



## Training teachers in data-driven learning: Tackling the challenge

*Agnieszka Leńko-Szymańska, University of Warsaw*

### Abstract

*The aim of this article is to assess the effectiveness of a semester-long pre-service teacher-training course on the use of corpora in language learning and teaching. This is achieved by an analysis of 53 corpus-based projects prepared by the participants. First, the aims and the design of the course are briefly presented. The main section of the article examines the trainees' projects, which involved compiling small English for specific purposes corpora, analysing them and preparing corpus-based lessons. Specifically, the topics of the projects, the corpora that the participants built, the types of the analyses conducted, and the corpus-informed and corpus-based activities created by the trainees are examined both quantitatively and qualitatively. The projects' outcomes reveal that the competencies developed by the pre-service teachers during the course were not sufficient. The participants seemed to have mastered only the basic technical skills of manipulating corpora, and they lacked autonomy in corpus-linguistic skills and pedagogical skills which were necessary for successful exploitations of corpora in language education. This observation—supported by earlier research—calls for the development and empirical validation of a model of effective teacher training in corpus-informed and corpus-based instruction.*

**Keywords:** *Data-Driven Learning, Teacher Training, English for Specific Purposes, Student Project*  
**Language(s) Learned in this Study:** *English*

**APA Citation:** Leńko-Szymańska, A. (2017). Training teachers in data-driven learning: Tackling the challenge. *Language Learning & Technology*, 21(3), 217–241. Retrieved from <http://llt.msu.edu/issues/october2017/lenko-szymanska.pdf>

### Introduction

Language teachers' reluctance to use corpora in their instruction—frequently acknowledged in literature and attested in several surveys (e.g., Mukherjee, 2004; Römer, 2009; Tribble, 2015)—can be partially explained by high demands on hardware and computing skills as well as the scarcity of easily available, inexpensive, reliable, relevant, and user-friendly resources. However, the main reason is probably teachers' lack of knowledge about the different ways in which corpora can be exploited in the classroom and the necessary skills for the application of this knowledge. The implementation of data-driven teaching requires at least three kinds of competencies (Callies, 2016; Mukherjee, 2006). First, teachers need to have good *technical skills*, which means that they have to be fairly competent corpus users who are acquainted with available resources and tools and can handle them proficiently. Next, teachers also need to be versed in *corpus linguistics*, which means they should know what kind of information they can search for in a corpus, how to search for it, and how to interpret the results of these searches. Finally, teachers should also possess *pedagogical skills* allowing them to implement corpus data in the teaching process by designing and creating suitable and pedagogically-sound materials that are both corpus-informed (i.e., based on information retrieved from corpora) and corpus-based (i.e., using corpus data), combining them well with other teaching techniques and incorporating them in the instructional context. It can be argued that pre-service and in-service teachers need to be trained in order to develop these three sets of skills. The development of a model of effective teacher training in data-driven learning is a challenge that has not been adequately tackled in the literature to date.

## Review of the Literature

There are a range of books devoted to the applications of corpora in language teaching (e.g., Anderson & Corbett, 2009; Bennett, 2010; Flowerdew, 2012; O’Keeffe, McCarthy, & Carter, 2007; Reppen, 2010). They are good introductions to corpus linguistics, and all offer a fairly comprehensive overview of various possible applications of corpora in language instruction. Another source of inspiration can come from numerous articles describing corpus-based teaching projects (for recent collections of articles, see Boulton, Carter-Thomas, & Rowley-Jolivet, 2012; Frankenberg-Garcia, Flowerdew, & Aston, 2011; Kübler, 2011; or Leńko-Szymańka & Boulton, 2015; however, there are more articles published in academic journals). These articles report on projects around the world where teachers and learners have access to various types of corpora and utilize them to improve learners’ control of language systems and skills or their professional expertise as translators. They showcase how corpora can be exploited in and outside the classroom.

However, there are several problems with the relevance of these descriptions for many language teachers working in a variety of educational settings. One of the problems is the profile of the learners targeted in these courses. They are frequently university students in language-related programmes, who are therefore more advanced and more interested in how language works in comparison with other types of learners. Thus, the ideas used in these classes often cannot be copied directly to other instructional settings. Moreover, these courses are usually heavily based on corpora, whereas many teachers may want to use data-driven materials and tasks as a supplement to other materials and techniques rather than the principal teaching aid. Finally, the resources exploited in these projects are often available only for in-house use.

The problems in direct transfer or adaptation of other instructors’ experience with DDL points to the need for institutionalized teacher-training courses for pre-service and in-service teachers. Such courses should enable participants to carry out their own analyses of linguistic features relevant for their students and to design their own corpus-informed and corpus-based teaching materials suitable for their classrooms. This need has already been voiced by several researchers (e.g., Mukherjee, 2004; Römer, 2010). Descriptions of such projects have started to appear in the literature, adopting different approaches to the role and place of corpora in teacher training. In some courses, trainees exploit corpora in their foreign language or language awareness classes (Amador Moreno, O’Riordan, & Chambers, 2006; Callies, 2016; Coniam, 1998; Farr, 2008; Heather & Helt, 2012; Zareva, 2016); thus, they use corpora as language learners rather than as teachers. However, as Breyer (2009) notes, “recognising that there is a significant difference between learning and teaching with corpora, as well as providing student teachers with the required skills, is of great importance” (p. 156). Some other projects (e.g., Breyer, 2009, 2011; Hüttner, Smit, & Mehlmauer-Larcher, 2009; Leńko-Szymańka, 2014, 2015) go beyond training pre-service teachers in the use of corpora as language learners and introduce teacher-training elements directly addressing various uses of corpora in the classroom. These courses include discussions of advantages and disadvantages of particular materials and DDL activities in language instruction in general or in a particular teaching situation and coach trainees in preparation of corpus-informed and corpus-based materials and tasks.

Since the accounts of both models of teacher training in direct uses of corpora are still scarce, it is impossible to establish with certainty the extent to which such courses are effective and bring about a real change in participants’ attitudes, skills, and behaviours for both language learners and instructors. In particular, research has not yet addressed the questions of which sets of skills discussed above (i.e., technical, corpus-linguistics, and pedagogical) are particularly hard to master, and of what is a suitable teacher training model to facilitate their successful development.

The effectiveness of teacher-training courses focusing on or featuring corpora has so far been evaluated mainly through questionnaires (Breyer, 2009, 2011; Farr, 2008; Leńko-Szymańka, 2014, 2015; Zareva, 2016) or reflective essays (Breyer, 2009, 2011; Heather & Helt, 2012; Lin & Lee, 2015). These methods are useful in revealing future teachers’ attitudes towards corpora and their place in language education, as well as their perceptions of their strengths and limitations in this area. However, the results they produce say little about the actual skills that the trainees have acquired. A more suitable method in this respect might

be scrutiny of activities and materials prepared by the participants during or at the end of the course. A rigorous and systematic analysis of students' projects has proven to be a useful method of evaluating their acquired knowledge and skills, which has been demonstrated by several researchers in the context of courses on English as a second language or corpus linguistics (e.g., Boulton, 2011; Charles, 2015; Lee & Swales, 2006). Such an attempt in the area of teacher education was made by Hüttner et al. (2009), but they presented an analysis of only a single project developed by one successful participant, more as an exemplification of the potential of corpus-based analyses rather than an evaluation of the effectiveness of teacher training. Breyer (2011) and Heather and Helt (2012) used the case-study methodology to analyse projects prepared by pre-service teachers in a teacher-training course featuring corpus use. Their accounts are very informative as they give an insight into trainees' real skills in applying corpus-informed and corpus-based materials, rather than their perceptions of these skills. However, they are understandably based on a small number of subjects: three and six, respectively.

The study reported in this article is a contribution to the discussion on the effectiveness of teacher-training courses promoting the use of corpora in language education. Applying the methodology proposed by Boulton (2011), Breyer (2011), Charles (2015), and Heather and Helt (2012), it presents an analysis of projects completed by pre-service teachers at the end of five editions of a course on corpora in language teaching. The project required the participants to demonstrate the acquired skills in practice. In particular, the study examines if and to what extent the trainees mastered the three sets of skills (i.e., technical, corpus linguistics, and pedagogical) necessary for independent use of corpora in language learning and instruction. The article first presents the course itself: its objectives and design as well as the end-of-the-course project used for the evaluation of its participants. This is followed by a thorough examination of the participants' projects. The discussion of the results leads to more general conclusions about the model of teacher training in respect to corpus-informed teaching and data-driven learning.

