universität wien

MASTERARBEIT

Titel der Masterarbeit

„Using Copora for Designing Teaching Material for ESP Coures in the Social Security Sector “

Verfasserin

Mag.[a] Anches Holzweber

angestrebter akademischer Grad

Master of Arts (MA)

Wien, 2012

Studienkennzahl lt. Studienblatt:	A 066 812
Studienrichtung lt. Studienblatt:	English Language and Linguistics
Betreuerin:	Univ.-Prof. Dr. M. Evelien Keizer

TABLE OF CONTENTS

1 INTRODUCTION

Over the last 50 years, the world has changed considerably in terms of politics, social systems and economics. Boundaries have shifted, borders have disappeared, and English - even more so than before - has become the language used for efficient global communication. Especially against the background of the European Union, where English is one of the official languages, where people with many different mother tongues have become one people, the demand for English for Specific Purposes (hereinafter referred to as ESP) has increased enormously, and people from all fields and businesses are faced with the challenge to improve their English skills in their specific areas, ranging from technical production departments to customer service or medical jobs, to name just a few. This development has, of course, not only confronted learners with new challenges, but English teachers as well, and ESP teaching has become a large sub-area of teaching English.

ESP has developed out of the basic insight that language should be seen as a tool for communication rather than as sets of phonological, grammatical and lexical items and has led to the development of learning programs that reflect the different communicative needs of different groups of students. It is mainly concerned with the practical part of language and a demand for special teaching material. Teaching an ESP course very often is an encounter of specialists of two fields; on the one hand, the ESP teacher, who is the language specialist and, on the other hand, the students who are experts in their respective fields. Even though it is practically impossible for ESP-teachers to be familiar with the language of all the different specialist areas, they need to design curricula that contain specific technical terminology as well as patterns of interaction between people involved in the communication processes in this field.

Corpora are said to be very useful tools to find out about the discourse used in specific areas. Corpus research goes back to the 1920s when material was counted and classified manually to get an idea of syntax or semantics or how

words were distributed. Today, corpus research is conducted with the help of computers and an enormous amount of material can be collected, stored and then be exploited for linguistic purposes. Corpus Analysis in an ESP setting seems to match the teachers' and students' requirements quite well as it highlights recurrent features of language. Corpus Research might not be able to answer all linguistic questions and it might hold some disadvantages, but it is unquestionably incredibly valuable in helping (ESP) teachers to decide what is worth giving attention to in the first place, and subsequently, to design lessons with the help of the samples provided by the corpus.

This thesis is divided into two large parts and its aim is, on the one hand, to introduce the readers to ESP as well as Corpus Linguistics, cover some aspects of the theoretical background and critically discuss this branch of linguistics and this specific research method, respectively. On the other hand, the second part presents and discusses the individual steps and results of the empirical study that was performed in the course of the thesis to prove the usefulness of Corpus Research for material designing in ESP, specifically for English courses in the Social Security sector.

Part One will be concerned with theoretical information about ESP and Corpus Linguistics. It will start with providing some different definitions of ESP in Chapter 2, in an attempt to introduce the reader to the topic. The chapter will continue with a brief history of the development of ESP and provide the background against which this branch of applied linguistics emerged. The challenges ESP teachers have to face and the process of designing ESP curricula and syllabi will be covered in the following sub-chapters.

Chapter 3 will then provide a brief definition of Corpus Linguistics, introduce different kinds of corpora, explain the appropriate set-up of a corpus and the questions that can be answered with the help of Corpus Analysis, and finally highlight its advantages and disadvantages, with a special emphasis on the usefulness for ESP teachers.

Part Two of the paper will focus on the results and interpretation of the study conducted with two home-made corpora with texts from the area of social

security, and will present potential teaching topics for "English for Employees in the Social Security Sector", based on the findings of the corpus research.

2 ENGLISH FOR SPECIFIC PURPOSES (ESP)

This first chapter will start with a presentation of different definitions of ESP and attempts to find a general definition for the term. Then, the development of ESP and the different branches mentioned in the relevant literature will be discussed. The last two sub-chapters will deal with the challenges ESP teachers have to face and what they have to consider when creating syllabi and curricula.

2.1 What is ESP?

The term ESP itself suggests that we are dealing with a special kind of English which is distinctive to a certain extent. It could be argued that its distinctiveness is caused by the specific principles and procedures of certain institutional activities followed by the speakers of a certain professional community. Widdowson (1998: 3) argues that all uses of English are specific and serve particular purposes, since whenever people communicate they want to be as effective as possible, and therefore they are being as specific as they can about the point they want to make. So in his opinion, all communication in English is in general ESP, as language is always used in a specific sense to serve pragmatic purposes. Widdowson definitely has a point here, but, as a starting point it is probably necessary to narrow the term ESP a little down and start from a more straightforward and simpler definition by Dudley Evans & St. John (1998, quoted in Gavioli 2005:5), according to which ESP is "the study of the English language in specialized contexts and fields of knowledge, such as medicine, engineering, business and the like."

If one considers the prospering global economy in the last five or six decades and the subsequent spreading of English as the most frequently used language in business, science, technology and academia, it becomes obvious that the need for being able to speak English at different workplaces has risen, too.

Language learners in an ESP setting do usually not attend courses out of a personal interest but rather learn the foreign language out of a professional need. Their proficiency level is generally already quite high, as this is expected by, and is an indicator of good performance for, employers. Nevertheless, however advanced the learners' linguistic proficiency might be, they very often "lack communicative or cultural proficiency in the work-specific context" (Kim: 2008, 1).

Typical environments that have a need for ESP courses would be companies that work internationally and whose workforce comprises people of many different nationalities, where meetings or conferences are usually held in English, contracts are drawn up in English and day-to-day communication quite often happens in English. Staff, therefore, need to be able to understand, read, write and speak in English on a certain level, depending on their positions and responsibilities. This can vary from receptionists who need to be able to politely welcome visitors, to secretaries who have to answer or connect telephone calls in English, give information on the telephone, write and answer e-mails and make reservations or entertain visitors from other countries, to executive staff who need the skills for negotiating or giving presentations on a professional level, but also need to be able to socialize with their business partners.

Doctors, who must keep up with the latest developments in medical science and technology to offer state-of-the-art treatments to their patients, need to attend international conferences - with English being the official language - where they have to give speeches and listen to talks on their field of expertise, a rather challenging project if one's language skills do not have a certain level.

Another typical environment for the use of ESP would be governmental institutions across the European Union, where English is one of the official languages for employees from many different cultural backgrounds. In this context especially, there is a big demand for legal English, as it is one of the European Union's objectives to standardize and harmonize laws and regulations so that they can be applied in all member countries.

However, also other institutions need to communicate with their respective counterparts in other countries, which need not necessarily be in Europe, such as

the Main Association of Austrian Social Security Institutions (Hauptverband der Österreichischen Sozialversicherungsträger, henceforth Hauptverband, the Austrian Health Funds (Österreichische Gebietskrankenkassen) or the Austrian Employment Service (Arbeitsmarktservice, henceforth 'AMS'), which have to communicate with their liaison offices around the globe.

Apart from these examples for the widespread use of English in specialized businesses and the consequent need for ESP, we must not forget that ESP is also taught in English speaking countries, where immigrants whose proficiency is either limited or nonexistent at all are being prepared for the 'world of work'. However, this paper is only concerned with ESP courses in non-English speaking countries.

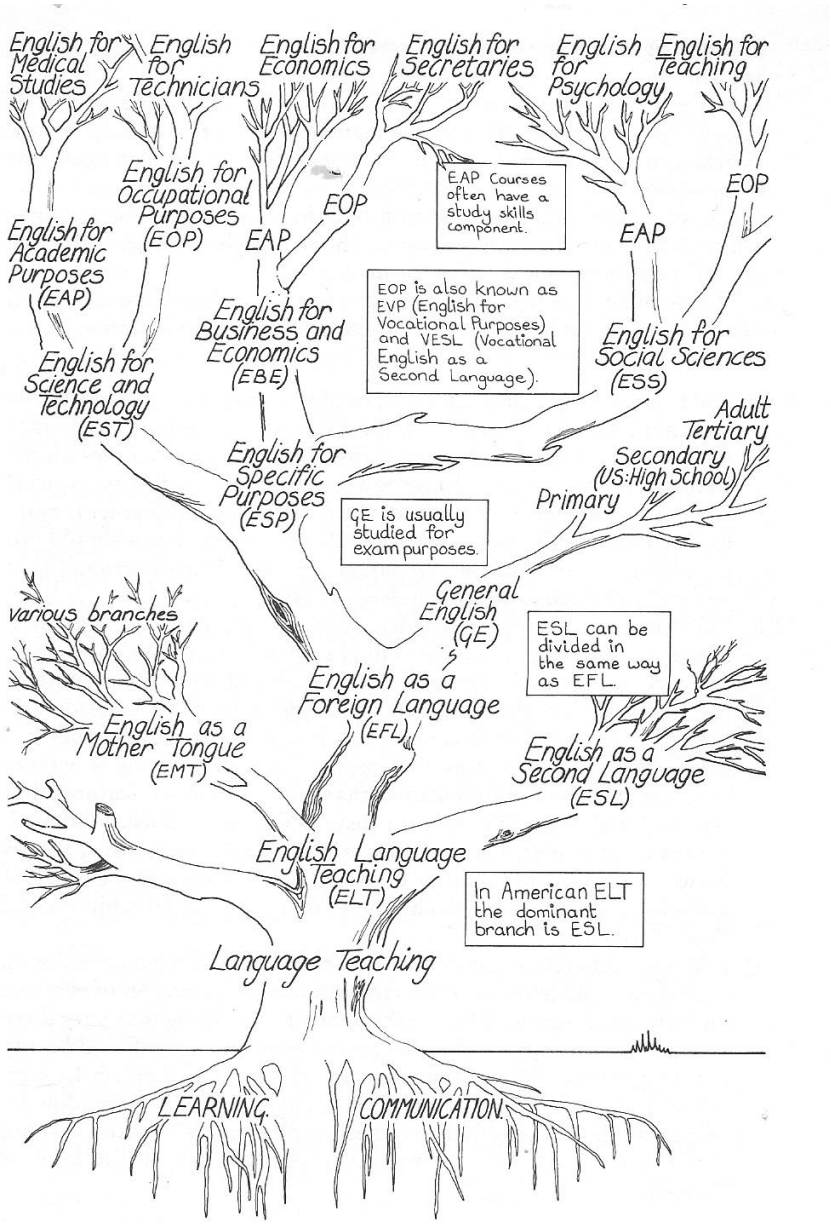
Different researchers on the topic have formulated different definitions, but they all seem to have one thing in common, namely that sometimes it is difficult to distinguish between ESP and English for General Purposes (EGP), because they are both concerned with a variety of registers, text types and discourse types, and speakers of general English actually speak for all kinds of special purposes every day. As in so many other fields, boundaries are fuzzy and it is quite hard to find the one correct definition for such an extensive area.

When we go back to Dudley Evans and St. Johns' definition of ESP mentioned above, one may argue that it is quite superficial and that there is something missing. And, indeed, sometimes it might be necessary to broaden our idea of Language for Specific Purposes (LSP) or how Gavioli (2005: 5) calls it, the "LSP umbrella", especially in the "light of increasing European political and economic integration" (Gavioli 2005: 5), where there is undoubtedly a general need for cross-cultural or cross-linguistic communication.

Also Barnard and Zemach (2003: 306-7 quoted in Basturkmen 2010: 3) argue correctly that in many cases it might be quite difficult to draw a line between EGP and ESP, as some ESP courses will, due to the students' level of proficiency of English, overlap with EGP. They conclude that in ESP courses the aims are generally more focused and the range of topics is not as wide as in EGP curricula.

Hutchinson and Waters (1987: 19) provided one of the first definitions, and they emphasize that both EGP as well as ESP are concerned with communication and language and they advise that one should not consider ESP as a special kind of language product but rather a special approach to English teaching "which is directed by specific and apparent reasons for learning". With their renowned analogy of a tree (see Figure 1 below) they show where ESP stands in relation to other branches of language teaching such as English as a Foreign Language (EFL) or English as a Second Language (ESL) and with this visualization at least they give us an idea what ESP is not.

