DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

"Corpora in Austrian Classrooms: Familiarity, Barriers and Use of Corpora by Austrian Teachers of English as a Foreign Language"

verfasst von / submitted by

Simone Kainzner

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magistra der Philosophie (Mag.phil.)

Wien, 2016 / Vienna, 2016

ACKNOWLEDGMENTS

ABSTRACT

This thesis investigates the use of corpora and their implementation in language teaching. In particular, the study aims to examine the familiarity of Austrian teachers of English as a Foreign Language (EFL) with corpus linguistics as well as with corpora for language teaching. Moreover, the actual use of corpora for language teaching and learning purposes by Austrian teachers is illustrated and the barriers that prevent teachers from using corpora in the EFL classroom are investigated. The study also aims to draw attention to possible improvements and developments of corpora that could initiate more teachers to use corpora for language teaching purposes. In terms of methodology, online questionnaires addressing the issues of familiarity, barriers and future improvements were sent to EFL teachers at secondary institutions of Austria. The findings reveal that there is a discrepancy between the familiarity of teachers with corpus linguistics and the familiarity with the use of corpora for language teaching. In addition, a lack of familiarity and teaching materials, time restrictions and insufficient teacher training resulting in a lack of skills and competencies on the part of the teacher were identified as major barriers that challenge the implementation of corpora in the EFL classroom. The thesis highlights the importance of teacher training that focuses on corpus analysis as well as on the practical implementation of corpus use in language teaching and learning. The thesis also serves as a means of spreading information about possible applications of corpora, their limitations as well as the benefits of corpus use in the EFL classroom.

TABLE OF CONTENTS

List of abbreviations

BoE	Bank of English
BNC	British National Corpus
Brown	Brown University Standard Corpus of Present-Day American English
CCEC	Collins Cobuild English Course
CEC	Cambridge English Corpus
COBUILD	Collins Birmingham University International Language Database
COCA	Corpus of Contemporary American English
DDL	Data-Driven Learning
DIY	Do-It-Yourself (Corpus)
DPU	Delayed Pedagogical Use (Corpus)
EFL	English as a Foreign Language
ELT	English Language Teaching
ICAME	International Computer Archive of Modern English (Conference)
ICE	International Corpus of English
ICLE	International Corpus of Learner English
IPU	Immediate Pedagogical Use (Corpus)
KWIC	KeyWord-In-Context
LeaP	Learning Prosody in a Foreign Language
LSP	Language for Special Purposes
MICASE	Michigan Corpus of Academic Spoken English
OED	Oxford English Dictionary
PALC	Practical Applications in Language Corpora (Conference)
POS	Part-Of-Speech
SEU	Survey of English Usage Corpus
TaLC	Teaching and Language Corpora (Conference)

List of figures

List of tables

1. INTRODUCTION

The potential of computer corpora for language description has been recognized for decades now and the popularity of corpus applications in language education can be inferred from the multitude of conferences devoted to this topic as well as the large number of publications of corpus-based dictionaries, grammars, academic journals, articles or books. Corpus-based research has had a great impact on the production of dictionaries (such as the corpus-based *Collins COBUILD English Language Dictionary*), reference grammars (such as the *Collins COBUILD English Grammar*), on syllabus design and the creation of teaching materials. Apart from that, various publications offer suggestions about the use of corpora, corpus-linguistic methods and the implementation of corpora in language teaching and learning (cf. e.g. Aston, Bernardini & Stewart 2004; Sinclair 2004; Tribble & Jones 1990; Wichmann, Fligelstone, McEnery & Knowles 1997).

The increasing popularity of corpora in the teaching and learning context among researchers has led to the perception that corpus use is equally important to language practitioners. There are a number of researchers who believe that corpora have already entered the classroom to a large extent and that their use among teachers increases steadily (cf. e.g. Meunier & Gouverneur 2009; O'Keeffee, McCarthy & Carter 2007; Sinclair 2004). However, corpus research does not necessarily translate into teaching practice (Boulton 2010: 129), and more and more researchers doubt that corpora are widely used by language teachers yet (cf. e.g. Aijmer 2009; Breyer 2006; Kaltenböck & Mehlmauer-Larcher 2005; Mukherjee 2004; Römer 2011). Up until now, there have only been a few studies investigating the extent to which language teachers have been using corpora in their English lessons. Tribble (2015), for instance, conducted a survey in which researchers as well as language practitioners were asked whether they used corpora, and in what contexts. In addition, Thompson (2006) examined the use of corpora of teachers at tertiary institutions in the United Kingdom while Breyer (2011) focused on teacher trainers at tertiary institutions in Germany. Mukherjee (2004), on the other hand, studied the extent to which teachers from secondary institutions in Germany know of and use corpora. However, there are no such studies about the familiarity and use of corpora by secondary school teachers in Austria yet.

For this reason, the following thesis focuses on English as a Foreign Language (henceforth referred to as EFL) teachers of secondary institutions in Austria and their knowledge and use of corpora in language teaching and learning. Based on the example of

Austria, the gap between the advances of corpus linguistic research and the actual teaching practice regarding corpora will be illustrated. In particular, the central aim of the thesis is to investigate the extent to which Austrian EFL teachers are familiar with corpus linguistics and corpora in language teaching, and whether they make use of corpora themselves. As a consequence, the first two research questions are as follows:

1. How familiar are Austrian EFL teachers with corpus linguistics and with the use of computer corpora for language teaching and learning?

2. How many Austrian EFL teachers have used corpora for language teaching or other purposes related to teaching?

While these questions focus on the familiarity and use of corpora by teachers, the following two research questions aim to illustrate the attitudes of teachers towards corpus use as well as the challenges they face when using corpora for teaching purposes. Therefore, the third and fourth research questions are the following:

3. What are the attitudes of Austrian EFL teachers towards the use of corpora in language teaching?

4. What are the barriers that prevent Austrian EFL teachers from working with corpora in EFL classes?

Especially the fourth research question implies that teachers face certain challenges when having decided to work with corpora in language teaching. Breyer (2009: 156) suggests that it is necessary to understand these challenges in order to allow "a more widespread application of corpora in language teaching". This thesis aims to call attention to those challenges that most likely prevent teachers from using corpora in order to facilitate a more widespread implementation of corpora. Moreover, those factors that need to be further improved and developed in order to convince more teachers to make use of corpora in the language classroom are highlighted. The last research question, therefore, is as follows:

5. What improvements and developments of corpus use for language teaching and learning could initiate EFL teachers to use corpora more frequently?

This thesis not only aims to investigate the familiarity with and use of corpora, the barriers as well as possible improvements, but it also serves as a means of 'spreading the word' (Römer 2009: 95) about corpus use. In order to achieve this, the first part of this paper deals with some theoretical issues with regard to corpora while the second, empirical part focuses on answering the research questions. To be more precisely, the organization of the paper is as follows:

After this introduction to the thesis, Chapter 2 is concerned with general aspects of corpora. A definition of 'corpus' will be given, followed by information about the developments or history of corpus linguistics and the application of corpora. Then, the most common types of corpora will be presented before the potential as well as limitations in terms of corpus applications, annotation and representativeness are discussed.

Chapter 3 outlines some theoretical aspects with regard to corpora in the context of language teaching and learning. First, an overview of corpus applications in language pedagogy is given. This subsection is divided into the indirect and direct applications of corpora with special focus on the direct applications of corpora by teachers as well as language learners. This is followed by the advantages and challenges of corpora in language pedagogy. In particular, the issues of authenticity, autonomy, the role of learners and teachers as well as the availability of corpora and teaching materials are addressed. Last, the gap between the advances in corpus research and the actual implementation of corpora in the teaching practice is highlighted in the last subsection, which is also the transition to the empirical part of the thesis.

Chapter 4 focuses on the survey, which was conducted as part of this thesis. The aim, data and method of the survey are illustrated first. Then, the results regarding the familiarity with corpora, the attitudes towards and use of corpora as well as possible improvements are presented. A discussion of the results including recommendations or possible implications for future developments and research is provided in the last subsection. Chapter 5 briefly summarizes the thesis.

2. CORPORA

This first part of the thesis is concerned with general considerations of corpora. First of all, a definition of the term 'corpus' will be suggested in the following subsection. Then, a brief overview of the historical development of corpus use will be presented, followed by an outline of the most common types of corpora. The last subsection discusses the potential as well as some limitations of corpora.

2.1. DEFINITION

In the course of time, different definitions of the term 'corpus' or 'corpora' (pl.) have been proposed. In particular, a distinction between corpora in a traditional sense and corpora in a modern understanding has been made. In a more traditional and literal meaning, a corpus is a collection of texts; or in the terms of McEnery & Wilson (2001: 29), "the term 'corpus' is simply the Latin for 'body', hence a corpus may be defined as any body of text". Especially before the advent of computers, a corpus was associated with a body of works, i.e. all the works of one author, for instance (O'Keeffee, McCarthy & Carter 2007: 1). However, in a more modern understanding of the term and in particular in the context of modern linguistics, corpora are usually associated with their machine-readable form (McEnery & Wilson 2001: 31).

One of the pioneers of the first machine-readable corpus (*Brown Corpus*, cf. section 2.2.), W. Nelson Francis, defined 'corpus' as "a collection of texts assumed to be representative of a given language, dialect, or other subset of a language, to be used for linguistic analysis" (Francis 2007 [1982]: 285). A more extensive definition of a corpus is presented by Anderson & Corbett (2009: 4) who suggest that "a modern corpus is a sample of naturally occurring language, in electronic form, which has been designed to represent a language, language variety, register or genre". Similarly to this definition, McEnery & Wilson (2001: 32) claim that

> a corpus in modern linguistics, in contrast to being simply any body of text, might more accurately be described as a finite-sized body of machine-readable text, sampled in order to be maximally representative of the language variety under consideration.

According to this last definition, a corpus is characterized by four typical features. First, a corpus is usually presented in a machine-readable form, which allows researchers to quickly search for particular language items, and to manipulate and adapt the search easily. With the

help of corpus software, large amounts of data can be accessed, displayed, searched and analyzed with a few clicks in very little time (O'Keeffee, McCarthy & Carter 2007: 2; McEnery & Wilson 2001: 31). This was only made possible by the development of computers, which have the advantage of being fast, reliable when it comes to statistical analyses; and they can accurately replicate findings (Kennedy 1998: 5). Prior to the use of computer corpora, linguists had to manually search for particular language items, read through printed texts and record their findings. Now, computer corpora allow linguists to expedite this process and save time (Cook 2003: 73). Strictly speaking, however, a corpus itself is not able to do anything because it is only a collection of texts. What is therefore needed in order to analyze corpus data is special corpus software which accesses and manipulates the data. Moreover, the interpretation of the findings is still left to the researcher (Hunston 2002: 3).

Second, the definition given above also implies that texts are sampled with a specific purpose in mind. Besides McEnery & Wilson (2001), many other linguists refer to a corpus as a 'principled' collection of text, i.e. a collection designed for a particular – usually linguistic – purpose (cf. e.g. Anderson & Corbett 2009; Hunston 2002; Meyer 2008; O'Keeffee, McCarthy & Carter 2007). Kennedy (1998: 3), however, notes that not all corpora were necessarily designed for linguistic purposes in the first place but for different purposes which had an influence on the size and nature of the corpus. Nevertheless, it is the purpose with which a corpus was collected that distinguishes corpora from other text collections such as archives. While the compilation of a corpus is usually planned and structured having a particular purpose in mind, an archive is an unstructured, opportunistically compiled text repository (Kennedy 1998: 4). Also, archives are not collected in order to be representative of a larger language variety or text genre, in contrast to corpora (Anderson & Corbett 2009: 4).

Third, corpora are characterized by their aim to represent a language, variety, genre or other subset of language (Anderson & Corbett 2009: 4). A thorough discussion of the issue of representativeness as well as of a fourth characteristic of corpora that is closely connected to the issue of representativeness, i.e. size, will follow at a later stage (cf. section 2.4.3.). However, at this part of the thesis it has to be mentioned that with the exception of monitor corpora (cf. section 2.3.), a corpus is usually limited in its size because, on the one hand, the creation of a corpus usually entails a detailed plan that also defines a grand total of words to be included (McEnery & Wilson 2001: 22-23). On the other hand, corpora have a finite size because the "totality of general language cannot be known" and all instances of language or a subset of a language cannot be captured in a corpus (Breyer 2011: 23).

Another characteristic of corpora that is often mentioned in various definitions is the fact that a corpus presents a sample of 'naturally occurring language' (cf. Anderson & Corbett 2009: 4; Breyer 2011: 7; Tsui 2004: 39-40). This means that corpora provide insights into how language has really occurred or is really used rather than how people think that language is used (Anderson & Corbett 2009: 2; Tsui 2004: 39). In other words, a new perspective on language has been gained: While linguistic description used to depend on introspection as well as native-speaker intuition and elicitation, corpus data provide linguists with new insights into the way how language is really used and present features that are typical of or frequent in that language or subset of language (Hunston 2002: 3, 20; Tsui 2004: 39-40; Mauranen 2004: 89).

2.2. HISTORY

While corpora are usually associated with the use and existence of computers for corpus analysis, corpora in the sense of mere text collections have been used in linguistics for centuries. For instance, the language use of the bible has been analyzed since the 13th century (cf. Mukherjee 2009: 12-13). In the 18th century, concordances of the *King James Bible* were created by Alexander Cruden who not only included content words in his concordances, but also some function words (Kennedy 1998: 14). Because of this type of analysis, the bible can be regarded as an early corpus in the broadest, more traditional sense (Mukherjee 2009a: 12-13).

Other early corpus linguistic projects before the emergence of computers encompass early dictionaries such as the *Oxford English Dictionary* (OED) – completed in 1928 – or grammars such as Charles Carpenter Fries' *American English Grammar* (1940), which is based on Fries' own corpus data. The OED is based on text collections, which included examples of typical uses of words occurring in real language (Mukherjee 2009a: 13-14). The entries of the dictionary were restricted to written texts produced from 1250 to 1858. In addition, the samples tended to come from upper classes, making the dictionary rather unrepresentative of the English language as a whole (Meyer 2008: 8). Fries' analysis, on the other hand, aimed at grammatical structures, uses and frequency of spoken and written language (Mukherjee 2009a: 13-14).

In the pre-electronic corpus linguistics time of the 1950s, many linguists working with pre-electronic corpora focused in their linguistic analysis on observable data rather than on introspection (Breyer 2011: 7). A minority of researchers also believed that language is finite,

i.e. sentences of a naturally occurring language are limited and can be enumerated (McEnery & Wilson 2001: 7), and that a large corpus "can contain the totality of a language" (Breyer 2011: 7). This view, however, was soon challenged by Noam Chomsky.

During the 1960s, corpus linguistics faced great criticism particularly by the linguist Noam Chomsky, who suggested that linguistics had to focus on modeling language competence rather than language performance in order to be able to describe the generative rules of a language (Breyer 2011: 8; McEnery & Wilson 2001: 6). Therefore, Chomsky (1965: 4) suggested

> [making] a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations). [original emphasis]

According to him, performance could only mirror competence under ideal conditions but in reality it cannot reflect competence because natural speech is characterized by deviations from rules, false starts or the like (Chomsky 1965: 4). Also, performance can be influenced by different factors such as short-term memory restrictions or external factors such as alcohol consumption (cf. McEnery & Wilson 2001: 6). Since the aim of linguistic theory is to model language competence and corpora can only contain performance data, Chomsky argued that corpora are an insufficient basis for linguistic analysis and a poor basis for modeling competence (cf. e.g. Breyer 2011: 8; McEnery & Wilson 2001: 6). Chomsky also claimed that introspection and native speaker intuition are essential for judging whether utterances are grammatical or not (cf. Breyer 2011: 9).

Despite Chomsky's criticism, corpus linguistics and the compilation of corpora continued during the 1960s. In particular, the emergence and development of modern computer technology helped to establish the field of corpus linguistics as it is known today (Bernardini 2000: 104; Breyer 2011: 10). Moreover, the development of the first machine-readable corpus, the *Brown University Standard Corpus of Present-Day American English (*known as *Brown Corpus)*, had a strong influence on the promotion of corpus linguistics. In building this corpus, the pioneers Henry Kučera and W. Nelson Francis ignored Chomsky's criticism and created the first computer corpus for linguistic analysis (Kennedy 1998: 23). Furthermore, the corpus texts were collected in a systematic way and the compilation was carefully documented by Kučera and Francis who made the corpus available to other linguists worldwide (Johansson 2008: 38).

Only two years prior to the beginning of the compilation of the *Brown Corpus* in 1961, the last major pre-electronic corpus, the *Survey of English Usage (SEU) Corpus*, was built by

Randolph Quirk at the University College London. The development of this corpus, which was especially compiled for grammatical description, can be considered as transition from pre-electronic corpus linguistics to modern, computerized corpus linguistics (Kennedy 1998: 17).

With further advances in computerization, "corpus studies boomed from 1980 onwards, as corpora, techniques and new arguments in favour of the use of corpora became more apparent" (McEnery & Wilson 2001: 24) and the status of corpora for linguistic analyses improved (Breyer 2011: 11). The rise in number and variety of texts for corpus analysis can on the one hand be associated with the advances in computer technology, which made its applications more user-friendly and cheaper (Johansson 2008: 33). On the other hand, the increase of corpus studies since the 1980s can also partly be linked to linguists' interest in 'naturally occurring language' (cf. section 2.1.) rather than in "rules for generating ideal sentences" (Halliday 1982: 11).

The increase of computer corpora had a tremendous influence on various areas of linguistic research, in particular on the writing of dictionaries and grammars that are based on corpora. One of the last major dictionaries of the English language that was created without any computers was *Webster's Third New International Dictionary of the English Language* (Breyer 2011: 12). The dictionary, which was published in 1961, was based on a corpus consisting of about 10 million citation slips, which had been manually analyzed (Kennedy 1998: 15). Then in the 1980s, the creation of the *Collins Birmingham University International Language Database (COBUILD)* project, which was led by John Sinclair, changed the way of writing dictionaries and grammars. The *Collins COBUILD English Language Dictionary* was the "first dictionary based entirely on the analysis of a corpus, in this case the *COBUILD Corpus*, which later became the *Bank of English (BoE)*" (Breyer 2011: 13). Also grammar books based on the *COBUILD Corpus* have been published from the 1990s onwards and many publishers continue to base their dictionaries, grammars or teaching materials on corpus analyses (Breyer 2011: 15).

2.3. TYPES OF CORPORA

Depending on the linguistics purpose, corpora vary in terms of their size, content and design (cf. Mukherjee 2009a: 41). The following section introduces the most common types of corpora, i.e. sample (or reference) corpora, monitor corpora, specialized corpora, spoken corpora, parallel corpora, and learner corpora.

Sample or reference corpora include large numbers of texts across all registers and genres of a language and are supposed to mirror the general use of the language. The texts of these corpora are very heterogeneous and the content of these corpora is finite (Breyer 2011: 26-27). In other words, reference corpora are static and aim to illustrate language at a certain point in time (Kennedy 1998: 60). Popular examples of reference corpora are the *Brown Corpus*, the *British National Corpus* (BNC*)*, or the *International Corpus of English* (ICE) (Breyer 2011: 26-27).

Monitor corpora, on the other hand, are dynamic in the sense that they 'monitor' language and its "changing patterns of usage over time" (Kennedy 1998: 61). They are adapted continuously as language changes in the course of time. While new data is added, also outdated data may be removed. One of the most famous monitor corpora is the *Corpus of Contemporary American English* (COCA) (Breyer 2011: 27-28).

Specialized corpora usually differ from reference or monitor corpora in terms of size, as they are much smaller. They comprise only specific genres, varieties of language, or text types, e.g. legal texts, newspapers, specific dialects etc. Also spoken corpora can be regarded as specialized corpora since they specialize on spoken language and exclude written texts (Breyer 2011: 29). Specialized corpora are often compiled by researchers in order to examine a specific subset of language of which these corpora are supposed to be representative (Hunston 2002: 14).

Spoken corpora, as the name indicates, consist of spoken materials that are orthographically transcribed. When analyzing spoken corpora, linguists are usually interested in spontaneous as opposed to scripted speech. Actual recordings may not always be included in the corpus but the orthographic transcriptions that also indicate typical features of speech such as pauses or discourse markers allow researchers to analyze grammatical or lexical features of spoken language (Wichmann 2008: 188-189). The drawback of spoken corpora is that they are hardly ever freely available because the compilation, transcription and annotation of spoken material is expensive and time-consuming (Breyer 2011: 29).

Parallel corpora "consist of a source text and its translation into one or more languages" (Aijmer 2008: 276). These corpora are useful for finding equivalent expressions in different languages; therefore, translators as well as language learners can greatly benefit from them (Hunston 2002: 15; Mukherjee 2009a: 53).

Learner corpora are defined as "electronic collections of texts produced by language learners" (Granger 2008: 259). This type of corpora allows linguists and pedagogues to identify learner needs, and over- or underuse of certain language items can be discovered

(Breyer 2011: 30). Furthermore, learner corpora can provide insights into the theory of Second Language Acquisitions and can contribute to the development of methods and tools for language pedagogy (Granger 2008: 259). The comparison of learner texts with native speakers' language can be particularly helpful for error analysis and allows linguists to compare over- and underuse of certain language features by language learners. One popular example of a learner corpus is the *International Corpus of Learner English* (ICLE), which consists of subcorpora with texts from learners with first languages such as German, Finnish, French, Russian or Spanish (Mukherjee 2009a: 56-57).

2.4. POTENTIAL AND LIMITATIONS

After having addressed some general issues such as the definition of 'corpus' and the historical developments as well as types of corpora, the following section focuses on the potential and limitations of corpora. In particular, typical applications of corpora are addressed, followed by information on corpus annotation. Last, the issue of representativeness of corpus data is discussed.

2.4.1. Applications

Frequency Lists and Keyword Analysis

Corpora can provide useful information with regard to the frequency of specific words. In particular, frequency lists based on corpus texts can be generated, which present all the words of a corpus in alphabetical or in frequency order, and provide information on their number of occurrence (Breyer 2011: 35). With the help of frequency lists, the lexis of a particular corpus or specific texts can be established and it can be analyzed whether this vocabulary deviates from the norm of, for instance, a reference corpus (Wynne 2008: 728). On the basis of frequency lists, particular words can be chosen for further analyses and the significance of particular words can be investigated by comparing the list with lists of other language varieties, corpora or texts (Barnbrook 1996: 47; O'Keeffee, McCarthy & Carter 2007: 11). However, comparability between two different corpora, language varieties or other subsets of language is only possible if the corpus data are of equal size. When comparing data of unequal size, normalized frequencies have to be calculated first. Normalized frequencies in

contrast to raw (or absolute) frequencies indicate the occurrence of a word per thousand, per ten thousand or per million words (Anderson & Corbett 2009: 30).

Apart from frequency lists, corpora are also used for keyword analyses. In O'Keeffee, McCarthy & Carter's (2007: 12) terms, keywords are those words "whose frequency is unusually high in comparison with some norm. Key words are not usually the most frequent words in a text (or collection of texts), rather they are the more 'unusually frequent'". Keyword analyses can be used to compare a specialized corpus with a general corpus in order to investigate the use and frequency of special features or words because highly frequent words of specialized corpora may differ significantly from general corpora (Hunston 2002: 68).

Frequency and keyword analyses are two common ways of quantitative analysis which is one type of "empirical analysis of electronically stored naturally occurring language data" (Breyer 2011: 7). A quantitative analysis allows researchers to classify and count specific language phenomena and to find out about the frequency and distributions of particular words or phrases (Anderson & Corbett 2009: 22). Quantitative data can, in principle, be generalized in order to describe how language works and they provide insights into the rareness or frequency of specific items from which their "relative normality or abnormality" can be measured (McEnery & Wilson 2001: 76-77). Despite this advantage, there is also the danger that in quantitative research corpus data "are often idealised in order to solve ambiguity and [they] marginalise rare occurrences even further" (Bernardini 2000: 121). This means that rare occurrences of a specific word form, for instance, can remain undiscovered or are intentionally disregarded.

In this case, qualitative analyses can be helpful. In qualitative analyses, the researcher identifies and describes specific language features and makes inferences based on the findings. The frequency of particular items has a minor role in qualitative analysis and rare items get at least as much attention as more common language items. The aim of qualitative analysis is to provide detailed and precise descriptions of a language or of specific language features, whereas the aims of quantitative analysis are quantification and generalizable results (McEnery & Wilson 2001: 76-77; Bernardini 2000: 121). Although these two approaches have different aims, they are not necessarily incompatible with each other. Since quantitative research does not provide any explanations of why particular language phenomena are used more or less frequently, for instance, many linguists add a qualitative description to the quantitative analysis of a specific language feature (Anderson & Corbett 2009: 22).

Concordancing

It is quite obvious that frequency lists and keyword analyses have to be generated in some way. The typical way of doing so is referred to as 'concordancing'. As a tool of corpus linguistics, concordancing allows researchers to find all occurrences of a specific item (word or phrase) by using corpus software (O'Keeffee, McCarthy & Carter 2007: 8). A distinction has to be made between 'concordancing' and the programs used by linguists for corpus analysis, i.e. 'concordancers'. Concordancers access language data and display particular words or phrases that have been searched for. The main functions of concordancers are the creation of frequency lists, keyword analyses and concordances (Breyer 2011: 31). The findings, i.e. concordances, are usually displayed in a 'Key-Word-in-Context' (KWIC) format, in which the search word or 'node' is presented in the middle of the concordance line. This search word is accompanied by seven or eight words on the left and on the right side (Bernardini 2000: 121; Breyer 2011: 31).

The concordance lines are arranged in the way that they present all the search words on a vertical axis. That way, the concordances can be scanned vertically at first, which has advantages but also a major disadvantage. O'Keeffee, McCarthy & Carter (2007: 8) call attention to the fact that concordance lines force the reader to "read in an entirely new way, vertically, or even from the centre outwards in both directions", which might be challenging for people of Western cultures who are used to reading from left to right. However, this vertical presentation has two major advantages. First, there is no need to read and search every concordance line for the node. Second, since each node is displayed beneath the other, re-occurring language patterns that precede or follow the node can easily be detected and analyzed (O'Keeffee, McCarthy & Carter 2007: 9). In other words, this way of presenting the node facilitates the detection of patterns which might have remained undiscovered without electronic concordancing (Breyer 2011: 1).

2.4.2. Annotation

As already discussed in the previous section, the concordancer is a valuable tool for finding and displaying all occurrences of a particular word or phrase. However, what a concordancer is unable to do is to distinguish between different grammatical functions or word classes of certain words such as homographs or polysemes (Breyer 2011: 36). When searching for the word *light*, for instance, a concordancer will display any occurrence of *light* as a noun, verb

and adjective. If, however, researchers intend to analyze only one word class, corpus enhancements such as annotation become necessary (Breyer 2011: 37).

Anderson & Corbett (2009: 8) define 'annotation' as the "process of adding information to a corpus, so that it becomes possible to search for features that lie below the surface of language". In general, corpora can exist in their unannotated form, that is, a raw form of plain text. However, corpora can also be annotated, which means that the text is supplemented with extra linguistic information. This linguistic information which has been made explicit by annotation facilitates the analysis of corpus data (McEnery & Wilson 2001: 32).

Texts can be annotated on various levels and the term 'annotation' usually comprises different forms such as tagging or parsing. The most common types, i.e. part-of-speech tagging, parsing and lemmatization, are discussed below. However, it should be noted that there are also various other types of annotation such as error-tagging (in learner corpora), semantic tagging, prosodic tagging or discourse tagging (cf. e.g. Meunier & Gouverneur 2009; Anderson & Corbett 2009; Breyer 2011).

One of the most common types of annotation is part-of-speech (POS) tagging in which a label (tag) that indicates the word class is attached to every word in the corpus. In the case of the word *light,* this means that the word is tagged as noun, verb or adjective depending on its use (Hunston 2002: 80). POS tagging allows to, for example, search for *light* only used as a noun, excluding other uses of the word. Therefore, unwanted results can easily be eliminated which contributes to efficient and accurate searches and results (Breyer 2011: 38).

POS tagging can be carried out manually or automatically by a computer. Manual tagging tends to guarantee up to 100% accuracy; however, it is expensive and time-consuming, which consequently restricts its use to very small corpora (Anderson & Corbett 2009: 8). Automatic tagging, on the other hand, is also claimed to be correct in about 90% of the tagged words because 'tag-programs' (taggers) are based on two principles: rules and probability. As a first step, particular rules that identify the word class of a word are applied (Hunston 2002: 82-83). The word class of a particular word is usually "highly predictable from the surrounding context" or more precisely from the co-text (McEnery & Wilson 2001: 50). In the case of Hunston's (2002: 80-83) example of the word *light*, for instance, the rule is that *light* may be a noun or adjective rather than a verb when following the determiner *a*. If no rule can be applied in order to identify the word class, then the principle of probability is applied. If, for instance, *light* is more frequently used as noun than as adjective or verb, the word will be tagged as a noun (Hunston 2002: 82-83).