## The Course

The course on Corpora in Foreign Language Teaching is offered annually in the winter semester, as an option for graduate students at the Institute of Applied Linguistics, University of Warsaw. In addition to linguistics and translation subjects, the MA programme includes teacher-training classes focusing on adolescent and adult learners in a range of educational contexts. To date, the course has been run six times and completed by the total of 98 students. It is taught in English, which the participants speak at advanced level. The trainees have a good knowledge of linguistics and language teaching methodology, but no prior experience in corpus linguistics.

The aim of the course is twofold: (a) to introduce participants to the concept of a corpus and its analysis and (b) to acquaint them with various applications of corpora in language education. In particular, trainees are expected to learn how they can exploit corpora for the development of courses, teaching materials, and class activities in English for specific purposes (ESP). Since the participants are exposed to comprehensive teacher training in other classes, the focus of this course is solely on the presentation of various corpora, corpus-based resources, tools and techniques, and their potential in language education. However, DDL activities presented in the course are usually accompanied by other types of tasks such as reading texts or discussions in pairs to demonstrate that corpus-based instruction must be integrated with other teaching techniques and not become an end in itself.

The design of the course followed several criteria, the most important being the presentation of a range of corpora, corpus-based materials, and tools. The resources used had to be free, easily available, and stable, as well as relatively simple and user-friendly. Different functions of the chosen software were introduced, which at the same time could exemplify various types of corpus analyses (e.g., studying genre differences, extracting collocations, etc.). The examples of possible corpus explorations were always set in the context of language learning activities, rather than pure linguistic analysis. During the classes, participants were asked to complete a variety of language learning tasks at their level. On several occasions, trainees were also assigned the task of creating their own teaching materials or activities using corpora. Finally, class

discussions were held focusing on the pedagogical value of the resources and tools presented.

The course covered thirteen to fifteen 90-minute classes.<sup>1</sup> It consisted of three thematic modules, each encompassing three or more meetings. The first module featured an introduction to corpora and types of corpus analyses along with an overview of their applications in language education. The second concentrated on the exploration of large general corpora for teaching language systems (i.e., vocabulary, phraseology, grammar, discourse) and skills (i.e., writing and reading, with some ideas for listening and speaking). The third was devoted to designing and compiling small ESP corpora and to exploring their uses in syllabus design and the development of language teaching materials and activities.

Classes were taught by the author in a computer lab with 20 computers connected to the internet, and were run from the Moodle platform. Except for four classes featuring the instructor's presentations, the meetings had a workshop format. Moodle tasks guided the trainees in studying selected language points and included detailed step-by-step instructions on how to manipulate software. Participants could work individually or in pairs to complete all the tasks and submit them through Moodle. The last 10–15 minutes of each session were devoted to class discussions of the activities with a focus on their pedagogical benefits.<sup>2</sup>

## Project

At the end of the course, participants prepared a project which formed the main part of the assessment. The project consisted of three parts: (a) a compilation of a small corpus of specialized language (ca. 30,000 words), (b) its linguistic analysis, and (c) the preparation of one ESP lesson including corpus-based materials or activities. The focus on ESP was deemed most suitable for the project for two reasons. First, it linked well with other teacher-training and translation classes the participants took. Second, it offered the best opportunity for the trainees to demonstrate their autonomy in corpus analysis and corpus-based language teaching and for the instructor to check and evaluate the widest range of technical, corpus-linguistic, and pedagogical skills acquired by the participants during the course.

Trainees were encouraged to select their own topics for the project. No specific recommendations were provided, but some examples are discussed. Participants were also instructed to select a target learner or a group of learners for their projects. These could be either real learners that they were teaching, had taught, or would teach, or hypothetical learners. Trainees were given guidelines regarding the format of their corpora (clean .TXT files); the analysis of their corpora, which had to include the list of key terminology characteristic of the selected area (using the wordlist or keyword function in a concordancer); a list of interesting collocations (also retrieved with corpus tools); and a discussion of at least one characteristic feature found in the corpus and that they found interesting and relevant for teaching purposes. Some ideas for individual explorations of self-compiled corpora were provided. Finally, trainees were required to prepare a state-of-the-art language lesson for their target learner(s) and include a worksheet with materials and tasks based on their corpus. [Figure 1](#) below presents the instructions for participants posted on the Moodle platform.

The project was assigned at the end of Class 8, which completed the 3-week course module devoted to using corpora for teaching ESP. This class was a workshop in which participants were taken step by step through the first two stages of the project. During the class they had a chance to build a mini-corpus of weather forecasts using detailed instructions and analyse it using AntConc (Anthony, 2014).<sup>3</sup>

Participants could work on the tasks individually or in pairs and were given four weeks to complete the projects. The last classes were devoted to their presentations and class discussions of the projects. If they wished, they could revise their assignments according to the comments received from the instructor and in-class discussions.

**Project**

At the end of the course you will be required to submit a project which can be completed in pairs. It will consist of the following tasks.

1. Compilation of a small ESP corpus (ca. 30 thousand words) on a chosen topic.
2. Basic analyses of the corpus: a list of terminology and frequent collocations, other characteristic features
3. Preparation of one lesson plan including activities based on the compiled corpus (corpus-informed and corpus-based materials and/or classroom concordancing)

**Format of the project**

- Clear title page
- List of corpus sources
- List of terminology
- List of collocations
- List/description of other characteristic features
- State-of-art lesson plan
- Teaching materials (exercises, texts used for the class)
- The corpus on a CD

Please be environment-friendly and save paper. Your project has to be submitted in the traditional paper form (except for the corpus itself) but when preparing your word lists, print them in several columns on one page.

*Figure 1.* Project instructions

**Course Modifications**

As could be expected in the case of a course placed in an authentic educational context and repeated several times over an extended period, some modifications were implemented to the syllabus along the way. This was done in response to the results of two questionnaires, administered after the third and fourth course editions ending in 2013 and 2014, respectively (Leńko-Szymańska, 2014, 2015). The outcome of the initial scrutiny of 32 projects conducted in 2014 was also taken into consideration. In the next two editions of the course, ending in 2015 and 2016, the DDL activities presented in the course and featuring grammar and discourse were slightly altered to give participants more experience with searching and analysing grammatical patterns and discourse features in corpora. In addition, the importance of the project's section on special features was emphasized and more time was spent discussing it when assigning the project. Furthermore, in the last edition of the course (ending in January 2016), the projects were rejected if they did not contain this part of the analysis and had to be resubmitted. [Table 1](#) summarizes the modifications introduced to the original syllabus.

*Table 1.* Modifications to the Original Syllabus of the Course

Course Edition	Syllabus	Course Content	Project
2011, 2012, 2013, 2014	Original		
2015	Modification 1	Activities featuring grammar and discourse altered and extended	Emphasis on the analysis of special features when assigning the project
2016	Modification 2	Activities featuring grammar and discourse altered and extended	The analysis of special features in the project made obligatory

## Analysis of the End-of-Semester Projects

The effectiveness of the classes and the participants' attitudes toward the application of corpora in language teaching were first assessed through two questionnaires conducted after two editions of the course (Leńko-Szymańska, 2014, 2015). The surveys engaged the trainees in self-evaluation of their technical, corpus linguistics, and pedagogical skills acquired during the course. However, a more objective assessment of their competences could be achieved through an analysis of their end-of-semester projects, which gave them an opportunity to demonstrate the whole range of their skills in practice. The exploration of the projects was guided by the following research questions pertaining to the three sets of skills:

1. Do the trainees acquire sufficient technical skills to perform their autonomous explorations of specialized language with the use of corpora and to exploit corpus resources for teaching? Specifically, can the trainees build their own corpora and operate software for their analyses, such as all-purpose concordancers, including their various functions such as extracting collocations?
2. Do the trainees develop corpus-linguistic skills necessary for independent linguistic research which can inform their teaching of ESP? Specifically, can the trainees do the following:
  - build ESP corpora which include a balanced selection of representative texts;
  - identify language points which are interesting for analysis and relevant for teaching, and can they explore them using corpus data and methodology; and
  - demonstrate equal competence and confidence in the analysis of all language systems (lexis, phraseology, grammar, discourse)?
3. Do the trainees master pedagogical skills which would enable them to exploit corpora in their own teaching successfully? Specifically, can the trainees do the following:
  - relate their corpus-based projects to the demands of real-life ESP instructional settings;
  - prepare methodologically sound and varied language teaching materials, and are they well motivated by lesson aims and well integrated with other non-corpus activities;
  - create corpus-based materials and tasks that target all language systems (lexis, phraseology, grammar, discourse) and skills (writing, reading, speaking, listening); and
  - use corpora for the development of both traditional teaching materials (corpus-informed activities) and DDL tasks (corpus-based activities), and which corpora (general or ESP) do they use for these two purposes?