Figure 1: The tree of ELT (Hutchinson and Waters 1987: 17)



The conclusions they draw are that "ESP is not a matter of teaching 'specialized varieties' of English" (Hutchinson and Waters 1987: 18), and that using language for a specific purpose does not automatically mean that the form of the language is special and differs from other forms. Further, it is not just a matter of specialized words and grammar for specialists in a specific field, as communication happens not only on the surface, and there is much more to it than just what we read and hear. Lastly, they do not consider ESP teaching as different in kind from any other form of language teaching as it also needs to be based on the principles of efficient and effective learning. They are of the opinion that there is no such thing as particular ESP methodologies, and see ESP rather as an approach to language teaching and learning which depends solely on the reasons why learners need to acquire English skills and their very specific and individual needs.

Widdowson (1998) considers ESP not only a matter of teaching or learning the language, but strongly defends the view that ESP is to a great extent about communication and mastering communicative skills. He reminds us that communication may not only fail due to disparity of linguistic knowledge but also due to cultural disparity, when the people involved in communication are not members of the same discourse community within a certain discourse community.

> Communication, then, is not simply a matter of issuing semantic tokens of fixed meaning but of using the resources of the language code indexically, to indicate (point out, invoke) shared schematic knowledge of ideational patterns of conceptualization and interpersonal patterns of communication, shared assumptions, beliefs, values, conventions of behaviour, that is to say aspects of a common culture which define particular discourse communities. If we think of communication in this way (in these terms) certain issues arise of relevance to the use of English for institutional purposes and indeed of relevance to the very nature of English as a language of international communication. (Widdowson 1998: 9)

However, it could be argued that each aspect of language teaching deals with mastering communication and communication skills, and it cannot only be ascribed to ESP, but is true for EGP as well. Ultimately, one learns a new language to be able to communicate, so this is not an exclusive feature of ESP.

Another important aspect of ESP is that is not only concerned with the language aspect but also with culture, and Widdowson (1998: 13) advises us strongly to consider such questions as what kind of culture we are dealing with, what its role and responsibility is, how it adapts to different circumstances and what the business is for and the purposes actually are. Having answered these questions, an ESP program can be designed "to instruct students in the effective use of the language which serves to further their professional purposes" (ibid: 13). Widdowson also distinguishes between the 'mastery' and the 'mystery' of a language, both terms deriving from the French word for 'craft' or 'trade', *metier*. For insiders and practitioners of a specific field, the jargon they use is a 'mastery language'. For outsiders, on the other hand, it is a 'mystery language'. According to him, the need for ESP originates from the difference in proficiency between the 'mastery language' and the general proficiency in English.

Finally, Robinson (1991: 1) gets to the nub of the issue when she points out that it is almost impossible to offer a definition of ESP that can be applied to any situation. According to her (1991: 2-5), ESP is goal directed, which means English is a necessity for the students rather than a personal interest, and they need it either for their studies or for work. She emphasizes the importance of basing ESP courses on needs analyses, which are performed to specify in as detailed a way as possible in what situations learners have to use the English language. Usually, ESP courses are held within a given time frame, which asks for a very thorough and precise course preparation. To achieve the goals within the limited time frame, close collaboration between organizers, course designers, teachers and learners is indispensable.

Robinson also introduces an interesting point of view when she argues that specialist language (especially terminology) and specialist content are not always a necessary part of an ESP course, but rather the activities students engage in are the decisive factor (Robinson 1991: 4). She also argues that what makes a course an ESP course is the special need rather than the special material, and one should plan the curriculum with the particular students and their personal preferences in mind. This might be right to certain extent, because obviously it is not only the

material that makes an ESP course an ESP course. However, if teachers did not use special exercises containing specialist language on the respective topic, it would be inappropriate to call the course an ESP course. So consequently, it could be argued that the exact opposite of Robinson's contention is true, namely that it is especially the teaching material, which is - of course - adapted to the special needs of the students, that makes the course an ESP course and not the learners' needs. Robinson further suggests that "what we are really involved in as ESP practitioners is not so much teaching English for specific purposes but teaching English to specified people" (1991: 5), and Widdowson (1998: 11) seems be of the same opinion as she is when he argues that "[…] ESP is concerned with sustaining […] discourse communities […]".

However, if one teaches general English to lawyers, why should one call it an ESP course? The lawyers' needs of legal English would not make it an ESP course. On the other hand, if one teaches legal English to a group of shop assistants, it would still be an ESP course. Having mentioned the 'specified people' and 'discourse communities' above, it is important to clarify who they actually are. Even though in some countries - such as in Austria - ESP is taught in vocational secondary schools, ESP course participants are more likely to be adults in tertiary education or experienced members of the workforce than secondary school pupils.

Basturkmen (2010: 1-3) summarizes different definitions of ESP and distinguishes between absolute and variable characteristics of ESP, which were also conveyed by Garcia Mayo (2000: 24) (see Table 1)

One thing that might be missing in the absolute characteristics of this table is 'multidisciplinarity', as the intertwining of language instruction and the respective specialist area such as business, medicine or science are a core quality of ESP, and, as argued above, the use of special materials is crucial and should also be included as one of the absolute characteristics.

In conclusion it can be said that even though "[…] specificity to purpose is a general pragmatic condition for effective communication" and "specificity is as much a matter of the process as the purpose of learning" (Widdowson 1998: 11) and, therefore, one could argue that ESP concerns all parts of communication in

English, the term ESP is generally used to define language which happens within specific environments, between participants with specific needs, who need to reach specific goals. Holme (1996: 3-4, quoted in Basturkmen 2010: 3) calls it "simply a narrowing down of [the] needs spectrum" of different students. Unfortunately, the importance of the special material that should be used for effective teaching seems to be widely ignored.

Table 1: Absolute and variable characteristics of ESP as listed by Garcia Mayo

Absolute characteristics of ESP	Variable characteristics of ESP
• designed to meet specific needs of learners • makes use of the underlying methodology and the activities of the discipline it serves • centered on the language (grammar, lexis, register), skills, discourse and genres appropriate to these activities	• may be related to or designed for specific disciplines • may use a different methodology from that of EGP • likely to be designed for adult learners, either at a tertiary level institution or in a professional work situation; however, it could be used at secondary school level • generally designed for intermediate to advances learners; courses mostly assume a basic knowledge of the language, but it might be used with beginners

2.2 The development of ESP

A specific date for the beginning of ESP is quite difficult to establish, and different researchers suggest different points of time. However, they seem to agree that the dramatic changes that took place in science, technology as well as society after World War II can be held responsible for establishing the "foundations for the development of ESP" (Garcia Mayo: 2000, 21).

Garcia Mayo (2000: 21 pp) mentions four factors that appeared at approximately the same time and, therefore, might have been involved in giving birth to the development of ESP. The first phenomenon is what she calls the 'information explosion' due to scientific research. Secondly, she mentions the fact that after the end of World War II, the Unites States of America were the nation that was most industrialized, with a booming educational and technological infrastructure. They not only contributed most to the pool of information, they also were the most frequent users of this information. The third factor she considers important is that the end of World War II and the invention of the computer took place at approximately the same time and "[b]y political agreement, the Allies captured the languages of information". The fourth, and probably most important, reason for the emergence of ESP Garcia Mayo discusses, is the enormously high number of English speakers worldwide, not only in the 'classic' English speaking countries such as the United Kingdom, the United States of America and Australia, but also in countries such as Nigeria or India, where English is spoken by well-educated citizens. She also points out that ESP cannot be seen as a monolithic phenomenon, but that its development has taken place at different speeds in different parts of the world

When Teaching English to Speakers of Other Languages (TESOL) started to appear, English for Academic Purposes (EAP) used to be the major area in the field. Due to the economic changes and the expansion of international business, English for Business Purposes (EBP) became more and more important, and by the 1980s ESP had become a vital and innovative field in the TESOL movement (Howatt 1984: 345).

The development of ESP was also influenced by a shift in applied linguistics from the focus on language form to that of language use, and finally, thanks to developments in educational psychology, learners gained center stage in teaching.

According to Hutchinson and Waters (1987: 9-15) we can distinguish between five phases of development in ESP:

1) Register Analysis in the 1960s and early 1970s, which was especially concerned with the form of language, with language at the sentence level.

2) Rhetorical/Discourse Analysis, with its belief that the students' difficulties rather derive from an unfamiliarity of the use than an unfamiliarity of the form, shifted the attention to a level above the sentence. Whereas Register Analysis was concerned with the grammar in the sentence, Discourse Analysis focuses on the correct combination of sentences to produce meaning.
3) Target situation analysis, as the primitive version of needs analysis established processes that made it possible to relate language analysis more closely to the learners' reasons for learning and, therefore, set the existing knowledge on a scientific basis.
4) The 'skill centered approach' was concerned with the cognitive processes that underlie language use and introduced the theory that learning is more effective if meaningful context is provided, because learners cope with new material by referring to existing knowledge and experience.
5) The 'learning-centered approach' has developed out of the realization that the first four stages were quite flawed. It is an approach to ESP that is based on understanding of language learning which recognizes the variety of learning styles of different learners.

Basturkmen (2010: 1) captures the essence when she says that ESP has developed out of the basic insight that language is above all a tool for communication and not about stringing together phonological, grammatical and lexical items. This insight lead to the development of learning programs that reflect the different communicative needs of different groups of students. ESP has its own approaches to curriculum development, materials design, pedagogy, testing and research.

Summarizing, it can be said ESP developed out of a relatively sudden emergence of a need for specific English language that had not been there before global economy started to prosper after WW II, business started more and more to cross the borders, and speaking a common language became a necessity. Based on this need, the development of the linguistic discipline, which is now called ESP, started and it has continued ever since.

2.3 Current trends in ESP

Education, training, and naturally also ESP have been greatly influenced by the advancement of (computer) technology and information science. Computer aided language learning (CALL) as well as multimedia language learning has enabled learners as well as teachers to access all kinds of learning or teaching material, respectively.

The internet has proved to be an invaluable (re)source for material of various kinds, genres and fields of expertise. It is quite easy to find appropriate and authentic listening material such as speeches or interviews in numerous different accents and registers on all kinds of topics; specialist texts can be found and downloaded or at least printed out, interaction with other learners around the world has become quite easy and an important part in the learners' lives, be it via e-mail, instant messengers or Skype. The possibility of exploiting the world wide web obviously serves not only the personal interests of the students but also the trend towards diversity and cross-cultural communication, another newly recognized part of ESP.

2.4 The challenges of an ESP teacher

ESP teachers are often faced with the widespread belief amongst learners that any native speaker of English qualifies as a good English teacher. One might argue that this can be discounted as nothing more than a popular misconception, as English teachers, specifically ESP teachers, and language teachers in general, need to fulfill a broad range of requirements. Being a native-speaker is one of the less important ones and can probably be completely disregarded.

As any other teacher of any subject, regardless of the students' age-group, gender or profession, an ESP teacher needs many more skills than the required English skills. Apart from a very good command of the language, ESP teachers need pedagogical skills, communicational skills, knowledge of different teaching methodologies and they should have broad general knowledge. Moreover, they need to be interested in numerous different topics and willing to do research on them, they need to be able to design curricula and thoroughly plan syllabi, access

different materials, and they also need to be open-minded and sensitive to processes that happen within a specific (speaking) community.

As mentioned before, ESP is not only concerned with teaching or learning a language in the sense of producing single, grammatically correct, sentences. ESP teachers do not simply teach grammar and vocabulary, they are in fact teachers of language, communication and culture, or at least cultural awareness. With 'culture' not only religions or habits in different geographical regions are addressed, in fact it can also refer to corporate culture or the culture of a specific profession. ESP teachers might not be, and very often are not, familiar with the respective culture before they start teaching a course, but it is their responsibility to get acquainted with it, to do research on it, call the students' attention to the idiosyncrasies of their discourse community and to raise their awareness for certain processes and principles they might not have detected and realized themselves.

Considering the fact that each learner group has different requirements, expectations and goals, it becomes obvious that an ESP teacher, then, can very seldom make use of ready-made material but regularly has to do research and tailor-make materials on the basis of the results.