Part-of-speech tagging often forms the basis for other types of annotation such as parsing. Parsing refers to the analysis of syntactic functions on a word-, phrase-, or sentence-level. The individual words including their tags are grouped into different syntactic parts (Mukherjee 2009a: 81). A parser thus indicates various constituent parts and labels those parts as, for example, 'adverbial clause' or 'prepositional phrase', allowing a detailed analysis of the sentence structure. Parsing, like POS tagging, can be done automatically, but the accuracy of parsers is rather low. Therefore, corpora are often also manually edited (Hunston 2002: 84).

The third type of annotation that is addressed here is lemmatization. If a corpus is lemmatized, it is "annotated so that related word-forms are all treated as instances of a lemma" (Anderson & Corbett 2009: 196). In other words, every form of a word is linked to a headword (lemma). Therefore, a lemmatized corpus allows researchers to retrieve, for instance, all word forms of the word *walk,* i.e. *walks, walked, walking* (Fitschen & Gupta 2008: 552). That way, there is no need to search for every single word form (Anderson & Corbett 2009: 8).

2.4.3. Representativeness

Having addressed some basic aspects of annotation, the following subsection focuses on a central issue of corpora, i.e. representativeness. As has been indicated in section 2.1., one characteristic of corpora is their aim to be representative of a language or subset thereof. Representativeness refers to the objective of collecting corpus data that mirror the language use of a certain community, time, genre, variety, etc. in a valid and reliable way (Kennedy 1998: 62). If corpus analysis aims to derive generalizable findings from a corpus, the data must either contain a language or subset of language exhaustively, or it must contain samples that represent the whole language or subset of language (Breyer 2011: 23). Thus, if researchers intend to make generalizations about a language based on corpus data, a relationship between the target language and the corpus data has to be assumed. In the words of Bernardini (2000: 107), the assumption is that "a finite entity can be assumed to be representative of an infinite one".

Representativeness, however, is difficult to achieve and can sometimes only be guaranteed in exceptional cases. For instance, the language of a specialized corpus that covers all the research articles of a particular journal can be analyzed in its entirety since only a finite and limited amount of language is under investigation. Here, the corpus data is completely

representative of this particular subset of language, i.e. the articles of a particular journal (Breyer 2011: 23). However, it is impossible that general corpora capture the totality of language and they are therefore never entirely representative of a language (Renouf 1997: 257-258).

The issue of representativeness of corpora has been widely discussed among researchers (cf. e.g. Barnbrook 1996; Bernardini 2000; Hunston 2002; Kennedy 1998; Renouf 1997). However, it is again Noam Chomsky's criticism that is cited by most researchers dealing with this topic. Chomsky argues the following:

> Any natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still others because they are impolite. (1962: 159, cited in Breyer 2011: 8)

Chomsky criticizes that corpora could never cover the whole language in its entirety because certain – more common – utterances will be left out either intentionally or by chance. Other utterances, which are quite rare might, on the other hand, be included in the corpus, contributing to the 'skewedness' of the corpus (McEnery & Wilson 2001: 30). At this point, it has to be mentioned that Chomsky's criticism has to be contextualized since he expressed his criticism at a time when corpora contained only very small samples of language due to the lack of computerization. Of course, corpus size cannot ensure representativeness but it is still one considerable factor in enhancing the representativeness of a corpus. Despite the developments in terms of computerization and corpus size, Chomsky's criticism that corpus data is skewed and thus unrepresentative is still relevant today (McEnery & Wilson 2001: 77-78).

The problem of representativeness plays a role particularly at the design stage of a corpus. O'Keeffee, McCarthy & Carter (2007: 1) suggest that the "design criteria of a corpus allow us to assess its representativeness". These criteria usually include considerations about the structure, the size of the corpus and the balance of samples which all depend on the intended purpose of the corpus. A research plan, which defines the purpose of the corpus, its size and samples, is thus a prerequisite for establishing criteria of representativeness (Breyer 2011: 23-25).

Still, the question of representativeness and balance remains, especially with large general corpora in which balance cannot be achieved due to over- or underrepresentation of specific genres or text types that are easier, faster and cheaper to collect. However, also specialized corpora face the problem of balance in that way that certain topics, genres etc. are greater represented than others (Kennedy 1998: 62-63). In addition, there are texts types such

as diaries or intimate conversations that can hardly be represented in corpora because of their private nature. Anderson & Corbett (2009: 6) address this issue accordingly:

> How do we deal with the fact that some texts are very widely read or heard, over either a long or a short period of time (for example, the Bible, the Queen's speech or the Presidential address, the fiction of J.K. Rowling, the front-page newspaper articles), while others are read or heard by few people, or small groups of people (for example, specialised scholarly monographs, horse-racing reports, sermons in a small village church).

While the problem of representativeness still remains, there have been various suggestions of how to deal with this issue as well as with balance. Hunston (2008: 162), for instance, argues that one possible response to the problem of representativeness is to present the design criteria to the corpus user who then assesses the corpus' representativeness himself/herself. Other approaches encompass the following: One possibility is to disregard the concept of representativeness altogether. In this case, texts from various registers and genres are included in a corpus without claiming to be balanced or fully representative of a language or subset thereof (Hunston 2008: 162). Concerning balance, a corpus can be compiled with disregard to the criterion of balance at an initial stage. Here, as many texts as possible from various genres are included; however, no attempts at balancing the genres are made at first. When the corpus is of considerable size, more data from underrepresented genres are collected in order to achieve greater balance between the genres. An alternative to this approach is to list all the variables of the design criteria and to collect approximately equal amounts of data from different registers, genres or other subsets of language (Hunston 2002: 29).

In connection with representativeness, the issue of corpus size arises. Generally, it is argued that a large corpus reduces some of the difficulties connected with representativeness (Hunston 2008: 165). In lexicography, for instance, a fairly large amount of lexis is necessary in order to conduct meaningful analyses. However, size is also dependent on the type of corpus as well as on the purpose (Anderson & Corbett 2009: 7). O'Keeffee, McCarthy & Carter (2007: 4) argue that "[f]or corpora of the spoken language, anything over a million words is considered to be large; for written corpora, anything below five million is quite small". Moreover, the small size of a corpus can be particularly beneficial for specific users. For instance, small corpora may be more useful for teachers who work with corpora in the classroom (cf. Anderson & Corbett 2009: 7).

In sum, it has been shown that representativeness of corpora has been widely debated and is generally difficult to achieve. It is essential to be aware of this issue especially when making inferences from corpus data that are represented disproportionally. Also, the issues of

balance and size are closely related to representativeness and greatly depend on the type and purpose of the corpus.

3. CORPORA AND LANGUAGE TEACHING

While the previous section of the thesis has addressed some general aspects of computer corpora, the following section will deal with corpora in the context of language teaching and learning. In particular, the first subsection will outline indirect corpus applications as well as possibilities of direct uses of corpora by language teachers and learners. This is followed by a subsection on the advantages and challenges of corpus use in the context of language pedagogy. Last, the gap between corpus research and the actual teaching practice will be illustrated.

3.1. APPLICATIONS

Despite the challenges of corpora especially in terms of representativeness and balance (cf. section 2.4.3.), corpus linguistics as a branch of linguistics has been well established by now. The potential of corpus-based research has been widely acknowledged and corpus research was soon also extended to the field of language pedagogy (Chambers 2007: 249). Especially during the 1980s and 1990s, the idea of incorporating corpora into language teaching emerged and research conferences focusing especially on this topic were convened. During the International Computer Archive of Modern English (ICAME) conference in 1992, which was concerned with research developments in corpus linguistics for instance, the use of corpora in language teaching and learning was addressed (Leech 1997: 1). Furthermore, the potential of corpora for language teaching received special attention at the Teaching and Language Corpora (TaLC) conferences and in particular at the conference in 1996 where increased optimism about the use of corpora for language pedagogy was expressed (Granath 2009: 47). Conferences such as the TaLC allow researchers to share their work, suggestions and ideas about this topic and to present new tools; they also promote the use of corpora in language pedagogy (Johansson 2007: 26). Other conferences in this field such as the Practical Applications in Language Corpora (PALC) are an indication of the increasing perception that corpus use and corpus research can contribute to the developments in language pedagogy (Braun 2005: 47). Apart from conferences, the first quantitative studies regarding the use of

concordances in the language classroom were conducted in the 1990s, followed by qualitative studies (Chambers 2007: 249).

The conferences that are usually accompanied by the publication of books, papers and journals about the use of corpora in language teaching strongly indicate that the legitimacy of corpora "as useful pedagogical aids has now been established" (Meunier & Gouverneur 2009: 180). This can also be attested by the variety of corpus applications in language pedagogy. On the one hand, corpora can of course be directly used in language teaching and learning. On the other hand, corpus findings can also serve as a basis for language description and, consequently, for the development of corpus-based dictionaries, reference grammars, syllabi or teaching materials, as the following section will show.

3.1.1. Indirect Applications

As already briefly mentioned in section 2.2., the development of corpora and corpus-based research has had a substantial impact on the publication of dictionaries and reference grammars (Römer 2008: 114). In this context, the COBUILD project which resulted in the writing of the first English dictionary based entirely on a computer corpus should be mentioned again (cf. Breyer 2011: 13; Leech 1997: 13). However, also other dictionary publishers such as Longman followed COBUILD's lead and based their works on corpora as well (Leech 1997: 13-14). This is rather unsurprising since the advantages of corpus-based lexicography are manifold, as Leech (1997: 14) observes: "[C]omputer corpora can be searched quickly and exhaustively, can provide frequency data, can be easily processed to provide updated lists of words, [and] can provide authentic examples for citation". Besides lexicography, the COBUILD project also contributed to the publishing of reference grammars (cf. *Collins COBUILD English Grammar*), which are also based on corpora (Leech 1997: 14).

In connection with the COBUILD project, advances in the development of corpus-based English Language Teaching (ELT) syllabi were made. One of the most prominent examples of corpus-based syllabi is the *Collins COBUILD English Course* (CCEC), which is a 'lexical syllabus' that contains the most common words and phrases of the English language (Römer 2011: 208). Since "700 of the most frequent words in English make up approximately 70% of the English language" (Breyer 2011: 45) and language consists to a great extent of repeated combinations of words (Römer 2011: 208), it is argued that learners should focus on understanding the most frequent words, their patterns of use as well as their typical combinations with other words (Breyer 2011: 45). However, there is also the objection that

the most frequent words are not necessarily the most important words that learners ought to learn. When it comes to syllabus design, frequency is only one out of many criteria that determine what language features should be taught and learnt at a specific point of the learning process (Leech 1997: 16). It is not only frequency but also difficulty that needs to be taken into account. Granger (2007: 62), for instance, argues that also difficult words have to be included in a language syllabus because otherwise learners are likely to avoid studying difficult words and their meaning altogether. In sum, it can be said that corpus findings concerning frequency and difficulty of words or lexico-grammatical patterns have had a great influence on what to teach and when to teach it, i.e. on syllabus design (Breyer 2011: 44-45).

Besides syllabi, corpora can also serve as a source for the design of teaching materials such as textbooks. Compared to the advances in the writing of dictionaries and grammars as well as syllabus design, however, teaching materials for the English language classroom have rarely been based on corpus findings so far (Römer 2008: 116). Some exceptional cases include the *CorpusLAB series* by Barlow & Burdine (2006), which contain information on and exercises of phrasal verbs of American English. Moreover, other corpus-based textbooks are *Exploring Grammar in Context* by Carter, Hughes & McCarthy (2000), which focuses on grammar, and *Touchstone* by McCarthy, McCarten & Sandiford (2005). The latter draws on data from the *Cambridge English Corpus* (CEC), focuses on the four skills of reading, writing, speaking and listening, and additionally "introduces unique 'conversation management' strategies" (McCarthy, McCarten, & Sandiford 2005). The *Touchstone* series particularly demonstrates how textbooks can benefit from corpus research (O'Keeffee, McCarthy & Carter 2007: 22).

While textbooks that are entirely based on corpus data are still a quite rare phenomenon, there is a variety of studies comparing the contents of existing textbooks with findings from corpus research (Biber, Conrad & Cortes 2004; Meunier & Gouverneur 2007, 2009; Römer 2004). What these studies reveal is that the content and language use of textbooks differ significantly from authentic language use as captured and illustrated in corpora (Breyer 2011: 46). These findings are an indication for the need to base more textbooks on corpus data, i.e. on actually occurring language.

While the considerations above assume that indirect corpus applications are widely based on general corpora, it has to be noted that findings from specialized corpora can also contribute to syllabus or textbook design. One type of these specialized corpora is the Language for Special Purposes (LSP) corpus which focuses on language that is needed in particular specialized fields. A LSP corpus that, for instance, contains English business letters

or legal documents, can help to focus on words and expressions that learners will need in this field (Römer 2008: 117). Another type of specialized corpora is the learner corpus (cf. section 2.3.), which provides insights into language learners' needs and is thus a valuable source for materials and syllabus design. So far, applications of learner corpus findings are largely restricted to dictionaries. Monolingual learners' dictionaries such as the *Longman Essential Activator Dictionary* have incorporated learner corpus data in order to emphasize common mistakes by language learners, which are highlighted in special 'help boxes' (Granger 2007: 64). According to Granger (2007: 64), a further step in the development of indirect applications of learner corpora is to incorporate learner corpus data in bilingual dictionaries.

3.1.2. Direct Applications

Of course, the use of corpora in the field of language pedagogy is not restricted to indirect applications as discussed above, but corpora have found their ways into direct applications by language teachers and learners as well. Direct applications are more "teacher- and learner focused" because "language learners and teachers get their hands on corpora and concordance tools themselves and find out about language patterning and the behavior of words and phrases in an 'autonomous' way" (Römer 2011: 211). In other words, language practitioners do not have to rely on researchers providing learners and teachers with corpus findings or materials but make use of corpora themselves.

Breyer (2011: 49) observes that "from an early stage language practitioners and applied linguists have recognized the potential of employing corpus resources and tools directly in the classroom." From the 1980s onwards, direct applications have been suggested by linguists such as Tim Johns (1986, 2007 [1991]) who coined the term 'data-driven learning' (DDL). The advances of direct corpus applications were facilitated by the shift from the use of costly mainframe computers to the use of micro computers that were more affordable and accessible by researchers as well as language practitioners. Moreover, the role of the computer as tool and aid for language teaching and learning started being debated during this time (Breyer 2011: 49-50).

Not only was the role of computer applications discussed, but corpus studies have also had an impact on language teaching methodology. The direct use of corpora by teachers and learners has influenced language learning procedures, techniques of instruction, the role of the teacher as well as the learner, and the content of language learning (Kennedy 1998: 281-282). Prior to the interest in corpus research, mainstream language teaching methodology shifted

from a greater focus on the form of language to the focus on learning through communication as a means and goal of language learning. More emphasis was placed on the role of interaction and fluency and this approach can be seen as a response to the "dissatisfaction with the teaching of a language as an unapplied system based on grammatical descriptions" and "principled ways of learning vocabulary" (Kennedy 1998: 281). Despite these valuable contributions to language teaching and learning, increasing recognition of the relevance of accuracy in language pedagogy can be observed again. According to Kennedy (1998: 281), this change took place at a time when corpus-based research started affecting the content of language teaching and learning. Corpus research places considerable emphasis on the role of lexis and grammar as well as their interplay, and provides insights into language patterns, frequency of occurrence of particular language items and their usefulness for effective communication. Therefore, it can have a major influence on language teaching theory (Kennedy 1998: 280-282).

The impact of corpora on teaching methodology and the rediscovered value of the focus on form, lexis and grammar can clearly be linked to the advances in corpus-based techniques such as concordancing. The use of this technique in the language classroom as in data-driven learning (DDL) had a great impact on the way language could be learned and taught, that is, inductively by forming generalizations. However, a deductive approach is also possible (Breyer 2011: 51-52). This technique of directly using corpus data in the classroom will be discussed in the following section. Prior to the discussion of the use of corpora for learners including DDL, various uses of corpora by teachers will be examined.

3.1.2.1. Uses for Teachers

With regard to one of the aims of this study, namely to investigate the use of corpora by Austrian English as a Foreign Language teachers, it is of special interest to have a look at the possibilities of how teachers can make use of corpora in and outside the classroom. This subchapter will give an overview of how teachers can profit from corpora by using them as a reference tool, as a basis for the design of individual syllabi and teaching materials as well as for demonstration.

Corpora are an invaluable source of reference for teachers in many ways. They can provide teachers with information about specific language items or about variation in language, for example (Granath 2009: 50), and are thus useful as a reference tool supplementing grammars. That grammar rules of traditional grammar books sometimes differ

from authentic language use is shown by Granath (2009), for instance. In her study, she reports of discrepancies between grammars and corpus data when investigating the verb forms that are used with collective nouns. Granath (2009: 51) states, "the grammar says that whereas the singular form of the verb is used with collective nouns in American English, in British English they are often treated as plurals". Granath (2009) compared this rule with the findings from British and American newspapers presented in concordances which revealed that the collective noun couple is used as a plural word in both American and British English. With this result, which was repeatedly found when the search was conducted in a much larger corpus, Granath (2009) exemplifies how grammar rules can deviate from corpus findings. Similarly, Tsui (2004) illustrates that corpora can reveal instances of particular language items that are not covered by grammar rules or even deviate from grammar rules. Also, teachers can benefit from rather odd corpus findings that differ from rules in traditional grammars insofar as to understand that there are exceptions to rules and occasions in which rules do not apply. Corpus searches enable teachers to check whether "the rules and generalizations indeed capture how language is actually used rather than how language is perceived to be used, and whether they reflect the dominant patterns of use" (Tsui 2004: 56). This means, by searching certain linguistic items in a corpus, teachers can make decisions about whether naturally-occurring text examples mirror a grammatical rule in the majority of cases or whether alternatives and exceptions are acceptable as well. Consequently, language teachers can become more independent by testing grammatical rules instead of solely relying on grammar books, but they also become more sensitive about generalizing rules and interpreting alternative findings (Tsui 2004: 56-57).

Corpora can be used for reference purposes in and outside the classroom. In the classroom, corpus evidence proves useful for answering learners' questions. As language teacher herself, Granath (2009: 53), for instance, reports that she provides learners with an answer to a language issue based on her intuition and experience right away. Afterwards, she tests her suggested answer by conducting a corpus search and then presents the findings to her learners in the next session (Granath 2009: 53). Outside the classroom, corpora can be consulted as a reference tool for correcting homework or exams, for instance. The potential of corpora for reference purposes is in the possibility to search for complex language constructions: "Corpus searches can provide answers to many language-related questions, especially questions about the combinability of words and the appropriateness of collocations" (Römer 2009: 93-94), which traditional reference books often fail offer. Thus, corpora provide information on common phrases, word combinations, word patterns, typical

uses of language items and phraseology. While grammar and dictionaries are rather unsuccessful in solving complex language issues, corpora facilitate the search for complex items. Teachers can check unusual word combinations before correcting, find more appropriate or preferred patterns by native speakers and base their decisions on these corpus findings (Römer 2009: 92-93).

Besides the advantage of providing information on complex language issues, corpora can also be consulted for rather simple searches such as collocations. In contrast to traditional reference works, corpora immediately offer a wide range of examples of the language item in question. Since, however, the meaning of the word or phrase is not given in a corpus, O'Keeffee, McCarthy & Carter (2007: 3) suggest combining the corpus search with the consultation of a dictionary or grammar.

Corpora can be valuable tools for reference purposes; their use as a basis for teaching materials should also be mentioned briefly. As discussed above (cf. section 3.1.), studies have proven that the content of non corpus-based textbooks can differ from native speaker language of corpora. On the other hand, textbooks that are entirely based on corpus findings are hardly available (with some exceptions as addressed in section 3.1.1.). One possible solution to this problem is that teachers create appropriate teaching materials based on corpus findings that complement traditional materials such as textbooks. These materials can provide text samples from corpora that mirror native speaker usage, and the selection and sequencing of linguistic items can be based on frequency information (Kennedy 1992: 366). While the challenge of this task is the selection of items and texts, the advantage is that the materials are based on naturally occurring language. Kennedy (1992: 366) addresses the issue of selecting appropriate texts:

> Texts selected without awareness of how typically they represent salient features of the language can present a chaotic picture of the language, while invented examples can present a distorted version of typicality or an over-tidy picture of the system.

The challenge, as illustrated by Kennedy (1992: 366), is that on the one hand, corpus examples can present a specific language item including all its exceptions and alternative possibilities, which could be too complex and demanding for language learners. On the other hand, contrived examples often present a simplified picture. Therefore, careful attention has to be paid to the selection of corpus texts or examples. However, the question whether a certain item should be included in corpus activities can be based on the findings of corpus studies that inform about the difficulty, frequency or relevance of certain items for language learners. Since the design of teaching materials can be quite time-consuming, though, the choice of

items should be made cautiously (Hunston 2002: 178-179). Also the choice of the appropriate corpus is of importance as it determines the quality of the exercises. Tribble & Jones (1990) give an overview of how to create teaching materials, in particular on the basis of concordances, and present a range of exercises and applications. Generally, these materials contain concordances which are enriched with "a heading at the top and some questions at the bottom" in order to create an appealing and supportive worksheet (Tribble & Jones 1990: 36). The types of exercises that can be generated with corpus materials are manifold. Concordances serve as basis for exercises in which the meaning of the keyword is to be deduced. Grammatical features as well as homonyms and synonyms could be investigated. Also group work activities, gap-fill and matching exercises are possible examples of corpus exercises (Tribble & Jones 1990: 35-55). As Tribble & Jones (1990) illustrate, the worksheets introduce the learners with the basic concept of concordances and provide information about their content, followed by the selected concordance lines as such. The essential part of the worksheet are the specific questions about the linguistic item that serve as a guidance for the learner in order to notice relevant features (Tribble & Jones 1990: 35-55). In preparing concordances for learners, teachers have the possibility to edit particular concordance lines and select only those lines that fit the purpose of the exercise and that are appropriate in terms of difficulty so that learners do not get distracted from unwanted results (Hunston 2002: 177).

Teachers may not only profit from exercises based on native speaker corpora, but also learner corpora can provide beneficial insights into how learners use language. Thus, another application of corpora by teachers is the use of learner corpora in order to establish an understanding of frequent errors of learners and to improve language teaching. Granger (2009: 14) defines learner corpora as "electronic collections of foreign or second language learner texts assembled according to explicit design criteria". Learner corpora can be divided into corpora for delayed pedagogical use (DPU) and corpora for immediate pedagogical use (IPU). The former are compiled by researchers with the purpose of linguistic description, while the latter are compiled by teachers for teaching purposes. DPU corpus analysis provides insights into overuse, underuse or misuse of certain language features while IPU corpora have the advantage that learners can work on their own texts of which these corpora consist. Whether particular items are included in language teaching, however, depends on various factors such as the learners' needs or teachability (Granger 2009: 20-23).

In connection with IPU corpora, it has also been suggested that teachers should focus on the needs of their own learners by creating local learner corpora. Millar (2008), for instance, provides a description of how teachers can compile a corpus of learner texts that

they receive for marking. This corpus can then be used for different purposes. On the one hand, teachers can closely examine particular items that are constantly misused and find out why and how these items are misused. A keyword analysis can be conducted, for instance. That way, teachers get data about significant over- or underuse of certain lexical features which can prompt them to work on vocabulary expansion, for instance (Granger 2009: 26). On the other hand, comparisons of the development of a learner's texts over the course of time can reveal how writing skills have improved, which can be particularly encouraging to learners. In addition, teachers might become motivated to reflect on their own teaching if learner corpus analysis highlights certain problem areas of learners that repeatedly emerge (Millar & Lehtinen 2008: 65-67). The information gained by learner corpus analysis is thus beneficial to teachers because choices concerning the relevance of the content and materials used for teaching can be based on the insights from their own learner corpora. Furthermore, also learners can profit from learner corpus analysis: The comparison of learners' own language use with native speaker use ideally leads to the realization of a gap between the two, "initiating a process of restructuring of their linguistic knowledge" (Lee 2011: 170).

A last application of corpora by teachers that has to be discussed is direct corpus use in class. As these applications focus on the potential for learners rather than teachers, this type of corpus application is addressed in section 3.1.2.2., while only a brief overview of different applications of corpora in the classroom is given in the following paragraphs.

Corpora can facilitate the demonstration of particular language features in the classroom as they can provide genuine examples from genuine texts. Teachers have the possibility to offer their learners prepared concordance lines in which a particular feature is demonstrated or which is analyzed by the learners. Alternatively, teachers and learners consult a corpus individually or together in order to answer questions that were raised by learners (Hunston 2002: 170-171).

The range of language issues and linguistic features that can be analyzed and the aims of corpus use in the classroom are manifold. In a study by Farr (2008), student teachers were asked to list those uses of corpora that they will consider to make in the future. Besides the use of corpora for research purposes, the respondents stated that they would use corpora for language teaching in order to "explain subtleties of language, e.g. semantic prosody, connotations, collocations", "illustrate varieties of English", "illustrate grammatical features not clear from the texts", "show students the lexical and structural patterns to help them integrate these into their language use" and to promote learner autonomy by integrating corpus use in language teaching (Farr 2008: 38). Moreover, corpora are suitable for teaching

idioms, semantic prosody or phrasal verbs. The direct applications of concordance lines or corpus searches in the classroom, however, require that teachers have acquired some degree of 'corpus literacy' (Mukherjee 2009a: 173). Teachers should thus have experience in corpus analysis, in using corpora and interpreting concordances before teaching those things to learners. By learning how to deal with and interpret corpus findings and by being exposed to the language system through corpus searches, teachers consequently develop their own language skills as well (Breyer 2009: 155).

As has been illustrated, there are many possibilities of using corpora as a teacher, which are highlighted and summarized by Granath (2009: 49):

> Corpora are invaluable for teachers, in that they can employ them in a number of ways, such as, for example, to create exercises, demonstrate variation in grammar, show how syntactic structures are used to signal differences in meaning and level of style, discuss near-synonyms and collocations, and last (but not least) to give informed answers to student questions.

To conclude, teachers may find the value of corpora in their potential to be used as a reference tool, as a basis for creating teaching material or even as a basis for decisions about what language features to focus on. Corpora are also a valuable tool in the classroom in which learners have the opportunity to act as researchers themselves. These direct applications by learners are discussed in the following section.

3.1.2.2. Uses for Learners

Corpora and concordances are often claimed to be the most powerful tool for the language classroom (cf. e.g. Bernardini 2004: 31-32). The approach of using corpus data directly in the classroom is known as 'data-driven learning' or DDL. The term was coined by Johns (1986, 2007 [1991]) and is an umbrella term for learning with corpus data in the classroom (Breyer 2011: 51-52). The following section focuses on DDL and its applications by language learners.

'Data-driven learning' (DDL) (Johns 1986, 2007 [1991]) promotes inductive learning by exposing learners to corpus data from which they form generalizations about language. These inductive learning strategies entail that learners discover patterns of language, similarities and differences and that learners formulate generalizations on the basis of these patterns (Johns 2007 [1991]: 5). Learners practice their 'generalizing' and 'inferring skills' through data-driven learning and develop their skills for independent learning (Bernardini

2000: 136). DDL is process-oriented and focuses on the learners' development of their competencies by providing them with the opportunity to discover language themselves. DDL allows students to 'learn how to learn', which also fosters their autonomy in learning (Bernardini 2000: 136). Besides the inductive approach, DDL can also be applied deductively. In this case, a rule is learnt at first, and then it is applied or tested using corpus data (Breyer 2011: 52-53).

One distinctive characteristic of DDL is its primary role of the corpus and the corpus data as well as the role of the learner as researcher. First of all, corpora serve as a "special type of informant" (Johns 2007 [1991]: 3). That is, as opposed to a teacher, the informant provides information on a question asked by learners who then try to interpret this information or response. A teacher, on the other hand, usually asks learners questions in order to figure out whether learning has happened. The learners then try to answer these questions (Johns 2007 [1991]: 3). DDL thus changes the process and the direction of acquiring information and asking questions. Secondly, learners take on the role of researchers in DDL. Johns (2007 [1991]: 4-5) has famously claimed that "research is too serious to be left to the researchers", suggesting that learners should act as researchers "whose learning needs to be driven by access to linguistic data". Learners should discover language in a similar way to how researchers discover language (Bernardini 2004: 16). Following the procedure of identifying – classifying – generalizing as suggested by Johns (2007 [1991]: 7), learners adopt approaches similar to those used by researchers (Bernardini 2000: 135).