In order to answer the research questions, four elements of the participants' projects were examined: the topics, the ESP corpora, the analyses, and the lessons. The methodology applied in this study involved a detailed scrutiny of each project and coding the contents with several categories of relevant information addressing the research questions. The codes included (a) the format of the corpus, (b) the kinds of analysis conducted (i.e., how the lists of terms and collocations were compiled and what other analyses were performed), and (c) the type of activities proposed in the lesson plan (vocabulary, phraseology, grammar, or discourse; reading, listening, writing, or speaking; corpus-informed or corpus-based tasks; full texts or concordance lines; etc.). The coded information was recorded in a spreadsheet for quantitative analysis. The projects were also scrutinized qualitatively to see if, for example, the proposed lessons were varied and pedagogically sound.

In most analyses, the projects were treated as a coherent collection, irrespective of the year in which they were created. However, in places where the modifications of the syllabus were expected to produce a (positive) influence on the quality of the projects, the examination was performed chronologically, to account for a possible effect of the syllabus alterations. The chronological analyses pertained to the trainees' scrutiny of special features found in the corpus as well as to the elements of their corpus-based lessons related to grammatical patterns or discourse features.

## Data

Altogether, 53 projects were gathered from five editions of the course offered in five consecutive academic years, with the last collection in January 2016.<sup>4</sup> The projects consisted of three main parts: the corpus, its analysis, and a lesson plan accompanied by teaching materials. Unfortunately, for our research purposes, sometimes not all the components were made available by the participants. [Table 2](#) presents the data available for scrutiny.

*Table 2. Data Analysed in the Study*

Type of Data	Original syllabus			Mod. 1	Mod. 2	Total
	2012	2013	2014	2015	2016	
All Projects	11	14	7	10	11	53
Corpora	7	13	7	10	11	48
Corpus Analysis	9	8	5	10	11	43
Lesson Plans Plus Teaching Materials	10	12	7	10	11	50
Complete Projects	6	7	5	10	11	39

## Topics

First, the topics of all 53 projects were analysed with an attempt to find common themes. This was done in order to establish to what extent these pre-service teachers could relate the content of their projects to real-life demands. [Table 3](#) summarizes the results of this analysis. It indicated that the topics chosen by the trainees were quite varied. The projects differed in their focus—some topics were very broad (e.g., Medical English or Economics) and some very specific (e.g., Resuscitation or Gothic Architecture). In the majority of cases, the choice of topics was not motivated by particular instructional settings that the participants were experiencing or had experienced as ESP teachers or learners or of which they were aware as existing in various institution and companies or among different vocational and professional groups. Only two trainees reported having selected their topics in view of their private tutoring classes: In one case, a beautician wanted to improve her ability to speak to her foreign clients and to recommend and discuss the details of different treatments (Beauty). In the other case, a psychologist and psychotherapist was involved in academic work at a university (Psychology and Psychotherapy). In two cases, the projects were inspired by the participants' background: one who was also studying engineering prepared a corpus on civil engineering, and a former medical student created a corpus of medical English. The most popular theme among the participants was cooking and recipes. Its attractiveness could be linked to a relative ease with which it can be delineated both as a topic and as a specific genre, coupled with availability of suitable texts and relatively small demands on background knowledge and familiarity with technical jargon. However, other popular themes (Environment, Medicine, and Sport) were more challenging in these respects, and yet the trainees selected them willingly. This analysis demonstrates that, although the participants may have had suitable theoretical knowledge on teaching specialized language from their other courses, they had no experience in this area, which could have had an effect on their projects, in particular on the choice of relevant topics and language points for analysis and instruction.

Table 3. Common Themes and Project Topics

Common Theme	N	Project Topics
Cooking	7	Cooking (4), Recipes, Cuisine, British Cuisine
Environment	5	Environment (2), Weather (2), Environment and Natural Disasters
Medicine	5	Medical English, Neurology, Cardiology, Cardiovascular System, Resuscitation
Sport	5	Fitness (2), Football (2), Extreme Sports
Art/Architecture	3	History of Art, Gothic Architecture, Modern Architecture
Economics	3	Economics, Economy, Stock Exchange
Healthy Lifestyle	3	Healthy Eating, Healthy Lifestyle, Dietetics
Law	3	Tax Law, Code of Commercial Companies (in German), Crime and the Legal System
Shopping	3	E-Commerce, Kids' Clothes, Furniture
Beauty	2	Makeup, Beauty
Computers	2	League of Legends, Windows 8
Drugs	2	Drugs, Cannabis
Politics	2	International Relations, Political Speeches
Miscellaneous	8	Astronomy, Flamenco, Easter, Civil Engineering, Human Resources, Photography, Psychology and Psychotherapy, Volunteering

## Corpora

The first part of the project involved the compilation of an ESP corpus of around 30,000 words and supplying it with appropriate documentation, consisting of a list of sources of texts selected for inclusion in the database. The basic guidelines for corpus compilation presented in the course included selecting representative texts, storing the texts in .TXT format, and cleaning them of irrelevant material such as figures, tables, references, bibliographical lists, dates, and so forth. The trainees were also made aware of the advantages of saving texts in separate files. The analysis of the corpora made it possible to verify to what extent the participants had acquired these technical and linguistic aspects of building a corpus.

Less than half (42%) of the 48 corpora analysed in the study stayed within the limits set up for the project, which was 30,000 words  $\pm$  10%. Over one third (38%) of the projects exceeded the requirement by 10–30%. Only one participant prepared a much smaller corpus: Crime and the Legal System (19,838 tokens). Four groups compiled corpora which were above 60,000 tokens: Weather\_1 (60,285 tokens), Football\_2 (65,556 tokens), Economics (71,411 tokens), and Code of Commercial Companies (in German; 98,081 tokens). Figure 2 presents the variation in corpus sizes.

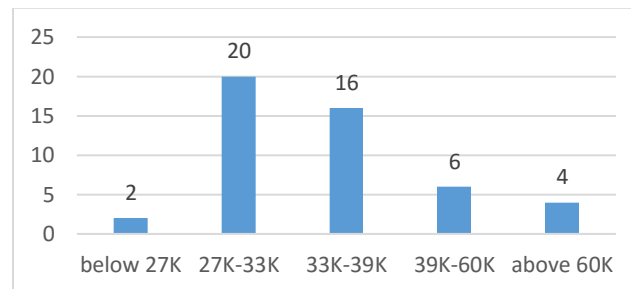


Figure 2. Corpus sizes



The documentation submitted by the participants showed that all the texts included in the corpora were retrieved from a variety of Internet sources: newspaper and magazine sites (e.g., Environment, Crime and the Legal System), academic journals (e.g., Neurology), specialized sites (e.g., Cooking, Weather), promotional materials (e.g., Beauty), product descriptions (e.g., Windows 8, Drugs), or official institutional sites (e.g., Political Speeches, which contained transcripts of EU parliament debates). On the whole, the texts were fairly representative of the genres selected by the trainees for the analysis and teaching.

17 corpora consisted of several files (from 2 to 60). However, in most corpora, all texts were stored in a single file. Seven collections were submitted as .DOC files, two of which were not cleaned at all and still contained pictures and irrelevant material. All the other corpora were submitted in .TXT format. Eight of these collections had undergone a very low level of cleaning and contained, for example, figure captions or authors' names. Yet the analysis revealed that 38 corpora (79%) were fairly clean. The information about the formats of the corpora is summarised in [Table 4](#).

**Table 4.** *Corpus Formats*

<b>Information</b>	<b><i>N</i></b>
Total Corpora Collected	48 (100%)
Number of Files	
Several Files (2 to 60)	17 (35%)
One File	31 (65%)
File Format	
.TXT File	41 (85%)
.DOC File	7 (15%)
Cleaned of Irrelevant Content	
Yes	38 (79%)
No	10 (21%)

The analysis of the ESP corpora shows that, in most cases, the trainees met the criteria of corpus compilation presented during the course. It can be thus concluded that, on the whole, they mastered the technical and corpus linguistic skills related to the compilation of a corpus and understood how to select and prepare texts for inclusion in a database.