As ESP considers the learner as someone who needs to pursue either academic, professional or occupational goals and objectives, the ESP teacher has to ensure that the curriculum is in accordance with these various objectives. They have often to deal with the fact that they have no experiences whatsoever in the field of expertise or the subject they are supposed to teach. Consequently, ESP teachers sometimes might be faced with the fear that they may not cope with their students' specialist area. Non-native speaking teachers of English, due to the critical attitude of some learners', might additionally have doubts about their competence in the language. However, Tudor (1997: 91, quoted in Basturkmen 2010: 8) argues that ESP is sometimes concerned with "domains of knowledge which the average educated native speaker could not reasonably be expected to be familiar with". Accordingly, not being a native speaker should not irritate teachers, because teaching an ESP course usually is an encounter of specialists of two fields; on the one hand, the ESP teacher, who is the language specialist - no

matter whether native or non-native - and, on the other hand, the students who are experts in their respective area. Thus, ESP teachers are challenged to put much effort into preparation work, to become familiar with the field of expertise and they should be prepared to deal with the fact that there is no ready-made teaching material available on many fields of expertise is part of the job.

If a teacher is hired to teach a course in the field of social security, for example, it is not necessarily a prerequisite for them to be an expert of the social security system. However, it is expected that they acquire some basic knowledge to be able to engage students in meaningful discussions about and activities on their day-to-day work. They need to design a curriculum that covers important aspects of social security, social security terminology, patterns of communication that is typical for the students' working lives and language that is typically used within this discourse community. This is naturally not possible without doing any research beforehand, and during ESP courses, teachers often become learners, as a constant exchange of knowledge happens between the trainers and their students.

Hunston (2002, quoted in Gavioli, 2005: 14) argues that "in ESP […] language teachers […] are often not experts in the discipline in question and need to find a 'linguistic path' to gain access to both language conventions and concepts that are entailed in the specialized discipline and construct its own meanings." Finding this linguistic path constitutes one of the 'true' challenges for the ESP teacher.

Robinson (1991: 1) mentions another very important aspect of ESP when she refers to ESP as an 'enterprise' that involves, apart from language and pedagogy, the learners' specialist areas of interest. She recognizes the manifold roles of ESP teachers as researchers, course designers, material writers, testers, evaluators as well as classroom-teachers. Due to this variety of responsibilities an ESP teacher has, she prefers the term 'ESP practitioner'.

No matter how you look at the role of an ESP teacher, to get optimum results in a course it is inevitable to consider the learners' needs, which are defined with the help of a 'needs analysis', research the specialist area in general,

but also the specific discourse by means of a 'language analysis'. Based on the results of those analyses, the syllabus and appropriate teaching materials can be designed.

The 'needs analysis' as well as the 'language analysis' will be discussed in the chapter on syllabus- and curriculum design, another important point in preparation, the 'genre analysis', is discussed briefly in the following.

2.4.1 Genre Analysis

There are different definitions of the term 'genre' to be found in the literature, sometimes it is used synonymously to 'text type', sometimes it is used to talk about a whole discipline or domain (Robinson 1991: 25). Here it is used in the sense of sets of typical language features which are used in certain domains.

Genre analysis is an important part of ESP, as it strives to find out about "particular conventions for language use in certain domains of professional and occupational activity" (Widdowson 1998: 9). It deals not only with text but also with discourse, which means that it does not only analyze e.g. the "English of Social Security", but also the "conventions of thought and communication which define [this] area of professional activity, and how, incidentally these are given expression, or textualised, in English." So, according to Widdowson the requirements of the profession regulate a language, and the language changes accordingly should requirements or the culture change[1]. Consequently, if professions and occupations and their conventions change, also the purposes of language will.

The challenge for the ESP teacher regarding these changes in conventions and subsequently in language are self-evident. One has to constantly be aware of them and revise curricula and syllabi, respectively, if necessary.

Another important issue is that professional cultures, or as Widdowson calls them "secondary cultures", and generic conventions not only change over time, but they also differ at any one time. One cannot expect that the conventions for

[1] Following Widdowson in this context, it would not be correct to call it "English of Social Security", instead it should be "English for Social Security"

writing certain kinds of documents or negotiating are the same all around the world. They are strongly influenced by the 'primary cultures', the cultures of specific countries or regions. This is why ESP teachers also need to be aware of what culture their students are from, or who their students have to deal with in the course of their working lives. Dealing with partners in the United States might ask for different communicational approaches than dealing with Japanese people in the same field. Robinson (1991: 1) observes that "[w]hat is specific in and appropriate in one part of the globe may well not be elsewhere."

This is why generalizations are quite questionable; curricula need to be tailor-made for each and every group of learners, and English for a specified field of business should rather be understood as an approximation than the one and only truth.

2.4.2 The ESP teacher and authentic texts

As mentioned at the beginning of this sub-chapter, a native-speaker is not necessarily a good language teacher and a good language teacher does not necessarily have to be a native speaker. Nevertheless, there is one aspect of being a native speaker that cannot be neglected. Native speakers can use their intuition to tell students what is possible and sounds natural and what does not. Non-native speakers can undoubtedly also use their intuition, but they might be misled by interference from their mother tongue, which is unquestionably the reason why many ESP teachers always strive to find authentic texts to use in class.

The importance of material being authentic is always stressed in ESP teacher trainings, and the idea of 'inventing' texts that can be used for practice in class is almost demonized. But how important is it really to find authentic material and how does one decide whether a text is authentic or not? Widdowson (1998: 10) brings up a very interesting perspective when he argues that what is real or authentic to users is not necessarily authentic to learners, because they might not (yet) be part of the community. Therefore, authenticity as a matter of principle is not always necessary and does not always make sense. He further argues that consequently "the pedagogy of ESP […] requires us to devise different textual

versions of generic structure, beginning with realizations in simple language and gradually complexifying the direction of genuine, actually occurring language use." Widdowson addresses a quite crucial point here, namely that available authentic materials are more often than not unsuitable for language teaching, because they are too difficult. It is the responsibility of the teacher to virtually cut the material into bits and pieces, adapt it to the students' level of English and slowly lead them to the point where they - hopefully - are able to deal with the complete text or text type. He also advises ESP teachers not to uncritically believe in authenticity but to, additionally, show creativity in the creation of classroom conditions, because specificity is as much a matter of the process as the purpose of learning.

In this respect, Widdowson (1978: 79-80; quoted in Gavioli 2005: 19) also suggests a distinction between 'genuine language' and 'authentic language', the first one being instances of language that was actually produced, the latter language that was produced in a communicative context.

The question remains what is regarded as authentic, and it can probably be broken down to saying that even texts that are not left in their original form, but have not been explicitly produced for classroom use but only edited, can be regarded as authentic in a broad sense.

2.4.3 Responsibility of an ESP teacher

A very interesting and important aspect is the fact that belonging to an expert group, who uses the 'mastery' language, means having power and being exclusive. Thus, ESP teachers have the responsibility of making their students, who are expert in a specific field, aware of that power they have over the people who are not members of the community and the rise of power that correlates with the rise of expertise, because "[…] ESP is concerned with sustaining the discourse communities that are in control and in power, and with supporting their purposes, it is obviously more implicated in moral and political issues than is general English teaching" (Widdowson 1998: 11). Becoming more communicatively effective, means becoming more exclusive, and teachers have to some extent to advise their students on the cautious and respectful use of their power and exclusivity.

2.4.4 Students' motivation

There are two sides to offering or teaching highly specialized courses for teachers as well as for learners. On the one hand, students attending ESP courses very often need English for their work and are motivated to acquire new skills in the foreign language, which means that ESP teachers are in a much better position to maintain motivation than their colleagues in schools and vocational schools or even at university. Learners are eager and ambitious, because they can see the immediate relevance of what they learn for their jobs. On the other hand, as Basturkmen (2010: 57-58) mentions, there are also disadvantages to those specific courses, because even if we expect students to find them motivating, learners are not always as focused as teachers think they would be. This might be caused by possible plans of students to change their jobs or hopes for promotion in position in the not so far future, which makes ESP courses seem almost useless and a waste of time and money. Language that is taught in such courses is, obviously, very restricted, and the situations and circumstances this language could be used outside the classroom, is rather limited.

Regarding motivation, the ESP teacher is always challenged to keep it as high as possible by bridging the gap between specialist and general language so that students also acquire skills they can use in any other discourse community as well.

2.4.5 The ESP teacher and an unfamiliar field of expertise and work

ESP courses are needed in numerous lines of businesses and industries or professions. Offered courses range from English for all kinds of employees in a certain area, such as finance, medicine, law, technical production or insurance, to English for a certain kind of profession or occupation such as lawyers, insurance brokers, real estate agents or customer care agents, to name a few. Due to this diversity, the challenges teachers have to face, even if a thorough needs analysis is available, are numerous.

The first thing, which has already been discussed above, is the fact that having good English skills and being a good teacher are not sufficient for being an

ESP teacher, and that it is necessary to bridge the gap between the students' and teacher's expert knowledge. This can be done in many ways, such as asking the company for material that they consider important, reading, or doing careful research on the internet. The next step would be to conduct some language analysis. If one is not familiar with a certain industry, then it is highly likely that one does also not know about all the different genres and the specialist vocabulary that is used. Consequently, step two would be to determine what different text types the potential students are dealing with and what technical terms are used repeatedly. Doing research and, subsequently, designing a curriculum and materials, do not make the English teachers experts in the respective field, they still are lay-people and rely to a great extent on the good interaction between themselves and their students. However, considering the courses as 'joint-ventures' of specialists at eye-level, with their own field of expertise, who have the same goal in mind, might reduce their insecurity.

The English specialist will always be responsible for the organization of the course, can help with language issues such as grammar or lexis, provides appropriate exercises for practice and gives tips on learning strategies. The other 'expert group' is responsible for giving information on situations they have to deal with in their working lives or their industry in general. Sometimes they might even provide the teacher with specialist words, as many companies have their own terminology, which cannot be found anywhere, even if research is done very thoroughly.

For inexperienced ESP teachers, the fact that they are not 'superior' in each and every respect can constitute a problem and they have to get accustomed to the idea that leading an ESP course is not a typical teacher-pupil relation such as in school and they should consider the courses as an opportunity to learn themselves.

Summarizing, it can be said that the challenges ESP teachers have to face are numerous and manifold. They are English teachers, linguists, researchers, analysts, course-planners and business people. Additionally, they need to be perceived as teachers of the English culture, which can refer to the culture of a

specific geographical region (which makes it even more complicated as there is not an "English culture" per se), but - especially in the ESP context - refers to the culture of a specific discourse community.

2.5 Curricula, syllabi and teaching material

Most ESP teachers are either self-employed or work as free lancers for language schools and institutes. Sometimes, a syllabus might be prepared by the school, but these syllabi are very often tentative, because the people responsible for customer acquisition are hardly ever language specialists. However, a 'needs analysis' is often provided for the teacher. If the ESP teacher is self-employed, they are responsible for all steps of course preparation themselves.

A syllabus or curriculum is a 'course-plan' which lists the topics to be covered within the given timeframe and might also include planned activities. It functions as a guideline for the teacher, but it can also be quite useful to introduce the course participants to the syllabus and give them an idea what will be done during the course.

Before a curriculum can be created, some preparatory work has to be done that helps guarantee a sensible and effective course design.

Filtering out the participants' needs and identifying the specific language and skills the group of language learners need is obviously crucial before a course-plan can be developed; therefore, it is important to conduct a 'needs analysis'. Only through this can the teacher ensure that the curriculum is in accordance with the academic, professional or occupational goals and objectives that the learners strive to pursue.

2.5.1 Needs Analysis

According to Basturkmen (2010: 19), the following aspects need to be studied and following questions need to be answered in the needs analysis process:

1. Target situation:

   What tasks, activities and skills are and/or will the learners be using English

for?

What should the learners know and be able to do?

2. Discourse:

   What language is used in the target situation?

3. Present situation:

   What do the learners know and not know?

   What can the learners do and not do in relation to the demands of the target situation?

4. Learner factor:

   What is the learners' motivation?