In both cases in which learners access corpus data themselves or teachers prepare concordance lines for concordance-based activities, corpus-based learning is characterized by a number of advantages. First of all, inductive learning strategies of data-driven learning promote learner autonomy, making learners independent of the teacher who merely acts as learning/research coordinator or facilitator (Bernardini 2004: 16-17). Moreover, DDL stresses the "probabilistic nature of language use, training learners to live with the uncertainty and incompleteness of most statements about the language", and focuses on both meaning as well as form (Bernardini 2000: 139). Also, corpus data has the potential to illustrate the varieties in language. Learners can observe what combinations of words are probably acceptable and develop their own combinations (Sinclair 1997: 37). The potential of DDL for learners is also in the motivational factor. Learner autonomy and the access to genuine texts tend to be highly motivating for learners (Breyer 2011: 59-60). Furthermore, learners are actively involved in discovering language and get hands-on experience in working with corpus tools. DDL can be

especially rewarding if learners make original observations or come to new conclusions that are unknown to teachers or even to researchers (Leech 1997: 3).

However, before a corpus-based approach is introduced in the language classroom, the teacher has to choose and thus evaluate the corpus, the method and the users. Particularly the corpus as basis for DDL exercises plays a crucial role. Its design, size and content partly determine its usefulness for the classroom. Since most corpora have been designed for research purposes, their use in the language classroom requires special considerations (Breyer 2006: 158). Furthermore, in order to meaningfully implement DDL in the classroom, certain requirements in terms of the learners' skills and knowledge are inevitable. Gavioli (1997: 83) rightly remarks that "[s]imply giving students direct access to the data produced by the computer is not enough to make them research workers". Learners engaged in DDL exercises need to be equipped with specific skills in order to successfully learn from DDL activities. This also requires guidance by teachers. The prerequisite of successful data-driven learning by students is their knowledge about different word classes and how to identify those (Granath 2009: 49-50). It is required that learners acquire linguistic and metalinguistic knowledge in order to be able to identify patterns, regularities and to further categorize and generalize them (Gavioli 1997: 83-84). When working with an annotated corpus, learners need to be familiarized with linguistic categories and ways of searching for words of a particular category (Leech 1997: 9). They have to acquire the "ability to know how to search for words and phrases in a corpus, how to sort and interpret the output and make sense of concordance lines" (Ebeling 2009: 69-70). When working with unedited concordances, learners face the challenge of identifying those lines that are relevant in that particular case and they need to understand how to deal with variation in language (Granath 2009: 49-50). Moreover, the pitfalls of generalizing patterns have to be stressed by teachers. Finding recurring patterns in corpus data can suggest that regularities of use are found. However, these regularities do not necessarily need to match the rules that can be found in grammars or dictionaries. Learners need to be made aware that regularities found in a corpus with a limited range of data may not always be fully accurate or generalizable. Nevertheless, the value of these findings for the learning process has to be stressed (Aston 1997: 60). Furthermore, learners need to be made aware that corpus data cannot provide immediate answers to a question, but it provides data that must be interpreted. This interpretation has to be trained and practiced. Also, learners must understand that corpus data cannot provide insights into 'the language' in general, but only into the language contained in that particular corpus (Gavioli 1997: 84-90).

There is no need to say that DDL may pose a great challenge for novice learners. However, teachers can control the level of difficulty of DDL activities to a certain extent. Aston (1997: 62-63) describes how to adjust the levels of difficulty to suit learners' competencies. First, the data can be simplified by choosing a corpus with simple or modified texts or by compiling an appropriate do-it-yourself (DIY) corpus. Second, corpora containing data that are familiar or predictable to learners should be selected. Task difficulty can be reduced when texts of familiar topics are worked on and the variety of text types is restricted (Aston 1997: 62). Third, reducing the quantity of data may decrease the level of difficulty. A smaller corpus with homogeneous collections of texts facilitates the use of the data because the choice of texts and contexts is limited (Aston 1997: 62). Fourth, the task itself can be simplified. The type of analysis can be adjusted, for instance: instead of demanding a contextualized interpretation of texts, teacher could also ask learners to recognize particular patterns or categories that learners are already familiar with rather than assigning a label to unknown categories, for instance (Aston 1997: 62-63). In addition, teachers may choose from different approaches. On the one hand, learners may be provided with concordance print-outs in order to avoid direct confrontation with corpus software (Leech 1997: 10). This is a rather controlled way of confronting learners with concordance exercises, which allows teachers to pre-select relevant concordance lines in order to adjust the difficulty. On the other hand, learners may directly access corpus data themselves. They make use of corpus software and investigate certain language aspects by exploring the corpus and searching for specific words or phrases. In both ways, learners develop language awareness and learn by discovery (Römer 2009: 91).

There are many ways of how learners can directly apply corpora in the classroom. The following subsections present a selection of ways of how corpora can be used to foster learners' reading, speaking and writing skills as well as the acquisition of lexis and grammar.

Reading

With the help of concordances and corpus data, the learner's reading skills can be developed through a 'bottom-up' as well as a 'top-down' approach (cf. Aston 2001: 28-30). The former involves, for example, the interpretation of a small sample of a text presented in concordance lines with the aim of generating greater textual meanings from a short passage with limited context. Practice in bottom-up interpretation is particularly useful when texts are characterized by lexico-grammatical ambiguity: Learners can be asked to categorize

polysemous words that appear in a concordance line of a longer text (Aston 2011: 28). Another possible activity is to ask learners to identify boundaries of clauses in a text. According to Aston (2001: 28), bottom-up activities such as these practice the understanding of texts and train learners' skills needed for fluent reading. A bottom-up exercise can be supplemented with a top-down activity in which learners read a larger part of the text. Prior to reading whole texts, concordances of frequent words may support learners in understanding texts, their lexis and structure, and the contents of texts can be predicted (Aston 2001: 29-30). Moreover, concordances can be used for the analysis of literary writing. Tribble & Jones (1990: 72-83), for instance, illustrate how concordances about the main characters of a short story help learners in the description of these characters by listing the verbs and adjectives that are used when referring to them. In addition, concordance printouts are also used in order for students to identify the boundaries of the main parts of the short story or to create their own story based on concordance lines. Studies of the plot, theme, narrative progression or characters could be done with the help of concordances. Also, the creation of wordlists prior to studying texts is suggested as a first step in order to get an idea of the text and possible ways of studying the text (Tribble & Jones 1990: 72-83).

Speaking

Besides the acquisition of reading skills, corpus-based activities can also aid in the development of speaking skills, in particular in the development of pronunciation and the teaching of phonology. Gut (2006), for example, demonstrates the use of the *Learning Prosody in a Foreign Language* (LeaP) corpus for the purpose of teaching phonology and pronunciation to university students and advanced students at secondary institutions. According to Gut (2006: 71), neither learner corpora nor spoken corpora include transcriptions of typical features of speech such as stress, intonation or pronunciation. Therefore, the LeaP corpus was created in order to provide these. The LeaP corpus is a "fully text-to-tone aligned and extensively annotated speech corpus of learner English and learner German" (Gut 2006: 71) that includes various types of annotation, e.g. orthographic and phonemic transcriptions, POS tagging, semantic annotation or lemmatization. Furthermore, the beginnings and ends of phrases, syllables, words, consonantal and vocalic intervals were annotated manually (Gut 2006: 72). As suggested by Gut (2006: 75-81), the LeaP corpus can serve as a basis for creating exercises on individual sounds, intonation, stress or tone and it offers the possibility to work on phonetic transcription.

Spoken corpora can help learners in the acquisition of their speaking skills because they offer insights into natural features of speech. Corpus analyses of spoken corpora reveal features such as discourse markers, hedges, ellipsis, and show false starts, repetitions or hesitation devices (O'Keeffee, McCarthy & Carter 2007: 22). Also, with regard to lexis and grammar, spoken language differs considerably from written language. Consequently, those differences as well as spoken grammar and spontaneous speech have to be taught to learners in the EFL classroom. Mukherjee (2009b: 214-215) stresses the importance of teaching features of speech such as discourse markers because of their frequency in native speaker usage and their contribution to learners' overall fluency. Concordances of particular discourse markers can raise learners' awareness about their frequency, their contexts in which they occur, their functions and forms. As a second step, Mukherjee (2009b: 215-220) suggests that teachers should then give learners the opportunity to use and practice them.

Writing

For the purpose of writing, corpora can be utilized by learners in order to "check whether a hypothesized realization is likely to convey the meaning desired, and to identify possible alternatives" (Aston 2001: 35). Concordances thus serve as a reference tool. Corpus data, however, can also provide insights into the form and function of repetitive and recurring chunks of language, which can serve as a basis for the creation of learners' own texts. Assuming that language is repetitive, Gavioli (2001: 126) argues that "speakers tend to reproduce multi-word chunks of language that they have memorized as such, rather than generating them through the application of grammatical rules". These chunks can aid learners in their language production while still giving them the choice to deviate from typical, recurrent chunks. Learners can benefit from corpus activities that present possible chunks of language. As suggested by Gavioli (2001: 126-127), learners might search for chunks for each move of an advertisement and use these for the production of their own text. Alternatively, learners could look at atypical chunks of language that deviate from conventional language use (e.g. the newspaper headline *First class male* instead of 'first-class mail') (cf. Gavioli 2001: 127). When learners are engaged in this type of creative writing, they need to be aware of existing deviations and understand how they work (Gavioli 2001: 127-129). A corpus-based approach to teaching writing is also suggested by Mukherjee (2009a: 170-173). He, however, describes a corpus-based 'genre teaching' approach in which whole texts of particular genres are analyzed with regard to their moves. The next step in this process would be to identify typical chunks of every move. Last, learners produce their own texts of that

genre but on a different topic while following the sequence of the moves and using appropriate chunks. This leads to the production of creative, innovative texts that nonetheless include conventional moves of the genre as well as appropriate language (Mukherjee 2009a: 170-171).

Lexico-Grammar

With regard to the investigation of lexico-grammar, the potential of corpus-based activities in the EFL classroom becomes apparent. The following section gives a brief overview of general applications of corpora for teaching/learning grammar and vocabulary followed by suggestions for investigating particular linguistic categories with corpora.

Corpus exercises can greatly facilitate grammar learning and teaching. Dodd (1997), for instance, describes two ways of how grammar items can be studied with the help of corpora. First, learners can gather information about a language item in reference books or textbooks, and then move to the corpus and use corpus data for comparison. The rules of reference grammars can, for instance, be tested by finding matching examples in the corpus data (Dodd 1997: 135). The second approach is to move from data to textbook, i.e. learners use corpus data in order to formulate their own grammar rules, which are compared with the rules in reference works afterwards. The comparison of learners' own rules and those from published grammars may allow them to discover aspects that are not covered in reference grammars. However, they might also notice some aspects in which the rules of reference grammars are more comprehensible or precisely formulated than their own (Dodd 1997: 136). This second approach is obviously more demanding but both approaches are likely to be motivating for learners since new discoveries are possible. Furthermore, both approaches are suited for collaborative exercises, which are also more likely to increase the learning experience (Dodd 1997: 135-136).

Another grammatical query is concerned with the grammaticality of certain structures. Teachers as well as learners may want to examine whether a structure is appropriate and acceptable. Kaltenböck & Mehlmauer-Larcher (2005: 74-75), for instance, illustrate how simple corpus searches of split infinitives or *if*-clauses (*If I was* vs. *If I were*) can provide answers in terms of acceptability.

While the investigation of grammatical features with corpora is rarely described in the literature, corpus searches regarding lexical items can be found more often (Kaltenböck & Mehlmauer-Larcher 2005: 71). One possibility of corpus use in terms of lexical queries is the investigation of synonyms. Tsui (2004: 44-50) reports that teachers often fail to explain the

difference of synonyms or near-synonyms. In the case of *tall* and *high*, for instance, it can be useful to examine the semantic environment or frequent collocations of these words in order to establish an understanding of their difference in meaning and/or use (cf. Tsui 2004).

As indicated, one of the most beneficial lexical searches of corpora is the investigation of collocations. In Sinclair's (1991: 170) terms, collocation "is the occurrence of two or more words within a short space of each other in a text". The study of collocations, or "the co-occurrence of words" (Kaltenböck & Mehlmauer-Larcher 2005: 72), can prove useful in language teaching, and corpus exercises can greatly facilitate the acquisition of common collocations. Collocations are often subtle, not readily apparent to learners and difficult to explain. However, corpora can display common collocational patterns and illustrate which co-occurrences of words are more probable or likely to occur (Breyer 2011: 18-19). In addition, many collocations emerge from the semantic proximity of their collocates (e.g. *school* and *children* frequently collocate) while others are rather unpredictable (Mukherjee 2009a: 101). In this case, information about common collocations of a particular word can be used to detect semantic fields of the word, which may provide insights into the meaning of the word. Also different meanings of one word can become apparent when considering the word's collocates. As shown by Hunston (2002: 76), the collocates of the word *leak* reveal that the word has a physical meaning when used with words like *oil, water* or *gas*; however, with collocates such as *document/s, information, news* or *details* the word is used in its metaphoric sense.

The analysis of the occurrence of words can also include the investigation of a word's semantic prosody, i.e. "whether it is used in a positive or negative verbal environment" (Kaltenböck & Mehlmauer-Larcher 2005: 73). A frequent example of a word with a strong semantic prosody is *cause* and its 'counterpart' *provide*. While *cause* typically occurs in negative verbal environments, *provide* is usually used in positive co-texts (cf. e.g. Kaltenböck & Mehlmauer-Larcher 2005: 73; Anderson & Corbett 2009: 60-62).

Apart from illustrating occurrences of words within their closest surroundings, corpus data can also be used to analyze the occurrences of words in their wider environment. Corpora comprising different genres, text types or modes of production (i.e. spoken or written) can reveal variation in usage patterns or meanings of words depending on the genres, mode or register of the texts in which they appear (cf. Anderson & Corbett 2009: 63-64; Kaltenböck & Mehlmauer-Larcher 2005: 71).

Other uses of corpora include the investigation of colligations. Anderson & Corbett (2009: 58) define colligation "as the tendency of a word to co-occur, not with another word or phrase, but with a grammatical category or construction". Hunston (2002: 151) describes

colligation as "collocation between a lexical word and a grammatical word such as a preposition". The investigation of colligations can encompass queries about appropriate prepositions following certain verbs or the 'verb plus verb agreement': This means that learners may have difficulties deciding whether a verb requires to be followed by an infinitive, a gerund or a finite verb clause (Partington 1998: 80). Corpora can be particularly useful for issues such as this one by showing tendencies of common usage.

In sum, this subsection above has given an overview of the possibilities of corpus use in the context of language teaching. It has been shown that corpora can be applied by teachers as well as learners; they can be used for developing various skills but also for clarifying lexical and grammatical questions. However, this overview is of course not a comprehensive summary of all the possible uses of corpora but it offers some insights into the variety of applications of corpora in the context of language pedagogy.

3.2. ADVANTAGES AND CHALLENGES

While the previous sections have outlined some of the applications of computer corpora in connection with language teaching and learning, the following subchapter gives a brief overview of the advantages and challenges of corpus use in the EFL classroom. These include the issue of authenticity, changing roles of teachers and learners, learner autonomy as well as the availability of corpora and teaching materials.

3.2.1. Authenticity

In the context of language teaching and learning, authenticity is generally regarded as a positive characteristic of materials or texts. In the teaching community, authentic materials in contrast to didactic materials tend to be seen as being more motivating, encouraging and interesting to learners because of their richness as well as their cultural and linguistic content (Mishan 2004: 219). Moreover, it is often argued that the use of authentic texts, i.e. naturally occurring native speaker language, "can facilitate discussion of cultural background" and even foster effective communication (O'Keeffee, McCarthy & Carter 2007: 26). Therefore, the use of authentic rather than contrived examples of language is often recommended. In this context, 'contrived' materials involve texts that are invented with the pedagogic purpose of exemplifying a particular language feature (O'Keeffee, McCarthy & Carter 2007: 26).

In the research community, on the other hand, the terms 'authentic' and 'contrived' have been widely debated. In language description, the concept of authenticity has a different meaning than in language pedagogy: collecting 'authentic' texts in language description means to collect language that is 'attested', i.e. language that has occurred. In language pedagogy, on the other hand, authenticity involves authentic responses of learners to texts. In the latter case, the origin of the texts is less crucial than what learners do with the texts in their learning process (Seidlhofer 2002: 220). Moreover, apart from learning materials, the concept of authenticity can also be applied to the learning situation, the learners, the purpose or the (corpus-based) activity (cf. Breyer 2011: 60-61).

The debate about the concept of authenticity of corpus-based activities in the language classroom, however, usually focuses on authenticity of corpora, i.e. texts or materials based on corpus data. In this debate, the question whether corpora present 'real' or 'authentic' language frequently arises. Widdowson (2000: 7), for instance, argues that texts that are contained in corpora are only partially real because the language of corpora is decontextualized. This means that corpora only display the co-text, i.e. the verbal environment, but not the context in which the language originally occurred. Since the circumstances of the message or language, the intended audience and the illocutionary intentions are invisible in language data included in corpora, this language is isolated from its context and only represents text rather than discourse (Lee 2011: 161). *Discourse*, in this case, describes the "the relationship between the speaker's intention and listener's interpretation" (Kaltenböck & Mehlmauer-Larcher 2005: 69), while *text* is the product of "language use isolated from any communicative situation" (Braun 2007: 32). According to Widdowson (2003: 98),

> [p]eople make a text real by realizing it as discourse, that is to say by relating it to specific contexts of communal cultural values and attitudes. And this reality does not travel with the text.

The "contextual conditions upon which the discourse realization depends" (Widdowson 2003: 98) are not given in language from corpus data which consequently means that corpora do not represent discourse and contain 'real' language only in a limited, partial sense.

Widdowson (2000, 2003) further distinguishes between the concepts of 'authenticity' as a characteristic of discourse interpretation and 'genuineness' as a characteristic of texts. Corpus language may be genuine because this language is produced with real communicative goals (Gavioli & Aston 2001: 240). However, corpus language cannot be authentic because of its decontextualization and isolation from discourse (Lee 2011: 161). The implication for

language learners is that corpus language has to be authenticated in order to allow learners to understand and interpret the text. Authentication in this sense means to appropriately relate texts to their context, i.e. re-contextualizing them (Braun 2005: 52; Widdowson 2003: 104). This can be achieved by the teacher constructing an appropriate context that makes the texts meaningful to the learners (Widdowson 2003: 104). Alternatively, learners themselves create or re-construct appropriate contexts in their minds. Braun (2005: 53) illustrates this:

> Context construction heavily relies on subjective knowledge. Therefore, each learner will construct their own individual context, and the construction process will be greatly facilitated if the topic is familiar to, and interesting for, the learner.

While the replication of the original contextual conditions of texts is impossible, learners can re-contextualize and thus authenticate texts for their own purposes; but they need the willingness and the ability for authentication (Mauranen 2004: 93). Usually, learners need to be enabled to authenticate texts and materials for themselves (Braun 2007: 32) and pedagogic mediation between the learners and the corpus materials, in which the reconstruction of the discourses of corpus data is achieved, becomes necessary (Braun 2005: 61).

Having argued that corpus materials have to be authenticated, the question arises how this authentication can be achieved. The most common approach to authenticate corpus materials is to use corpora as tools for discovery learning or DDL as discussed above. With regard to the methodological approach chosen to enable learners to authenticate corpus materials, Braun (2005: 54), for instance, suggests complementing the corpus-based approach with a discourse-based approach "which focuses on the analysis of linguistic means of expression in relation to their communicative (situational) and cultural embedding". According to this, corpora should be studied on the whole, that is, texts are studied in their entirety in order to familiarize learners with these texts. Braun (2005: 54) claims that the interpretation and in-depth study of concordances afterwards will be greatly facilitated by this preceding whole-corpus reading. A complementary discourse-based approach allows learners to study larger text passages, context-specific meanings, expressions and their discourse functions (Braun 2005: 54). A focus on the wider cultural and social context of the texts, however, also requires that the corpus contents are chosen carefully in order to meet the pedagogical needs of the learners (Braun 2006: 29).

Besides the choice of approach and content of corpora, pedagogic enrichment of corpora supports the authentication process of corpus materials by learners. As suggested by Braun (2006: 38), pedagogic enrichment of corpus materials facilitates learners' text comprehension and provides opportunities to practice and test learners' knowledge.

Pedagogic enrichment can be achieved by the inclusion of audiovisual, informative and illustrative materials as well as exercises and tasks. For instance, with the *ELISA Corpus*, which is based on video interviews of English native speakers, Braun (2006) demonstrates how the enrichment of the corpus with the above mentioned materials helps learners in their text comprehension and exploitation of the corpus. Audiovisual materials, i.e. the video clips that complement the transcripts of the interviews, allow learners to clarify ambiguous utterances. Also visual and gestural information that is included facilitates the understanding of spoken language (Braun 2005: 55; Braun 2006: 38). Moreover, the video clips can serve as listening comprehension tasks which can – if supplemented with comprehension questions or exercises – contribute to successful comprehension training (Braun 2006: 38). The informative and illustrative materials, on the other hand, can include translations of the interviews, simplified transcripts, explanations or ready-made word lists or concordances. The benefit of such ready-made data is that it can be used by learners as well as teachers who have not been familiarized with corpus techniques (Braun 2006: 39). Last, ready-made exercises and tasks that can range from awareness-raising tasks to detailed studies of language items or more 'global' explorative tasks including more than one text on the same issue can help learners to contextualize corpus materials (Braun 2006: 42).

3.2.2. The Role of Learners and Teachers

As already briefly addressed above, corpus data have to be authenticated by learners in order to be relevant for the learning process. One way of achieving this is by applying data-driven learning, which allows learners to take on the role of researchers. This has the effect that the role of language learners changes considerably since the goals of researchers and language learners usually differ greatly: Researchers use corpora in order to formulate language descriptions. Learners, on the other hand, "generally aim to learn a language, rather than to learn about it" (Breyer 2011: 97). Besides, while language learners are in the process of acquiring language proficiency, researchers are already highly proficient and have distinct research skills. In addition to this, researchers generally choose to use corpora in order to pursue their research. Therefore, the initial motivation for corpus use by researchers may differ considerably from language learners' motivation (Breyer 2006: 162-163). Because of these differences in motivation and objectives, however, the term 'researcher' seems to be inappropriate. A more suitable metaphor for the role of learners engaged in DDL activities may be the one of 'language detectives' (Johns 1997: 101). Instead of viewing language

learners as corpus linguists who produce accurate descriptions of language, they are regarded as detectives solving mysteries of language (Breyer 2011: 100).

Aside from these metaphorical considerations, the idea of DDL involves a change of the learner's role "from a passive consumer of the teacher's output into an autonomous researcher" (Götz & Mukherjee 2006: 50) or active observer or participant (Knowles 1997: 201-202). On the one hand, learners can take on the role of an observer who explores texts from an analytical perspective by, for instance, noticing and identifying language patterns. On the other hand, learners can also act as participants when interacting with texts as they carry out tasks which test the patterns that had previously been identified, for example (Gavioli & Aston 2001: 241; Kaltenböck & Mehlmauer-Larcher 2005: 79).

In both cases, learners are likely to need guidance from the teacher whose role also changes. The idea that learners increase their autonomy through DDL activities does not immediately imply that learners become absolutely independent of the teacher. Instead, the teacher's role shifts from being the source of knowledge to being a director of student research (Knowles 1997: 201). This 'new' role has been described in various ways and labels such as "coordinator of student-initiated research" (Johns 2007 [1991]: 5), "learning rather than language experts" (Aston 1997: 63), or "adviser and facilitator" (Leech 1997: 7-8) have been assigned to the teachers. The change of role, however, does not reduce the teacher's significance in providing a supportive environment for the learners' successful learning process since learners still need guidance and support from the teacher in handling corpus data (Knowles 1997: 201). Furthermore, it has to be remembered that it is essential that teachers acquire corpus skills before they can be taught to the learners. Therefore, the teacher remains a crucial part of the learner's learning process (Mauranen 2004: 100).

The skills that teachers are required to have prior to applying corpus-based activities in the classroom are manifold. On the one hand, teachers are expected to possess some degree of corpus literacy (cf. section 3.1.2.). In addition to that, teachers need the appropriate skills to support learners in corpus analysis and to select and present corpus data accordingly as well as to guide learners in how to deal with corpus data (Breyer 2011: 107). Additionally, teachers have to supply learners with relevant tasks but also with feedback on the learners' performance (Leech 1997: 7-8).

The change of the teacher's role is, however, often accompanied by some challenges. In particular, the perception of the 'traditional' teacher role started to be questioned with the introduction of corpus-based activities in the classroom. This can be illustrated by the

following quotation. Johns (2007 [1991]: 5) claims that if learners are given the opportunity to carry out corpus searches independently, then

> the teacher does not know in advance exactly what rules or patterns the learners will discover: indeed, they will often notice things that are unknown not only to the teacher, but also to the standard works of reference on the language. […] Once the concordancer becomes an important focus of activity in the classroom, many old certainties start to crumble (for example the central position of the syllabus and of the teacher's key at the back of the textbook).

The challenges here are of pedagogical as well as linguistic nature. First of all, teachers may have difficulties with admitting ignorance when it comes to corpus findings of specific linguistic items by learners. The fact that learners' direct corpus applications may result in outcomes unknown even to the teacher can be challenging for teachers and their authority. Moreover, discovering facts about language side by side with the learners and thus being on the same level (socially as well as linguistically) may not necessarily be welcomed by either the teacher or the learners (Breyer 2011: 107-108). In addition to this, the expert role in linguistic matters becomes less essential while the skills to create a suitable learning environment become more crucial (Bloch 2009: 59). Generally, this change of the teacher role may be challenging for many teachers and requires much confidence, but it is facilitated by high levels of language proficiency, teaching experience and corpus skills (Breyer 2011: 108).

3.2.3. Autonomy

Direct corpus applications by language learners are often said to promote learner autonomy which Holec defines as "the ability to take charge of one's own learning" (1981: 3). Corpus-based activities are thus suited for learner-centered language learning, which fosters learner autonomy. As already mentioned, learners take on the role of researchers or detectives and become active in the learning process in order to control and direct this process. However, learner autonomy has to be developed and this development can be "understood as a gradual process on a continuum" (Kaltenböck & Mehlmauer-Larcher 2005: 80). In this process of developing learner autonomy, the teacher plays a crucial part in fostering learner autonomy in so far as to provide opportunities in which learners take on an active part in learning and in which the awareness of this learning process is raised. Breyer (2011: 65) suggests that "[r]aising awareness in learners of the learning process and fostering learning strategies is an integral part of learner autonomy". Since language learners do not automatically take on

responsibility for their own learning, teachers have to provide opportunities to foster autonomy. Depending on the learner's language proficiency, previous knowledge, age and attitudes towards autonomous learning, corpus-based activities that foster autonomy have to be chosen and adapted according to these factors because different corpus activities require different degrees of teacher mediation (Breyer 2011: 65-66; Kaltenböck & Mehlmauer-Larcher 2005: 80-81). The teacher can manipulate the degree of learner autonomy by choosing corpus-based activities that are more teacher-controlled for instance. In this case, the teacher selects and presents particular concordance lines to the learners. But also more learner-centered exercises in which learners cope with all concordance lines or in which they even conduct their own searches or research autonomously can be chosen. The last option is, however, usually only suited for the most advanced and proficient students and requires a high degree of autonomy (Breyer 2011: 97).

Learner autonomy is generally regarded to be beneficial to learners because of several reasons. First, learners are given the chance of 'learning how to learn', to find their own way of learning as well as the way of how to "regulate their learning pace, [and] satisfy their learning requirements" (Bernardini 2000: 79). Second, learner autonomy is often claimed to enhance motivation. Little (2016: 1), for example, claims that "autonomous learners have developed the reflective and attitudinal resources to overcome temporary motivational setbacks". When learners accept their responsibility for their own learning, they draw on their intrinsic motivation because they have a feeling of control, a feeling of being in charge and of having at least some degree of choice (Bernardini 2000: 79-80). In sum, learner autonomy is generally regarded as being beneficial to learners and can be fostered with the help of corpus applications and teacher mediation.

3.2.4. Availability of Corpora and Teaching Materials

One problem of the implementation of corpora in language pedagogy is the limited availability of corpora and corpus-based teaching materials. First of all, many corpora are still restricted in their use, i.e. they are available to researchers for an annual fee, for instance (Anderson & Corbett 2009: 9). However, an increase in freely available and online corpora can be noticed. Anderson & Corbett (2009: 183) provide an overview of currently available corpora "of which at least a sample of texts are available free of charge online and which, in most cases, may be explored using integrated analysis tools". These include, for example, the *British National Corpus (BNC),* the *Corpus of Contemporary American English (COCA),* the

Michigan Corpus of Academic Spoken English (MICASE) or the *TIME Corpus of American English* (cf. Anderson & Corbett 2009: 183-187).