## **Analyses**

The instructions for the project required participants to submit a list of terminology and interesting collocations extracted from their ESP corpora. During the course, the trainees were taught how to generate a basic wordlist from a corpus and were shown the advantages of using more advanced corpus tools for this purpose (i.e., tagged corpora and keyword analysis). The participants also learned how to use the clusters and collocates functions in AntConc. Another project requirement was a systematic exploration of at least one other characteristic feature of the language which the trainees had identified by themselves and found relevant for teaching purposes. This was meant to verify whether they had developed an awareness of various lexical, grammatical, and discourse characteristics of specialised language, which can be explored in a corpus, and whether they had learned to perform a corpus-based analysis of these features including basic statistical analyses, retrieving recurrent patterns, and detecting contextual restrictions. These data were available for 43 projects.

As far as terminology was concerned, the lists varied in size from around 60 items to several pages of words in columns. As mentioned above, the trainees were encouraged to use the keyword function in AntConc for the extraction of the list, and were provided with a reference wordlist retrieved from Section B of the

F-LOB corpus.<sup>5</sup> However, in only six cases was it clear from the keyness statistics (British Cuisine, Fitness, Medical English, Modern Architecture) or the description of the procedure (History of Art, Economics) that this was what the trainees had done. While the size of the lists indicates that they could not have been compiled manually, the fact that they did not include more general items with unusual frequencies in the analysed ESP genre implied that the participants preferred producing a wordlist in AntConc and then sorting it manually, deleting topic-irrelevant items. Only one group of participants (Cooking\_4) divided their list of terminology into nouns, verbs, adjectives, and adverbs, and then further into ingredients, appliances, quantities, and so forth. But again, all the sorting was done manually, as reported during their presentation. In order to prepare the list of collocations, the trainees generally selected 5–10 items from the list of terminology and then looked for clusters with these words using the AntConc clusters function; some retrieved the most frequent 2- to 4-word clusters. Only two groups (Cardiovascular System, Football\_2) produced a list of collocations using the more sophisticated collocates function. An example of its result can be seen in Figure 3 below.

kick	
on right	on left:
free kick	kick-off
penalty kick	kick correctly
corner kick	
goal kick	
bicycle kick	
simultaneous kick	

Figure 3. An example of collocations retrieved using the Collocate function (Football\_2)

This indicates that the participants had developed basic technical skills which enabled them to operate standard software for corpus analysis, but that they were reluctant to use more advanced tools, such as keyword analysis or tagging, which were included in the syllabus. Alternatively, the trainees may not have developed sufficient corpus-linguistic skills which would have allowed them to appreciate the relevance of the linguistic information arising from these advanced instruments.

An examination of the sections of the projects containing the trainees' analyses of corpus special features was conducted in two steps, due to the changes that were introduced in the content of the course in its fifth and sixth editions and the influence these were expected to have on the content and the quality of the projects. The quantitative results are presented in Table 5, along with a discussion of the qualitative aspects below. The criteria for evaluating the standard of trainees' analyses as low, medium, or high were fairly flexible. Claims supported by any corpus evidence including individual citations were already counted as medium. Analyses classified as high were based on more exhaustive scrutiny of corpus data such as multiple citations, statistical information, and so forth.

The examination of the projects collected between 2012 and 2014 revealed that the trainees had attempted to describe self-selected characteristic features of their corpora in only eight cases. Unfortunately, the majority of these analyses (62%) were very disappointing as they consisted of very general statements unsupported by any evidence that clearly came from a systematic examination of the corpus using corpus linguistics methods (e.g., statistical analysis, automatic searching for patterns, sorting of concordance lines, etc.). In one case (Medical English), the trainees provided statistical facts about the corpus (i.e., the number of texts, paragraphs, sentences), but failed to comment on the relevance of this information. In two projects (Resuscitation, Windows 8), the participants made a comment about a large number of acronyms in the texts, but did not provide a list. One project attempted to address a grammatical feature (Cooking\_2), and three commented on discourse characteristics: text purpose (Beauty), text structure (Cooking\_2), and linking expressions (Windows 2008). Figure 4 presents an example of an analysis of the special features extracted from the Beauty corpus. As can be seen, the description is impressionistic and is not backed up by any evidence.

**Table 5.** *Quantitative Analysis of the Section on Corpus Characteristic Features*

	Original Syllabus	Mod. 1	Mod. 2	Total
Number of analyses collected	22 (100%)	10 (100%)	11 (100%)	43 (100%)
Analysis of self-selected features	8 (36%)	5 (50%)	11 (100%)	24 (59%)
Analysed features				
Vocabulary	8 (100%)	4 (80%)	5 (45%)	17 (71%)
Grammar	1 (13%)	5 (100%)	7 (64%)	13 (54%)
Discourse	3 (38%)	0 (0%)	1 (9%)	4 (17%)
Quality of analysis (evidence of corpus-based analysis)				
High	0 (0%)	2 (40%)	4 (36%)	6 (25%)
Medium	3 (38%)	2 (40%)	3 (27%)	8 (33%)
Low	5 (62%)	1 (20%)	4 (36%)	10 (42%)

#### 4. Other Characteristic Features of the Corpus

The corpus contains articles about beauty treatments (at first it was supposed to be the manicure and pedicure corpus, but then we realized that most of the beauticians perform a variety of treatments every day). The common feature of the articles is that they are advertising beauty treatments and trying to convince the reader that they are indispensable. Using such a corpus during lessons with beauticians will enable them to use similar language when talking to their clients and to advertise their services. Some of the articles contain beauty tips and some of them were taken from the websites of beauty salons, but they are all characterized by very emotional and emphatic structures and persuasive vocabulary as well as professional terminology.

**Figure 4.** An example of the trainees' analysis of special features in their corpus (Beauty)

A particularly thought-provoking observation was made when analysing the corpus on Medical English. Its list of collocations contained clusters generated for the word *body*, which included the items *body's* and *of the body*. These clusters in fact point to a more general grammatical problem of noun modification which could be potentially interesting for a further corpus-based exploration and instruction. However, the trainees did not pick up this thread in the section on special features.

In the next two editions of the course (2015, 2016), following two modifications to its content, 21 analyses were collected and only five of them (submitted in 2015) did not contain a section on special features. However, only six projects contained a thorough and detailed corpus-based analysis of a selected feature or features (for an example, see [Figure 5](#)). At the same time, five projects contained a discussion of discourse features that again were not based on a systematic scrutiny of the corpus, as evidenced by the lack of examples or statistical information drawn from the data.

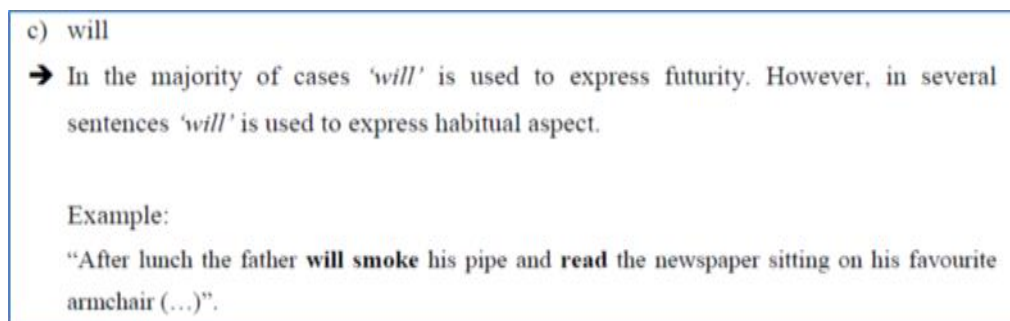


Figure 5. An example of a corpus-based analysis of a selected feature (British Cuisine)

Out of 24 analyses of special corpus features collected over five years, the majority ( $N = 17$ ) contained comments on vocabulary—often unsupported by corpus-based data—such as the use of terminology, vivid emotional vocabulary, verbs referring to actions, and acronyms. Grammar was tackled less frequently ( $N = 12$ ) and the analyses covered such points as modals (Cardiovascular System, British Cuisine), the present perfect (Cooking\_3), imperatives (Cooking\_2, Cooking\_3, Cooking\_4, Fitness), past participle forms (Fitness), and emphatic structures (Beauty, British Cuisine). Three projects contained comments of noun phrase modification (Neurology, Human Resources, Drugs) and two offered an observation on binomial structures (League of Legends, Tax Law). Except for four rather general observations (e.g., “Connectors such as: firstly, then, after etc. are common” [Cooking\_3] or “The common feature of the articles is that they are advertising beauty treatments and trying to convince the reader that they are indispensable” [Beauty]), no discourse or genre-related features were observed.