   How do they perceive their needs?

   How do they learn?

5. Teaching context:

   In which environment will the course take place and what factors can influence it?

   What can the ESP course and teacher realistically offer?

Hutchinson and Waters (1987: 55-58) suggest a subdivision of target needs into three aspects, namely the 'necessities' (what do the learners need), 'lacks' (what do the learners lack) and 'wants' (what would the learners additionally like to learn). It is vital to know what the learners already know and find out what necessities the learners lack. The proficiency aimed at needs to be matched against the proficiency existing to get an idea of what has to be covered during a course. This gap is what they call the 'learners' lacks'. They also strongly recommend including the 'wants' of the prospective students into the planning process, as students often perceive what they need totally differently from the other parties that are involved, such as sponsors, course designers or teachers. This conflicting view of needs should be taken into consideration and compensated accordingly by creating appropriate material and using the right methodologies to keep motivation amongst learners as high as possible. Like Basturkmen, Hutchinson and Waters also stress the importance of investigating

the ways people learn to do what they do with language and include it in the needs analysis.

A needs analysis is essential to establish the 'what' and the 'how' of a course. It consists of a present situation analysis, which makes it clear where students stand at the moment regarding their language proficiency, and a target situation analysis, which determines the activities the learners have to carry out in English. It helps also to determine how many lessons will be necessary to get from the present situation to the target situation or what has to be prioritized if there is only a limited number of lessons budgeted.

2.5.2 Language Analysis

The second step towards effective course design is the linguistic analysis of the specialist discourse.

As participants of ESP courses need to learn how to communicate effectively in their special work- or study area, it is essential that research is performed on how language is used in this business. Basturkmen (2010: 36) mentions that "[l]anguage content of ESP courses is pivotal in ESP course design. Many courses are strongly focused on language content." As it is a central premise in ESP to facilitate the students with very specialized communication skills, it is of utmost importance to base courses on "detailed, accurate and realistic description of how language is actually used in these areas" Basturkmen (2010: 36).

Conducting one's own empirical research into specialist language is very time consuming and results often do not justify the amount of time and effort spent. Before the ESP teacher starts doing this, they should find out if research on the needed specialist discourse has already been done and if so how the available data might be used. However, language teachers should be acquainted with different approaches of research if they need to create their own material. Basturkmen (2010: 43-47) suggests three methods for this: ethnography, genre analysis and corpus analysis.

2.5.2.1 Ethnography

A very time consuming approach is ethnography. This method is a way of performing a qualitative investigation, where the researcher spends a considerable amount of time in the respective environment to gather information about the context and understand the behavior of the group from the participants' perspectives.

2.5.2.2 Genre Analysis

As has already been discussed in the chapter on ESP teachers' challenges, there are more interpretations of 'genre' one can find.

Here it is used for describing different types of text that are typically used in specific domains and the language conventions of professional and occupational activities within these domains. These types can be categories such as telephone conversations, social e-mails or business letters.

The ESP teacher should collect samples of the different text types which are typical for the industry and culture the course participants are involved in to get a picture of the specific conventions and the discourse used.

2.5.2.3 Corpus Analysis

As the following chapter is on corpus analysis and the empirical part of this thesis is on a corpus study that has been conducted, it will only be described briefly here.

Corpora are collections of texts that are stored and can be accessed electronically.

The obvious advantage of Corpus Analysis is that the techniques allow researchers to concentrate on different specific language features, such as grammatical structures and collocations, or it can simply be used to find out which terms occur most often in the texts. It is a very useful tool to find patterns of language use in a body of data, and the authentic texts may either represent a language variety or a specific field. For specific research it is also possible to create a small corpus whose aim it is not to be representative for a language as a whole but rather focuses on the aims of ESP-oriented investigation.

2.5.3 Vocabulary

Another important part of course preparation is the research of vocabulary. Here, we can distinguish between three levels of vocabulary, the 'specialist vocabulary', the 'semi-technical vocabulary' and lastly the 'non-academic/general vocabulary'. The 'specialist vocabulary' are highly technical terms that are most likely not understood by speakers who do not belong to the discourse community that needs and uses these terms. 'Semi-technical vocabulary' can be found in a number of scientific or technical areas, not only in one.

As important as thorough research of vocabulary is, very often the specialist vocabulary is not what needs to be taught primarily, as the students very often know the specific terms anyway. The problem ESP course participants often have is to properly structure sentences around the specialist vocabulary. Hence, ESP teachers seem to agree widely that making teaching semi-technical and general vocabulary, which is used to structure texts and give them cohesion, should be an integral part of their syllabi.

However, this decision has to be made individually for each course. It has already been mentioned that the students' proficiency in general English is often quite high, which would be an argument for the need of teaching specialized vocabulary rather than general or semi-technical vocabulary. Also, especially in Europe, communicating in English is sometimes just a relatively novel development in the learners' working environment and they are still rather busy acquiring the technical terms.

In the context of the European Union, many (governmental) institutions that used to work only within the respective country, are suddenly faced with the requirement to communicate with their equivalent institutions in other member countries. This challenges the employees of these institutions in many ways. First, they are suddenly required to communicate in a foreign language with people who (often) also have a different mother tongue than the language that is used for communication. Second, they need to communicate about very specialized topics or explain concepts that might have been used in their country for a long time but do not exist in other countries in the same form. All of a sudden, the need for a

translation of 'specialist words' appears, but very often there is no direct translation possible or available.

Even more so must the ESP teacher be familiar with the 'specialist vocabulary' available or sources that provide them. Additionally, they must also know about the vocabulary that is not available and, therefore, must be 'created' or paraphrased.

2.5.4 Course design

Probably there is no universal recipe for designing course curricula, but obviously the basis for ESP course design is taken from ESL course design. It is mainly influenced by the factors already covered in this chapter: students' needs, working environment, existing (or non-existing) material, but also by the methodology the teacher wants to use.

Robinson (1991) suggests three bases for language syllabi: 'content-based', 'skill-based' and 'method-based' syllabi.

There are four different kinds of 'content-based' syllabi Robinson mentions. The first one is the 'language form' syllabus which consists of an ordered set of language items and has been considered unacceptable for some time, but is still rather powerful in ESP practice. The second one is the so called 'notional-functional' syllabus which is based on notions or concepts, such as time or space, or functions such as greeting or asking. The 'situationally organized' syllabus presents situations in a chronologically organized manner, and the 'topic-based' syllabus makes use of the content of the students' work or field of study.

'Skill based' syllabi can have two key aspects: the language skills and the learning skills. While one focuses on the four language skills writing, reading, speaking and listening, the other one is concerned with the cognitive skills. They can be used together or alone.

In the last type of syllabus, the 'method based' syllabus, it can be distinguished between the 'task-based' and the 'process-based' syllabus. The 'process based' syllabus emphasizes the method or process of learning, and contains an important element, the "redefinition of the authority relations in the

classroom" (Robinson 1991: 38). Robinson considers it is a matter of negotiation between students and teacher what actually happens in the classroom. Classes that are planned according to a 'task-based' syllabus are devoted to performing a task instead of consciously concentrating on language, but nevertheless have relevant subject content, on an appropriate and manageable linguistic level and an appropriate performance requirement.

Hutchinson and Waters (1987: 65-77) suggest a distinction between 'language-centered', 'skills-centered' and 'learning-centered' course design.

The 'language centered' approach identifies the target situation, goes on to identifying the linguistic features of a target situation and then a syllabus is created. However, it is quite inflexible as it takes neither the skills people need to cope in a target situation nor the learning situation into account. The 'skills centered' approach additionally considers what processes enable someone to perform, and the 'learner-centered' approach goes even further and is also interested in how someone acquires their competence.

Basturkmen (2010: 52-58) takes a slightly different path with her 'wide-angled' and 'narrow-angled' course design. For her the target group of learners is the center of attention, which determines the content of the course and what it will focus on. The wide-angled courses, then, would be for a wider target group, the narrow-angled course would be developed for a very specific target group.

Which is the best kind of syllabus, or the most effective? Which one should one choose for course preparation? As ESP is always focused on practical outcomes and concentrates on the study or work situation of the students, this is a decision that probably has to be made individually for each course. However, Swan's (1985, quoted in Robinson, 1991: 40) contention that it is not the issue of which syllabus is the best, but rather how you can integrate many syllabi into an effective teaching program to provide variety for the students and boost their motivation, seems to be very realistic.

The next step, after having designed the course program, is to find appropriate materials, adapt them if necessary or create some new material.

2.5.5 Materials

There is much ready-made teaching material available for EGP and EBP but, understandably, the amount of material for highly specific languages is quite limited. Even if one can find course books on law or medicine, they by necessity cover a broad range of subjects, as they are produced to appeal to as many learners and teachers as possible, which makes them more or less unsuitable for many ESP courses. Parts might be useful for the introduction of individual topics in a very general manner, as warm-up exercises for the students and to get them acquainted with the theme. However, the range of the topics covered in the majority of the ESP course books is too wide and the language too general to be of real use to a rather limited (speech) community. Consequently, ESP teachers have to design the greater part of the material themselves to ensure proper service of the learners' needs.

Hunston (2002, quoted in Gavioli 2005: 14) argues that "[…] in ESP 'what' is taught (as compared to 'how' it is taught) is a very crucial problem for language teachers", and ESP practitioners should plan the curriculum with the particular students and their personal preferences in mind.

It has already been mentioned that ESP teachers usually have a strong intention to find authentic material as a way to bring the "real world" into the classroom, and they put an enormous amount of time and effort into it. As important as the use of authentic material is, Widdowson (1998: 13) reminds us that authenticity should not be a matter of principle, and good course design does not only involve "the simple uncritical belief in authenticity", but also "the special and specific contrivance of appropriate conditions in classrooms".

What kind of materials teachers use or develop depends eventually on the students' needs, on the methods and kinds of syllabi that are used and ultimately on what the teachers considers useful. It is probably most effective and motivating if a good mix of carefully selected ready-made and well thought-through self-made materials is used.

In this chapter it has been argued that some groundwork has to be done before designing a course. The importance of the needs analysis, the most characteristic feature of ESP, which is a complex process taking not only into account what the students will have to do in a target situation but ideally also consider the learner's preferences, reasons for learning, the environment, time-frame and frequency of the course as well as the resources that are available, have been stressed.

Before starting to design a course or collect empirical data, teachers should ask themselves what skills, genres and features of a language learners really need to know, and they should investigate whether there is already some data available that might be useful for the course. If there is already data, teachers have to decide if and to what extent it can be used for the course design. If not, they need to collect and investigate data themselves. This can be done using different approaches such as ethnography, genre analysis or corpus analysis, which will be discussed in detail in the following unit.

3 CORPUS LINGUISTICS

The following part will take a look at what corpus linguistics is, it will provide definitions of important terms, briefly comment on the development and introduce different kinds of corpora and their field of application. The question why Corpus Analysis should be used at all will also be tried to answer, and the advantages (with an emphasis on the ESP environment) as well as the disadvantages of the use of Corpus Analysis in linguistic research will be discussed.

3.1 What is Corpus Linguistics?

3.1.1 Corpora and Corpus Linguistics

Modern corpora are, simply put, collections of texts, written and spoken, that are stored and accessed electronically. Some corpora are tagged - coded for word classes -, and others are parsed - completely analyzed in regards to syntax. There

are different types of corpora available, such as general corpora of main varieties (e.g. American or British English), specialized corpora (e.g. for spoken academic language), diachronic corpora, newspaper corpora, literary corpora, learner corpora, translation corpora, 'D.I.Y.' corpora for specific purposes, and last but not least the most extensive and largest corpus: the world wide web.

The term 'Corpus Linguistics' first appeared in the 1980s, but the system had been used long before. Indeed, corpus research goes back to the 1920s, when material was counted and classified manually to get an idea of syntax or semantics, of how words were distributed, and the first corpora were not computer based but could have been collections of written and transcribed texts, stored in cardboard boxes. Even if this would not be considered 'representative' according to the requirements of modern corpus research, it still was authentic data that was useful in many respects.