In the context of language teaching and learning, however, the question arises whether corpora, which have been mainly compiled for research purposes, are also suitable for language learners. Braun (2005: 50), for example, claims that teachers and learners may easily be overwhelmed with the large size of mainstream corpora, with their content or with the amount of results they produce. Therefore, it has been suggested that pedagogically relevant corpora should be used in the teaching context. Breyer (2011: 113), for instance, states that pedagogically relevant corpora "should be manageable in size, contain texts relevant to the respective learning context and the curriculum, and ideally be accompanied by pedagogical enrichments, such as learning materials and ready-made concordances". However, pedagogically relevant corpora that are freely available are a rare phenomenon. One exception is the *ELISA Corpus* (cf. section 3.2.1.) which, however, also requires users to register. A last possibility of using pedagogically relevant corpora is that teachers themselves create DIY corpora that perfectly fit their learners' needs (cf. section 3.1.2.1.).

When it comes to the availability of teaching materials, similar limitations can be observed. Breyer (2011: 105) explains the lack of corpus teaching materials with the fact that "direct corpus applications are neither a standard component in curricula nor in textbooks". As briefly mentioned in section 3.1.1., there are some textbooks that include corpus-based materials such as the *CorpusLAB series* by Barlow & Burdine (2006), *Exploring Grammar in Context* by Carter, Hughes & McCarthy (2000) and *Touchstone* by McCarthy, McCarten & Sandiford (2005). Besides the limited availability of textbooks, there are hardly any online materials that would also meet the learners' needs (Breyer 2011: 106). While Tribble & Jones (1990) provide detailed descriptions of corpus activities for teaching literature, many other research publications include examples of what could be taught with corpora but fail to offer ready-to-use teaching materials (cf. e.g. Aston, Bernardini & Stewart 2004; Sinclair 2004; Wichmann, Fligelstone, McEnery & Knowles 1997). Of course, teachers have the possibility to create their own teaching materials; however, this is generally very time-consuming and often accompanied by many challenges (Breyer 2011: 106).

3.3. GAP BETWEEN RESEARCH AND PRACTICE

Having discussed some theoretical issues concerning various possible applications of corpora in language teaching and learning and the potentials as well as challenges that accompany these applications, this section will deal with the discrepancies between theory and practice of corpus applications in language teaching and learning. In particular, the gap between the advances of research on corpora for language teaching/learning and the reality of corpus use in the language classroom will be illustrated.

The popularity of corpus applications for teaching and learning purposes can be inferred from the increase of publications with regard to *what* should be learnt or taught as well as *how* language items should be learnt (cf. Aijmer 2009; Aston, Bernardini & Stewart 2004; Bernardini 2000; Breyer 2011; Gavioli & Aston 2001; Sinclair 2004; Wichmann, Fligelstone, McEnery & Knowles 1997). In addition to this, papers about the developments of corpora and corpus tools indicate great interest into this field (cf. e.g. Braun 2006, 2007; Breyer 2006; Meunier & Gouverneur 2009). While the potential and advantages for language teaching tend to be given greater attention in the literature, also the challenges of corpus use in the language classroom are increasingly addressed (cf. e.g. Breyer 2011; Farr 2008; Kaltenböck & Mehlmauer-Larcher 2005; Lee 2011; Römer 2011).

This popularity of corpora for language teaching and learning in the research community has, however, lead to the perception of many that the use of corpora is equally popular among language practitioners. The announcement of the second TaLC conference in Lancaster in 1996, for example, declared the following:

> While the use of computer text corpora in research is now well established, they are now being used increasingly for teaching purposes. This includes the use of corpus data to inform and create teaching materials; it also includes the direct exploration of corpora by students, both in the study of linguistics and of foreign languages. (cited in Stewart, Bernardini & Aston 2004: 1)

This rather optimistic evaluation of the use of corpora for language teaching and learning purposes is also shared by Sinclair (2004: 2):

> Now corpora, large and small, are seen by many teachers as useful tools, and are being put to use more and more every day. Access has become fairly easy on standard small computers, user-friendly software is available for most normal tasks, websites are accumulating fast, and corpora are almost part of the pedagogical landscape.

Also Meunier & Gouverneur (2009: 179) consider the large number of publications in this field as indication for the existence of corpora in the 'learning and teaching scene' and claim that the legitimacy of corpora "as useful pedagogical aids has now been established". With regard to DDL and teaching materials for teachers, O'Keeffee, McCarthy & Carter (2007: 24) argue that the existence of various homepages which provide links to resources and teaching materials are "evidence of the popularity of DDL among language teachers, many of whom post their materials online and conduct action research into the classroom application of these materials". Moreover, Hunston (2002: 1) regards corpora and their application in language teaching as a revolution in the study of language.

There is no doubt that researchers recognize the potential of corpora for language teaching, and its use in language pedagogy is welcomed with much enthusiasm. Kaltenböck & Mehlmauer-Larcher (2005: 66) remark that, "the integration of computer corpora into language teaching is often seen as a development well on its way, a general trend happening before our very eyes." However, more and more researchers recognize and admit that this 'revolution' in language teaching has not yet occurred in reality. In this context, Breyer (2006: 157-158) notes that "doubts have emerged more recently about to what extent teachers and learners are actually using corpus technology in their classrooms." Granath (2009: 48), while being more optimistic about the spread of corpora in language teaching, admits that this spread has not been as extensive as expected. Braun (2005: 48), Lee (2011: 160) and Götz & Mukherjee (2006: 50) agree that the integration of corpora in the language classroom is still a rare phenomenon. Also Aijmer (2009: 1) observes this discrepancy between the researchers' enthusiasm and the teaching practice and claims that it is the teachers' reluctance or their lack of corpus skills that prevent them from using corpora. Römer (2011: 206), on the other hand, argues that few teachers are aware of the availability of corpora as useful tools. With regard to the incorporation of corpora into language teaching and learning, she remarks that "much work still remains to be done in bridging the gap between research and practice" (Römer 2011: 206).

Despite growing awareness among researchers that the implementation of corpora in the language classroom has not developed as expected decades ago, only few studies provide data about the actual use of corpora in language teaching and learning. One of these few studies was conducted by Mukherjee (2004) who analyzed the use of and the awareness about corpora of teachers at secondary schools in Germany. The results of the survey revealed that nearly 80% of the participants had not heard about corpus linguistics prior to the survey and the *test workshop,* which introduced some key concepts about corpus linguistics as well as the

application of corpora in language teaching. After the introductory test workshop, almost 97% believed that teachers and learners could benefit from corpus data. However, out of these 97% only 13% thought that both teachers and learners may profit from corpus data. The remaining 84% regarded the use of corpus data beneficial only for teachers, not for learners. This perception highlights the gap between applied corpus-linguistic research and the average teacher's view. Mukherjee (2004: 242) remarks:

> [W]hile in applied corpus linguistics, there is an increasing tendency to focus on corpus-based activities carried out by increasingly autonomous learners […], most teachers think that corpus data are particularly useful for themselves.

The participants would focus on teacher-centered rather than learner-centered activities when using corpora in the classroom. Moreover, the efforts of applied corpus linguistics do not necessarily have direct effects on the actual teaching practice. As opposed to advances in applied corpus linguistics where a rapid development can be noticed, the use of corpora in language teaching (in Germany) has not become mainstream yet, as Mukherjee's (2004) study shows.

In addition to Mukherjee, also Thompson (2006) conducted a study in order to examine the extent to which corpora are used by teachers of English for Academic Purposes (EAP) at tertiary level in the United Kingdom. Questionnaires asking about the access to and use of corpora were sent out to member institutions of the British Association of Lecturers of English for Academic Purposes. The results have shown that a large part of these institutions participating in the survey either do not have access to corpora at all or do not use corpora in any way. Thompson's findings indicate that the use of corpora in EAP at tertiary levels is "clearly limited, and in many cases non-existent" (2006: 14).

Similar results were obtained in a survey by Breyer (2011: 121) who investigated "to what extent teacher trainees of English in Germany encounter corpus-based language learning". For practical reasons, teacher educators at tertiary institutions were asked about their awareness of corpora and their corpus use. Regarding the familiarity with corpus linguistics, more than half of the participants stated that they would be 'vaguely familiar', and about 26% stated to be 'not at all familiar' with corpus linguistics. This means that only a small percentage (i.e. about 15%) of the participating teacher educators were 'very familiar' with corpus linguistics. Furthermore, out of those teacher trainers having at least some knowledge about corpora less than half actually made use of corpora in their teaching. Breyer (2011: 152) infers from this data that "consequently only a small portion of teacher *trainees* encounter corpora in language practice or teaching methodology courses" [my emphasis].

All of these studies indicate that despite the enthusiasm of researchers about the development of corpora for language teaching, in reality a far smaller proportion of teachers is aware of and familiar with corpora than expected, and even less are actually using them. In other words, a huge gap between research and the teaching practice can be assumed. Römer (2009) suggests bridging this gap by finding out about the current situation and the needs of teachers. After identifying the wishes and problems of teachers, namely "better teaching materials, support in creating materials, native speaker advice, and more reliable reference tools" (Römer 2009: 89), Römer concludes that "many of the problems teachers have could be solved, at least partially, if they were introduced to some basic corpus resources and received more support from corpus researchers" (2009: 95). Bridging this gap means to 'spread the word' about corpora which is, according to Römer (2009: 95), a task that corpus researchers have to accomplish. On the one hand, teachers have to be informed about the availability of corpora, corpus tools, corpus exercises, etc. On the other hand, teachers have to be convinced that they and their learners can profit from learning with corpora. In addition to this, Römer (2009: 95) recommends to carry out more research 'at the interface' to language teaching in order to learn more about the teachers' situation, needs and wishes. Also Mukherjee (2004) addresses the need for popularization of corpora among language practitioners and stresses the importance of appropriate teacher training as prerequisite for the implementation of corpora in the language classroom (for the discussion about teacher training see section 4.5.).

What this brief overview has shown so far is how the efforts of applied corpus linguistics differ from the actual teaching practice. While some researchers seem to believe that the corpus revolution of corpus use has equally taken place in corpus research as in the teaching practice, studies such as those from Mukherjee (2004), Thompson (2006) and Breyer (2011) indicate the opposite, i.e., only few teachers actually know of and use corpora in their teaching. As already mentioned, these studies focus on teachers from Germany and the United Kingdom, while there is a lack of research concerning the familiarity with and use of corpora by teachers in Austria. Therefore, the following survey, which is conducted as part of this diploma thesis, focuses on Austrian teachers of English as a Foreign Language at secondary institutions. The aim, method and results of this survey are presented in the following section.

4. SURVEY

4.1. AIM

The previous section has shown how various researchers assume that the use of corpora in the EFL classroom had increased just as much as in corpus research. On the other hand, more and more researchers start recognizing that the reality does indeed look different. In the real teaching practice, hardly any teachers seem to be familiar with the use of corpora for language teaching, not to mention their actual application by teachers and learners (cf. Breyer 2011; Mukherjee 2004).

While the studies by Breyer (2011) and Mukherjee (2004) investigated the awareness about and use of corpora by German secondary school teachers and tertiary teacher trainers, a comparable study including Austrian secondary school teachers is (as far as I am informed) non-existent. Therefore, the following study focuses on EFL teachers working in secondary schools in Austria. When considering Lee's (2011: 160) claim that the application of corpora in secondary education "has been a rare occurrence" and Kaltenböck & Mehlmauer-Larcher's (2005: 65-66) statement that "comparatively little has been said about the actual use of computer corpora in an English as a Foreign Language (EFL) or English as a Second Language (ESL) classroom", the need for a closer investigation of corpus use in Austrian EFL classrooms becomes apparent. This study, thus, aims to provide data about the teaching practice in Austrian EFL classrooms with regard to corpus applications. At the same time, the study tries to find evidence for what many researchers claim, i.e. that corpora are still absent in EFL classes of secondary institutions (cf. section 3.3.); in this case of secondary institutions of Austria. Once again, attention will be drawn to the question whether a gap between the advances in research about corpus applications for language pedagogy and the reality of the actual teaching practice indeed exists.

As Mukherjee (2004) has illustrated, only a fraction of German teachers had been familiar with corpus linguistics prior to their introduction in a workshop complementing the survey. This survey also aims to investigate the familiarity with corpus linguistics, however, in this case of Austrian teachers. In other words, one essential aim of the present study is to examine the extent to which Austrian EFL teachers are familiar with corpora as well as with the use of corpora for language teaching and learning. As a next step, the actual teaching practice with regard to corpus use is explored: Since being familiar with corpora does not necessarily imply that corpora are actually being used in or outside the classroom, another aim of the study is to examine the actual applications of corpora by teachers. In other words, the

number of Austrian secondary school teachers who have used corpora with regard to teaching and learning purposes is determined.

Apart from illustrating the actual teaching practice, another objective of this study is to outline the teachers' attitudes towards the use of corpora. Moreover, researchers usually stress the potentials of corpora for language teaching and learning, whereas only few also draw attention to their limitations (cf. e.g. Kaltenböck & Mehlmauer-Larcher 2005; Lee 2011). These limitations of corpus use as proposed by researchers, however, do not necessarily have to correspond with the actual challenges that teachers face when applying corpora in the classroom. Therefore, the study focuses in particular on possible challenges and barriers that prevent teachers from using corpora for language teaching purposes.

These insights into the most prominent barriers preventing teacher from corpus use may also shed light on possible improvements that have to be made in the future in order for more teachers to make use of corpora in the EFL classroom. Assuming that various factors hinder teachers from directly applying corpora, this study not only aims to detect these factors, but also aims to determine and list possible improvements and developments based on the needs of language teachers.

In general terms, the study aims to provide data about the current teaching situation in Austrian EFL classrooms with regard to the use of corpora. In addition to this, by conducting the study and approaching Austrian EFL teachers, the study tries to 'spread the word' (Römer 2009: 95) about corpora. While a number of teachers participating in the study may have already come across corpora, others are introduced to the use of corpora by participating in the survey. In both cases, the aim remains the same: to spread information about the existence of corpora and ways of corpus application among Austrian EFL teachers at secondary institutions.

4.2. DATA

In order to gain data for the study, a questionnaire covering relevant questions about familiarity, use and barriers as well as improvements of corpora was created (for a detailed description of the questionnaire see section 4.3.). The questionnaire was designed, distributed and answered online, which had several benefits: First, designing the questionnaire layout was facilitated by the pre-built templates as provided by an online software package provider.[1]

---

[1] Software used for the questionnaire: SoSci Survey: Leiner, D. J. (2014). SoSci Survey (Version 2.5.00-i) [Computer software]. Available at http://www.soscisurvey.com

Individual questions could easily be designed and changed if necessary. Second, the online survey allows researchers to distribute the questionnaire via the internet. This means that the questionnaire is available online; participants receive the link to the questionnaire and can then immediately start answering the questions. Moreover, online surveys save time and costs compared to paper questionnaires that are distributed via mail. The only requirement in order to successfully complete the survey is an active internet connection on the part of the participants. In addition, the researcher immediately receives the data from the questionnaire when a participant finishes it. The data is available online, can be downloaded and processed with programs such as *Microsoft Excel* or *SPSS*.

After the design phase, the questionnaires were sent out to the target audience. Primarily, the target audience consisted of teachers of English as a Foreign Language (EFL) teaching at Austrian secondary schools. On the one hand, these schools included *Berufsbildende Höhere Schulen* (BHS) such as *Handelsakademien* (HAK) or *Höhere Technische Lehranstalten* (HTL); on the other hand, the questionnaires were also sent to teachers at *Allgemeinbildende Höhere Schulen* (AHS). As this school type encompasses upper levels (9th – 12th grade) as well as lower levels (5th – 8th grade), it is possible that a minor percentage of these respondents have solely been teaching at lower secondary levels.

The questionnaires were sent to randomly selected AHS and BHS in Vienna, Lower Austria, Upper Austria, Salzburg, Tyrol and Styria in order to get a wider distribution and a higher return rate. Nevertheless, equal distribution among these Austrian states cannot be guaranteed. Most requests asking teachers to participate in the survey were directly sent to the schools; some teachers were contacted directly who then forwarded the questionnaire to their colleagues. All in all, 65 questionnaires had been at least started. Out of these 65, however, 45 were fully completed and were taken into account for the analysis of the results. Regarding the respondents' sociodemographic data, it can be said that 32 women and 13 men were participating. The age of the respondents was divided into six age groups: <20, 20-29, 30-39, 40-49, 50-59 and >60 years. There was no respondent younger than 20 years; however, the majority of 18 people were between 20 and 29 years. Age group 30-39 consisted of 8 respondents, the age groups 40-49 as well as 50-59 each included 7 participants, while only 5 people were 60 years or older. With regard to the respondents' teaching experience, it has to be mentioned that 21 participants have been teaching for five or fewer years. The highest number is 40 years and the lowest number is one year. While the median is 10 years, the arithmetic mean of the years of teaching experience is 12.9 years. Additionally, the respondents were asked about their educational institution, i.e. the place of where they had

earned their teaching degree. The majority of 25 people finished their studies at the University of Vienna, followed by the University of Salzburg (with 11 respondents), the University of Innsbruck (with 6 respondents) and the University of Graz (with 2 respondents). One participant had earned his/her degree from a foreign university in the United Kingdom.

4.3. METHOD

The study was conducted by using a quantitative approach. Quantitative research is concerned with quantifiable information, i.e. numbers, and asks questions about how much or how many there is/are of whatever somebody is interested in (Rasinger 2008: 10). Because questionnaires can be a useful tool in obtaining relevant data, this tool was selected for the study at hand. In addition to the quantitative analysis of the questionnaire data, a qualitative element in the form of three open-response questions was added to the study. While the response to these questions was optional, the questions allowed the participants to express their opinions or share their thoughts, impressions or suggestions at different stages of the survey. These responses are classified as qualitative data, which "deals with the question of *how* something is, as opposed to *how much/many*" (Rasinger 2008: 11). Since these responses were optional, they primarily serve as additional information and provide further insights into the participants' attitudes. Although these open-ended questions are still restricted in their 'openness' and length since "[q]uestionnaires are not the right place for essay questions", they can still be very valuable in case the questionnaire fails to cover all possible response options (Dörnyei 2003: 47).

The questionnaire was constructed according to various principles as suggested by Dörnyei (2003). On the one hand, the questionnaire is supposed to have a four-page limit but should definitely not exceed six pages, and on the other hand, the participants should not have to spend more than 30 minutes on answering the questions (Dörnyei 2003: 18). As can be seen in the appendix, the questionnaire did not exceed the six-page limit. Furthermore, the average time spent on answering the questions was 5:54 minutes (median = 5:36; max. = 10:31; min. = 3:41).

Dörnyei also suggests some guidelines regarding the main parts of a questionnaire (2003). Besides essential components such as the title at the beginning of the survey, a questionnaire has to give informative instructions. General instructions following the title provide the respondents with information about the content of the study, about confidentiality and express the researcher's gratitude for participation. Specific instructions, on the other

hand, are explanations of how the respondents should answer the questions or how different types of rating scales work (Dörnyei 2003: 25-28).

At the heart of the questionnaire are the actual items, which are divided into different sections. The first section is concerned with the respondents' familiarity with corpus linguistics and corpora for language teaching and learning. For this purpose, multiple choice questions were designed, which asked the respondents about their knowledge of corpus linguistics, their familiarity with a list of specific corpora as well as their past behavior regarding the use of corpora. For instance, one of the questions was as follows, "How familiar are you with the use of corpora for language teaching?" The participants were then asked to answer the question by choosing one of the following options: "Very familiar", "Vaguely familiar" or "Not at all familiar".

The next section provided the respondents with the opportunity to try corpus exercises themselves. These exercises were optional and served the purpose of giving insights into the possibilities of how corpora could be used in the EFL classroom. Also, if the respondents had not been familiar with the use of corpora, they would get a first impression of how corpora could be used as an additional tool in language teaching and learning. Additionally, by trying these exercises the participants might also get an idea of potential barriers of corpus use. Each of the two exercises consisted of a task and step-by-step instructions. The second task, for instance, asked the participants, "What is the difference between *although* and *despite* in terms of sentence structure? Use the link below (COCA) and look for the structures that follow *although* and *despite*." The instructions, then, gave precise information on how the task should be completed:

> Use the 'KWIC' display and type in 'although'. Click on 'search'. Do the same with 'despite'. When you pay attention to the structures that follow *although* and *despite*, you may notice that *although* is usually followed by a clause, while *despite* is only followed by a noun or gerund. (cf. Appendix I.)

As can be seen in this example, a common interpretation of the results based on the *Collins COBUILD English Language Dictionary* as well as the *Practical English Usage* (Swan 2005) was given beforehand. This would allow the participants to check whether their own findings were similar to those provided. In order to access a corpus without any problems and to ensure that the findings would not deviate from the given results when using different corpora, the link to the Corpus of Contemporary American English (COCA) was provided below the exercises. The COCA was primarily chosen because of its free online access. After 10 queries, however, users are required to sign up. Also, the COCA is a rather large corpus

containing over 520 million words (cf. COCA, available online at http://corpus.byu.edu/coca/, 24/05/2016) and is rather simple to use because of its clear layout. The exercises were followed by an optional open-response question giving the participants the opportunity to share their experience, opinions or thoughts about the use of corpora.

The third section of the questionnaire was dedicated to the barriers of corpus use in language teaching and learning. The respondents were asked to agree or disagree with twenty statements such as, "Using corpus activities in my lessons is too time-consuming". The responses were based on Likert scales where the participants are required to "make an evaluative judgement of the target by marking one of a series of categories organized into a *scale*" (Dörnyei 2003: 36). The response options for the statements included "Agree", "Slightly agree", "Slightly disagree", "Disagree" as well as "Don't know". Although controversially debated (cf. e.g. Dörnyei 2003; Rasinger 2008), the neutral category "Don't know" was included in order to give respondents the opportunity to answer questions that might be unpleasant. Furthermore, there was the possibility that respondents were not able to answer certain questions because of a lack of knowledge about this particular issue.

Regarding item wording, Dörnyei (2003: 52-54) suggests to keep the language simple and natural, to avoid ambiguous words, modifying words (such as 'just', 'only') or nonspecific adjectives or adverbs (such as 'good' or 'sometimes'). Moreover, negative constructions containing words such as 'no' or 'not' are to be avoided because answering those can be problematic and confusing. However, the questions should include positively as well as negatively worded items so as to "avoid a response set in which the respondents mark only one side of a rating scale" (Dörnyei 2003: 55). Besides these considerations for item wording, attention had also been paid to the grouping of items (cf. Dörnyei 2003: 32-35). The statements were formulated having more general problems or 'problem areas' in mind in which the statements could be grouped. In particular, this means that two to four statements that have a different wording ask about one and the same barrier. For instance, the statements "Using corpus activities in my lessons is too time-consuming" and "When using corpora, too much time is spent on too specific language issues" are both concerned with the challenge of time restrictions. The other challenges aside from 'time' are 'competence', 'changing teacher and learner roles', 'availability of/access to corpora/materials', 'familiarity' and 'relevance' (see section 4.4.2.). In the case that only the result of combined statements was given, a test for internal consistency reliability (Cronbach Alpha) was conducted (Dörnyei 2003: 112-113).

Another question, which addressed the barriers of corpus use, was added. However, this question explicitly asked the respondents what they regarded as the greatest obstacle of

corpus use in language teaching. While the participants were again asked to give their judgment on a rating scale, this time a semantic differential scale was chosen. For answering the question on this scale, the respondents were "asked to indicate their answers by marking a continuum […] between two bipolar adjectives on the extremes" (Dörnyei 2003: 39). In this case, the opposite extremes were 'small obstacle' which was assigned the number '1' and 'great obstacle' which was assigned the number '4'. Two more options (2 and 3) were added for ratings in between the two extremes and the option "Don't know" was included again.

The fourth section focused on possible improvements of corpus use. At the same time, the respondents' future behavior with regard to corpus use was implicitly predicted. The participants were asked to express agreement or disagreement with statements such as the following: "I would use corpora in the future if there was more time in class". The response options were "Agree", "Slightly Agree", "Slightly Disagree", "Disagree" and "Don't know". In addition to these statements, another open-ended question was incorporated in order to give the respondents the opportunity to share their own ideas about future improvements or developments of the application of corpora in the teaching and learning context.

In the last section of the questionnaire, sociodemographic data about the respondents were collected. These included information about gender, age, years of teaching experience, educational institution, computer skills and whether the participants had come across corpora during their studies at university. As already indicated, the age of the respondents was divided into the age groups <20 years, 20-29, 30-39, 40-49, 50-59 and >60 years. Regarding the years of teaching experience, the respondents were asked to give the exact number of years and in terms of the respondents' computer skills they could choose from the options "Basic", "Intermediate" or "Advanced".

Prior to the analysis of the results, it was assumed that the respondents' age group as well as their educational institution would play a crucial part in terms of familiarity with corpora. In order to test this, the correlations between different variables were calculated with a chi-square test, which determines the relationship between two or more categorical variables (cf. e.g. Rasinger 2008: 144-145; Müller-Benedict 2011: 191-199). Since, however, the results that are obtained with the chi-square test are not standardized, another coefficient, i.e. the contingency coefficient C, is presented in the results section as well. This coefficient converts the results of the chi-square test into values between 0 and 1 (cf. e.g. Hagl 2008: 84-85; Müller-Benedict 2011: 199).

With regard to the reliability of the results and the sample size, it has to be mentioned that there is disagreement about the effect that the sample size has on survey data quality (cf.

e.g. Breyer 2011: 122). Dörnyei (2003: 74), for example, suggests that a sample should have a normal distribution and include 30 or more participants. Statistical significance, according to Dörnyei (2003: 74), could be achieved when the sample includes around 50 participants. Regarding the reliability of chi-square test results, Rasinger (2008: 148) also addresses the question whether a minimum count of 5 in each cell is required for a reliable result. He, however, believes that this minimum count does not matter (Rasinger 2008: 148). Sticking to Rasinger's view, the study at hand presents correlations that were calculated without a minimum count of 5 in each cell. What this possible limitation and the sample size of 45 imply is that generalizations based on the findings from the data have to be made with caution.

4.4. RESULTS

In this part of the thesis the findings of the survey are presented. The section is divided into three main parts that are structured according to the research questions as presented in the introduction (section 1.). Accordingly, the first subchapter discusses the findings regarding the familiarity of the respondents with corpora. The second subchapter deals with the barriers of corpus use in the EFL classroom and provides insights into the teachers' attitudes towards corpus applications. Finally, the third subchapter illustrates the findings about potential improvements and developments necessary for a wider and more frequent application of corpora in the classroom.

4.4.1. Familiarity with Corpora

As already mentioned in section 4.3., the first part of the questionnaire was created with the aim of gathering information about the familiarity of Austrian EFL teachers with corpora as well as about their usage. In order to achieve this, two questions about the familiarity with corpus linguistics and corpora for language pedagogy were asked. Furthermore, one question investigated the familiarity with specific corpora and a fourth question focused on the use of corpora for different purposes.