On the whole, the results of the examination of the part of the projects devoted to the analysis of ESP corpora demonstrate that the participants had mastered basic skills in corpus-based linguistic analysis. Yet, they lacked intuition about what could be analysed with these skills. The trainees were specifically instructed to provide a list of terminology and collocations and they could complete the task successfully, even if they employed only simple tools for the purpose. However, once the instruction was less specific and required selecting a feature that was characteristic or relevant for teaching, the participants were at a loss as to what to analyse and how to do it. In particular, they seemed less confident when scrutinizing grammatical features and patterns and presented few insights as far as discourse features were concerned.

### Lesson Plans and Teaching Materials


As the last part of the project, the trainees were required to prepare an ESP lesson that would address chosen linguistic elements or skills with corpus-inspired or corpus-driven activities. Throughout the course, the students were given opportunities to perform and design corpus-based and corpus-driven tasks catering for all the language elements (i.e., vocabulary, phraseology, grammar, and discourse) as well as for reading and writing. They were also provided with examples of how such activities could be combined with other types of tasks addressing the same linguistic needs. The evaluation of this section of the projects made it possible to establish whether the participants were capable of designing a pedagogically-sound ESP lesson (irrespective of whether it was inspired, based on, or driven by corpus material). More importantly, it demonstrated whether these future teachers had acquired sufficient skills in exploiting corpora in their teaching. The largest data set ( $N = 50$ ) was available for this part of the project. An example of this segment, drawn from the Cooking\_4 project, is provided in the [Appendix](#). On the whole, the lesson plans varied greatly in the degree of grounding in the earlier corpus-based analyses of the self-compiled ESP corpora and the use of corpus data for developing teaching materials. [Table 6](#) presents the list of language systems and skills addressed in the lessons. Since the modification to the course content (introduced in the course editions ending in 2015 and 2016) only related to corpus-based activities addressing grammatical and discourse points, the results for these two features exclusively are broken down to relevant periods.

**Table 6.** *Lessons Prepared by the Participants Addressing Language Systems and Skills*

	2012-2014	2015-2016
Language Systems	50 (100%)	
Vocabulary	48 (96%)	
Collocations	30 (60%)	
Grammar	3 (8%)	4 (19%)
Discourse	0 (0%)	1 (5%)
Language Skills	42 (84%)	
Reading	25 (50%)	
Writing	18 (36%)	
Listening	11 (22%)	
Speaking	11 (22%)	
DDL Activities	18 (36%)	
General Corpora	10 (20%)	
ESP Corpora	10 (20%)	
Total Lessons	29 (58%)	21 (42%)

The great majority of lessons ( $N = 48$ ) included tasks devoted to vocabulary (mainly terminology of the relevant discipline) and 60% ( $N = 30$ ) included tasks devoted to collocations. In 60% of the lessons ( $N = 30$ ), the vocabulary and collocations were first introduced without context, through matching or sorting activities (e.g., match words with a picture or a definition, sort words into groups, match two words). In all but two cases, these activities were followed by gap-filling tasks or other context-based activities based on corpus data. Alternatively, the vocabulary and collocations were introduced in texts. Figures 6 to 9 illustrate such activities. In most cases ( $N = 46$ ), the vocabulary and collocations introduced and practised during the lessons were drawn from the list of terminology retrieved from the ESP corpora. The sentences and texts providing context had the same origin (for examples, see Figure 6 and Figure 7).

1. Words in the box concern two sides of photography: the artistic and the more technical one. Work in pairs, first discuss the vocabulary and then try to divide them into those two groups.



CAMERA SHOT MOOD TO CAPTURE THE PERSONALITY FLASH  
LIGHT EXPOSITION DIGITAL NEGATIVE SHARP TO SHOOT FOCUS  
BLUR SHUTTER ZOOM TRIPOD TOOLS FILTER OUT OF FOCUS  
PERSPECTIVE

**Figure 6.** An example of a vocabulary activity (Photography)

2. Fill in the gaps using the vocabulary from the exercise number 1.
1. The photo evokes a certain ..... and often carries a specific message.
  2. However, when taken in the wrong ....., black and white images may look washed out.
  3. I faithfully put the ..... and slides in protective, archival sleeves inside notebooks. Then I dated them and stored them in a dry, safe place. To this day I can find ..... and slides from years ago within five minutes. */the same word/*
  4. During the shoot, double check that the ..... head is locked down securely to prevent any movement taking place and causing unwanted ..... */two different words/*
  5. The goal of portraiture is to ..... of the subject or group of subjects on film.

Figure 7. An example of a context-based vocabulary activity (Photography)<sup>6</sup>

However, in the case of two projects, it could be doubted whether it was the analysis of the ESP corpus that formed an inspiration for vocabulary and collocation activities. For example, in the project on Cardiology, there is no activity that was clearly prompted by the corpus analysis or created using corpus data, as some of the target words in Exercise 1 and the text used in Exercise 2 do not even appear in the ESP corpus (see Figure 8 and Figure 9).

**Ex. 1**

**Match the names of the heart parts with the picture.**

right atrium, tricuspid valve, pulmonary valve, right ventricle, septum, aorta, left atrium, mitral valve, aortic valve, left ventricle

Figure 8. An example of a vocabulary activity (Cardiology)

**Ex. 2**

**Read the definition below. Look at the bolded words, try to find their equivalents in the Polish definition. Write them down.**

“**Cardiology** is a medical specialty dealing with disorders of the **heart** (specifically the human heart). The field includes diagnosis and treatment of **congenital heart defects, coronary artery disease, heart failure, valvular heart disease and electrophysiology**. Physicians who specialize in this field of medicine are called **cardiologists**”.

„Kardiologia to jedna ze specjalności medycznych, która zajmuje się badaniem i leczeniem chorób serca (zwłaszcza ludzkiego serca). Dziedzina ta obejmuje diagnostykę i leczenie wad

Figure 9. An example of a context-based vocabulary activity (Cardiology)<sup>7</sup>

Only three projects prepared in the first three editions of the course contained activities devoted to grammar points: one to plural verb forms following collective nouns (Football\_1) and two to prepositions (Beauty and Resuscitation). In the next two years, following the changes introduced to the content of the course, four more examples of grammar activities were found in the projects: prepositions (Tax Law), *will* versus *going to* (Weather\_1), the present perfect (Cooking\_3), and the gradation of adjectives (Weather\_2). Three of these seven lessons in total involved gap-filling based on sentences retrieved from the ESP corpora. Another three classes including grammatical elements entailed studying examples drawn from the corpora in order to guess the rule. Finally, one lesson featured a traditional activity (gradation of adjectives), requiring learners to provide comparative and superlative forms of adjectives and had not been informed by corpus analysis. Only one project (Cooking\_4) contained three activities devoted to discourse structure (time sequencers). These activities are available in the [Appendix](#).

The great majority of the lesson plans ( $N = 42$ ) included activities designed to promote language skills. Tasks developing reading comprehension were the most frequent ( $N = 25$ ), always based on texts retrieved from the ESP corpora. Writing tasks ( $N = 18$ ) were generally based on model texts explored in previous activities (e.g., for useful expressions). Interestingly, the lessons included 11 listening tasks, though only one was drawn from the ESP corpus, containing transcripts of the EU debates (Political Speeches). All the remaining listening activities were based on supplementary materials found on the Internet. There were also 11 instances of speaking activities based on earlier reading or listening tasks. [Figure 10](#) provides examples of a speaking and writing activity drawn from the Beauty project.