Fortunately, modern technology has made the collection, storage and subsequent exploitation of material for linguistic purposes possible. McEnery, Xiao and Tono (2006: 4) define the modern corpus as "a collection of sampled texts, written or spoken, in machine readable form which may be annotated with various forms of linguistic information". The concept of sampling, which will be discussed later in the chapter, is very important, and one needs to keep in mind that corpora are never a random collection of texts but texts are always collected systematically to serve a special need or intention. Leech (1992: 116) highlights that

> computer corpora are rarely haphazard collections of textual material. They are generally assembled with particular purposes in mind, and are often assembled to be (informally speaking) *representative* of some language or text type.

A corpus can be defined in different ways, but the term should always be considered as something quite vague. Nevertheless, researchers seem to agree on a small number of qualities corpora have or should have: they can be defined as collections of texts which are, 'machine readable', 'authentic' (including transcripts of spoken data) and 'sampled' so they can be considered 'representative' of a particular language or language variety. Some linguists suggest that corpora should be defined in linguistic terms such as the distribution of words or other

linguistic patterns, whereas Mc Enery, Xiao and Tono (2006: 5) propose that also non-linguistic parameters, which will be discussed later in this unit, should be used for the definition. They suggest that we should consider corpus linguistics "as a methodology with a wide range of application across many areas and theories of linguistics". (2006: 8)

3.1.2 Corpus Analysis

Corpus Analysis is research based on a corpus, and it

> looks at how words frequently combine in actual language performance; such combinations can be of a lexical, grammatical and textual type and are linked to conventions depending on registers, genres or text-types. (Gavioli 2005: 35)

Of course one needs to base corpus analysis on theoretical premises, but "performing actual corpus work helps enrich our knowledge of both such theoretical premises and of new theoretical implications" (ibid: 35).

Corpus analysis has proved to be a very useful tool for various kinds of studies, such as translation studies or comparative studies, but also for teachers who need to create materials and prefer to make use of existing texts. Whatever one tries to achieve with the help of this kind of research, there are many different kinds of corpora to choose from, which will be introduced in the following.

3.2 Different kinds of corpora

For conducting Corpus Analysis, different ready-made corpora with different features are available. Hunston (2002: 14-16) provides a good overview, and she distinguishes, amongst others, between general- or reference-, specialized-, comparable-, parallel-, historical- or diachronic- and monitor corpora. She argues that corpora are always designed for a special purpose, and that the decision on which of the types is used for research will to a large extent depend on the research question(s).

3.2.1 General- or reference corpora

General corpora comprise texts of many different genres of a main variety, including written and/or spoken texts on a wide variety of topics. They are usually very large and "often used as a baseline in comparison with more specialized corpora" (Hunston 2002: 15).

Well-known general corpora are the British National Corpus (BNC) which contains approximately 100 million words and covers a period from the 1970s to 1993; the much larger Corpus of Contemporary American English (COCA) comprises approximately 450 million words and covers a time span from 1990 to the present. The Corpus of Historical American English (COHA) is a collection of approximately 400 million words from the 1810s to the 2000s, giving the same text categories as the COCA. The categories of the texts entered into these three corpora are divided into: spoken, fiction, magazine, newspaper and academic. All three corpora can be accessed through the internet on http://corpus.byu.edu/.

3.2.2 Specialized corpora

Specialized corpora are representative of a particular type of text and they contain a particular kind of language. "They are specialized relative to the general corpora" (McEnery, Xiao and Tono 2006: 60) and differ from the general corpora in so far as that there are limits as to the kinds of texts included. Texts could for example be limited according to the time period or the social setting in which they were produced, or to the topic they refer to.

Examples of specialized corpora are the Corpus of Academic Spoken English (MICASE)[2], a corpus of about 1.7 million words, transcribed from nearly 200 hours of recording or the Cambridge and Nottingham Corpus of Discourse in English (CANCODE)[3], a 5 million word corpus of informal British English collected from 1995 to 2000.

[2] access on: http://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple

[3] for more information go to: http://universal.elra.info/product_info.php?cPath=42_43&products_id=1871

3.2.3 Comparable corpora

We refer to comparable corpora if there are two or more corpora in different languages or different varieties of a language, which use the same sampling techniques. The sampling frame is essential, and the proportion, genre, domain and sampling period of the texts should be the same. According to McEnery, Xiao and Tono (2006: 49), they are clearly useful for extraction of terminology, but also a useful resource for contrastive- and translation studies, especially when they are used together with parallel corpora.

3.2.4 Parallel corpora

The term parallel corpus is sometimes used interchangeably with comparable corpus, depending on the researchers. Here, parallel corpora are introduced as 'translation corpora', where either two or more corpora in different languages contain texts which have been translated to the other language(s) or the texts have been produced parallel in different languages. Like the comparable corpora, they are quite useful to investigate differences and similarities between languages.

3.2.5 Historical or diachronic corpora

This type of corpus is a collection of texts from different time periods, which is used to trace the changes in and development of language. The probably best-known corpus of this type is the Helsinki Corpus[4] that comprises approximately 1.5 million words and covers texts from 700 to 1700.

3.2.6 Monitor corpora

In contrast to so-called 'sample corpora', which are set up once and stay constant in size, 'monitor corpora' are regularly updated and new texts are added. However, the proportion of the texts which are entered need to remain constant.

[4] for more information go to http://icame.uib.no/hc/

Similarly to the diachronic corpora they are set up to find changes in language, but they usually do not cover such long time spans.

3.2.7 D.I.Y. corpora

Apart from the numerous ready-made corpora introduces above, another important kind of corpus is the 'DIY corpus' which is usually specialized and much smaller than general corpora. They have the advantage that they can be created according to one's very specific needs. However, even if created just for the individual use, one should try to maintain certain standards. In the following the important factors for setting up a corpus will be discussed.

3.3 How to set up a corpus

This unit will try to concentrate on the factors one needs to consider when setting up a corpus according to linguistic standards and in such a way that the research question(s) can be answered.

As already mentioned at the beginning of the chapter, corpora should contain language that is representative for a specific and certain kind of language or variety thereof. To guarantee this representativeness it is crucial to give consideration to balancing and sampling the texts that should be included. However, first of all it is important to find out what representativeness of the texts in a corpus actually means.

3.3.1 Representativeness

According to McEnery, Xiao and Tono (2006: 13), a corpus is "essentially a *sample* of language or language variety (i.e. population)". The broader the range of variability in a population the more representative the corpus is. Representativeness is achieved if what holds for the sample can also be said for the general population. It is widely influenced by the range of genres that are included in a corpus and the selection process for each piece of text. However, the careful choice of text types is not the only factor that contributes to the representativeness of a corpus. Hunston (2002: 30) claims that corpora should be

updated on a regular basis, otherwise they might rapidly become unrepresentative.

Depending on whether a researcher represents a dynamic or static view, they will make use of either 'monitor corpora' or 'sample corpora', respectively. 'Dynamic corpora' can track changes in language quite well, whereas 'monitor corpora' are useful for presenting a status-quo in language.

In connection with representativeness it should also be mentioned that the selection of texts for a general corpus is different to the same process for a specialized corpus. A general corpus should understandably comprise as many domains and genres as possible, as it is meant to represent a language as a whole (such as the COCA or the BNC). Specialized corpora tend to represent a specific domain, such as biology or psychology, or genre, such as newspaper articles or abstracts, and also texts of a domain which is rather restricted need to be balanced within this same domain.

3.3.2 Balance

As it became evident above, balance is a factor which influences representativeness to a great extent. According to McEnery, Xiao and Tono (2006: 16),

> [a] balanced corpus usually covers a wide range of text categories which are supposed to be representative of the language or language variety under consideration. These text categories are typically sampled proportionally […].

Nevertheless, they also emphasize the fact that there is no reliable measure for corpus balance and that the process of balancing is heavily influenced by intuition and best estimates. As corpora that reach substantial size, such as the COCA, will in effect balance themselves, balancing has become less important to builders of monitor corpora. Building a sample corpus, in contrast, involves more careful selection of a wide range of text types.

3.3.3 Sampling

What should be achieved with sampling is that one sample represents the characteristics of a population as closely as possible. Samples are drawn from sampling units, such as books, periodicals or newspapers within the boundaries of the population or sampling frame.

McEnery, Xiao and Tono (2006: 20) mention two sampling methods. The first is the 'simple random sampling', where "all sample units within the sampling frames are numbered and the sample is chosen by use of a table of random numbers". The disadvantage of this method is that rare items might be disregarded even though they might be useful and interesting to researchers. To avoid this phenomenon, they suggest a second method, the 'stratified random sampling', where the population is first divided into relatively homogenous groups (strata) and each of the strata is then sampled randomly.

Another important aspect of sampling is the decision that has to be made on the size of the samples. McEnery, Xiao and Tono (2006: 20) remind us that, although for linguistic purposes it would be extremely use- and helpful to use full texts, this could lead to copyright problems and they advise researchers to sample text segments.

Depending on purpose and research question the decision on sampling method, proportion and number of samples for each text category has to be made individually.

3.3.4 Corpus mark-up

To change raw corpora into useable tools for linguistic research, data has to undergo some processes. Firstly, as discussed above, data has to be made machine readable, secondly it has to be coded electronically as to provide data about the data, so called meta-data. This coding is what is called 'corpus mark-up'. Mark-up helps to "restore the context and to enable us to relate the specimen to its original habitat" (McEnery, Xiao and Tono: 2006: 22), it enables the researcher to retrieve textual information, which in turn "adds value to a corpus and allows for a broader range of research questions to be addressed" (ibid: 23).

It also functions as placeholder for omitted text parts, insertion for editorial comments or is used for denoting quotations in foreign languages. It provides information on the components of a corpus as well as on the textual structure of each text, which is objectively verifiable.

McEnery, Xiao and Tono (2006: 23-24) introduce several available schemes for corpus mark-up, such as COCOA, DCMI, OLAC, ISLE or TEI[5] , which vary in complexity. Some of them provide only information on authors, titles and dates, whereas others offer detailed language descriptions.

Small specialized corpora are normally not marked-up, and if a corpus is marked-up the choice of the scheme depends on the research question and the level of details that is required to answer it.

3.3.5 Corpus annotation

In order to extract linguistic information from a corpus, it is necessary to encode some linguistic information in the corpus first. Corpus annotation is closely related to corpus mark-up and similarly enables researchers to extend the scope of research questions that can be answered with the help of the respective corpus.

However, corpus annotation adds linguistic information to the corpus. It can either refer to the encoding of textual and contextual information or in a narrow sense simply to the linguistic encoding such as the part-of-speech, or POS, tagging and syntactic parsing, of a text. 'Tagging' means to code each word of a corpus to indicate the part of speech. 'Parsing' is a process that analyzes sentences in a corpus into their components.

McEnery, Xiao and Tono (2006: 30) emphasize the value that is added to a corpus through annotation. A salient advantage is that annotation makes it much easier to retrieve relevant information, as POS tagging helps avoiding the extraction of unnecessary information and also accelerates the process of retrieval. For example, *back* can be used as an adverb, adjective, verb or noun. Searching for the verb *back* in a raw corpus can, therefore, be rather tiresome, as

[5] Mark-up schemes will not be discussed in detail here. For further information see: McEnery, Xiao and Tono 2006: 22-28

all instances of *back* will be listed, due to the same orthographic form. This means, the search will result in a list of *back* as a noun, an adjective, an adverb and a verb. In a POS tagged corpus, the researcher can specifically search for all instances of *back* as a verb, and even in all forms such as gerunds or in different tenses. Other advantages that McEnery, Xiao and Tono mention are reusability and multi-functionality; however, it can be argued that also raw corpora can be reused and are multifunctional.

There are some points of criticism to be found in literature, such as the creation of cluttered corpora through annotation and the contention that annotation might impose linguistic analysis on a user. Other points mentioned are the difficult accessibility, updateability and expandability (Hunston 2002: 92-93) due to 'overvaluation' of a corpus and lastly, the fact that annotation is not always accurate and consistent.

Even though annotation is a very time-consuming, difficult and, consequently, expensive process, advantages of annotation - especially of large corpora - seem to outweigh disadvantages as annotation makes research and analysis much easier.