First of all, the respondents were asked to rate their familiarity with corpus linguistics. The answers to the question "How familiar are you with corpus linguistics" as illustrated in Figure 1 show that the majority of the participants have at least some knowledge of corpus

linguistics. To be more precisely, 15.6% of the respondents claim to be 'very familiar' and even 55.6% claim to be 'vaguely familiar' with it. By contrast, 28.9% of the respondents are 'not at all familiar' with corpus linguistics (cf. Figure 1). In connection with this question, the respondents were also asked whether they were familiar with any specific corpora such as the British National Corpus (BNC). The findings suggest that four corpora are widely known by the participants. These corpora encompass the British National Corpus (BNC) (known by 51%), the Corpus of Contemporary American English (COCA) (44%), followed by the Vienna-Oxford International Corpus of English (VOICE) (31%) and the Corpus of Historical American English (COHA) (17%). While the Brown Corpus, the Lancaster-Oslo-Bergen Corpus (LOB) as well as the Michigan Corpus of Academic Spoken English (MICASE) were chosen occasionally, the Freiburg-Brown Corpus of American English (FROWN), the Freiburg-LOB Corpus of British English (FLOB) and the Michigan Corpus of Upper-Level Student Papers (MICUSP) are known by none of the participants. In sum, 60% of the respondents picked at least one of the corpora, which means that the majority has at least heard of or used some of these corpora before.
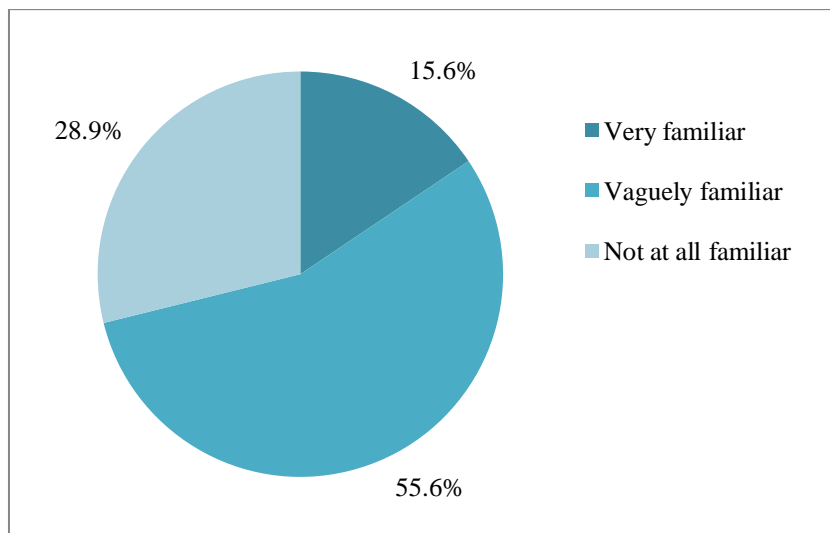

Figure 1: Respondents' familiarity with corpus linguistics

8.9%

■ Very familiar

■ Vaguely familiar

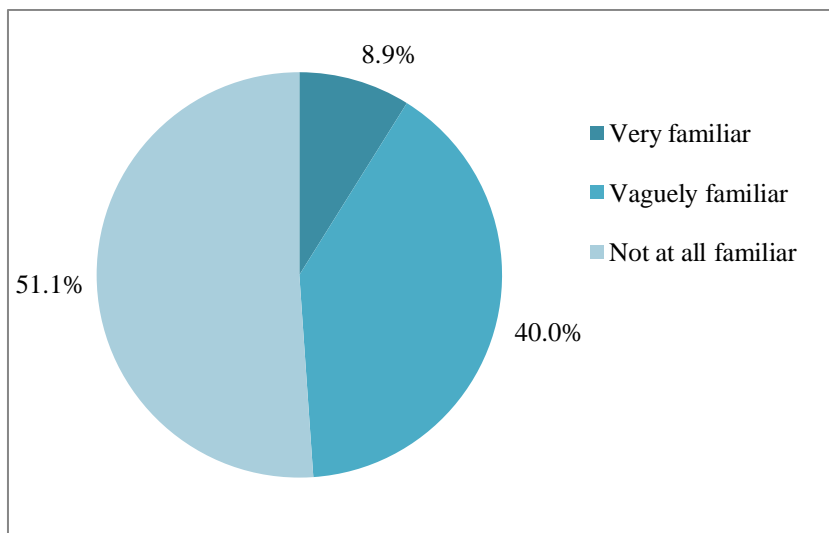■ Not at all familiar

51.1%

40.0%

Figure 2: Respondents' familiarity with corpora for language teaching and learning

The next step was to analyze the respondent's familiarity with corpora in connection with language pedagogy. Compared to the familiarity with corpus linguistics, the results of the question concerning the respondents' familiarity with corpora for language teaching and learning show a clear decrease in familiarity. As the pie chart in Figure 2 illustrates, only 8.9% of the respondents state that they are 'very familiar' with corpora for language teaching, whereas 40% claim to be 'vaguely familiar'. On the other hand, the percentage of those being 'not at all familiar' with corpus applications in language pedagogy has increased to 51.1%. Overall, a comparison between the familiarity with corpus linguistics and the familiarity with corpora for language pedagogy clearly demonstrates that the knowledge about the use of corpora in the teaching context is not as widespread as the knowledge about corpus linguistics.

A further step in the investigation was to test whether there is a relationship between the familiarity with corpus linguistics or corpora in language pedagogy and the age of the participants. In other words, the question was whether younger teachers are more/less familiar with corpora than elderly teachers. In order to investigate this, a contingency table presenting the distribution of answers of categorical data was created (Rasinger 2008: 145). The following contingency tables illustrate the number of participants of each age group who chose the response options 'very familiar', 'vaguely familiar' and 'not at all familiar' with corpus linguistics (Table 1) and with corpora for language teaching (Table 2).

	20-29	30-39	40-49	50-59	>60	Total
Very familiar						**7**
Count	4	0	0	1	2	
% of age group	22.2%	0.0%	0.0%	14.3%	40.0%	
Vaguely familiar						**25**
Count	13	6	3	2	1	
% of age group	72.2%	75.0%	42.9%	28.6%	20.0%	
Not at all familiar						**13**
Count	1	2	4	4	2	
% of age group	5.6%	25.0%	57.1%	57.1%	40.0%	
Total						**45**
Count	**18**	**8**	**7**	**7**	**5**	
% of age group	100.0%	100.0%	100.0%	100.0%	100.0%	

Table 1: Respondents' familiarity with corpus linguistics according to age groups

	20-29	30-39	40-49	50-59	>60	Total
Very familiar						**4**
Count	1	0	0	1	2	
% of age group	5.6%	0.0%	0.0%	14.3%	40.0%	
Vaguely familiar						**18**
Count	10	2	3	2	1	
% of age group	55.6%	25.0%	42.9%	28.6%	20.0%	
Not at all familiar						**23**
Count	7	6	4	4	2	
% of age group	38.9%	75.0%	57.1%	57.1%	40.0%	
Total						**45**
Count	**18**	**8**	**7**	**7**	**5**	
% of age group	100.0%	100.0%	100.0%	100.0%	100.0%	

Table 2: Respondents' familiarity with corpora for language teaching according to age groups

As Table 1 displays, the vast majority of respondents from age group 20-29 are either 'very familiar' (4 respondents) or 'vaguely familiar' (13 respondents) with corpus linguistics, leaving only 1 person who is 'not at all familiar'. Also in age group 30-39 most respondents (6 individuals) are 'vaguely familiar', while the remaining two are 'not at all familiar' with corpus linguistics. In contrast, the age groups 40-49 as well as 50-59 both include more participants being 'not at all familiar' with corpus linguistics. A shift can again be noticed when having a look at the respondents from age group >60 years: two of the five individuals claim to be 'very familiar' with corpus linguistics.

A first glimpse at the data suggests that the familiarity with corpus linguistics tends to be in relation to the age of the respondents. In order to test this, a chi-square test was conducted with the result of $\chi^2 (8) = 15.99$, p < 0.05. Since the critical value at this level is 15.51 and the χ^2 value exceeds the critical value, the result can be considered significant (Rasinger 2008: 148). In other words, there seems to be a correlation between the variables

'familiarity with corpus linguistics' and 'age' of the respondents. However, since the chi-square coefficient is not standardized and increases with a growing number of observations, the contingency coefficient C which standardizes the results to values between 0 and 1 was calculated. The result of C = 0.512 indicates that there is a correlation of moderate strength (Eckey, Kosfeld & Türck 2008: 164-165; Müller-Benedict 2011: 199).

Similar results were obtained with the analysis of the relation between 'familiarity with corpora for language pedagogy' and the respondents' age according to the age groups. The results χ^2 (8) = 13.86, p < 0.1 and C = 0.485 show that there is a relation between the two variables; however, this relation is not as strong as the previous one. When comparing the results from the two tables (Table 1 and 2), it is observable that the familiarity with corpus linguistics compared to language pedagogy changes only in the first two age groups. Conversely, the distribution of the results in the age groups 40-49, 50-59 and >60 remains the same. This suggests that the respondents from these age groups are either 'very familiar', 'vaguely familiar' or 'not at all familiar' with both corpus linguistics and corpora for language teaching. On the other hand, a shift from 'familiarity' to 'unfamiliarity' between corpus linguistics and corpora for language teaching in the age groups 20-29 and 30-39 can be noticed. In other words, while the respondents (20-29 and 30-39) tend to be quite familiar with corpus linguistics, they seem to be less familiar with the use of corpora in the teaching context.

The decrease in familiarity with the use of corpora for language teaching and learning is also reflected in the particular uses of corpora by the respondents. The answers to the question whether the respondents have used corpora for their own research, as a reference tool, for designing teaching materials or for teaching (e.g. for explorative learning of pupils in the classroom) are visualized in Figure 3.

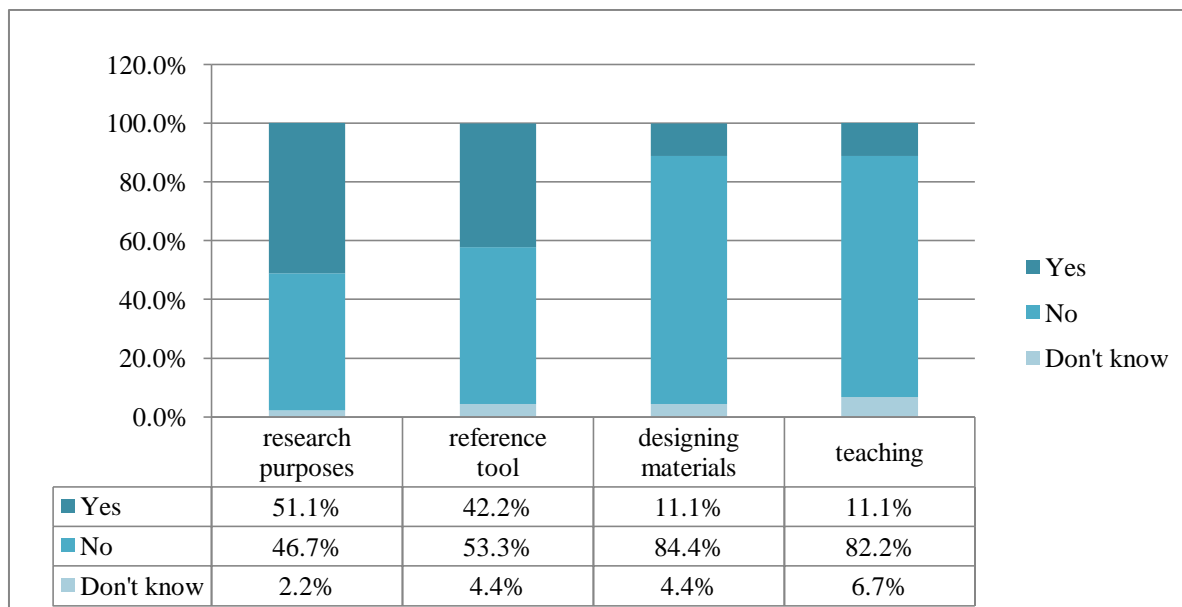|	research purposes	reference tool	designing materials	teaching
■ Yes	51.1%	42.2%	11.1%	11.1%
■ No	46.7%	53.3%	84.4%	82.2%
■ Don't know	2.2%	4.4%	4.4%	6.7%

Figure 3: Respondents' uses of corpora according to types of use

As can be seen in Figure 3, 51.1% of the respondents have used corpora for their own research and 42.2% have used them as a reference tool for correcting or as an alternative to a grammar book or dictionary for instance. A clear difference in terms of the frequency of use between these results and the last two types of uses, i.e. designing teaching materials and teaching as such, can be identified. In the case of the design of teaching material as well as teaching, only 11.1% of the participants claim to have used corpora for these purposes. Again, these findings are in accordance with those from above: A lack of familiarity with the use of corpora for language teaching is mirrored in the findings that corpora are indeed hardly ever used for purposes such as designing teaching materials or teaching as such. On the other hand, a greater familiarity with corpus linguistics is reflected in a greater use of corpora as reference tool or for research purposes.

In order to find evidence for the assumption that participants who are 'very familiar' with corpora for language teaching are more likely to use corpora than those who are 'vaguely familiar', the correlations between the variables 'familiarity with corpora for language teaching' and each type of use were tested with a chi-square test. Table 3 lists the results of these calculations and illustrates that there are relations between each of the types of uses with the variable 'familiarity'.

	Chi-square	Contingency Coefficient C
Research Purpose	$\chi^2 (4) = 15.58$, $p < 0.01$	0.507
Reference Tool	$\chi^2 (4) = 18.60$, $p < 0.001$	0.540
Designing Materials	$\chi^2 (4) = 35.49$, $p < 0.001$	0.664
Teaching	$\chi^2 (4) = 20.37$, $p < 0.001$	0.558

Table 3: Correlations between the familiarity with corpora in language teaching and types of uses

Regarding the use of corpora for research purposes as well as the use of corpora as a reference tool, the test results indicate that there is a correlation of moderate strength between each variable and the variable 'familiarity with corpora for language teaching'. In other words, those respondents who are 'very familiar' with corpora for language teaching are more likely to have used corpora for research purposes or as a reference tool than those being 'vaguely familiar' or 'not at all familiar'. A slightly stronger correlation can be noticed with the use of corpora for the design of teaching materials and for teaching on the one hand and the respondents' familiarity on the other hand (cf. Table 3).

Besides the age and types of uses of corpora in relation to familiarity, also the factor 'familiarity with corpus linguistics' in connection with the place of where the respondents had earned their teaching degree was analyzed. Again, a chi-square test about the familiarity with corpus linguistics and the postsecondary educational institutions of the respondents was conducted which resulted in the following: $\chi^2 (8) = 16.35$, $p < 0.05$ and C = 0.516. Similarly, the relation between 'familiarity with corpora for language teaching' and 'educational institution' was calculated: $\chi^2 (8) = 16.86$, $p < 0.05$ and C = 0.522. It can be said again that there is a correlation (of moderate strength) between those variables.

	Vienna	Graz	Innsbruck	Salzburg	Other	Total
Very familiar						
Count	2	1	2	2	0	**7**
% of place	8.0%	50.0%	33.3%	18.2%	0.0%	
Vaguely familiar						
Count	19	1	2	2	1	**25**
% of place	76.0%	50.0%	33.3%	18.2%	100.0%	
Not at all familiar						
Count	4	0	2	7	0	**13**
% of place	16.0%	0.0%	33.3%	63.6%	0.0%	
Total						
Count	**25**	**2**	**6**	**11**	**1**	**45**
% of place	100.0%	100.0%	100.0%	100.0%	100.0%	

Table 4: Respondents' familiarity with corpus linguistics according to educational institution

	Vienna	Graz	Innsbruck	Salzburg	Other	Total
Very familiar						
Count	0	0	2	2	0	**4**
% of place	0.0%	0.0%	33.3%	18.2%	0.0%	
Vaguely familiar						
Count	12	2	3	1	0	**18**
% of place	48.0%	100.0%	50.0%	9.1%	0.0%	
Not at all familiar						
Count	13	0	1	8	1	**23**
% of place	52.0%	0.0%	16.7%	72.7%	100.0%	
Total						
Count	**25**	**2**	**6**	**11**	**1**	**45**
% of place	100.0%	100.0%	100.0%	100.0%	100.0%	

Table 5: Respondents' familiarity with corpora for language teaching according to educational institution

The distribution of answers according to the familiarity and the respondents' postsecondary educational institution is given in Table 3 and Table 4. What is most striking about this distribution is the insight that respondents who had studied in Vienna (in sum 84%) as well as the respondents from Graz (100%) seem to be quite familiar with corpus linguistics. However, since only two participants had studied at Graz, the latter results are rather unrepresentative. More interestingly, the familiarity with corpus linguistics among participants from Salzburg seems to be much smaller (36.2%) when compared with the results of Vienna. Table 4 shows that, the familiarity with corpora for language pedagogy generally decreases when compared with the familiarity with corpus linguistics. The only exception is Innsbruck (16.7% being 'not at all familiar'). However, the comparison of Vienna and Salzburg, for instance, shows that corpora are less known by respondents from Salzburg (72.7% 'not at all familiar') than from Vienna (52.0% 'not at all familiar').

The last calculation of association was concerned with the correlation between the respondents' familiarity with corpus linguistics and whether the respondents had come across corpora during their studies at university. Also in this case, a moderate association could be found, which is displayed by the following results: $\chi^2 (4) = 22.00$, $p < 0.001$ and C = 0.573. This indicates that also in the case of the variables 'familiarity with corpus linguistics' and 'educational institution' a correlation of moderate strength can be evidenced.

4.4.2. Attitudes towards Corpora and Barriers of Corpus Use

After having examined the familiarity of teachers with corpus linguistics and corpora for language pedagogy as well as their usage, the following section discusses those barriers of corpora that most likely prevent teachers from applying corpora in the EFL classroom. However, it has to be mentioned that only a minority of participants has experience with the application of corpora in EFL lessons. Therefore, the answers to many of these questions most likely reflect the respondents' attitudes towards the potentials and barriers of corpus use rather than their hands-on experience.

As discussed in section 4.4.1., the respondents' familiarity with corpus linguistics may be regarded greater than the familiarity with corpora in connection with language pedagogy. This discovery can again be confirmed after having analyzed the answers to two questions regarding the teachers' competencies in using corpora.

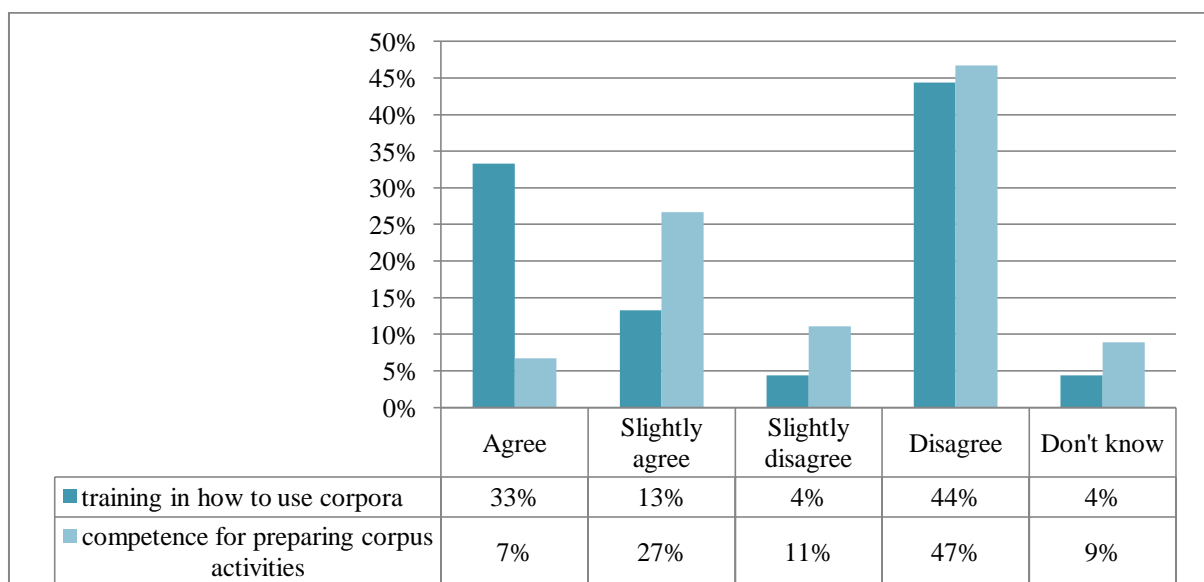|	Agree	Slightly agree	Slightly disagree	Disagree	Don't know
■ training in how to use corpora	33%	13%	4%	44%	4%
■ competence for preparing corpus activities	7%	27%	11%	47%	9%

Figure 4: Respondents' agreement/disagreement with the statements 'I have had training in how to use corpora' and 'I know how to prepare relevant corpus activities'

As Figure 4 shows, 33% of the respondents agree with the statement "I have had training in how to use corpora", while only about 7% agree with the statement "I know how to prepare relevant corpus activities". On the other hand, more than 40% disagree with both statements. These findings suggest again that there is a gap between the familiarity of teachers with corpus linguistics on the one hand and the familiarity with the possibilities of applying corpora for language teaching and learning on the other hand. Moreover, it comes as no surprise that 71% of the respondents disagree with the statement "I have been looking for

suitable corpora for teaching" while only 11% agree with it. These results can easily be explained by the fact that the majority of respondents are not familiar with the use of corpora for language teaching and learning.

Apart from a lack of familiarity with corpus applications in the context of teaching, potential barriers in terms of insecurities with the use of corpora were investigated. Four statements that tested whether the participants would feel uncomfortable with the changes that corpus use may entail compared to 'traditional' teaching were presented to the participants. These statements are: "I am uncomfortable answering technical questions that might be raised during corpus activities"; "Unpredictable outcomes of corpus activities make me feel insecure"; "I am uncomfortable answering linguistic questions that might be raised during corpus activities" and "The absence of an answer key to corpus activities bothers me". The answers to these four questions were collected and the arithmetic mean of the mode of each response option was calculated (cf. Figure 5). In order to guarantee internal consistency reliability of the four items, the Cronbach Alpha coefficient was calculated, which gave the result of α = 0.74. According to Dörnyei (200394-95), the coefficient should reach or exceed 0.70 in order to ensure internal consistency of the items.
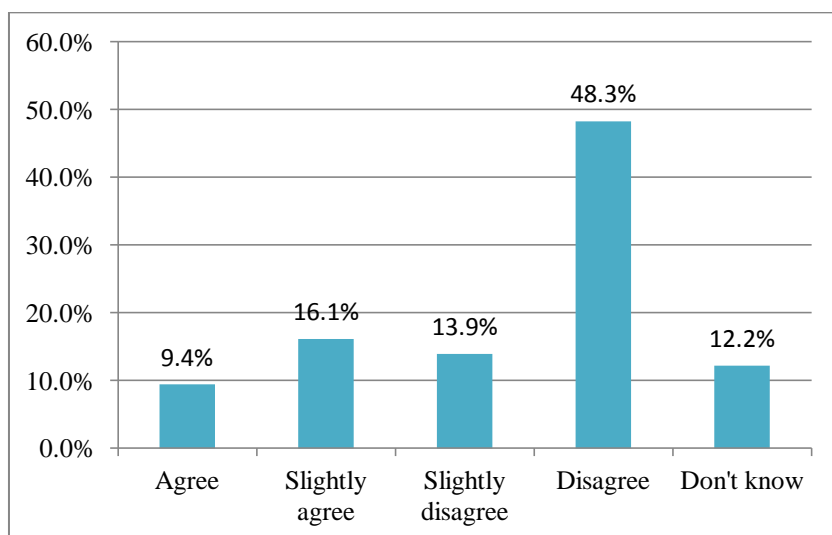

Figure 5: Respondents' agreement/disagreement with the questions of having technical, linguistic or pedagogical insecurities when using corpora

As Figure 5 demonstrates, about 48% of the participants disagree with the statements that unpredictable questions or incidents during the use of corpora would make them feel insecure or uncomfortable. In contrast, an average of about 9% of the respondents agrees that problematic questions raised in class would be uncomfortable to them. In particular, technical

questions seem to be challenging: here, about 13% of the respondents agree to feel insecure with these kinds of questions. In general, however, the majority of the participants seem to barely have a problem with situations in which unpredictable questions concerning technical, linguistic or pedagogical issues are raised. By implication this also means that the fact that the use of corpora in the EFL classroom could question the teachers' role as source of knowledge (cf. discussion in section 3.2.2.) does not seem to play a crucial role. In other words, teachers do not seem to be intimidated by unpredictable outcomes, questions or incidents that may occur and even challenge the teacher's competencies when using corpora with learners.

When it comes to the learners' competencies, however, the respondents seem to be unsure whether students can master the use of corpora. Although only 7% of the participants agree with the statement "Teaching and learning with corpora is too demanding for pupils", a relatively balanced result of the remaining response options (24% slightly agree, 29% slightly disagree, 24% disagree and 16% do not know) suggests that there is no clear agreement whether corpora are too demanding for learners. In particular, those 16% of the respondents who do not know whether students can manage the use of corpora indicate that it can be challenging for teachers to assess the requirements and skills that learners need in order to successfully use corpora. The results can also be due to the differentiation between learners from upper and lower levels who have different proficiency levels and skills and are more or less able to manage the use of corpora. The discussion of the differentiation between learners of upper and lower levels will be continued in section 4.5.

In addition to the learners' competencies, also the teachers' attitudes towards the relevance of corpus activities to language teaching and learning were examined. The statement "Corpus activities are relevant enough to be taught at school" was presented to the respondents; their agreement and disagreement is visualized in Figure 6.
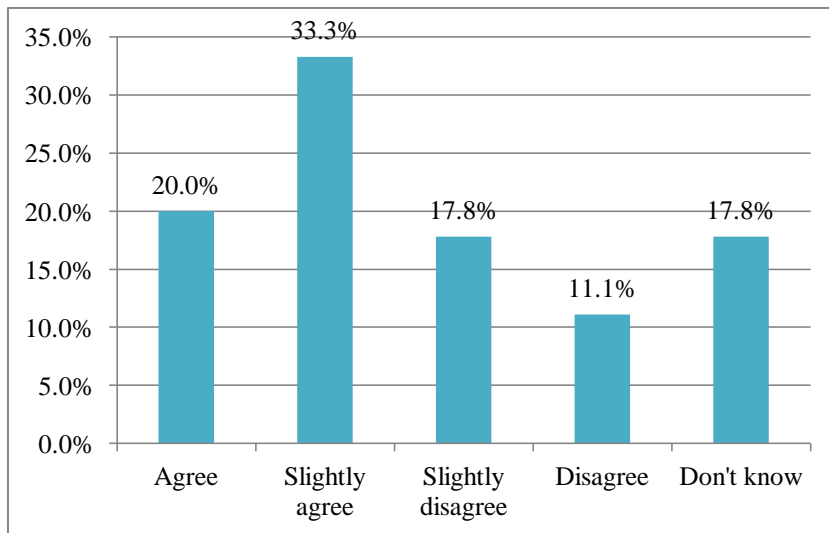
Figure 6: Respondents' agreement/disagreement with the question 'Corpus activities are relevant enough to be taught at school'

The answers as illustrated in Figure 6 show that 20% of the respondents agree that corpus activities are relevant for language teaching while the majority of about 33% slightly agrees. On the other hand, also a substantial number of participants do not feel that corpus activities are relevant enough for learners to be engaged with during EFL lessons. Moreover, more than 17% do not know whether to agree or disagree with the statement which can be due to the fact that many respondents are unfamiliar with the potential content and implementation of corpus-based activities. In general, these results indicate that the majority (in sum 53.3%) considers corpus activities relevant (or somewhat relevant) for language teaching and learning. However, since almost 29% of the respondents are of the opinion that corpus activities are irrelevant to some extent, a perceived lack of relevance of corpus activities can be regarded as a small but existing barrier of corpus use in the EFL classroom.

With regard to the effect that corpus activities can have on the roles of the teacher and learners as discussed in section 3.2.2., the following statements were shown to the participants in order to analyze the respondents' attitudes towards a possible change of teacher and learner roles: "I like the idea that pupils become researchers themselves when working on corpus activities" and "When pupils become researchers themselves, my role as a teacher becomes less important". Figure 7 and Figure 8 present the results that were obtained.
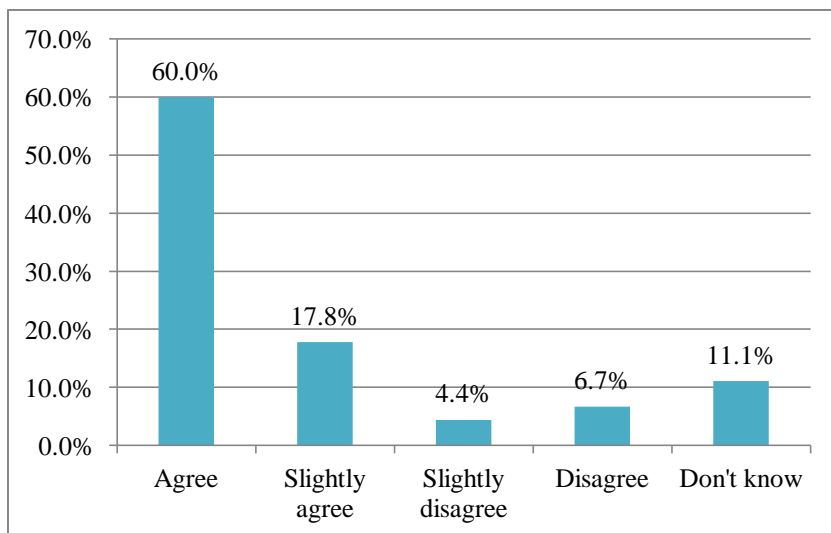
Figure 7: Respondents' agreement/disagreement with the statement 'I like the idea that pupils become researchers themselves when working on corpus activities'
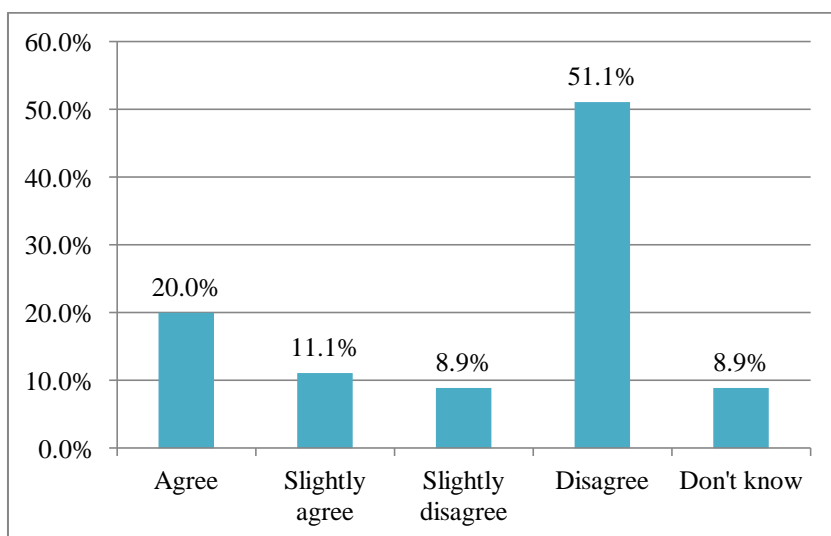

Figure 8: Respondents' agreement/disagreement with the statement 'When pupils become researchers themselves, my role as a teacher becomes less important'

As Figure 7 demonstrates, 60% of the participating teachers like the idea that learners take on the role of researchers when they are engaged in corpus activities. In addition to that, about 18% slightly agree with the statement while in sum roughly 11% do not like the idea that the learners' role changes during corpus-based activities. This clearly indicates that a change of the learner's role also in terms of gaining greater learner autonomy is perceived as a positive side effect. Apart from a change of the learner's role accompanying corpus-based activities, a change of the teacher's role is possible. Therefore, it was tested whether the respondents believe that their role in the EFL classroom will lose importance when corpus activities are implemented. Figure 8 illustrates the answers and demonstrates that the vast majority of about

51% of the respondents disagree with the statement that their role as a teacher becomes less important when learners are engaged in corpus activities and take on the role as researchers. On the other hand, 20% do agree and 11% slightly agree. However, it remains unclear whether a less crucial role of the teacher is perceived as a challenge for teachers or rather a chance for learners to take on the central part of learning. In any case, the results generally suggest that the majority of respondents do not believe that corpus activities have a negative effect on the importance of the teacher's role.