**Activity 5**  
**Speaking**

*Acting-out: The teacher plays the client. The task of chosen students is to make a small talk with the client, just like before a beauty treatment. They should ask about the client's problems and offer suitable treatments with accurate descriptions.*

**Activity 6**  
**Homework: Writing**

*The students' task is to write a short advertisement of a chosen beautician technique. It should answer the following questions: Who is it suitable for? How does the treatment look like? How much time does it take? Is it painful? How many treatments are required to see beneficial results? Are there any risks or side-effects?*

Figure 10. Examples of a speaking and a writing activity (Beauty)

Only 18 projects contained activities involving direct learner interaction with corpora: in 10 cases, the **BNC** or **COCA**, and in 10 more, the self-compiled ESP corpora (in two cases the activities involved both types of collections). The hypothetical learners were instructed to consult a corpus to discover or verify meanings of new words, to look for characteristic collocations, or to look for example sentences of a word or a collocation. With one exception (Football\_1; see [Figure 11](#)), these activities never involved searching for patterns of use. On the whole, the direct corpus consultations were not well motivated and seemed a bit forced, creating an impression that their main function was to meet the instructor's implicit expectations related to DDL. [Figure 12](#) presents an example of such an activity—one that is unrelated to other tasks in the lesson and that seems to have little pedagogical value on its own.

### Activity 3

Use the BNC corpus to study sentences concerned with the football teams (e.g., Arsenal, Chelsea, etc.). Can you see anything unusual? Find several sentences and comment on the use of verbs in them.

*Figure 11.* An example of a DDL activity on grammatical patterns (Football\_1)

### HANDOUT 2

Look at the words below. They are taken from the keyword list.

product problem complaint complain company service item

purchased defective upset return

Type them into the concordancer. Can you see any other phrases which are frequent in complaints?

*Figure 12.* An example of a DDL activity (E-commerce)

[Figure 13](#) presents the only example of printed concordance lines found in the teaching materials. The projects contained many activities involving working with citations drawn from the corpus, but they contained full sentences and were carefully pre-selected and edited. This is the only case where the KWIC format was used.

### 1. Study the sentences below. What other words could you use instead of the word „obese“?

t AA men with higher income are more likely to be obese than those with low income.27 our findings on low s immediate as well as long-term consequences for obese individuals and society as a whole cannot be over tive sleep apnea (OSA), occurring in up to 60% of obese children. OSA, which is diagnosed using the gold- ance as early recognition and treatment of OSA in obese children are likely to result in the reduction of ult in the reduction of cardiometabolic burden in obese children. This paper summarizes the current state ology, cardiometabolic burden, and management of obese children and adolescents with OSA. 1. Introducti ase, and obstructive sleep apnea [9]. Further, in obese individuals, the clustering of dyslipidemia, hype rden in adulthood given the knowledge that 75% of obese children will become obese adults [12, 13]. Addit adults [12, 13]. Additional complications of obes children and adolescents, and the second is the h first is the observed high prevalence of OSA among obese. Thus it appears that both conditions can coexist he higher proportion of children with OSA who are obese children and adolescents can be as high as 60% [2 . It is believed that the prevalence of OSA among obese Journal of Nutrition andMetabolism 3 Children und cents can be as high as 60% [26]. In one study of obese children undergoing evaluation for symptoms of sl A [23]. 5.Mechanisms for Increased Risk of OSA in obese children and Adolescents There are multiple facto t to significantly increase the risk of OSA among obese children and adolescents. Similar to nonobese chi hypertrophy is a fairly common cause of OSA among obese children [31-33] affecting approximately 45% of a ildren [31-33] affecting approximately 45% of all obese children with OSA [34]. However, alarmingly, foll adenotonsillectomy, OSA persists in about 50% of obese children [35] which is significantly higher than

*Figure 13.* An example of a DDL activity with printed concordance lines (Dietetics)



As far as the trainees' more general teaching skills relating to ESP instruction were concerned, it was observed that, in most cases, they demonstrated good pedagogical skills and prepared good lessons from the methodological point of view; although taken as a collection, they lacked variety. The lessons consisted of clear and logical stages built around the *presentation, practice, production* paradigm. As far as the exploitation of corpora in language instruction was concerned, most of the lessons leaned towards introducing and practicing new vocabulary and phraseology retrieved from the ESP corpora and developing reading comprehension; very few tackled grammatical points, and almost none were related to exploring discourse properties. The participants preferred using corpus resources as a source of authentic texts and sentences for their self-developed traditional language activities (e.g., gap-filling) and were reluctant to engage hypothetical learners in DDL tasks based on both general and ESP corpora.

## Discussion and Conclusion

It is worthwhile to return to the discussion of different approaches to teacher training in corpus-informed and corpus-based instruction mentioned above. As postulated by some researchers (Callies, 2016; Mukherjee, 2006), teachers need three kinds of skills in order to exploit corpus data in their classes: technical skills, corpus-linguistic skills, and pedagogical skills. The results of this study demonstrate that if the training in corpus use and corpus-driven teaching is limited solely to a one-semester course that attempts to cater to all three types of skills simultaneously, the future teachers do not reach satisfactory levels of attainment in any of the required competencies. It appears that such a limited contact with corpus analysis and DDL is simply not sufficient for trainees to develop their technical, corpus linguistic, and corpus-related pedagogical skills to a satisfactory degree. The analysis of the end-of-course assignments demonstrated that the participants had mastered the technical skills of manipulating corpora only at a basic level, but that they were not adept at applying more advanced functions and tools. The trainees had also developed only basic corpus-linguistic skills. They did not know how to autonomously select language points that could be analysed with corpora. Apart from searching for purely lexical and phraseological information, they were uncertain of how to use corpora for studying and finding other linguistic facts, such as register differences. As far as corpus-related pedagogical skills were concerned, the trainees could use corpora to design and develop their own teaching materials—albeit addressing mainly vocabulary and collocations and not any other language features. Unfortunately, it was impossible to establish if this focus on lexis and phraseology was a result of their lack of confidence in analysing corpus data for other features or their lack of awareness of the varied needs of ESP learners. In addition, the lesson plans demonstrated that the participants had not mastered the expertise in designing hands-on DDL activities for their learners.

The types of problems faced by future teachers, uncovered by the present analysis, have already been hinted at in earlier studies based on questionnaires and reflective essays. Challenges related to technical skills of operating corpus-analysis software were voiced by pre-service teachers in studies by Farr (2008), Breyer (2011), and Zareva (2016). In a survey administered to one cohort of trainees following the course described in this article (Leńko-Szymańska, 2014), the participants indicated that they found it difficult to identify and interpret language points that they could study autonomously with the use of corpora. The same observation was made by the trainees described in Farr (2008) and Zareva (2016). The results of another questionnaire completed by a different cohort of course participants (Leńko-Szymańska, 2015) pointed to grammar and discourse features as more problematic than lexical and phraseological information in both independent corpus analyses and the design of new teaching materials. This perception was also shared by the pre-service teachers surveyed by Breyer (2011), who considered DDL activities as a valuable extension in teaching vocabulary to a much larger extent than in teaching grammar or as an aid in writing.

The same insight was also offered by earlier analyses of trainees' projects, which resulted from teacher training courses in corpora in language instruction. Breyer's (2011) examination of three DDL activities prepared by her participants indicated that "creating a simple task aimed at lower to intermediate learners, that is linguistically relevant and successful, proved to be very difficult indeed" (p. 206). Her trainees experienced problems with locating suitable corpora and underestimated the importance of specific skills

and prior knowledge for successful completion of their tasks by their hypothetical learners. Heather and Helt (2012) found that their participants focused on lexical items and their meanings in their projects and did not demonstrate sufficient skills in utilizing concordance output to identify lexico-grammatical patterns. They also experienced problems in “organizing and presenting concordance data in ways that lead more clearly to autonomous learning for their students” (p. 436).