As with corpus mark-up, there are different types of corpus annotation. POS tagging is the most widespread form of annotation, but there are phonetic/phonemic annotation, prosodic annotation, morphological annotation, lemmatization, semantic annotation or parsing, to name just a few[6]. All forms of annotation can be achieved either fully automatically, semi-automatically or entirely manually.

3.3.6 Corpus Size

As many other factors of research do, corpus size also depends on research purposes and contents. Specialized corpora, which are set up for the purpose of designing a sample of specialized language such as law, medicine, or philosophy, will naturally not be as large as general corpora, which are designed to give a sample of a whole language or one of its regional varieties.

[6] Details on corpus annotation are to be found in McEnery, Xiao and Tono (2006: 29-45)

3.4 What can a researcher find out with the help of a corpus?

As Hunston (2002: 3) correctly states: "a corpus by itself can do nothing at all, being nothing than a store of used language". However, there is software available that helps us to rearrange this stored information, which, consequently, enables the researcher to get a new perspective, make observations and draw conclusions. A corpus does not offer any new information, but it can present the existing information in numerous different ways. This flexibility allows exploitation of a sometimes limited number of texts to answer various research questions. According to Hunston (2002: 3), the favored software programs help exploring corpora in regards to three important aspects, namely frequency, phraseology and collocation, which will be discussed below.

3.4.1 Frequency

A corpus can readily provide information on frequency of all the words of all the texts stored in the corpus. Frequency lists can be produced in ascending- or descending order, meaning the most frequently used word is shown in the first position, the second most frequently used word in the second position, etc, down to the end of the list, to the least frequently used word. Such frequency lists enable us to get information on quantity at one glance.

The importance of the frequency lists lies in the assumption that words which are used more frequently than others are probably more important, too. For teachers that can be of enormous help, as the results might influence the decision of which terms are worth teaching and which are not.

3.4.2 Phraseology / Concordance

Most researchers use so-called 'concordancing tools' to access a corpus, which list many instances of the use of a word or phrase in context, "allowing the user to observe regularities in use that tend to remain unobserved when the same words or phrases are met in their normal contexts" (Hunston 2002:9). Concordances are simply lists of occurrences of a word or phrase in a corpus. The 'key word in context' (KWIC) format lists recurrent combinations of a 'keyword', the word one

has searched for, positioned in the middle and 40-50 characters of context around it. Having all instances of a word on one page allows the observation of regularities that might never be noticed if one reads whole texts, where words or phrases reappear at irregular intervals. Carefully investigating these concordance lists might give some indication of how certain words are used.

3.4.3 Collocation

The statistical tendency of words to co-occur is referred to as 'collocation'. Collocation lists give similar information to concordance lines, with the difference that collocation lists deliver single words (collocates) with the number of occurrences, as opposed to parts of sentences.

3.5 Advantages and disadvantages of using corpora

3.5.1 Advantages

So why and what for should one use computers in linguistic research? The many advantages of modern computerized corpora are apparent. First and foremost, computerized texts are machine readable and therefore much faster in regards to processing and manipulation, which, as a consequence, makes linguistic research at minimal cost possible. Furthermore, the use of computers can guarantee accurate und consistent processing of material as well as exclusion of human bias, which leads to much more reliable and objective results than material that has been analyzed by humans. Moreover, corpora offer authentic language, the examples are contextualized and they give us a good indication of frequency of occurrence in general or in different text types. Another obvious advantage of corpora is that they can be used in many different ways and there is the possibility of further automatic processing. If one does not need a D.I.Y. corpus, the opportunity of accessing different corpora freely on the internet is also a benefit. Many corpora, such as the COCA or the BNC are available for everyone.

3.5.2 Special advantages for the ESP teacher

Corpus analysis in an ESP setting seem to match the teachers' and students' requirements quite well as it highlights recurrent features of language. Teachers of ESP classes who, as already emphasized in the preceding part of the paper, are "often split between the need to be both an expert in the (foreign) language and an expert in the specialized discipline" (Gavioli, 2005: 5) might find corpora of specialized texts an especially "useful instrument in isolating and providing indications about key lexical, grammatical or textual issues to deal with in ESP classes" (ibid: 5). Analyzing the texts in a corpus with concordancing software can suggest relevant lexico-grammatical or textual items that are worth dealing with in the ESP class and the way they are used.

In regards to frequency, it seems plausible that words that are used more frequently than others are probably more important, too, which is helpful in making the decision of which terms one should presented in class, and which ones might not be as important. McEnery, Xiao and Tono (2002: 98) suggest that syllabi should be created based on empirical data and frequency of usage should be made a priority in teaching. Also Kennedy (1992: 335) argues that

> corpus linguistics has held potential relevance for the teaching of languages because responsible language teaching involves selecting what is worth giving attention to.

Gavioli (2005) emphasizes the importance of frequency lists from a slightly different point of view. She sees another advantage of corpora of specialized texts, namely that they enable us to very quickly generate items that may characterize the texts in question. These items, that are mainly responsible for the character of the texts in question, should undoubtedly be included in ESP syllabi. She also highlights that concordances which have been generated from any corpus can very much support a teacher's intuition and additional integration of descriptions from dictionaries and grammars to provide more satisfactory explanations would nicely complete syllabus design and, subsequently, the teaching process (2005: 25).

Especially D.I.Y. or 'home-made corpora', as Gavioli (2005) calls them, can be very helpful to teachers in two main respects: Firstly, they are an invaluable indicator for selecting items to be included in the course syllabus; secondly, they provide teachers with items not adequately dealt with in traditional, ready-made teaching materials. Even though a teacher cannot limit syllabus design on corpus material, it may be a very interesting and practical resource for complementing the syllabus of an ESP course.

As (ESP) teachers are challenged to produce exercises that use the language in the respective specialist environment, and corpora provide many different examples, creating material is much easier with this tool. Instead of inventing sentences and situations, one can access the corpus and search for specific words or constructions and can retrieve information about the context they were used in. In this connection, Sinclair (1990 Collins Cobuild English Dictionary, quoted in Gavioli 2005: 21) states that "[…] real examples have a communicative value that the invented one lacks". So it seems that corpus-based material offers a kind of 'communicative interaction', which invented material does not.

Another salient advantage has to be mentioned in connection with non-native ESP teachers. As already discussed in the first part of the paper, they naturally lack native-speaker intuition, and corpus material might help them to verify or falsify their intuition used for material production.

3.5.3 Disadvantages of Corpus Analysis

As many advantages corpus analysis might have, one must also see the disadvantages and limitations of corpora. Clearly, results of this kind of research, especially of analysis of specialized corpora, only represent one certain aspect of language or present a snapshot in time, and they merely give information on the status quo. It also has to be mentioned that, obviously, not all kinds of research questions can be answered with the help of Corpus Analysis.

In context with material designing, Widdowson (2000, quoted in McEnery, Xiao and Tono 2005: 132 pp) suggests that to make the material based on corpus

research effective for the learners, the texts have to be 'recontextualized' to make them real for them. He states that

> [t]he textual findings of frequencies and co-occurrences have to be contextually reconstituted in the classroom for their reality to be realized, and this reconstruction must obviously be based on very different contextual conditions than those which activated the texts in the first place.

Widdowson might have a point here, but if one considers the efforts teachers have to make when creating material, using a corpus most likely makes their lives much easier.

Another disadvantage of corpora is that they cannot provide negative evidence (McEnery, Xiao and Tono 2006), which means that they cannot tell us what is actually possible or not possible in language. What they can tell us is what has been used and what is typically used. However, whatever findings they provide, they are only valid for the specific corpus or corpora, and explanations have to be produced using other methods.

Cook (1998: 61, quoted in Mc Enery, Xiao and Tono 2002: 101), criticizes the preference of what is "simply frequent" to other, possibly more effective or salient, expressions.

A further downside is that the setting up of a corpus needs a lot of preparatory work and, thus, time and expertise. If teachers want to create a specialized corpus for a course, they should ask themselves beforehand whether the results are worth the effort and time spent. In comparison, the intuitive approach is an instant solution because intuition is readily available, but intuition can be influenced by dialects, sociolects or the mother tongue, and what seems acceptable and correct to one person might appear utterly inacceptable and incorrect to another.

Even if some disadvantages have been mentioned above, it is assumed that corpus analysis is a very time-saving and practical method to answer a specific kind of research question and that there is no other method available that comes even close in terms of efficiency.

3.6 Summary

In this part, Corpus Analysis was introduced, explanations of some basic terms were provided and a number of different corpora, such as specialized and general corpora, were shortly defined.

Then, three important concepts of corpus linguistics were discussed: representativeness, balancing and sampling. A corpus is regarded representative, when it mirrors the language or language variety it is supposed to represent. Representativeness can be achieved through thorough balancing and sampling of the texts included in the corpus. However, one should not forget that there are no adequate measures for representativeness, and researchers must make use of their intuition as well. The helpful techniques of corpus mark-up and corpus annotation were also described, and their helpfulness for the researcher was highlighted.

The last part examined the advantages and disadvantages of corpus analysis, and summarizing it can be said that corpus analysis seems to be a methodology that proves pricelessly helpful for language studies but also for teaching because corpora can reveal "facts about language behavior which are not […] accessible to intuition" (Widdowson 2000, quoted in McEnery, Xioa and Tono 2005: 132). Depending on the research question, different kind of ready-made corpora are available, but sometimes it might be necessary to set up a new specialized corpus with either specialist texts or texts that were produced in a specific time period or area. Independently of the kind of corpus that is necessary to answer the research questions, it has to be ensured that texts are balanced and sampled accordingly to guarantee representativeness.

Even though one might find some difficulties in the work with corpora, the advantages seem to clearly outweigh the disadvantages. In very little time it is possible to get much information on the frequency of occurrences of words, on regular patterns, textual facts, concordances and collocational co-occurrence. A reality of language usage is revealed "about language usage which was hitherto not evident to its users" (Widdowson 2000, quoted in McEnery, Xioa and Tono 2005: 132).

The invaluable merit of a corpus is that it is a collection of texts that have already been produced and, therefore, have already been regarded acceptable and possible.

4 EMPIRICAL STUDY

Sinclair (2004: 2) states that "corpora, large and small, are seen by many teachers as useful tools, and are being put to use more and more every day." He continues with emphasizing the fairly easy access of corpora on standard computers and describes corpora as "part of the pedagogical landscape." However, Kaltenböck and Mehlmann-Larcher's (2005: 65) contention that "[t]he application of computer corpora for language teaching is still a rather neglected area" seems to be closer to reality. Sinclair might be right with his statement in regards to courses in linguistics, translation- or language studies, but this only concerns a very specialized target group of learners. Considering course design for English classes in secondary education, ESP, English for General Business Purposes (EGBP) or EBP, corpora are definitely not part of the pedagogical landscape, on the contrary, they are hardly used by teachers as many of them are not even aware that corpora exist. The scarce use of corpora is also expressed by Kennedy (1991: 365) when she says that "contemporary language teaching has not always been able to take advantage of this […]".

The purpose of this empirical study is to test the usefulness of corpora for material- and course design in ESP and to find out whether there are any major disadvantages of this method.

This part of the thesis will describe the study that was conducted to find out whether and in how far, corpora are an appropriate tool to create teaching materials for ESP courses. It will, first of all, explain how the idea materialized and why this kind of research was considered important and worth conducting. Then, it will go on with a short description of the fields of work and responsibilities of the students who would work with the materials created on the basis of the research results. The next part will deal with the problems that were expected to emerge;

this will be followed by a detailed description of the study set-up and the texts used for the study. Then, an overview of AntConc, the software used for research, and its different tools for data retrieval will be presented. The results that were obtained as well as an interpretation thereof will be discussed in the next part, which also includes suggestions for grammar topics and terminology to be taught in an ESP course.

4.1 Background

The idea to research 'Social Security Language' was born from a very practical need. I have worked as an ESP trainer in the Austrian Social Security sector, where I have trained employees of the AMS and the Hauptverband in different positions and with manifold responsibilities, for almost ten years. Even though Social Insurance has become an independent business branch in many countries and has become more and more important over the last decades, there are no suitable ready-made teaching materials or any specialized dictionaries available. The AMS provides an internal dictionary, which is available on their intranet, but it is only a glossary that does not give any examples of usage and unfortunately it offers partly wrong translations. The Hauptverband does not even offer such a word list to its employees.