In order to investigate the issues of access to computers and availability of corpora as well as corpus materials, four questions addressing these issues were asked. Concerning the first questions, "I have limited access to computers at school", the majority of 40% slightly agrees and roughly 18% agree to have limited access to computers while about 29% of the participants have full access to computers at school. Generally, it can be said that occasionally teachers still face the problem of limited access to computers at school.

Regarding the availability of corpora, the statement "I have found suitable corpora for teaching" was presented to the participants. The majority of the respondents, that is 36%, state that they 'don't know' whether they have found suitable corpora which can be explained by the fact that 71% of the respondents have not been looking for suitable corpora for teaching, as already mentioned above. Therefore, it comes as no surprise that besides these 36% there are another 33% disagreeing with the statement that they have found suitable corpora. About 13% express slight disagreement, 7% slight agreement and only 11% of the participants have been successful in finding suitable corpora for language teaching and learning.

A relatively frequent choice of the response option 'Don't know' can also be found when looking at the availability of teaching materials and relevant topics for corpus activities. The results of the respondents' agreement/disagreement with the statement "There is limited availability of ready-to-use corpus teaching material" are presented in Figure 9. Most respondents (42.2%) state that they 'don't know' whether there is enough teaching material available. Again, this may be due to the fact that most respondents have not been looking for suitable corpora. Therefore, it can be inferred that most respondents have not been looking for appropriate teaching materials. On the other hand, 20% agree and about 22% slightly agree that ready-to-use teaching materials are only available to a limited extent.
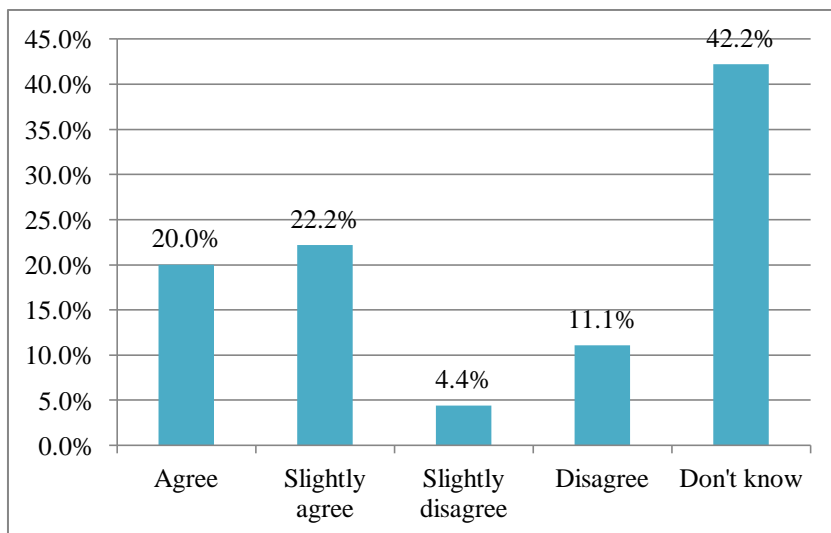
Figure 9: Respondents' agreement/disagreement with the statement 'There is limited availability of ready-to-use corpus teaching material'
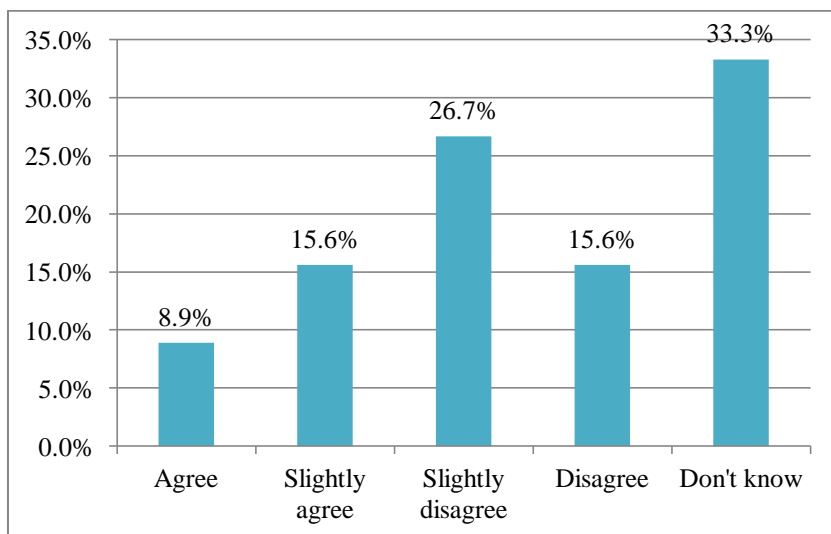

Figure 10: Respondents' agreement/disagreement with the statement 'It is easy to find appropriate topics for corpus activities'

Besides the availability of teaching materials, also the perceived availability of appropriate topics that are suitable for corpus activities was examined. For this reason, the participants were asked to respond to the statement "It is easy to find appropriate topics for corpus activities". Figure 10 summarizes the results and highlights that the most frequently chosen response option is 'Don't know' (33.3%). The reasons for this are likely to be the same as for the previous question concerning the availability of teaching material. In sum, the respondents tend to believe that it is difficult to find and identify appropriate topics that can be taught and learnt with the help of corpus activities.

Another barrier of corpus use in the EFL classroom under consideration was time. In order to test whether the factor time is an issue when teachers decide for the appropriate teaching activities such as corpus-based exercises in their lessons, the respondents were asked to agree or disagree with the following three statements: "Using corpus activities in my lessons is too-time consuming", "When using corpora, too much time is spent on too specific language issues" and "Creating appropriate corpus teaching material takes too much time". Despite the difference in wording, the three questions aimed at investigating the extent to which the respondents feel that corpus activities and their preparation require too much time. For this reason, the arithmetic mean of the mode of each response option was calculated and is presented in Figure 11. The calculations of the Cronbach Alpha coefficient gave the result of $\alpha = 0.69$.
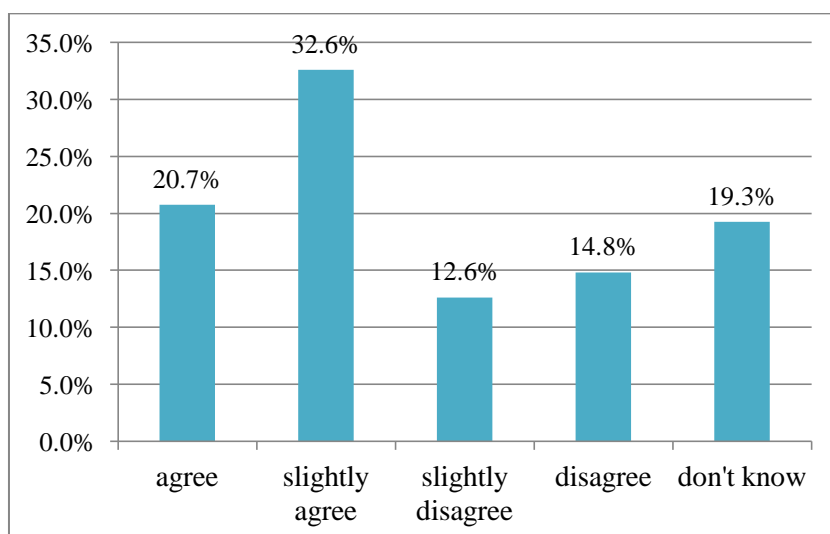

Figure 11: Respondents' agreement/disagreement with the question whether corpus activities are too time-consuming

The results as visualized in Figure 11 suggest that the majority (20.7% agree and 32.6% slightly agree) is of the opinion that too much time is spent on the design of corpus materials or on corpus activities when implemented in EFL lessons. On the other hand, more than 12% slightly disagree and roughly 15% disagree that using corpus activities is too time-consuming and that the creation of appropriate materials occupies too much of the teacher's time. Again, nearly 20% of the respondents state to have no knowledge on this matter. Nevertheless, the findings indicate that the factor 'time' is a crucial barrier that has to be considered before and during the application of corpora in language teaching.

The issues concerning a lack of time for the preparation or the implementation of corpus-based activities was analyzed even more closely in the following section of the questionnaire, which specifically focused on a selection of barriers. The question "In your opinion, what are the greatest obstacles of corpora use for language teaching?" was presented to the respondents, who were then asked to rate whether a 'lack of time in class' and a 'lack of time for preparation' are considered minor or great obstacles. The answers to these questions as presented in Figure 12 show that the respondents generally make a distinction between the expenditure of time on lesson/material preparation and on the implementation of corpus-based activities in class. With regard to the time spent on the preparation of lessons or teaching material, it can be seen in Figure 12 that a lack of time for preparation purposes is regarded a great obstacle (37.8%) of corpus use. On the other hand, roughly 18% believe that a lack of time for preparation is only a minor barrier of corpus use by teachers. In sum, it can be argued that a lack of time for preparation is considered a substantial obstacle of corpus use by the majority of respondents. However, the distribution of answers to the question whether a lack of time in class constitutes a barrier of corpus use is rather balanced when compared to the previous question. The results of the extremes 'small obstacle' (15.6%) and 'great obstacle' (28.9%) are not highly divergent from the middle categories (24.4% and 20.0%). In addition, also the response option 'Don't know' yields a result of 11.1%. These data are an indication that the respondents are in disagreement whether there is a lack of time in class that can be regarded a barrier of corpus use.
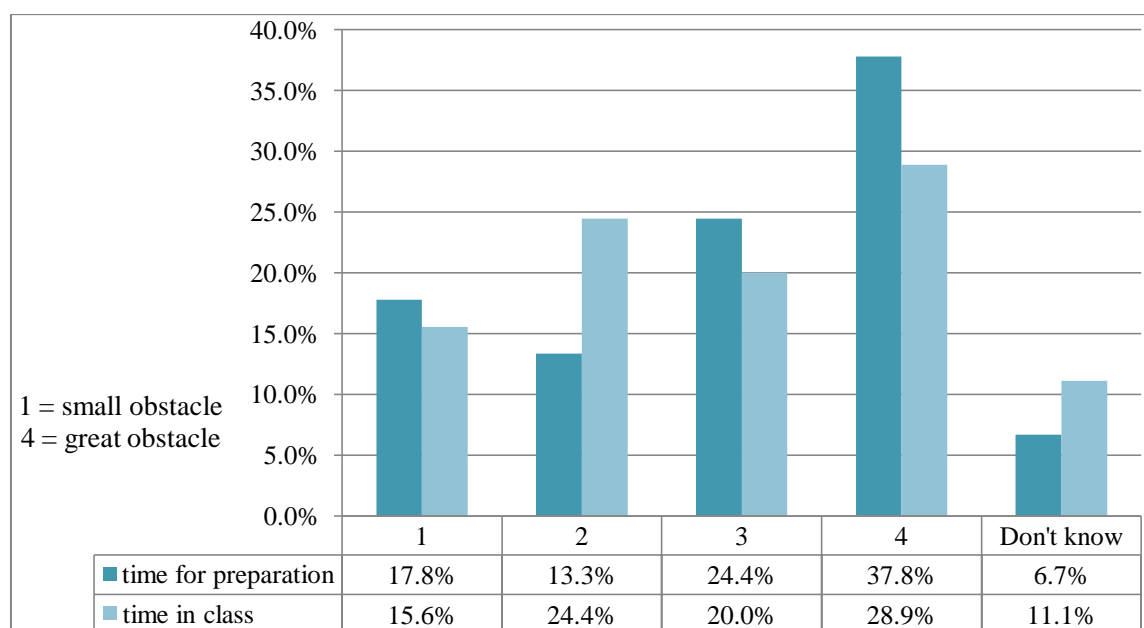

1 = small obstacle 4 = great obstacle	1	2	3	4	Don't know
■ time for preparation	17.8%	13.3%	24.4%	37.8%	6.7%
■ time in class	15.6%	24.4%	20.0%	28.9%	11.1%

Figure 12: Respondents' evaluation of the barrier time

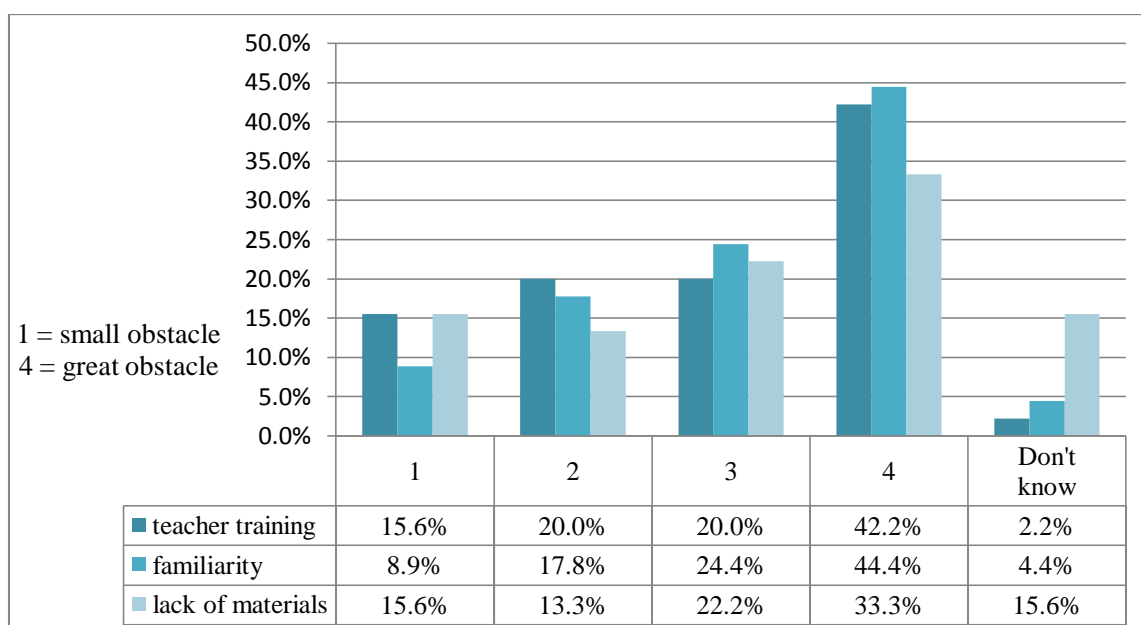|	1	2	3	4	Don't know
■ teacher training	15.6%	20.0%	20.0%	42.2%	2.2%
■ familiarity	8.9%	17.8%	24.4%	44.4%	4.4%
■ lack of materials	15.6%	13.3%	22.2%	33.3%	15.6%

Figure 13: Respondents' evaluation of the barriers 'lack of teacher training', 'lack of familiarity' and 'lack of materials'

Apart from the issue of time, three other potential barriers were taken into account. The participants were also asked to rate whether they perceived a 'lack of teacher training', 'a lack of teaching material' and 'a lack of awareness about the potential of corpora for language teaching' as minor or great obstacle. The findings to these ratings are presented in Figure 13, which illustrates the tendency of perceiving the above mentioned issues as great barriers of corpus use in the EFL classroom. Specifically, a look at the results of the 'lack of teacher training' reveals that more than 42% of the respondents believe that the absence of appropriate teacher training focusing on the implementation of corpora in language teaching constitutes a great barrier. In contrast, only 15.6% disagree with this view. Even more distinct results were obtained with regard to the lack of familiarity and awareness about the existence and the potentials of corpus use for language teaching. In this case, even 44.4% of the participants believe that a lack of familiarity with corpus use for language teaching prevents teachers from using corpora. About 9% regard this as a minor barrier, however. Similar findings can be noticed with the last barrier, i.e. the lack of teaching materials, even if the distribution of answers is not as extreme as with the previous question. While about 33% find that a lack of teaching materials is a great obstacle that prevents teachers from using corpora in language teaching, roughly 16% do not share this view. Compared to the two previous questions, the response option 'Don't know' obtains a rather high result of almost 16%. In sum, the findings to these three results indicate that a lack of teacher training, teaching

70

materials and a lack of familiarity of teachers with the use of corpora can be considered substantial barriers of corpus use.

In sum, these findings as visualized in Figure 12 and Figure 13 confirm that there are certain obstacles of corpus applications in the EFL classroom that can be challenging for teachers. Whether these barriers need to be improved in the future in order to convince more teachers to use corpora more frequently will be discussed in the following section.

4.4.3. Improvements

As already mentioned, the last part of the questionnaire was dedicated to the necessary improvements and developments of corpora and the respondents' possible future behavior concerning the use of corpora in or outside the classroom.

First of all, three statements regarding technical issues of corpora and computers were presented to the participants who were asked to agree or disagree. The results of each of the three questions deviate greatly from each other as can be seen in Figure 14.
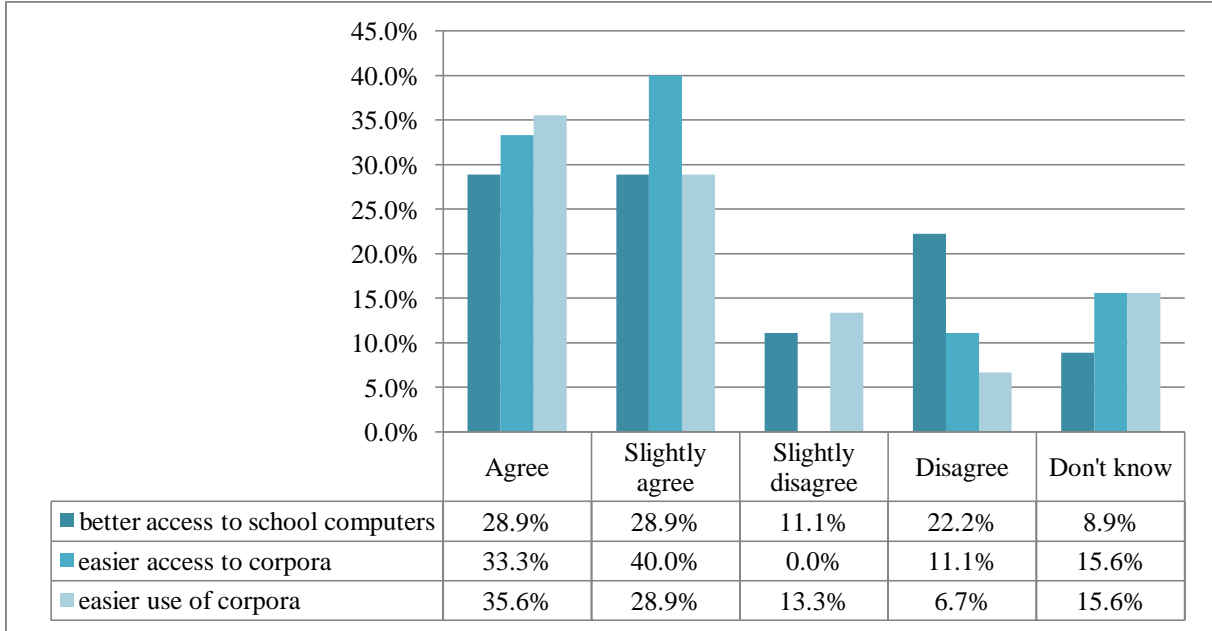

	Agree	Slightly agree	Slightly disagree	Disagree	Don't know
■ better access to school computers	28.9%	28.9%	11.1%	22.2%	8.9%
■ easier access to corpora	33.3%	40.0%	0.0%	11.1%	15.6%
■ easier use of corpora	35.6%	28.9%	13.3%	6.7%	15.6%

Figure 14: Evaluation of technical improvements and respondents' future behavior in terms of corpus use

About 22% disagree with the statement, "I would use corpora in the future if there was better access to computers at school", which reflects the results from the question concerning limited access to computers at school (see section 4.4.2.). As already mentioned, roughly 40%

of the respondents do have access to computers at school which serves as one explanation for the great disagreement with the former statement. However, the majority of respondents agree (28.9%) or slightly agree (28.9%) that they would use corpora in the future if there was better access to computers at school. In the case of the question, "I would use corpora in the future if free online corpora were easier to access", wide agreement (33.3% agree and 40.0% slightly agree) can be noticed. However, there are also nearly 16% who chose the option 'Don't know'. This can be due to the fact that many of these respondents have no knowledge about how to access corpora, or whether there are corpora that are relatively easy to access online and/or for free. Regarding the third statement, "I would use corpora in the future if free online corpora were easier to use", Figure 14 shows that the overall agreement (64.5%) is slightly smaller when compared with the previous question. Nevertheless, the respondents show much enthusiasm about their future implementation of corpora. Overall, it can be said that with the exception of those who have unproblematic access to computers and those who are undecided, wide agreement with the statements could be identified. By implication this means that the majority of respondents consider using corpora if improvements in terms of technical issues are made in the future.

Once again, it was investigated to what extent time restrictions and a lack of ready-to-use teaching materials would play a role for teachers when they consider using corpus-based activities in their lessons. The answers to the question, "I would use corpora in the future if there was more time in class" can be found in Figure 15.

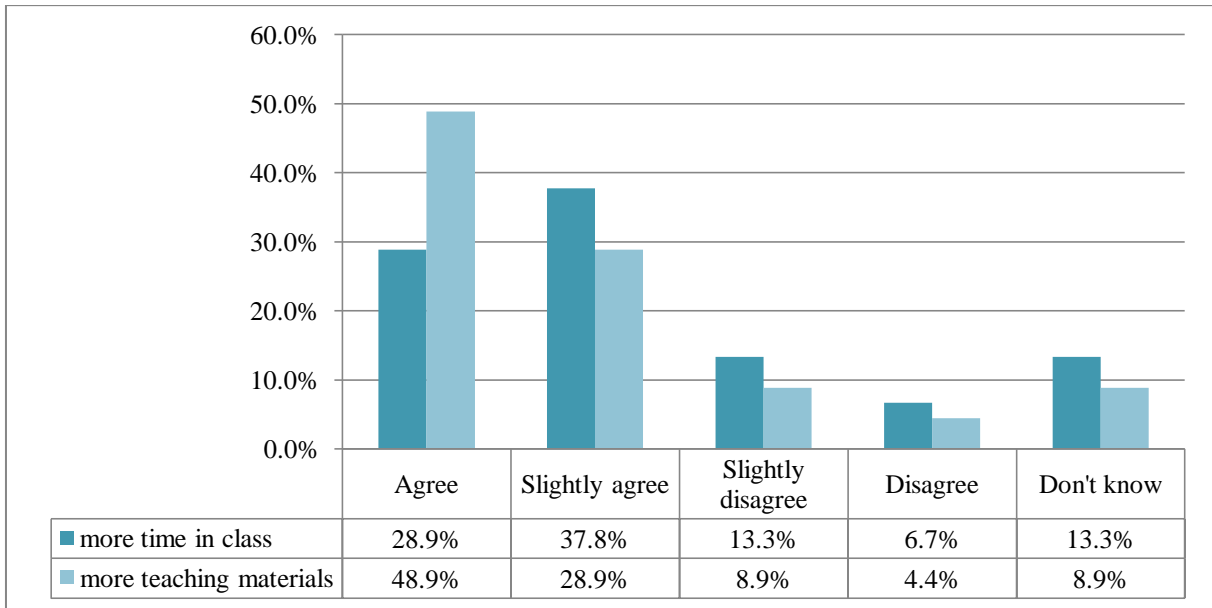|	Agree	Slightly agree	Slightly disagree	Disagree	Don't know
more time in class	28.9%	37.8%	13.3%	6.7%	13.3%
more teaching materials	48.9%	28.9%	8.9%	4.4%	8.9%

Figure 15: Evaluation of time and material improvements and respondents' future behavior in terms of corpus use

As apparent from this figure, the majority of respondents (in sum 66.7%) consider using corpora in the future if more time could be invested in corpus activities. Despite these large numbers, also roughly 13% are undecided whether time is an issue. A clearer result was obtained when the respondents were asked whether the existence of ready-to-use materials mattered (cf. Figure 15). Almost 49% of the participants agree and nearly 29% slightly agree with the statement, "I would use corpora in the future if there was more ready-to-use teaching material". These high results indicate that there is a need for relevant materials, which teachers could use in their EFL lessons and which facilitate the implementation of corpus-based activities.

As a last aspect under consideration, the improvements and developments concerning the teachers' competencies and skills in using corpora were analyzed. Figure 16 presents the answers to the questions, "I would use corpora in the future if there was more teacher training on how to use corpora" and "I would use corpora in the future if I had a clearer idea of how they can be used for teaching".

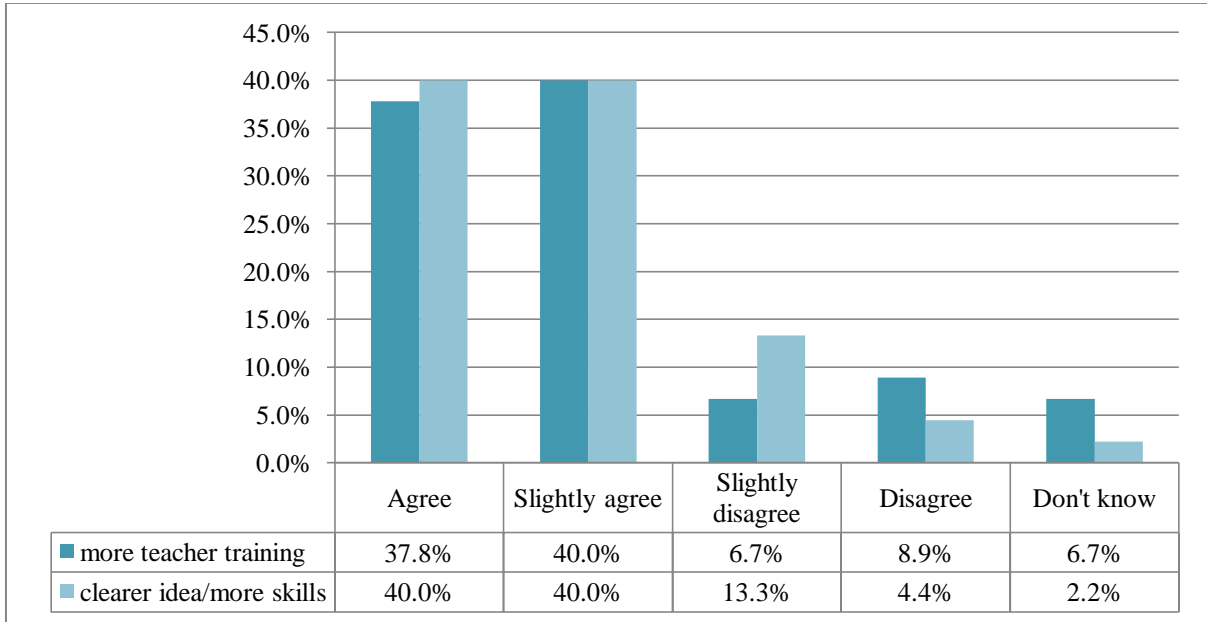|	Agree	Slightly agree	Slightly disagree	Disagree	Don't know
more teacher training	37.8%	40.0%	6.7%	8.9%	6.7%
clearer idea/more skills	40.0%	40.0%	13.3%	4.4%	2.2%

Figure 16: Evaluation of improvements (teacher training/ skills) and respondents' future behavior in terms of corpus use

Very similar results were obtained for both questions: Regarding the need for appropriate teacher training, nearly 78% of the respondents are in agreement that more/better training in how to appropriately use corpora in the classroom would increase the chance of actually making use of corpora in EFL lessons. Only a minority of respondents are not sure whether teacher training would initiate them to use corpora in the future. The answers to the second

questions show quite similar results. In sum even 80% of the participants agree or slightly agree that they would need to have a clearer idea of how corpora could be used before actually applying corpus-based activities in their EFL lessons. In conclusion, it can be inferred from these results that it is crucial for teachers to have appropriate competencies and skills in how to meaningfully use corpora prior to actually using them in the classroom. These aspects and in particular the importance of teacher training will be further discussed in the following section.