The results of this study indicate that a one-semester, self-contained course does not give future teachers sufficient opportunities to develop the technical, corpus-linguistic, and pedagogical skills adequately. This has also been suggested by other researchers (Breyer 2011; Heather & Helt, 2012). Thus, the model of teacher training in corpus literacy and DDL pedagogy based on short dedicated workshops and courses, as frequently reported in the literature, has not proven to be satisfactorily effective. It does succeed in familiarizing trainees with relevant issues, and although it creates mostly positive perceptions towards corpus-based techniques, it does not produce autonomous corpus users and corpus-using teachers. Therefore, as already advocated by Farr (2010) and Breyer (2011), a better solution may be to give future teachers a much wider opportunity to interact with corpora in their teacher training programmes from two perspectives: as learners and as teachers. Trainees should get a chance to gradually develop their technical and corpus-linguistic skills and get to know various corpus resources in their foreign language, language awareness, and linguistics classes that are almost always part of teacher training (at least for pre-service teachers), but that are not teacher training per se. Such an approach has already been applied by several researchers (e.g., Amador Moreno et al., 2006; Callies, 2016; Coniam, 1998; Farr, 2008; Heather & Helt, 2012; Zareva, 2016). It can guarantee that trainees not only develop computer and cognitive skills related to corpus exploitations but also realise the potential of corpora for learning, as observed by Breyer (2011):

If teacher trainees can discover the potential of corpora for their own learning, then this may foster intrinsic motivation to make use of corpora in their profession as teachers. It also allows teacher trainees to explore and address the challenges that such an approach entails. (p. 230)

However, the training cannot stop at language and linguistics classes—being versed in technology and corpus linguistics will not automatically educate future teachers in how to use corpus data in their instruction. The development of corpus-based pedagogical skills also has to be explicitly catered to in teacher training seminars (Breyer, 2009, 2011). It remains to be established whether such training should just focus on corpora as learning resources and teaching aids and overview their various applications to language pedagogy. The present study, as well as the research carried out by Breyer (2011), demonstrated that pre-service teachers, even if well-versed in language teaching methodologies and techniques, may lack sufficient pedagogical experience to design suitable and varied corpus-informed and corpus-based teaching materials that develop different language elements and skills. It may thus be argued that teacher training seminars should instead begin with particular teaching contexts, and that future teachers should be given a chance to discuss if and, if so, how corpus information or DDL materials (along with other instructional procedures and techniques) can support teaching in these settings. But such a model of DDL teacher training still needs to be verified and assessed empirically.

Thus, in order to tackle the challenge of implementing DDL in language education, a model for teacher training in corpus-based and corpus-driven instruction can be proposed. Rather than promoting courses devoted solely to the use of corpora in language teaching, as is frequently the case in pre-service and in-service teacher training programmes (if they address this issue at all), the model postulates two parallel strands of corpus training for future teachers. First, the trainees should be exposed to corpora in their language and linguistics classes. In addition, they should be presented with DDL methods and techniques along with other instructional procedures and techniques in general teacher training courses. Such a model is more likely to ensure that future teachers will not only have good technical and corpus-linguistic skills, but also have developed sound pedagogical skills necessary for methodologically-apt application of corpus-based tasks and materials in the classroom.

## Acknowledgements

I would like to thank my students for selecting my course and for submitting their work for this study. I would also like to thank the editors and anonymous reviewers for their invaluable comments on the earlier versions of this article.

## Notes

1. The number of weeks varied from one semester to another.
2. More details on the course and the resources used can be found in Leńko-Szymańska (2014, 2015). The content of the course is available on the [Moodle site](#) and at <http://corpora.blog.ils.uw.edu.pl/>.
3. In previous editions of the course, earlier versions of AntConc were used.
4. The projects were not collected in the first edition in 2011.
5. The [Freiburg-LOB Corpus](#) (F-LOB) was compiled by Christian Mair of Albert-Ludwigs-Universität Freiburg and released in 1999.
6. Some sentences were truncated.
7. The Polish definition was truncated for the reasons of space.

## References

- Amador Moreno, C. P., O’Riordan, S., & Chambers, A. (2006). Integrating a corpus of classroom discourse in language teacher education: The case of discourse markers. *ReCALL*, 18(1), 83–104. doi: 10.1017/S0958344006000619
- Anderson, W., & Corbett, J. (2009). *Exploring English with online corpora: An introduction*. New York, NY: Palgrave Macmillan.
- Anthony, L. (2014). AntConc (Version 3.4.3) [Computer software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/>
- Bennett, G. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Ann Arbor, MI: University of Michigan Press. doi: 10.3998/mpub.371534
- Boulton, A. (2011). Bringing corpora to the masses: Free and easy tools for interdisciplinary language studies. In N. Kübler (Ed.), *Corpora, language, teaching, and resources: From theory to practice* (pp. 69–96). Bern, Switzerland: Peter Lang.
- Boulton, A., Carter-Thomas, S., & Rowley-Jolivet, E. (Eds.). (2012). *Corpus-informed research and learning in ESP: Issues and applications*. Amsterdam, Netherlands: John Benjamins. doi: 10.1075/scl.52
- Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22(2), 153–172. doi: 10.1080/09588220902778328
- Breyer, Y. (2011). *Corpora in language teaching and learning: Potential, evaluation, challenges*. Frankfurt, Germany: Peter Lang.
- Callies, M. (2016). Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English. In R. Kreyer, S. Schaub, & A. Gulderning (Eds.), *Angewandte Linguistik in Schule und Hochschule* (pp. 391–415). Frankfurt, Germany: Peter Lang.

- Charles, M. (2015). Same task, different corpus: The role of personal corpora in EAP classes. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 129–154). Amsterdam, Netherlands: John Benjamins. doi: 10.1075/scl.69.07cha
- Coniam, D. (1998). A practical introduction to corpora in a teacher training language awareness programme. *Language Awareness*, 6(4), 199–207. doi: 10.1080/09658416.1997.9959929
- Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness*, 17(1), 25–43. doi: 10.2167/la414.0
- Farr, F. (2010). How can corpora be used in teacher education? In A. O’Keeffe & M. McCarthy (Eds.), *Routledge handbook of corpus linguistics* (pp. 620–632). New York, NY: Routledge. doi: 10.4324/9780203856949.ch44
- Flowerdew, L. (2012). *Corpora and language education*. Basingstoke, UK: Palgrave Macmillan. doi: 10.1057/9780230355569
- Frankenberg-Garcia, A., Flowerdew, L., & Aston, G. (Eds.). (2011). *New trends in corpora and language learning*. London, UK: Continuum.
- Heather, J., & Helt, M. (2012). Evaluating corpus literacy training for pre-service language teachers: Six case studies. *Journal of Technology and Teacher Education*, 20(4), 415–440.
- Hüttner, J., Smit, U., & Mehlmauer-Larcher, B. (2009). ESP teacher education at the interface of theory and practice: Introducing a model of mediated corpus-based genre analysis. *System*, 37, 99–109. doi: 10.1016/j.system.2008.06.003
- Kübler, N. (Ed.). (2011). *Corpora, language, teaching, and resources: From theory to practice*. Bern, Switzerland: Peter Lang.
- Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56–75. doi: 10.1016/j.esp.2005.02.010
- Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, 26(2), 260–278. doi: 10.1017/S095834401400010X
- Leńko-Szymańska, A. (2015). A teacher-training course on the use of corpora in language education: Perspectives of the students. In A. Turula & B. Mikołajewska (Eds.), *Insights into technology-enhanced language pedagogy* (pp. 129–144). Frankfurt am Main, Germany: Peter Lang.
- Leńko-Szymańska, A., & Boulton, A. (Eds.). (2015). *Multiple affordances of language corpora for data-driven learning*. Amsterdam: John Benjamins. doi: 10.1075/scl.69
- Lin, M. H., & Lee, J.-Y. (2015). Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*, 69(3), 264–274. doi: 10.1093/elt/ccv010
- Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 239–250). Amsterdam, Netherlands: Rodopi. doi: 10.1163/9789004333772\_014
- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art—and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 5–24). Frankfurt am Main, Germany: Peter Lang.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge, UK: Cambridge University Press.

- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge, UK: Cambridge University Press.
- Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 83–98). Amsterdam, Netherlands: John Benjamins. doi: 10.1075/scl.33.09rom
- Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present, and future. In M. C. Campoy-Cubillo, B. Bellés-Fortuño, & L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 18–35). London, UK: Continuum.
- Tribble, C. (2015). Teaching and language corpora: Perspectives from a personal journey. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 37–62). Amsterdam, Netherlands: John Benjamins. doi: 10.1075/scl.69.03tri
- Zareva, A. (2016). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal*, 71(1), 69–79. doi: 10.1093/elt/ccw045

## Appendix. Lesson (Cooking\_4)

Note. The student's original wording and spelling are retained in the Appendix.