To get more ideas for material design, Corpus Research appeared to be a suitable method.

In the following, an overview of the prospective learners' responsibilities is given which helps determine what topics need to be covered in an English course.

4.1.1 Potential areas where the materials will be used

4.1.1.1 Info Zone /AMS:

Employees in this department are in direct contact with clients and jobseekers and need to give general information on the labor market, vacant positions and professions but also on different possibilities of schooling, education or training.

4.1.1.2 Service Zone/ AMS:

In this department, jobseekers are interviewed about their personal qualifications and career aspirations, but also informed about their opportunities and the general situation of the labor market by the potential course participants. Service Zone employees also need to give information on available benefits or financial support that is available to jobseekers for a short period of time.

4.1.1.3 Counseling Zone / AMS:

This is the department where the jobseekers or unemployed are supervised and guided most intensively. Counselors offer intensive individual help, support their clients in their job-seeking process, and try to offer a customized package of measures that helps them on their way back to work. These measures include financial benefits, but also training programs or professional retraining courses which jobseekers and unemployed can attend.

4.1.1.4 Legal Department / AMS:

The Legal Department is, amongst other areas of responsibility, involved in communication with its European and international counterparts. Especially in the context of the European Union, where the 'freedom of movement for workers' is a fundamental right, which allows workers from one member state of the European Union to work in any other member state, the Legal Department of the AMS is constantly in contact with employment services of different other EU member states. The responsibilities of this department reach from discussing critical cases to giving presentations on the Austrian social system or participating in the process of designing new contracts. Additionaly, they are also contacted by clients or claimants who need help with provisions and laws regarding the Austrian social security system.

4.1.1.5 Department for International Affairs / Hauptverband

This department of the Hauptverband functions as a liaison office and is in permanent contact with other liaison bodies of international and European Social Security Institutions. On the one hand, they communicate with institutions in

countries with which bilateral social insurance contracts already exist, and, on the other hand, they also need to conduct negotiations with prospective agreement partners.

There are three 'teams' within the Department for International Affairs, the 'SRS (Strategy Legal Service) Team', the 'Reimbursement Team' and the 'Organizational Team'. The 'SRS Team' members regularly attend meetings of the European Commission, where they - even though there are German interpreters present – need to be able to speak about social security issues in English with their colleagues of the other attending countries. Other tasks include reading of legal texts, laws or statutes, but also giving legal advice to people who are insured in Austria or would like to move here and need information about the system – either on the telephone or by e-mail. In the 'Reimbursement Team', English is needed for reading meeting minutes or decisions, but also for helping English speaking clients who call to inquire about their rights regarding medical treatment in Austria and reimbursement of the doctor's fees. The English used in the 'Organization Team' is less technical and specific than in the other two teams. It is used for the preparation of international or 'intereuropean' meetings, which includes the writing of invitations, telephone- and e-mail contact with administrative staff in the other countries, set-up and translation of the agenda, preparation of documents and sometimes hotel- and flight bookings.

In all five departments mentioned above, employees either have to inform people who do not speak German, and whose mother tongue is very often also not English, and do not know the Austrian Social System about their rights and responsibilities, or they have to communicate about Social Security with other specialists on the topic, who mostly are also not native English speakers.

After years of using very complicated and time-consuming material-designing-methods, such as explaining the concept behind certain terms used in these institutions to native English speaking friends who in return tried to give an adequate English term, or researching the internet for hours and hours, I was introduced to Corpus Linguistics and the advantages and many possibilities of

Corpus Analysis. The various uses Corpus Analysis offers seemed to be perfect to solve the problem of material designing.

The purpose of this study was to set up a corpus with American and British texts from the Social Security sector as well as a comparable corpus with German texts from this business, to find out about terminology in this specific area and, subsequently, use the results for the design of teaching materials. Moreover, it was expected that in the course of analyzing terminology, ideas for other teaching contents might appear that had not been apparent before.

One might, rightly, argue that especially for the employees who are in contact with non-native English-speaking clients, terminology is not what they need most. If a Hungarian woman inquires about unemployment benefit in Austria it might make more sense to tell her that she needs the 'Antrag auf Arbeitslosengeld' and not the 'claim form for unemployment benefit' or the 'application for jobseekers' allowance'. However, teaching terminology does not only mean providing translations for a word, but also involves finding different English definitions for a word and challenging the students to explain the concept of the term in question. Moreover, it is crucial for experts to be able to use correct English terminology in their communication with other experts, and it is important to incorporate specialist terms, even if these terms are in German, into authentic English sentences.

4.2 Anticipated Problems

There were different kinds of problems to be anticipated. Some of these problems would have to do with the topic of social security while in addition it was anticipated that difficulties in connection with availability of texts and language itself might appear. However, there was another obstacle likely to emerge: complications in handling the software and possible difficulties in connection with clear and unambiguous interpretation of the results, which might ask for a new strategy of data retrieval.

The first problem to be considered had to do with the area of business this study dealt with, in particular the difference between the social security systems in

question, namely the systems in Austria, the United Kingdom and the United States.

In Austria, the very first forms of insurance which provided financial help in the case of accident, disease, invalidity or death date back to as far as the Middle Ages. Due to the high accident risk for the miners and the manifold occupational diseases in this time, the first miners' health insurances were founded. In the middle of the nineteenth Century, the Allgemeine Berggesetz [General Miners' Act] was adopted, and at the end of the nineteenth century, rudimentary forms of social security were introduced for other occupational groups. Workers, employees, civil servants, farmers, and the self-employed each paid into a different social security plan. The main thrust in the development of the Austrian social security system was the creation of a unified social insurance policy for all occupational groups in the twentieth century (Steiner, 2011).

The system of the United Kingdom has its roots in the National Insurance Act 1911, which introduced the concept of benefits based on contributions paid by employed persons and their employer. The scheme, which came into effect in 1912, was created to insure working people against illness and unemployment. All wage-earners aged between 16 and 70 were obliged to pay contributions and employers and the state paid contributions as well. In return, workers were entitled to a level of free medical care and advice, and a 'dole' for up to 15 weeks per year in the event of unemployment.

The system of the United States of America appeared only in the 1930s, when during the Great Depression poverty rates among senior citizens exceeded fifty percent, and President Franklin D. Roosevelt introduced a limited form of Social Security as a measure to implement "social insurance". It was an attempt to limit the harmful effects of what were seen as dangers in the modern American life, including old age, poverty, unemployment, and the burdens of widows and fatherless children[7].

Due to the different eras and situations in which the systems emerged, their concepts and types of offered insurances are different. The Austrian system has developed for centuries, the American system only for some decades.

[7] for more information on the Social Insurance in the USA see: Rogne et al (2009)

Whereas Austria prides itself on being a social welfare state, the United States were reluctant to introduce such systems for a long time, as they mirror the socialist ideas to a certain extent, and socialism is regarded related to communism, which does not fit into the capitalist philosophy of the USA. So it can be expected that the Austrian system offers many more 'products' than the American one, and that there are many terms in German for which no American equivalents exist. However, the British system might provide similar concepts, and, therefore, might also offer some adequate translations.

Another problem that was anticipated was that it might be difficult to find appropriate texts in American English and German texts that really correspond to the English ones and that the language of the texts might not serve the purposes I had had in mind.

Apart from those factors, problems with the handling of the software seemed possible, which, consequently, would make it necessary to rethink and repeat parts of the study.

The last obstacle that was considered was interpretation of results. It was clear from the beginning that the corpus study would provide much material which would have to be interpreted in a logical, reasonable and as objective as possible way.

4.3 Corpus Set-up

4.3.1 Sampling, representativeness and balancing

As already mentioned in Chapter 3, a corpus should be a 'representative sample' of a language or its variety - also called the 'population' - and it needs to be balanced, i.e. it needs to include different genres of the language. This "is to secure a sample which, subject to limitations of size, will reproduce the characteristics of the population, especially those of immediate interest, as closely as possible" (Yates 1965:9, quoted in McEnery, Xiao and Tono 2006: 19). However, McEnery, Xiao and Tono (2006: 5) argue that

> not all corpora are balanced. Specialized corpora serve a different yet important purpose from balanced corpora […]. If specialized corpora

> which are built using a different sampling technique from those for balanced corpora were discounted as 'non-corpora', then corpus linguistics would have contributed considerably less to language studies.

Even though that, depending on purpose and research question, the decision on sampling method, proportion and number of samples for each text category has to be made individually, one should not forget that to some extent also a specialized corpus needs balancing.

In the course of this thesis it did not seem sensible to collect too many texts containing too many words, but to find texts that are useful and challenging for the ESP courses and their students, who should ultimately be provided with efficient learning material. The only clear reference to corpus size was found in a paper by Flowerdew (1993: 232), in which he describes a study where he used a corpus which consisted of slightly more than 100,000 words. So it was concluded that the 120,000 words contained in the English corpus of the current study should be sufficient. It has to be mentioned that, especially regarding text collection, the many years of teaching experience in the sector proved to be an enormous advantage, because I had first-hand information about what learners actually use and need most urgently.

4.3.1.1 First steps / English and German corpus

The very first step of the study was the search for adequate texts that could be collected for setting up the sample corpora. Due to the easy and quick access and 24-hour availability, the internet was chosen as a (re)source for finding respective texts, and the first sites that were browsed, to get an overview of available materials, were different homepages of American and British Social Security Institutions, such as the British National Insurance Government Services[8] or the official website of the U.S. Social Security Administration[9], both of which provided either relevant documents or further links to relevant institutions which then offered suitable downloadable documents. In this first step, texts were not immediately downloaded, but the aim was rather to get an impression of what

[8] http://www.nidirect.gov.uk
[9] http://www.socialsecurity.gov

was available and would prove helpful and equally challenging. Moreover, balancing and sampling had to be taken into consideration. It was important to identify different text-types, their lengths and their potential audience, to finally make a decision regarding which of the text should be used for 'filling' the English corpus.

German texts were not considered at this point of the study as it was expected that finding comparable German texts would be easier, because of the wide range of 'products' the Austrian Social Security System offers.

Having found the relevant homepages that promised to offer interesting material, the more general texts, which were addressed to customers, patients or claimants and provided manifold information on the respective social system, were searched for. It very soon became apparent that, surprisingly, there were more texts available than had been expected and that it would be necessary to carefully choose between them to ensure a representative sample of the specific language in question.

What should the decision about representativeness be based upon, and would leaving out subjects mean that the corpus is not representative? As there is no universal answer to this question and literature tells us that ultimately the decision depends on the research question, the decision was based on the very subjective feeling of what could be relevant for students.

In view of the teaching environment in which the materials created with the help of this corpus analysis should be used, it was decided to choose a number of topics that would seem relevant for inclusion in an English course for employees of this sector, and to look for documents that would cover these topics. Some of the themes were chosen because prospective course participants often need them, and they were: Old Age Pension, Medical/Health, Unemployment Benefit, Social Security Number and Disability Benefits. Other subjects were chosen not only because of this reason, but also because it was suspected that they might not exist in the same form in the English speaking countries, at least not in the U.S.A., namely: Child Benefit, Attendance Allowance and Maternity (Benefit).

The texts that were selected are all addressed to claimants, patients or people who need information about or are interested in the respective part of social security. Therefore, language used in the documents is informative and the audience is mainly addressed directly. However, the style used is quite formal, which is reflected by the use of formal grammar and vocabulary.

While searching for texts covering the respective areas, the assumption that not all of the parts of the Austrian social security exist in the other two countries, was confirmed. Whereas the British system seems to be quite similar to the Austrian one, the American System is far less extensive than the Austrian one. For some of the chosen topics, U.S. American texts were not available, and it was decided to use only British texts in such cases.

An interesting discovery in the searching process was to realize that looking for suitable texts without having a good knowledge of the social security system would most probably be quite problematic. This supported the claim made in Chapter 2 that being an ESP teacher is much more than being a teacher of English, and that teaching ESP courses effectively asks for in-depth research on the domain beforehand. The decision on the relevance of the texts could only be made, because I was already familiar with English Social Security terminology and because I know what the students in this sector most often need.