4.5. DISCUSSION

As the results from the previous sections have shown, the majority of respondents have some knowledge about corpus linguistics; especially younger participants from the age groups 20-29 and 30-39 as well as those participants who have studied at the University of Vienna and the University of Innsbruck. The chi-square test results provide evidence that there is a (moderate) correlation between the age of the participants and their familiarity with corpus linguistics as well as between the educational institution and the participants' familiarity with corpora. Moreover, it could be shown that most respondents who are somehow familiar with corpus linguistics have acquired their knowledge during their studies at university. In addition, popular and large corpora such as the BNC and the COCA are known by a large number of participants which is an indication that many teachers know about the existence of corpora and the field of corpus linguistics. Surprisingly, a substantial number of respondents are also familiar with VOICE, which is compiled at the Department of English at the University of Vienna. This can be of course due to the high number of participants who had studied at the University of Vienna. However, this result also shows that the Department of English at the University of Vienna has succeeded in promoting corpus linguistics and/or VOICE among its students.

With regard to the familiarity of teachers with corpus linguistics, the findings as presented in this thesis seem to deviate from previous findings of researchers in Germany. As already mentioned in section 3.3., Mukherjee (2004) detected a much smaller familiarity of English teachers from North Rhine-Westphalia with corpus linguistics. His findings have shown that only 10.9% are familiar with corpus linguistics, 9.7% have heard of corpus linguistics while 79.4% 'don't know anything about corpus linguistics' (Mukherjee 2004: 241). The findings from the study at hand, however, correspond with the results obtained by Breyer (2011) (15.6% 'very familiar', 58.0% 'vaguely familiar', 26.4% 'not at all familiar')

who interviewed teacher educators at tertiary institutions in Germany. What this could be suggesting is that the proximity of time (and place) of teachers to tertiary institutions is determinative for their familiarity with corpus linguistics. This can be inferred from the correlations as evidenced in section 4.4.1. and Granath's (2009: 63) claim that corpus use is most widespread at universities. The findings from the study at hand also suggest that the degree of familiarity is partly also depending on the educational institution of the respondents. Although one has to be cautious when generalizing data that contains only a small sample, like in the case of the study at hand, there is still reason to believe that students from the University of Vienna and the University of Innsbruck were more likely to acquire knowledge about corpus linguistics than students from the University of Salzburg.

When turning to the familiarity of respondents with corpora in connection with language teaching, the findings clearly show that the participants are generally less familiar with the use of corpora in the teaching context than with corpus linguistics. Moreover, the results also indicate that a shift from 'familiarity' to 'unfamiliarity' between corpus linguistics and corpora for language teaching is only noticeable within the age groups 20-29 and 30-39. In the remaining age groups there is the tendency that those who are familiar/not familiar with corpus linguistics are also familiar/not familiar with corpora for language teaching. This tendency and the fact that the majority of respondents being familiar with corpus linguistics have acquired their knowledge during their university studies imply that (some) Austrian universities do include corpus linguistics in their curriculum while at the same time neglect to focus on the potential of corpora for language teaching. One comment by a respondent confirms this assumption:

> I have come across corpus linguistics during my studies but have never learned how to apply my knowledge to teaching! (*30-39; Vaguely familiar; Not at all familiar*)[2]

Generally, it is very likely that students have encountered corpus linguistics during their studies; however, it seems that teaching methodology classes fail to inform students about the application of corpora in language teaching in order to familiarize future language teachers with the potential as well as the challenges.

It is not surprising then that only a minority of 11% of the respondents have used corpora in connection with teaching or for designing teaching materials. On the other hand, a greater percentage has used corpora for their own research projects or as a reference tool, which again emphasizes the discrepancy between the familiarity with corpus linguistics and

---

[2] The information in brackets includes information about the respondents: the age groups, familiarity with corpus linguistics and familiarity with corpora for language teaching.

the familiarity with corpora as a language teaching tool. It can be assumed that many of the participants have used corpora during their university studies, for instance in linguistics courses, for term papers or for quickly checking collocations instead of looking them up in a dictionary. However, the actual use as a teaching tool has been neglected by most participating teachers – partly due to a lack of awareness about the potential of corpora in language teaching or due to a lack of knowledge about how to meaningfully apply corpora. A closer look at these assumptions below will confirm this.

Prior to the discussion about the barriers, however, the general attitudes towards corpus use have to be mentioned briefly. Several comments from the open-response questions reveal that the participants generally have a positive attitude towards corpus use. These comments encompass statements such as the following:

Great tool in cases of doubt ...! (*50-59; Very familiar; Very familiar*)

[S]eems to be helpful when correcting texts (*50-59; Not at all familiar; Not at all familiar)*

Seems interesting and a reliable source (*>60; Not at all familiar; Not at all familiar*)

Although many respondents show a positive attitude and even gratitude for the opportunity of having gained insights into this 'new' teaching method, there are also comments expressing reservations about the usefulness of implementing corpora in the EFL classroom. These comprise issues such as a lack of time, constraints in using different teaching materials or access to well-functioning computers. While these and more barriers will be discussed in the following paragraphs, it has to be emphasized again that in general, the respondents showed much enthusiasm and a positive attitude towards the use of corpora for language teaching.

Having addressed the familiarity with corpora and the respondents' attitudes towards corpora, it is now time to discuss the barriers that prevent teachers from using corpora in their EFL classes. First of all, the competencies of teachers concerning the use of corpora were investigated. As the findings have shown, about a third of the respondents have had training in how to use corpora, which can be again explained by the fact that a substantial number of respondents have encountered the use of corpora at their universities where they have most likely had the opportunity to learn how to use corpora. On the other hand, a slightly higher percentage has never had training in corpus linguistics. Considering the respondents' competencies concerning corpora in the teaching context, however, it becomes obvious that the majority of the participating teachers lack the appropriate skills for preparing relevant corpus activities. Once more, this result reflects the previous findings that there is a gap between the respondents' familiarity and competence with corpus linguistics and their

familiarity and competence with the use of corpora for language teaching. This may be again explained by the universities' neglect to include particular teaching methodology courses on corpus applications for language teaching in their teacher training programs.

In connection with the respondents' competencies, it was tested whether the participating teachers could feel insecure with unpredictable incidents during corpus activities and whether these insecurities could be the reason why the teachers are unwilling to make use of corpora in EFL classes. The results to this question are unambiguous. A vast majority of respondents disagree to feel insecure with certain (technical, linguistic or pedagogical) issues that could possibly arise during corpus activities. As reported in the literature (and already discussed in section 3.2.2.), there seems to be the assumption that the teacher has to be the source of knowledge while the change of this perceived role of the teacher "requires a lot of confidence" (cf. Breyer 2011: 108). Johns (2007 [1991]: 3) even argues that "many old certainties [such as the answer key at the back of the textbook] start to crumble" with the implementation of corpus activities in the EFL classroom. For example, corpus activities can present results that are unknown and unexpected even by the teacher. Breyer's study (2009: 166), for instance, revealed that students participating in a teacher training program felt the "fear of losing control of the teaching process" and a lack of control due to unpredictable results of corpus activities. However, the findings from the study at hand demonstrate that the majority of respondents do not consider this a great problem. To be more precisely, the majority would not feel insecure if technical or linguistic questions were raised or if the respondents had to deal with unpredictable outcomes. Examining the reasons for these results would go beyond this thesis; however, it can be said that unexpected outcomes or questions accompanying corpus activities can be regarded a minor issue or challenge to the participating teachers.

Another potential issue was a change in the teacher's and learners' roles. While the majority of participants believe that the teacher's role does not become less important with corpus-based activities, there are also a substantial number of respondents who do think that the teacher's role becomes less important. However, there is room for interpretation whether these respondents consider this 'less important role' as a positive or negative side effect. For instance, a less important teacher role could imply that the learner's role becomes more crucial in the learning process. Regarding the learners' roles, the results are unambiguous. Most participating teachers (60%) like the idea that the learners take on the role of researchers when doing corpus activities. As discussed in section 3.2.3., this can also imply that students

become more autonomous learners – depending on the activity and the teacher's extent of guidance and mediation. There was one comment dealing with this issue:

> I am curious how future language skills are acquired and I would appreciate it if we teachers succeeded in empowering our students to become more autonomous learners (*>60; Not at all familiar; Not at all familiar*)

As apparent from this quote and the results above, there is no doubt that many teachers want their students to be autonomous learners. However, whether corpus-based activities are the appropriate tool to work on and achieve learner autonomy is unclear to many respondents. To many, corpus activities seem to be too demanding for learners. To be more precisely, there is no clear tendency of whether the respondents believe that learners can handle corpus activities and can realize their full potential. On the one hand, these results could stem from fact that many respondents have problems with evaluating what skills learners need before they can make the best of corpus activities. On the other hand, it may be difficult to generalize whether corpora are too demanding for all learners at all proficiency levels. In particular, the question whether one has to differentiate between learners at upper and lower levels has come up several times with the open-response questions. An immediate response to the corpus exercises presented in the questionnaire states the following:

> You get a lot of context which is great; advanced learners might benefit and find it useful; I see confused faces of beginner or intermediate students though (*40-49; Not at all familiar, Not at all familiar)*

Whether learners at beginning levels can also profit from corpus activities is undoubtedly a legitimate question. Granath (2009: 63), for instance, claims that

> advanced students definitely benefit from working with corpora. […] Intermediate students can also profit from hands-on exercises using the computer. For less advanced students, more teacher guidance is needed, but if the exercises are on the right level, corpus work can help raise their awareness of structures.

Granath is in line with Römer (2011: 214) who argues that beginners are likely to be overwhelmed with too difficult and complex texts or large amounts of native-speaker texts. However, Römer (2011: 214) also claims that simple concordance exercises created by the teacher in order to fit the learners' needs can be very useful and meaningful.

In terms of the corpus activities' relevance for language teaching and learning, it can be noticed that most respondents regard the use of corpus activities in the EFL classroom relevant, at least to some extent. However, a considerable percentage claims to be unable to

assess its relevance. The most obvious explanation for this result is – again – a lack of familiarity with the use, potential and challenges of corpus activities. A somehow greater and more immediate issue is the access to computers and availability of corpora. While the majority of participants slightly agree to have limited access to computers, one comment by a respondent illustrates a related problem:

> In the next section you assume that pupils can work on corpora themselves. I would not have thought about that, because the computers hardly ever work without problems, so I would select the interesting texts/sentences, and only then work with them. (*30-39; Vaguely familiar; Vaguely familiar*)

Apparently, teachers and learners not only face the problem of limited access to computers at school, but also the problem that these may not always function properly. In addition to the access of computers, the question about the availability of suitable corpora for language teaching reveals that the overall majority of respondents have never been looking for corpora that could be suitable for language teaching. It comes as no real surprise then that only a fraction of respondents has been successful in deciding on a suitable corpus.

What the question about the limited availability of ready-to-use corpus teaching materials reveals is that the majority of respondents have no knowledge about whether enough materials are available. This clearly reflects previous findings, which have demonstrated that many participants have not dealt with the implementation of corpus-based activities in the classroom. Therefore, many respondents have neither been looking for corpus-based teaching materials. On the other hand, many respondents tend to think that there is limited availability of ready-to-use materials. This lack can thus be considered a great barrier and the results also show that the majority of respondents would think about using corpus-activities in the future if there were more ready-to-use materials available. This suggests that there is a need for more corpus-based materials, which is also confirmed by Römer's (2011) results. In her study, Römer (2011: 89) found that the teachers participating in her survey seemed to need and want "a wider range of better teaching materials with more interesting, longer and genuine texts, and with more exercises". As she suggests, a response to this demand could be the creation of corpus-based materials by teachers themselves (Römer 2011: 89). In this case, the materials can be designed in order to perfectly fit the learners' needs. However, language teachers would also need training in how to use corpora and how to create appropriate teaching materials. Lenko-Szymanska (2014: 272), for instance, shows that her respondents admitted needing more guidance in how to create appropriate teaching materials and exercises. On the other hand, Römer (2011: 92) also argues that "the creation of suitable materials should […]

not be left entirely to the teacher". Instead, teachers should also be able to rely on support from the research community who ideally promotes the development of suitable corpus-based teaching materials (Römer 2011: 92). Additionally, Breyer (2011: 153) demonstrates in her study that a lack of ready-to-use materials contributes to a lack of corpus application in the classroom. The call for more support in the creation of teaching materials is also reflected in the following comments by respondents from the survey at hand:

> I became a great fan of corpus use in the early Nineties, when I saw its great potential for translating specialized texts and dictionary development. I have been using corpora for writing teaching materials ever since but most of these are for my courses at tertiary level, where I can't rely on ready-made course books. I'd see the key step forward at this stage in motivating SB-authors [schoolbook authors] to incorporate corpus-informed materials into their English books for secondary students, esp. those at intermediate and upper intermediate levels. In the initial stage, it might make sense to offer these corpus-informed tasks/modules as additional online materials. (*>60; Very familiar, Very familiar*)

> [A] school-based introduction and a platform where teachers could SHARE their material (at OUR school) may be ONE good idea when integrating this tool into every day classrooms (*40-49; Vaguely familiar; Vaguely familiar*)

What the first comment suggests is that corpus-based materials should be integrated into schoolbooks more frequently. It also indicates that additional material is often created in case textbooks do not offer sufficient materials and exercises. On the other hand, the second comment also shows that collaboration in terms of sharing materials with colleagues may facilitate the implementation of corpus-based activities. Generally, it can be said that a lack of ready-to-use teaching materials is a barrier that prevents the implementation of corpus use in language pedagogy.

Besides the issue of materials, also a lack of time can be regarded a deterrent factor for the implementation of corpus activities. The findings indicate that too much time would have to be spent on the preparation of corpus activities, of materials or on introducing learners to corpus activities. In particular, a lack of time for preparing corpus-based activities and/or materials is considered a great obstacle. It is also worth noting that the majority of respondents would, at least to some degree, consider using corpora if there was more time for the implementation of corpora in class. Similar results confirming this problem have been obtained by other researchers as well. For instance, Breyer (2011: 152) found that one of the main reasons why teacher trainers decided not to teach with corpora was a lack of time. Also,

Farr (2008) reports that the amount of time needed for corpus activities is one of several negative aspects that were listed by her research participants. She claims that "[t]ime is often cited as a difficulty with the use of technology in education in general, and specifically in the use of corpora" (Farr 2008: 36-37). Since teachers have to be competent in using corpora before they can teach it to learners, much time and training is needed in order for teachers to familiarize themselves with corpus use. In addition, learners need to be introduced to corpus activities which, of course, requires much time as well. Moreover, attention has to be paid to the preparation of corpus activities and materials as well as to the choice of topics that are suitable for corpus use. As the results of the present study demonstrate, the respondents seem to have slight difficulties with deciding on appropriate topics for corpus use. Apart from them, however, the majority of respondents have not been dealing with this question due to a lack of familiarity with the use of corpora for language pedagogy. Lenko-Szymanska (2014) also shows that the participants of her survey claimed that they needed more support in the selection of appropriate topics which could be addressed with corpus activities. Furthermore, she argues that teachers need to be competent corpus users and need linguistic as well as pedagogical skills in order to be able to teach with corpora, and that "these skills take time to develop" (Lenko-Szymanska 2014: 272). Therefore, the factor 'time' has to be kept in mind when calling for greater implementation of corpus activities in the EFL classroom.

As a last point, the importance of the teachers' competencies and skills as well as of effective teacher training with regard to corpus use in the classroom has to be addressed here. The findings from the study at hand suggest that, first of all, a lack of teacher training about the use of corpora constitutes a great barrier for the application of corpora in the classroom. Secondly, the results also indicate that the majority of respondents would use corpora in the future if more teacher training was offered and if the respondents had more skills and knowledge about the use of corpora for language teaching and learning. Therefore, the need for a comprehensive teacher training in the use of corpora for language pedagogy arises, which has been emphasized by various researchers. In a case study conducted by Breyer (2009), student teachers were trained to teach with corpora and experienced corpus use from the perspectives of the learner as well as the teacher. Breyer (2009) argues that it is crucial for teachers to have a thorough understanding of corpus analysis prior to teaching with corpora. However, also the implementation of corpus use in language teaching as well as the creation of materials has to be trained. Therefore, effective training is essential for teachers in order to implement corpora in language teaching. Breyer (2009: 156) argues further that (student) teachers are more likely to apply corpora in language teaching when they have experienced

the use of corpora from the perspective of both teachers and learners, and when they have discovered the potential of corpus use for themselves. Similarly to Breyer, Lenko-Szymanska (2014) introduced students to corpora and their applications in language teaching and came to the pessimistic conclusion that even a weekly 90 minutes course lasting for an entire semester is not sufficient for equipping future teachers with the skills needed for effective implementation of corpora in language teaching. Lenko-Szymanska (2014: 272) claims:

> Until corpora enter mainstream education in language departments and teacher-training institutions on a large scale, which will allow future teachers to encounter various resources, tools and methods of corpus analysis in a variety of language and linguistics classes, we cannot expect that corpora will find their way into language classrooms in other instructional settings.

Teacher training is widely considered a crucial step in promoting corpus use among language practitioners and in encouraging teachers to apply corpora in the language classrooms. Römer (2009: 91-92), for instance, suggests to offer courses taught by corpus linguists that focus on the teachers' needs and Granath (2009: 64) proposes offering courses in corpus analysis, which additionally focus on the implementation of corpus tasks in the language classroom. In sum, effective teacher training is crucial for the promotion of corpora among language teachers and it should encompass the training of the teacher's own corpus skills as well as the training of how to implement corpus use in language teaching and learning.

Future Developments / Recommendations

The findings from the present study confirm the claim by researchers such as Mukherjee (2004) that a gap between advances in corpus research and corpus use in language pedagogy exists. As a consequence, the question arises how this gap can be closed. I would argue that a first step is to take a closer look at the challenges teachers face when using corpora for language teaching. As suggested above, attention has to be paid to issues such as a lack of familiarity, time, teaching materials and limited access to computers or corpus software. Again, the call for more ready-to-use teaching materials has to be emphasized in order for teachers to have more time and be able to focus on more important issues such as the "social problems of learning: encouragement, motivation, and fostering group activities and learning through collaboration" (Wilson 1997: 117). In addition, Mauranen (2004: 100) highlights the need for "user-friendly program packages, with simple concordancers and good initiation exercises". On the one hand, applied corpus linguists need to be aware of teachers' needs and

the challenges they face when applying corpora in their teaching. This implies that more research on these needs and challenges have to be conducted in the future. Furthermore, Breyer (2011: 151) suggests that "teaching methodology has to drive and inform the linguistic developments that are relevant to the processes of teaching and learning a language" instead of the other way around. The focus has to be on the teachers' and learners' needs instead of the linguists' wishes and ideas (Breyer 2011: 151). Mauranen (2004: 99) rightly remarks the following:

> No teaching method can become an important innovation, whatever its potential, if it does not make its way to the normal classroom where teachers and students can use it as part of their everyday routines, with not too much extra hassle.

This implies that existing challenges and barriers that prevent teachers from using corpora in their teaching have to be tackled and resolved with the help of future research.

Even more importantly, however, teachers have to be familiarized with the existence, the potential and challenges of the use of corpora in language pedagogy. In particular, this involves teacher trainers but also other colleagues from linguistic and language departments at tertiary institutions who have to 'spread the word' and use corpora in their own teaching (Lenko-Szymanska 2014: 273). Teacher training is an essential prerequisite secondary school teachers have to have gone through in order to effectively use corpora in their teaching. Appropriate training that enables teachers to competently use corpora as well as teach with them will have to focus on corpus analysis and on the implementation of corpus use for language teaching. The findings from the present study demonstrate that there is some knowledge about and familiarity with corpus linguistics while the application of corpora for teaching purposes has been neglected in teacher training so far. Therefore, teaching methodology classes will have to focus on practical issues of corpus implementation in the classroom and enable and encourage (future) teachers to competently and effectively use corpora in the EFL classroom.

Limitations of the Study

There are some limitations to this study, particularly due to its small sample size. As mentioned in section 4.2., the sample encompasses a majority of teachers from Vienna and participants of the age group 20 – 29 years. A more balanced sample could have resulted in different findings. Moreover, since many respondents are not at all familiar with corpora in the teaching practice and therefore have no knowledge about the challenges they would have

to face when applying corpora, there is the risk that the results to some of the questions concerning the barriers reflect the respondents' expectations rather than their actual experience. However, the response option 'Don't know' may have helped to balance and correct the results in order to obtain reliable results. Moreover, any generalizations refer only to the sample of this survey; generalizations of the findings to all Austrian secondary school teachers have to be made with great caution. Nevertheless, the findings from this study can be regarded as tendency of Austrian teachers to be familiar with corpus linguistics and corpora for language teaching. Moreover, it presents those barriers that teachers are likely to face when applying corpora in language teaching and learning. As already mentioned, more research about EFL teachers' needs and wishes concerning the use of corpora seems necessary, and tertiary institutions need to include the teaching of corpus skills and the practical implementation of corpora in the teaching practice in the future.

5. CONCLUSION

This thesis has investigated a number of theoretical issues concerning corpora in general as well as corpora in the context of language teaching and learning. Besides a brief overview of the history of corpus use, the definition of 'corpus' and the most common types of corpora, the potential and limitations of corpus use have been demonstrated. In particular, the possibilities of frequency counts, concordancing and keyword analyses as well as the possibility to annotate texts are the most apparent advantages of corpus technology. It has also been shown, however, that the issue of representativeness of corpus data has been widely debated. With regard to the discussion of corpora in language pedagogy, it has been shown that the applications of corpora in the teaching and learning context are manifold. Indirect corpus applications encompass the production of dictionaries and reference grammars that are based on corpus findings. Also syllabi and teaching materials including textbooks have been influenced by corpus research. Direct applications, on the other hand, comprise corpus use by teachers and learners. Corpora constitute a valuable tool especially for teachers who have the possibility to make use of corpora in as well as outside the classroom. One of the most prominent and meaningful applications is the use of corpora as a reference tool. Corpus consultation can be an alternative to traditional dictionaries or grammars and has the advantage of offering answers to complex questions that dictionaries or grammars fail to answer. Additionally, corpora can serve as a basis for the creation of teaching materials; also

do-it-yourself corpora based on learners' texts can be created. This allows teachers to detect common mistakes and facilitates the decision about what language items to focus on, which at the same time enhances teaching and learning. Furthermore, direct applications by learners encompass approaches such as data-driven learning (DDL). It has been shown that the possibilities of DDL are manifold and comprise reading, writing, speaking skills as well as the acquisition of lexis and grammar. However, also a number of advantage and challenges have been mentioned including the issue of authenticity, autonomy, the changing roles of teachers and learners as well as the issue of availability of corpora and teaching materials. Last, the gap between corpus research and the actual teaching practice has been addressed. In particular, it has been highlighted that only few studies have examined the extent to which language teachers use corpora for language teaching purposes. Moreover, the results from these studies contradict the common claim that the use of corpora in the language classroom has been widely established by now.

These common assumptions were also reason enough to conduct a survey as part of this thesis. The aim was to analyze the extent to which Austrian EFL teachers at secondary institutions are familiar with corpus linguistics as well as with the use of corpora for language pedagogy, and the extent to which they actually make use of corpora in the teaching context. Apart from this, the second aim was to examine those challenges and barriers that teachers face when applying corpora in the language classroom as well as discuss what improvements are possible and necessary in order to convince more teachers to use corpora in EFL lessons. In terms of methodology, questionnaires were sent to teachers at secondary institutions of Austria. These questionnaires addressed the respondents' familiarity with corpus linguistics, with corpora for language teaching, the use of corpora as well as perceived barriers and possible improvements. In addition, the respondents were invited to participate in two optional corpus exercises, which would give them the opportunity to get insights into possible uses of corpora.

With regard to the results and the first research question, it has been shown that the majority of respondents are, at least to some extent, familiar with corpus linguistics. It has been suggested that there is a correlation of moderate strength between familiarity and age as well as familiarity and educational institution even though these results have to be considered with caution. The results have also demonstrated that the familiarity with corpora for language teaching and learning is much smaller compared to corpus linguistics. This is also reflected in the usage patterns that indicate that corpora have been used for research projects or as reference tools rather than for designing teaching materials or teaching as such. To be

more precisely and to answer the second research question, approximately half of the respondents have used corpora for research purposes, about 40% as a reference tool, while only roughly 11% of the respondents have used corpora for the design of teaching materials and teaching, respectively. These findings have illustrated that a gap in familiarity between corpus linguistics and corpus application in language teaching exists and that it is not enough to be familiar with corpus linguistics in order to apply corpora in language teaching.

The respondents' attitudes towards corpus use were positive for the most part. Especially the comments in the open-response sections confirmed the respondent's interest in the field and in the acquisition of this 'new' teaching method. Nevertheless, some comments also called attention to everyday problems teachers have to face. The barriers that seem least problematic for the application of corpora in EFL lessons encompass the issue that teachers may have to deal with unpredictable situations or corpus results, and that the teacher's and learners' roles may change. Moreover, the respondents perceive corpus activities as relevant to a large extent while there is disagreement whether learners can manage corpus activities.

On the other hand, four major barriers that prevent or could prevent teachers from using corpora for language teaching were detected. These include the issue that many teachers have been unfamiliar with the use of corpora for language teaching. Then, a lack of ready-to-use teaching materials as well as a lack of time plays a great role. Last, most respondents agree that they lack the competencies and skills to meaningfully implement corpora in language teaching, which can be due to a neglect of focusing on corpus activities during teacher training. These challenges are at the same time those aspects that need to be improved in the future. Therefore, the findings to research question 5 demonstrate that improvements in terms of providing more materials and first and foremost offering more teacher training courses that help teachers to become familiar with corpora as well as their implementation in language teaching are urgently required.

At this part of the thesis it is worth mentioning again that teacher training is an essential prerequisite for teachers to become familiar with corpora and their application in language teaching. Corpora have much to offer for learners as well as for teachers. While the former depend on the latter to provide them with the opportunity to learn from and with the use of corpora, teachers have to be competent corpus users and know about the potential as well as the challenges that corpus use encompasses. Only when educational institutions include teacher training on the implementation of corpus use in EFL classes and future research focuses more on the teachers' needs and challenges, corpus use can become more widespread in Austrian EFL classrooms.

6. REFERENCES

Aijmer, Karin. 2008. "Parallel and comparable corpora". In Lüdeling, Anke; Kytö, Merja (eds.). *Corpus linguistics: an international handbook*. Berlin: Walter de Gruyter, 275-291.

Aijmer, Karin. 2009. "Introduction". In Aijmer, Karin (ed.). *Corpora and language teaching*. Amsterdam: John Benjamins, 1-12.

Anderson, Wendy; John Corbett. 2009. *Exploring English with online corpora: an introduction.* Hampshire: Palgrave Macmillan.

Aston, Guy. 1997. "Enriching the learning environment: corpora in ELT". In Wichmann, Anne; Fligelstone, Steven; McEnery, Tony; Knowles, Gerry. *Teaching and Language Corpora*. London: Longman, 51-64.

Aston, Guy. 2001. "Learning with corpora: an overview." In Aston, Guy (ed.). *Learning with corpora.* Bologna: CLUEB, 17-45.

Atkins, Sue; Clear, Jeremy; Ostler; Nicholas. 1992. "Corpus design criteria". *Literary and Linguistic Computing* 7(2), 1-16.

Barlow, Michael. 2004. "Software for corpus access and analysis." In Sinclair, John. M. (ed.). *How to Use Corpora in Language Teaching.* Amsterdam: Benjamins, 205-224.

Barlow, Michael; Burdine, Stephanie. 2006. Phrasal verbs and collocations: American English. Houston, TX: Athelstan.

Barnbrook, Geoff. 1996. *Language and computers: a practical introdcution to the computer analysis of language.* Edinburgh: Edinburgh University Press.

Benson, Phil. 2007. "Autonomy in language teaching and learning." *Language Teaching 40(1)*, 21-40.