### Lesson Plan

#### Student Information

- Age of student: 27
- Level: Intermediate (B1)
- No of students: 1
- Time: 60 min

#### Aims of the Lesson

- revising vocabulary for foods and actions in the kitchen
- developing writing skills (writing recipes)
- teaching time sequencing (using time markers *before, then, until, when, immediately*) in recipes
- practicing reading comprehension skills (recipes)
- developing social media skills

#### Materials Used

- a smartphone, tablet or laptop
- recipes for Zucchini Bread
  1. <http://smittenkitchen.com/blog/2007/07/summer-of-the-bats/>
  2. <http://www.nigella.com/recipes/view/zucchini-bread-4382>
  3. <http://www.jamieoliver.com/recipes/member-recipes/recipe-detail/203/#S554Us9IBxcSyTyt.97>
- Exercises 1, 2, and 3
- film: “Zucchini Almond Bread” on Facebook fanpage *Tasty*

Stage	Aim	Procedure (teacher's and learners' tasks)	Materials	Time
Lead-in	arising interest activating schemata for usage of English	T asks S who his favourite English-speaking famous cooks are. S names his cooking role models and explains why he likes this particular person.	_____	3 min
Presentation	introducing subject of the lesson practicing reading comprehension skills	T sends links to three different recipes for Zucchini Bread to the S. T asks questions: <i>What are the main ingredients of Zucchini Bread? Can you find any differences between the three recipes? How did the authors indicate the order of performing actions in the kitchen?</i> S goes to the websites, reads the recipes and answers the questions.	smartphone, tablet or laptop; recipes for Zucchini Bread	12 min
Presentation/ practice	presenting and practicing the rules of writing a recipe practicing speaking skills	T asks S what the components of a perfect and useful recipe are. S shares his recipes experiences, S & T discuss and evaluate provided recipes. S writes down the rules of writing a recipe in a clear and well-organised way.	paper; markers	10 min
Presentation	presenting time sequencing markers practicing reading skills	S is given Exercise 1, T asks him to read the sentences. T asks a question: <i>What is the common feature of these sentences?</i> S answers the question and reads the sentences in Exercise 1 once again, this time underlining words that indicate the order of actions performed in the kitchen when preparing a meal.	handout: exercise 2	5 min
Controlled practice	practicing using time sequencing markers in a sentence	S is given Exercise 2 and solves the gap-filling exercise.	handout: exercise 3	5 min
Controlled practice	practicing using time sequencing markers in a sentence	S is given Exercise 3 and matches clauses to create meaningful sentences.	handout: exercise 4	5 min
Controlled practice	practicing writing recipes developing writing skills developing social media skills	S goes to Facebook fanpage <i>Tasty</i> and searches the Zucchini Almond Bread recipe. T asks him to focus on the ingredients. He watches the film once again and writes down the ingredients. T asks S to use time sequencing markers he has learned. S writes a recipe basing on what he saw in the video. T helps the S if need arises.	smartphone, tablet or laptop;	12 min
Practice	practicing speaking skills	T asks S to present the recipe he wrote, as if he was performing on a TV cooking show. S stands behind a desk or a table and pretends to be cooking and presents the recipe.	_____	8 min

## Student Handout

### Exercise 1

Read the following sentences. They are fragments of authentic recipes. Each of them contains a word that marks the order of actions performed when preparing a recipe. Underline those words and note how they are used in the sentences.

1. Heat 2 tablespoons vegetable oil in a large frying pan, and then add the crumbed salmon slices.
2. The bubbles will be getting larger and larger then suddenly turning brown.
3. My favourite method is to wash the leaves and stems and chop. Then pan fry in a little oil with some garlic just until everything is tender.
4. Whisk butter and sugar until fluffy and well combined.
5. Simply using a wooden spoon, cream the softened butter, caster sugar and lemon zest until light and fluffy in a bowl.
6. Allow the pizza to cool for a minute or two before adding the basil on top (whole leaves, lightly torn, or thinly sliced). Serve immediately.
7. Soak the dried mushrooms in warm water. When they have softened, drain, adding their soaking liquor to the casserole.
8. Put the griddle on a medium heat and, when hot, add the batter, using a quarter-cup measure but only filling it two-thirds full.
9. When the pasta is done, drain it and toss it with a bit of olive oil.
10. Fifteen minutes or so before the end, start to prepare the fish.
11. We soaked two chicken cutlets in a half-batch of brine for 30 minutes, and seasoned them before frying them in a pan.

### Exercise 2

Now that you have seen how the time sequencing markers are used in context, fill in the gaps with the words listed in the box below.

before	then	until	when	immediately
--------	------	-------	------	-------------

1. Generally, \_\_\_\_\_ cooking pancakes, you turn them over when you see bubbles coming to the uppermost side, and while that still holds true, the bubbles are rather understated here.
2. Leave to cool in the tin \_\_\_\_\_ cutting into pieces. Fifteen minutes or so before the end, start to prepare the fish.
3. Drain your pasta, stir it through the sauce for a good minute and a half, leave it to rest for another 30 seconds or so \_\_\_\_\_ serve \_\_\_\_\_.
4. Peel the potatoes, slice into four lengthways, and \_\_\_\_\_ slice each piece into chips, roughly the same size.
5. Repeat \_\_\_\_\_ all the dough is used.
6. \_\_\_\_\_ the oven is ready, put the dough on a floured pizza peel, or pull the pizza stone out of the oven and carefully place the dough on it.
7. Smash the hard-cooked egg yolk with the garlic and salt \_\_\_\_\_ you make the vinaigrette.
8. Take the chips out of the oven, sprinkle over some Sarsons vinegar and serve \_\_\_\_\_ alongside the cooked salmon.

9. Bake for 20 minutes \_\_\_\_\_ flip the wings over with a pair of tongues and bake for a further 10 minutes.
10. By bread crumbing the salmon here, it cooks perfectly and remains very moist. \_\_\_\_\_ the addition of chopped sun dried tomatoes really adds flavour and colour to the crust.
11. \_\_\_\_\_ it gets too crispy, throw in the garlic and stir for 10-15 seconds.
12. In a saucepan, heat vegetable oil and sauté sweet onions, green beans, and mushrooms \_\_\_\_\_ tender.
13. I always mix sweet and sour so will use sugar or honey and \_\_\_\_\_ mix in something tangy, like Sarson's malt vinegar.

### Exercise 3

Match to create meaningful sentences. Focus on how the time markers have been used.

1	Bake the chips in the oven for 20 minutes, <b>then</b> ...	A	... preparing food
2	Bring to room temperature and rewhisk <b>before</b> ...	B	... using.
3	Cook on each side for 2-3 minutes <b>until</b> ...	C	... to make it less potent.
4	Cut an apple in half, fill with oats, cinnamon, and brown sugar, <b>then</b> ...	D	... the oats carry on drinking up liquid, and the pancakes will dry on standing – with the warm raspberry honey poured on top.
5	If desired, you can sauté the garlic in a bit of olive oil <b>beforehand</b> ...	E	... place in oven until oats are golden, and apple is tender for a sweet treat this season.
6	If I were cooking for a vegetarian and a carnivore I'd cook the mushrooms and sausages separately and <b>then</b> ...	F	... this is the recipe for you.
7	Leave to cool <b>then</b> ...	G	... poke with a skewer and squeeze one lemon over the cake so it soaks up more juice.
8	Many people don't take the time to thoroughly wash their hands <b>before</b> ...	H	... turn them over using a large fish slice and carry on baking for a further 20 minutes until golden and crispy.
9	Mostly <b>when</b> I'm cooking broccoli or cauliflower...	I	... a batter is formed.
10	Now, if you want a healthy sort of cake <b>then</b> ...	J	... golden and crispy all over and cooked on the inside.
11	Pour liquid into dry ingredients and mix well <b>until</b> ...	K	... it is 3mm in thickness.
12	Serve <b>immediately</b> – ...	L	... it was soft.
13	Sprinkle some extra flour onto a board or worktop and roll out half of the dough <b>until</b> ...	M	... I finely slice the stems and just treat them the same as the florettes.
14	This frosting will slightly harden but still tastes the same <b>when</b> ...	N	... but just a touch of yolk in the egg whites may make them unusable as the fat in the yolk will prevent the whites from whisking properly.
15	<b>When</b> separating eggs, remember that a little white left in the yolk doesn't matter...	O	... serve the beans and mushrooms and spinach for the vege option



## About the Author

Agnieszka Leńko-Szymańska is a Lecturer at the Institute of Applied Linguistics, University of Warsaw. Her research interests lie in the acquisition and evaluation of second language vocabulary and formulaic language as explored through learner corpora. She teaches courses in BA and MA programmes which cover foreign language teaching and second language acquisition as well as the applications of corpora in language teaching.

**E-mail:** [a.lenko@uw.edu.pl](mailto:a.lenko@uw.edu.pl)