacher Education* vol. 20, n. 4 (2012): 415-440.
- Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, Susan, Gill Francis. 2000. *Pattern grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
- Johns, Tim. «Should you be persuaded: Two samples of data-driven learning materials». *English Language Research Journal* vol. 4 (1991a): 1-16.
- . 1991b. «From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning». In *Perspectives on pedagogical grammar*, edited by Terence Odlin, 293-313. Cambridge: Cambridge University Press.
- Kennedy, Claire, Tiziana Miceli. «Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource». *Language Learning and Technology* vol. 14, n. 1 (2010): 28-44. URL: <<http://llt.msu.edu/vol14num1/kennedymiceli.pdf>> (open access).

- . «Cultivating effective corpus use by language learners». *Computer Assisted Language Learning* vol. 30, nn. 1-2 (2017): 1-24. DOI: <<https://doi.org/10.1080/09588221.2016.1264427>>.
- Leńko-Szymańska, Agnieszka. «Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education». *ReCALL* vol. 26, n. 2 (2014): 260-278. DOI: <<https://doi.org/10.1017/S095834401400010X>>.
- . «Training teachers in data-driven learning: Tackling the challenge». *Language Learning and Technology* vol. 21, n. 3 (2017): 217-241. DOI: <<https://doi.org/10.125/44628>>.
- Ma, Qing, Jinlan Tang, Shanru Lin. «The development of corpus-based language pedagogy for TESOL teachers: a two-step training approach facilitated by online collaboration». *Computer Assisted Language Learning* vol. 35, n. 9 (2021): 2731-2760. DOI: <<https://doi.org/10.1080/09588221.2021.1895225>>.
- Magyar Nemzeti Szövegtár* [Hungarian National Corpus]. URL: <mnsz.nyelvtud.hu>.
- MagyarOK nyílt pedagógiai korpusz* [MagyarOK open pedagogical corpus, levels A1 to B2 (old version)]. URL: <<https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fmagyarok>> (*open access*).
- McCarthy, Michael. «Assessing and interpreting corpus information in the teacher education context». *Language Teaching* vol. 41, n. 4 (2008): 563-574. DOI: <<https://doi.org/10.1017/S0261444808005247>>.
- McCarthy, Michael, Jeanne McCarten, Helen Sandiford. 2005-2011. *Touchstone*. Cambridge: Cambridge University Press.
- . 2012-2013. *Viewpoint*. Cambridge: Cambridge University Press.
- Meunier, Fanny, Céline Gouverneur. 2009. «New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material». In *Corpora and Language Teaching*, edited by Karin Aijmer, 179-201. Amsterdam: John Benjamins.
- Meunier, Fanny, Randi Reppen. 2015. «Corpus versus non-corpus-informed pedagogical materials: Grammar as the focus». In *The Cambridge Handbook of English Corpus Linguistics*, edited by Douglas Biber and Randi Reppen, 498-514. Cambridge: Cambridge University Press.
- Nation, Paul. 2013. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nesselhauf, Nadja. 2005. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.

- O’Keeffe, Anne, Michael McCarthy, Ronald Carter. 2007. *From Corpus to Classroom. Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- Pawley, Andrew, Frances Hodgetts Syder. 1983. «Two puzzles for linguistic theory: native-like selection and native-like fluency». In *Language and communication*, edited by Jack Richards and Robert Schmidt, 191-266. London: Longman.
- Poole, Robert. 2018. *Using Corpora: A Guide for English Language Learners*. Edinburgh University Press.
- . «“Corpus can be tricky”: revisiting teacher attitudes towards corpus-aided language learning and teaching». *Computer Assisted Language Learning* vol. 35, n. 7 (2022): 1620-1641. DOI: <<https://doi.org/10.1080/09588221.2020.1825095>>.
- Römer, Ute. «Corpus Research Applications in Second Language Teaching». *Annual Review of Applied Linguistics* vol. 37 (2011): 205-225. DOI: <<https://doi.org/10.1017/S0267190511000055>>.
- Schmidt, Richard. «The role of consciousness in second language learning». *Applied Linguistics* vol. 11 (1990): 129-158.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- . 2003. *Reading Concordances. An Introduction*. London: Pearson-Longman.
- , edited by. 2004a. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- , edited by. 2004b. *Trust the Text*. London: Routledge.
- Stefanowitsch, Anatol. 2020. *Corpus Linguistics. A Guide to the Methodology*. Berlin: Language Science Press.
- Szita, Szilvia. 2014. «Invent content, not language. Meaningful interaction and natural language use in the classroom». In *Kleine Sprachen, was nun? Studies on language and culture in Central and Eastern Europe*, edited by Rita Hegedűs and Tamás Görbe, 112-127. München: Kubon und Sagner Verlag.
- . «Korpuszépítés és korpuszhasználat alacsonyabb nyelvtudási szinteken [Creating and using corpora at lower proficiency levels]». *Hungarológiai Évkönyv* vol. 21, nn. 1-2 (2020): 173-179. URL: <<https://epa.oszk.hu/02200/02287/00021/pdf/>>.
- Szita, Szilvia, Katalin Pelcz. 2013-2019. *MagyarOK A1-B2*. Pécs: University of Pécs.
- . 2020-. *MagyarOK nyílt pedagógiai korpusz [MagyarOK open pedagogical corpus]*. Online as «*MagyarOK Teaching Materials for Hungarian, Levels A1 to B2*» (old version). URL: <<https://app.sketchengine.eu/#dashboard?corpname=preloade d%2Fmagyarok>> (open access).

- . (forthcoming). *A modellalapú nyelvtanítás kézikönyve* [The handbook of model-based language teaching]. Pécs: Institute of Model-based Language Teaching.
- Timmis, Ivor. 2015. *Corpus Linguistics for ELT Teachers*. London: Routledge.
- Tribble, Christopher. 2015. «Teaching and language corpora. Perspectives from a personal journey». In *Multiple Affordances of Language Corpora*, edited by Alex Boulton and Agnieszka Leńko-Szymańska, 37-62. Amsterdam: John Benjamins.
- Tribble, Christopher, Glyn Jones. 1990. *Concordances In The Classroom. A Resource Book For Teachers*. London: Longman.
- Tyne, Henry. 2012. «Corpus work with ordinary teachers: Data-driven learning activities». In *Input, Process and Product: Developments in Teaching and Language Corpora*, edited by James Thomas and Alex Boulton, 114-129. Brno: Masaryk University Press.
- Widdowson, Henry. 1978. *Teaching Language as Communication*. Oxford: Oxford University Press.
- . 2003. *Defining Issues in English Language Teaching*. Oxford: Oxford University Press.
- Willis, Dave. 2003. *Rules, Patterns and Words. Grammar and Lexis in English Language Teaching*. Oxford: Oxford University Press.
- Zareva, Alla. «Incorporating corpus literacy skills into TESOL teacher training». *ELT Journal* vol. 71, n. 1 (2017): 69-79. DOI: <<https://doi.org/10.1093/elt/ccw045>>.