Bernardini, Silvia. 2000. *Competence, capacity, corpora: a study in corpus-aided language learning.* Bologna: CLUEB.

Bernardini, Silvia. 2004. "Corpora in the classroom: an overview and some reflections on future developments". In Sinclair, John. M. (ed.). *How to use corpora in language teaching*. Amsterdam: Benjamins, 15-36.

Biber, Douglas; Conrad, Susan; Cortes, Viviana. 2004. "If you look at...: lexical bundles in university teaching and textbooks". *Applied Linguistics* 25(3), 371-405.

Bloch, Joel. 2009. "The design of an online concordancing program for teaching about reporting verbs". *Language Learning and Technology* 13(1), 59-78.

Boulton, Alex. 2010. "Learning outcomes from corpus consultation". In Moreno Jean, Maria; Serrano Valverde, Fernando; Calzada Perez, Maria (eds.). *Exploring New Paths in Language Pedagogy. Lexis and Corpus-based Language Teaching*. London: Equinox Publishing Ltd., 129-144.

Braun, Sabine. 2007. "Designing and expoiting small multimedia corpora for autonomous learning and teaching". In Hidalgo, Encarnacion; Quereda, Luis; Santana, Juan (eds.). *Corpora in the foreign language classroom: selected paper from the sixth International Conference on Teaching and Language Corpora, Granada, Spain, July 2004*. Amsterdam: Rodopi, 31-46.

Braun, Sabine. 2006. "ELISA: a pedagogically enriched corpus for language learning purposes". In Braun, Sabine; Kohn, Kurt; Mukherjee, Joybrato (eds.). *Corpus technology and language pedagogy: new resources, new tools, new methods.* Frankfurt/Main: Peter Lang, 25-48.

Braun, Sabine. 2005. "From pedagogically relevant corpora to authentic language learning contents". *ReCALL* 17(1), 47-64.

Breyer, Yvonne. 2009. "Learning and teaching with corpora: reflections by student teachers". *Computer Assisted Language Learning* 22(2), 153-172.

Breyer, Yvonne. 2006. "My concordancer: tailor-made software for language learners and teachers". In Braun, Sabine; Kohn, Kurt; Mukherjee, Joybrato (eds.). *Corpus technology and language pedagogy: new resources, new tools, new methods.* Frankfurt/Main: Peter Lang, 157-176.

Breyer, Yvonne. 2011. *Corpora in language teaching and learning: potential, evaluation, challenges.* Frankfurt am Main: Peter Lang.

Carter, Ronald; Hughes, Rebecca; McCarthy, Michael. 2000. *Exploring grammar in context: grammar reference and practice, upper-intermediate and advanced.* Cambridge: Cambridge University Press.

Chambers, Angela. 2007. "Integrating corpora in language learning and teaching". *ReCALL* 19(3), 249-251.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge: The Massachusetts Institute of Technology.

Chomsky, Noam. 1962. "Transformational approach to syntax". In Hill, A.A. (ed.). *Studies in American English: third Texas Conference on Problems of Linguistic Analysis in English (1958).* Austin, TX: University of Texas, 124-158.

Cook, Guy. 2003. *Applied Linguistics.* Oxford: Oxford University Press.

Dodd, Bill. 1997. "Exploiting a corpus of written German for advanced language learning". In Wichmann, Anne; Fligelstone, Steven; McEnery, Tony; Knowles, Gerry (eds.). *Teaching and language corpora.* London: Longman, 131-145.

Dörnyei, Zoltan. 2003. *Questionnaires in second language research: construction, administration, and processing.* Mahwah, NJ: Lawrence Erlbaum Associates.

Ebeling, Signe Oksefjell. 2009. "Oslo interactive English: corpus-driven exercises on the web". In Aijmer, Karin (ed.). *Corpora and language teaching.* Amsterdam: Benjamins, 67-82.

Eckey, Hans-Friedrich; Kosfeld, Reinhold; Türck, Matthias. 2008. *Deskriptive Statistik. Grundlagen, Methoden, Beispiele.* Wiesbaden: Gabler.

Farr, Fiona. 2008. "Evaluating the use of corpus-based instruction in a language teacher education context: perspectives from the users". *Language Awareness* 17(1), 25-43.

Fitschen, Arne; Gupta, Piklu Gupta. 2008. "Lemmatising and morphological tagging". In Lüdeling, Anke; Kytö, Merja (eds.). *Corpus linguistics: an international handbook.* Berlin: Walter de Gruyter, 552-563.

Francis, W. Nelson. 2007 [1982]. "Problems of assembling and computerizing large corpora". In Teubert, Wolfgang; Krishnamurthy, Ramesh (eds.). *Corpus linguistics I: critical concepts in linguistics.* Abingdon: Routledge, 285-298.

Gavioli, Laura. 1997. "Exploring texts through the concordancer: guiding the learner". In Wichmann, Anne; Fligelstone, Steven; McEnery, Tony; Knowles, Gerry (eds.). *Teaching and language corpora*. London: Longman, 83-99.

Gavioli, Laura. 2001. "The learner as researcher: introducing corpus concordancing in the classroom". In Aston, Guy (ed.). *Learning with corpora*. Bologna: CLUEB, 108-137.

Gavioli, Laura; Aston, Guy. 2001. "Enriching reality: language corpora in language pedagogy". *ELT Journal* 55(3), 238-246.

Götz, Sandra; Mukherjee, Joybrato. 2006. "Evaluation of data-driven learning in university teaching: a project report". In Braun, Sabine; Kohn, Kurt; Mukherjee, Joybrato (eds.). *Corpus technology and language pedagogy: new resources, new tools, new methods*. Frankfurt/Main: Peter Lang, 49-68.

Granath, Solveig. 2009. "Who benefits from learning how to use corpora?" In Aijmer, Karin (ed.). *Corpora and language teaching*. Amsterdam: John Benjamins, 47-66.

Granger, Sylviane. 2002. "A bird's eye view of learner corpus research". In Granger, Sylviane; Hung, Joseph; Petch-Tyson, Stephanie (eds.). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins, 3-33.

Granger, Sylviane. 2008. "Learner corpora". In Lüdeling, Anke; Kytö, Merja (eds.). *Corpus linguistics: an international handbook.* Berlin: Walter de Gruyter, 259-274.

Granger, Sylviane. 2009. "The contribution of learner corpora to Second Language Acquisition and Foreign Language Teaching: a critical evaluation". In Aijmer, Karin (ed.). *Corpora and language teaching*. Amsterdam: John Benjamins, 13-32.

Gut, Ulrike. 2006. "Learner speech corpora in language teaching". In Braun, Sabine; Kohn, Kurt; Mukherjee, Joybrato (eds.). *Corpus technology and language pedagogy: new resources, new tools, new methods*. Frankfurt/Main: Peter Lang, 69-86.

Hagl, Stefan. 2008. *Schnelleinstieg Statistik.* München: Rudolf Haufe Verlag.

Halliday, Michael. 1982. "Linguistics in teacher education." In Carter, Ronald (ed.). *Linguistics and the teacher*. London: Routledge, 10-30.

Holec, Henri. 1981. *Autonomy and foreign language learning.* Oxford: Pergamon Press.

Hunston, Susan. 2002. *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Hunston, Susan. 2008. "Collection strategies and design decisions". In Lüdeling, Anke; Kytö, Merja (eds.). *Corpus linguistics: an international handbook.* Berlin: Walter de Gruyter, 154-168.

Johansson, Stig. 2007. "Using corpora: from learning to research". In Hidalgo, Encarnacion; Quereda, Luis; Santana, Juan (eds.). *Corpora in the foreign language classroom: selected papers from the sixth International Conference on Teaching and Language Corpora (TaLC 6), 2004*, 17-30.

Johansson, Stig. 2008. "Some aspects of the development of corpus linguistics in the 1970s and 1980s". In Lüdeling, Anke; Kytö, Merja (eds.). *Corpus linguistics: an international handbook.* Berlin: Walter de Gruyter, 33-53.

Johns, Tim. 1997. "Contexts: the background, development and trialling of a concordance-based CALL program". In Wichmann, Anne; Fligelstone, Steven; McEnery, Tony;

Knowles, Gerry (eds.). *Teaching and language corpora*. London: Longman, 100-115.

Johns, Tim. 2007 [1991]. "Should you be persuaded: two samples of data-driven learning materials". In Teubert, Wolfgang; Krishnamurthy, Ramesh (eds.). *Corpus inguistics: critical concepts in linguistics 6*. London: Routledge, 3-22.

Johns, Tim. 1986. "Micro-concord: a language learner's research tool". *System* 14(2), 151-162.

Kaltenböck, Gunther; Mehlmauer-Larcher, Barbara. 2005. "Computer corpora and the language classroom: on the potential and limitations of computer corpora in language teaching". *ReCALL* 17(1), 65-84.

Kennedy, Graeme. 1998. *An introduction to corpus linguistics.* London: Longman.

Kennedy, Graeme. 1992. "Preferred ways of putting things with implications for language". In Svartvik, Jan (ed.). *Directions in corpus linguistics: proceedings of Nobel Symposium 82, 1991*. Berlin: Walter de Gruyter, 335-373.

Knowles, Gerry. 1997. "Using corpora for the diachronic study of English." In Wichmann, Anne; Fligelstone, Steven; McEnery, Tony; Knowles, Gerry (eds.). *Teaching and language corpora*. London: Longman, 195-210.

Lee, Shinwoong. 2011. "Challenges of using corpora in language teaching and learning: implications for secondary dducation". *Linguistic Research* 28(1), 159-178.

Leech, Geoffrey. 1997. "Teaching and language corpora: a convergence". In Wichmann, Anne; Fligelstone, Steven; McEnery, Tony; Knowles, Gerry (eds.). *Teaching and language corpora*. London: Longman, 1-24.

Lenko-Szymanska, Agnieszka. 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.

Little, David. 2016. "Learner autonomy and Second/Foreign Language Learning". *Centre for Languages, Linguistics & Area Studies.* Available online at https://www.llas.ac.uk//resources/gpg/1409 (08/03/2016).

Mauranen, Anna. 2004. "Spoken corpus for an ordinary learner". In Sinclair, John M. (ed.). *How to use corpora in language teaching*. Amsterdam: Benjamins, 89-108.

McCarthy, Michael; McCarten, Jeanne; Sandiford, Helen. 2005. *Touchstone: student's book 1*. Cambridge: Cambridge University Press.

McEnery, Tony; Wilson, Andrew. 2001. *Corpus linguistics: an introduction.* Edinburgh: Edinburgh University Press.

Meunier, Fanny; Gouverneur, Celine. 2009. "New types of corpora for new educational challenges: collecting, annotating and exploiting a corpus of textbook material" In Aijmer, Karin (ed.). *Corpora and language teaching*. Amsterdam: John Benjamins, 179-202.

Meunier, Fanny; Gouverneur, Celine. 2007. "The treatment of phraseology in ELT textbooks." In Hidalgo, Encarnacion; Quereda, Luis; Santana, Juan (eds.). *Corpora in the foreign language classroom: selected papers from the sixth International Conference on Teaching and Language Corpora (TaLC 6), 2004*. Amsterdam: Rodopi, 119-140.

Meyer, Charles F. 2008. "Pre-electronic corpora". In Lüdeling, Anke; Kytö, Merja (eds.). *Corpus linguistics: an international handbook.* Berlin: Walter de Gruyter, 1-14.

Millar, Neil; Lehtinen, Ben. 2008. "DIY local learner corpora: bridging gaps between theory and practice". *The JALT CALL Journal* 4(2), 61-72.

Mishan, Freda. 2005. "Authenticating corpora for language learning: a rroblem and its resolution". *ELT Journal* 58(3), 219-227.

Mukherjee, Joybrato. 2004. "Bridging the gap between applied corpus linguistics and the reality of English Language Teaching in Germany." In Connor, Ulla; Upton, Thomas (eds.). *Applied corpus linguistics: a multidimensional perspective*. Amsterdam: Rodopi, 239-250.

Mukherjee, Joybrato. 2009a. *Anglistische Korpuslinguistik. Eine Einführung.* Berlin: Erich Schmidt Verlag.

Mukherjee, Joybrato. 2009b. "The grammar of conversation in advanced spoken learner English: learner corpus data and language-pedagogical implications." In Aijmer, Karin (ed.). *Corpora and language teaching*. Amsterdam: John Benjamins, 203-230.

Müller-Benedict, Volker. 2011. *Grundkurs Statistik in den Sozialwissenschaften. Eine leicht verständliche, anwendungsorientierte Einführung in das sozialwissenschaftlich notwendige statistische Wissen.* Wiesbaden: VS Verlag Springer.

O'Keeffee, Anne; McCarthy, Michael; Carter, Ronald. 2007. *From corpus to classroom: language use and language teaching.* Cambridge: Cambridge University Press.

Partington, Alan. 1998. *Patterns and meanings: using corpora for English language research and teaching.* Amsterdam: John Benjamins.

Rasinger, Sebastian 2008. *Quantitative research in linguistics: an introduction.* London: Continuum.

Renouf, Antoinette. 1997. "Teaching corpus linguistics to teachers of English". In Wichmann, Anne; Fligelstone, Steven; McEnery, Tony; Knowles, Gerry (eds.). *Teaching and language corpora*. London: Longman, 255-266.

Römer, Ute. 2004. "Comparing real and ideal language learner input: the use of an EFL textbook corpus in corpus linguistics and language teaching". In Aston, Guy; Bernardini, Silvia; Stewart, Dominic (eds.). *Corpora and language learners.* Amsterdam: John Benjamins, 151-168.

Römer, Ute. 2008. "Corpora and language teaching". In Lüdeling, Anke; Kytö, Merja (eds.). *Corpus linguistics: an international handbook.* Berlin: Walter de Gruyter, 112-130.

Römer, Ute. 2009. "Corpus research and practice: What help do teachers need and what can we offer?" In Aijmer, Karin (ed.). *Corpora and language teaching*. Amsterdam: Benjamins, 83-100.

Römer, Ute. 2011. "Corpus research applications in Second Language Teaching". *Annual Review of Applied Linguistics* 31, 205-225.

Seidlhofer, Barbara. 2002. "Pedagogy and local learner corpora: working with learning-driven data". In Granger, Sylviane; Hung, Joseph; Petch-Tyson, Stephanie (eds.). *Computer learner corpora, Second Language Acquisition and foreign language teaching.* Amsterdam: John Benjamins, 213-234.

Sinclair, John. 1997. "Corpus evidence in language description". In Wichmann, Anne; Fligelstone, Steven; McEnery, Tony; Knowles, Gerry (eds.). *Teaching and language corpora*. London: Longman, 27-39.

Sinclair, John. 1991. *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Sinclair, John. 2004. "Introduction". In Sinclair, John (ed.). *How to use corpora in language teaching*. Amsterdam: John Benjamins, 1-14.

Stewart, Dominic; Bernardini, Silvia; Aston, Guy. 2004. "Introduction: Ten years of TaLC". In Aston, Guy; Bernardini, Silvia; Stewart, Dominic. *Corpora and language learners*. Amsterdam: John Benjamins, 1-20.

Thompson, Paul. 2006. "Assessing the contribution of corpora to EAP practice". In Kantaridou, Zoe; Papadopoulou, Iris; Mahili, Ifigenia (eds.). *Motivation in learning language for specific and academic purposes*. Macedonia: University of Macedonia [CD ROM].

Tribble, Chris; Jones, Glyn. 1990. *Concordances in the classroom: a resource book for teachers.* London: Longman.

Tribble, Christopher. 2015. "Teaching and language corpora. perspectives from a personal journey". In Lenko-Szymanska, Agnieszka; Boulton, Alex (eds.). *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam: John Benjamins, 37-64.

Tsui, Amy B. M. 2004. "What teachers have always wanted to know - and how corpora can help". In Sinclair, John (ed.). *How to use corpora in language teaching*. Amsterdam: Benjamins, 39-66.

Wichmann, Anne. 2008. "Speech corpora and spoken corpora". In Lüdeling, Anke; Kytö, Merja (eds.). *Corpus linguistics: an international handbook.* Berlin: Walter de Gruyter, 187-206.

Widdowson, Henry G. 2000. "On the limitations of linguistics applied". *Applied Linguistics* 21(1), 3-25.

Widdowson, Henry G. 2003. *Defining issues in English Language Teaching.* Oxford: Oxford University Press.

Wynne, Martin. 2008. "Searching and concordancing". In Lüdeling, Anke; Kytö, Merja (eds.). *Corpus linguistics: an international handbook.* Berlin: Walter de Gruyter, 706-737.

List of corpora

A Standard Corpus of Present-Day Edited American English (Brown Corpus). 1964, 1971, 1979. Compiled by Francis, W. N.; Kučera, H. Providence, Rhode Island: Brown University. Further information: http://clu.uni.no/icame/brown/bcm.html (24/05/2016).

Collins Birmingham University International Language Database (COBUILD). University of Birmingham. Further information: http://www.collins.co.uk/page/The+Collins+Corpus (24/05/2016).

Davies, Mark. 2004. *The British National Corpus* (BNC). Available online at http://corpus.byu.edu/bnc/ (24/05/2016).

Davies, Mark. 2008. *The Corpus of Contemporary American English* (COCA). Available online at http://corpus.byu.edu/coca/ (24/05/2016).

Davies, Mark. 2007. TIME Magazine Corpus. Available online at http://corpus.byu.edu/time (24/05/2016).

ELISA corpus. 2005. Compiled by Sabine Braun. University of Tübingen. Further information: http://www.uni-tuebingen.de/elisa/html/elisa_index.html or http://www.corpora4learning.net/ (24/05/2016)

Learning Prosody in a Foreign Language (LeaP). 2001., Project director: Ulrike Gut, University of Bielefeld. Further information: http://www.philhist.uni-augsburg.de/de/lehrstuehle/anglistik/angewandte_sprachwissenschaft/workshop/pdfs/LeapCorpus_Manual.pdf (24/05/2016)

Simpson, R. C.; Briggs, S. L.; Ovens, J.; Swales, J. M. 2002. *The Michigan Corpus of Academic Spoken English (*MICASE*)*. Ann Arbor, MI: University of Michigan. Available online at http://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple (24/05/2016).

The *Bank of English* (BoE): An International Language Corpus. Jointly owned by HarperCollins Publishers and the University of Birmingham. Further information: http://www.titania.bham.ac.uk/docs/svenguide.html (24/05/2016).

The *Cambridge English Corpus* (CEC). Cambridge University Press. Further information: http://www.englishprofile.org/cambridge-english-corpus (24/05/2016).

The *International Corpus of English* (ICE). Copyright: Survey of English Usage. University College of London, 1998. Further information: http://ice-corpora.net/ICE/INDEX.HTM (24/05/2016).

The *International Corpus of Learner English* (ICLE*)*. 2009. Project Director: Sylviane Granger. Université catholique de Louvain. Further information: https://www.uclouvain.be/en-cecl-icle.html (24/05/2016)

The *Survey of English Usage Corpus*. London: University College London. Further information: http://www.ucl.ac.uk/english/survey-of-english-usage (24/05/2016)

VOICE. 2009. *The Vienna-Oxford International Corpus of English* (version 1.0 online). Director: Barbara Seidlhofer; Researchers: Angelika Breiteneder, Theresa Klimpfinger, Stefan Majewski, Marie-Luise Pitzl. http://voice.univie.ac.at (24/05/2016).

List of dictionaries and grammars

Fries, Charles Carpenter. 1940. *American English grammar: the grammatical structure of present-day American English with especial reference to social differences or class dialects*. New York: Appleton-Century-Crofts.

Longman Essential Activator dictionary. 2006. (2nd edition). London: Longman.

The Oxford English Dictionary computer file. 1992. (2nd edition). [CD-ROM]. Oxford: Oxford University Press.

Sinclair, John (ed.). 1987. *Collins COBUILD English language dictionary*. London: Collins.

Sinclair, John (ed.). 2005. (2nd edition). *Collins COBUILD English grammar*. Glasgow: HarperCollins.

Swan, Michael. 2005. (3rd edition). *Practical English usage: easier, faster reference*. Oxford: Oxford University Press.

Webster's Third New International Dictionary of the English Language: unabridged. 1986 [1961]. Springfield, Mass: Merriam-Webster.

Willis, Jane; Willis, Dave. 1988. *Collins COBUILD English Course*. London: Collins.

APPENDIX I: Questionnaire

Welcome to this survey about the use of corpora in language teaching. The survey is part of my diploma thesis with the topic "The Use of Corpora in Austrian EFL Classrooms" in which the use of corpora and the attitudes of Austrian EFL teachers towards corpora are investigated.

Corpora are electronic collections of texts. These databases provide information about actual language usage; they can be used for research purposes, but they can also be used in language teaching.

(If you are not familiar with the use of corpora, don't worry! You get a chance to try it out if you want to.)

I would like to ask you for your support. Please answer the following questions. There are no 'right' or 'wrong' answers and the survey is anonymous so please give your answers sincerely.
Answering the questions will take you about 10-15 minutes.

Thank you for your participation!

1. How familiar are you with corpus linguistics?

○	Very familiar
○	Vaguely familiar
○	Not at all familiar

2. Are you familiar with any of the following corpora?

☐	British National Corpus (BNC)
☐	Corpus of Contemporary American English (COCA)
☐	Corpus of Historical American English (COHA)
☐	Brown Corpus
☐	Freiburg-Brown Corpus of American English (FROWN)
☐	Lancaster-Oslo-Bergen Corpus (LOB)
☐	Freiburg-LOB Corpus of British English (FLOB)
☐	Michigan Corpus of Academic Spoken English (MICASE)
☐	Michigan Corpus of Upper-Level Student Papers (MICUSP)
☐	Vienna-Oxford International Corpus of English (VOICE)
☐	None

Other, please specify:

3. How familiar are you with the use of corpora for language teaching?

○	Very familiar
○	Vaguely familiar
○	Not at all familiar

4. Have you used corpora for any of the following purposes?

	Yes	No	don't know
For own research purposes	○	○	○
As a reference tool (e.g. for correcting)	○	○	○
For designing teaching materials	○	○	○
For teaching (e.g. for explorative learning of pupils in the classroom)	○	○	○
Other, please specify:			

Optional exercises

If you are curious how corpora could be used, you might want to try out the following exercises:

Task 1: "If I were" – "If I was": Which one is correct?
Use the link below (COCA) and check whether there is a correct/incorrect version.

Instructions: Type in "if I were" in the box next to "WORD(S)" and click on "search". In the right column, you get the structure and the frequency. Do the same with "if I was".
You might notice that there is no major difference in frequency between these two structures.

Tip: If you click on the structure, you get the co-text of the structure on the bottom of the page.

Task 2: What is the difference between *although* and *despite* in terms of sentence structure?
Use the link below (COCA) and look for the structures following *although* and *despite*.

Instructions: Use the "KWIC" display and type in "although". Click on "search".
Do the same with "despite". When you pay attention to the structures that follow *although* and *despite*, you may notice that *although* is usually followed by a clause, while *despite* is only followed by a noun or gerund.

Link: COCA

(Corpus of Contemporary American English)

(link opens pop-up window)

5. If you want to, you can share your experience, your opinion or other thoughts about the use of corpora here:

6. In this part, please tell us how much you agree or disagree with the following statements.

	agree	slightly agree	slightly disagree	disagree	don't know
Using corpus activities in my lessons is too time-consuming.	○	○	○	○	○
I have had training in how to use corpora.	○	○	○	○	○
I have limited access to computers at school.	○	○	○	○	○
Teaching and learning with corpora is too demanding for pupils.	○	○	○	○	○
I know how to prepare relevant corpus activities.	○	○	○	○	○
When using corpora, too much time is spent on too specific language issues.	○	○	○	○	○
There is limited availability of ready-to-use corpus teaching material.	○	○	○	○	○
I am uncomfortable answering technical questions that might be raised during corpus activities.	○	○	○	○	○
I have been looking for suitable corpora for teaching.	○	○	○	○	○

7. In this part, please tell us how much you agree or disagree with the following statements.

	agree	slightly agree	slightly disagree	disagree	don't know
Creating appropriate corpus teaching material takes too much time.	○	○	○	○	○
It is easy to find appropriate topics for corpus activities.	○	○	○	○	○
Unpredictable outcomes of corpus activities make me feel insecure.	○	○	○	○	○
Corpus activities are relevant enough to be taught at school.	○	○	○	○	○
I have found suitable corpora for teaching.	○	○	○	○	○
I am uncomfortable answering linguistic questions that might be raised during corpus activities.	○	○	○	○	○
The absence of an answer key to corpus activities bothers me.	○	○	○	○	○
I like the idea that pupils become researchers themselves when working on corpus activities.	○	○	○	○	○
When pupils become researchers themselves, my role as a teacher becomes less important.	○	○	○	○	○

8. In your opinion, what are the greatest obstacles of corpora use for language teaching?
Please give them a mark from 1 to 4 –---- 1 being a small obstacle and 4 being a great obstacle.

	1	2	3	4	don't know
A lack of time in class	○	○	○	○	○
A lack of time for preparation	○	○	○	○	○
A lack of teacher training	○	○	○	○	○
A lack of teaching material	○	○	○	○	○
A lack of awareness about the potential of corpora for language teaching	○	○	○	○	○

9. In this part, please tell us how much you agree or disagree with the following statements.

	agree	slightly agree	slightly disagree	disagree	don't know
I would use corpora in the future if there was more time in class.	○	○	○	○	○
I would use corpora in the future if there was more ready-to-use teaching material.	○	○	○	○	○
I would use corpora in the future if there was more teacher training on how to use corpora.	○	○	○	○	○
I would use corpora in the future if there was better access to computers at school.	○	○	○	○	○
I would use corpora in the future if free online corpora were easier to access.	○	○	○	○	○
I would use corpora in the future if free online corpora were easier to use.	○	○	○	○	○
I would use corpora in the future if I had a clearer idea how it can be used for teaching.	○	○	○	○	○

10. The space below has been provided for you in case you would like to share your thoughts about possible improvements that could initiate more teachers to use corpora.

11. Gender

○ female

○ male

12. Age

○ <20

○ 20-29

○ 30-39

○ 40-49

○	50-59
○	>60

13. How many years have you been teaching?

[   ]

14. Education

○	University of Vienna
○	University of Salzburg
○	University of Graz
○	University of Innsbruck
○	University of Klagenfurt

Other, please specify: [   ]

15. Did you come across corpora during your studies at university?

○	Yes
○	No
○	Don't know

16. How would you rate your computer skills?

○	Basic
○	Intermediate
○	Advanced

17. The space below has been provided for you in case you would like to share any comments, thoughts or suggestions.

[   ]

Your answers have been saved.

You can close the window now.

In case you have any questions, you can contact me via email: a1001683@unet.univie.ac.at

Thank you very much for your cooperation!

APPENDIX II: Abstract in German

Diese Arbeit untersucht die Verwendung von Textkorpora und deren Anwendung im Fremdsprachenunterricht. Das Ziel dieser Arbeit ist es herauszufinden, inwiefern österreichische LehrerInnen von Englisch als Fremdsprache mit Korpuslinguistik bzw. der Verwendung von Textkorpora im Sprachunterricht vertraut sind. Außerdem wird untersucht, inwiefern LehrerInnen bereits von Textkorpora für Lehrzwecke Gebrauch machen und welche Hindernisse Korpora mit sich bringen, die LehrerInnen von deren Verwendung im Unterricht abhalten. In Verbindung dazu sollen jene Verbesserungen aufgezeigt werden, die LehrerInnen dazu bringen, Textkorpora häufiger im Unterricht zu verwenden. Zur genaueren Untersuchung dieser Themen wurde ein Online-Fragebogen an österreichische LehrerInnen der Sekundarstufe geschickt, der die Fragen der Vertrautheit mit Korpora, der Hindernisse sowie potenzieller Verbesserungen aufgriff. Die Ergebnisse zeigen eine Diskrepanz zwischen der Bekanntheit von Korpuslinguistik und der Bekanntheit der Verwendung von Korpora im Sprachunterricht. Zudem konnten ein Mangel an Bekanntheit und Unterrichtsmaterial, Zeitmangel sowie unzureichende LehrerInnenausbildung, die meist in einem Fehlen entsprechender Fähigkeiten und Kenntnisse resultiert, als größte Hindernisse identifiziert werden, die die Verwendung von Korpora im Sprachunterricht erschweren. Diese Arbeit betont die Bedeutung der LehrerInnenausbildung, die sich auf die Vermittlung von Korpuslinguistik sowie auch auf die praktische Umsetzung des Gebrauchs von Textkorpora im Sprachunterricht konzentrieren sollte. Die vorliegende Arbeit dient letztlich auch dazu, über mögliche Verwendungen von Korpora, deren Einschränkungen sowie über deren Vorteile für den Gebrauch im Sprachunterricht zu informieren.