Below find Tables 2 and 3 that show how many texts and words per topic the respective corpus contains.

Table 2: Texts in the English Corpus

Content	No. of texts included	No of Words
Pension	2	4,031
Maternity (Benefit)	1	19,022
Medical/Health	3	75,260
Child Benefit	1	3,141
Unemployment Benefit	2	7,303
Social Security Number	3	4,703
Disability	2	7,069
Attendance Allowance	1	2,865
Total	15	123,394

Table 3: Texts in the German Corpus

Content	No. of texts included	No of Words
Pension	2	4,646
Maternity (Benefit)	2	1,557
Medical/Health	2	23,683
Child Benefit	1	3,566
Unemployment Benefit	2	1,858
Social Security Number	0	0
Disability	0	0
Attendance Allowance	2	2,045
Total	11	37,355

The reason that there are no texts regarding 'Social Security Number' and 'Disability' in the German corpus are that in Austria it is not the responsibility of the insured persons to get a social security number because in the Austrian system with its compulsory social insurance, every person that wants to live and work here is automatically granted a number, and that 'Disability' partly overlaps with the category 'Attendance Allowance'.

4.4 Software

For analysis of the texts 'AntConc 3.2.4', a concordancing tool that is freely available and downloadable on the internet[10], was used. This freeware is designed to carry out linguistic research, and it can be installed with Windows, Macintosh and Linux systems. AntConc offers its users seven tools, which can be easily accessed by clicking on their 'tabs' in the tool window or by using the function keys F1-F7.

Before AntConc can be used, the selected texts, which were either .pdf files or copied into Word documents, needed to be converted into .txt files, as this is the only format which can be read by the program. To analyze the texts, AntConc has to be started and the .txt files have to be opened in the program. As soon as this is done, all seven tools, which will be briefly explained below, in order of their relevance for the current study, can be used.

[10] http://www.antlab.sci.waseda.ac.jp/

4.4.1 Different tools in AntConc

4.4.1.1 Wordlist:

The wordlist simply shows all words used in the corpus and presents them in a list that can be ordered by frequency, which is the standard setting, but also alphabetically and by word end. It shows the number of 'tokens', which is the absolute number of words used, and 'types', words that were used in the same form. If the word *insurance* is used 500 times, it would count as 500 'tokens' but only one 'type'.

This tool makes it possible to quickly find which words are used most frequently in the corpus.

4.4.1.2 Concordance:

This tool shows the results in the 'Keyword in Context' (KWIC) format, which displays selected words embedded in their context and one can choose how many characters to the left and right should be shown in the window. The whole sentence in the text can be seen by clicking on the word. This enables us to see how the words are repeatedly used in texts.

4.4.1.3 Collocates:

In this tab the collocates of the search term are shown and ordered by frequency, frequency on the left or right, word, word end as well as by the statistical measure between the search term and the collocate. The span of words to the left and right of the search term in which to find collocates can be determined by the researcher

4.4.1.4 Clusters:

The clusters tool allows finding regular expressions in a corpus. It shows clusters of a number of words whose minimum and maximum number of words can be determined beforehand.

4.4.1.5 Concordance Plot:

This tool displays the results in the form of barcodes that show us the position of the search terms in the texts.

4.4.1.6 File View:

With this option, the separate text files can be completely viewed, with the search term displayed in a different color.

4.4.1.7 Key Words:

This tool is used for analyses that compare different corpora of one language. It shows which words are characteristic in one corpus in comparison to a reference corpus.

The tools most frequently applied in this study were the wordlist, collocates, concordances and clusters tool. The keyword tool was not used at all as only one English corpus was examined.

The advantage of AntConc is that it is very easy to use and there is the possibility of exporting the results to a .txt file and save them. The disadvantage is that the colors that are used to make the different levels in the concordance search visible are not 'transferred', which makes analysis a little more difficult.

4.5 Results

4.5.1 Wordlists

The first tool that was used as a basis for the rest of the analysis was the wordlist tool. It showed that in the English corpus there were 4,941 word types and 123,394 word tokens. This means that 4,941 words were used once or more often to form a total of 123,394 words. The German corpus, which was about 70% smaller, contained 6,007 word types and 37,355 word tokens.

The large number of word types is a result of the "less criminatory" word count (Flowerdew, 1993: 233) than other types of word count. Computerized word counts include all forms of a word as separate items as well as singular and plural of all nouns or inflected verb forms. However, homographs are represented

as one item. Flowerdew suggests that to provide a valid comparison with other word counts (e.g. those based on dictionaries), and therefore get the 'real' number of words included, one must deduct about 55% of the total number of items, which would leave us with a figure of just over 2,220 for the English corpus, and about 2,950 for the German corpus. This means that there were approximately 2,220 or 2,950 individual words used in the two corpora.

Tables 4 and 5, below, show the 100 most frequently used words and tables 6 and 7 which show the 10 most frequently used words in the English as well as the German corpus.

As can be seen, the first ten words in both corpora are, as had been expected, grammatical words such as articles, personal pronouns, conjunctions and prepositions, which are very similar in both languages. In the English corpus the 10 most frequently used terms account for 30,655 tokens or almost 25% of the whole corpus, whereas in the German corpus the first 10 words on the wordlist account for only 16.95% of the corpus. This might be explained by the fact that in German *der, die* and *des* mean *the* and take three positions as compared to only one position in the English list. *You, and* and *or* are used 59%, 25% and 75% more than their German translations *Sie, und* and *oder*. *For* and *in* are used approximately as often as their German equivalents *für* and *in.* The prepositions *to* and *of* cannot really be compared to the German list, as there are more translations for them and it would be necessary to go through each individual sentence to determine the correct translation. Finally, the equivalents to the indefinite article *a* and the possessive determiner *your* do not appear at all in the German corpus.

Under the first 100 words listed from the English corpus, more than half seem 'important' and 'relevant' and would 'qualify' for being taught (if students do not know them yet), whereas in the wordlist of the German corpus fewer relevant words appear within the first 100.

Table 4: 100 most frequently used word in the English corpus

1	5376	the	35	515	information	69	262	their
2	4509	you	36	492	as	70	257	b
3	4117	to	37	491	by	71	254	www
4	2964	and	38	457	at	72	253	insurance
5	2633	of	39	455	it	73	249	claim
6	2542	a	40	453	work	74	249	which
7	2429	your	41	442	other	75	248	we
8	2202	for	42	436	an	76	246	all
9	2164	or	43	429	right	77	246	see
10	1719	in	44	394	coverage	78	244	state
11	1631	if	45	378	about	79	243	week
12	1378	medicare	46	370	employer	80	241	period
13	1077	is	47	351	more	81	236	number
14	1017	have	48	345	what	82	233	how
15	1003	are	49	342	allowance	83	231	disability
16	840	can	50	341	also	84	228	who
17	836	be	51	336	must	85	227	gov
18	825	that	52	332	drug	86	224	i
19	781	care	53	328	t	87	221	staff
20	774	health	54	322	any	88	221	these
21	761	nhs	55	319	benefits	89	221	under
22	743	this	56	311	do	90	219	service
23	741	on	57	306	call	91	215	before
24	654	get	58	306	social	92	211	constitution
25	634	may	59	299	benefit	93	206	doctor
26	623	with	60	294	should	94	202	medical
27	577	pay	61	283	help	95	201	prescription
28	577	will	62	280	out	96	200	local
29	570	s	63	275	when	97	199	patients
30	565	services	64	274	need	98	199	visit
31	559	plan	65	271	security	99	198	employment
32	553	part	66	267	smp	100	197	re
33	537	from	67	267	they			
34	521	not	68	264	child			

Table 5: 100 most frequently used word in the German corpus (case sensitive)

1	1295	der	35	138	als	69	74	müssen
2	919	die	36	136	können	70	74	wir
3	663	und	37	126	einem	71	73	B
4	538	für	38	122	sich	72	72	Versicherten
5	514	Sie	39	119	bzw	73	71	Antrag
6	483	in	40	115	einen	74	70	ja
7	411	des	41	115	Euro	75	70	Kinderbetreuungsgeld
8	370	oder	42	115	Kind	76	69	nein
9	351	von	43	113	Anspruch	77	67	diese
10	350	1	44	111	nur	78	66	e
11	295	auf	45	109	haben	79	65	Ihre
12	270	ist	46	106	wgkk	80	65	Kosten
13	268	den	47	105	Für	81	63	E
14	254	mit	48	104	durch	82	62	unter
15	252	dem	49	103	Seite	83	57	Achtung
16	248	im	50	102	eines	84	57	Informationen
17	240	bei	51	102	Familienbeihilfe	85	57	um
18	237	zu	52	99	vor	86	56	Bei
19	235	werden	53	98	aus	87	56	Höhe
20	234	das	54	96	zum	88	56	Kindes
21	234	eine	55	94	www	89	55	Wochengeld
22	219	sind	56	92	ab	90	54	gilt
23	200	nicht	57	92	Leistungen	91	54	Mail
24	194	einer	58	92	Telefon	92	54	pro
25	184	Die	59	92	über	93	53	dass
26	183	ein	60	90	Kinder	94	53	vom
27	179	wird	61	89	an	95	53	z
28	177	wenn	62	86	EUR	96	50	Jahre
29	172	zur	63	81	es	97	50	Personen
30	164	at	64	80	keine	98	50	siehe
31	163	WGKK	65	78	Das	99	49	dann
32	153	nach	66	78	Wenn	100	49	Ihnen
33	147	auch	67	77	Der			
34	140	bis	68	75	kann			

Table 6: 10 most frequently used English terms / percentages

	term	no. of occurrences	percentage
1	the	5,376	4.35
2	you	4,509	3.65
3	to	4,117	3.33
4	and	2,964	2.40
5	of	2,633	2.13
6	a	2,542	2.06
7	your	2,429	1.96
8	for	2,202	1.78
9	or	2,164	1.75
10	in	1,719	1.39
		30,655	24.80

Table 7: 10 most frequently used German terms / percentages

	term	no. of occurrences	percentage
1	der	1,375	3.68
2	die	1,105	2.95
3	und	680	1.82
4	für	645	1.72
5	Sie	561	1.50
6	in	523	1.40
7	des	419	1.12
8	oder	371	0.99
9	von	360	0.96
10	auf[11]	295	0.78
		6,334	16.95

[11] the number "1", which was originally in place 10 was ignored.

What was quite unexpected is the fact that even though the German corpus was much smaller than the English one, there were 5.5% more word types in the German texts. On the surface, this could suggest that the German language has more word types than the English language, which is not the case. What could be deduced from this is that German speakers either use a larger variety of words or that they make use of more different word types in 'technical' or 'semi-technical' texts. This would also be in accordance with Widdowson's suggestion that experts use their own 'jargon' or 'mastery language' (see 2.1. and 2.4.3.), and it would indicate that German speakers tend to make use of their 'jargon' more than English speakers do. Another reason for the higher word type count in German could be the more frequent use of compound nouns in German.

However, it might validate the contention of teachers at the Department for Translations at university that German texts are about 10% longer than English texts, which, consequently proves that German speakers need more word types to say the same thing as English speakers do.

Just as in Flowerdew's study, the 'relative frequency' (see Table 8, below) shows that the relation between frequency of occurrence and number of items occurring with a particular frequency is inverse. 1,540 items of the English corpus occur once, a little less than the half occur twice, 396 terms occur three times, etc.

Table 8: relative frequency of terms

Number of occurrences of items	Number of items occurring this many times
1	1,540
2	728
3	396
4	296
5	194
6	193
7	101
8	104
9	95
10	87

Flowerdew (1993: 234) suggests that this would give a course designer a strong indication of what terms to include in or exclude from a course plan, as

> each level of frequency offering a potential cut-off point for selection and grading of items (the number of words targeted by the syllabus designer, of course, depending on the entry level of learners and time available for teaching/learning).

Overall, the wordlist was different from what I had expected. On first appearance, the list did not contain as much 'specialist-vocabulary' as had been anticipated, but consisted of more 'semi-technical' and 'general' vocabulary. This was surprising, considering the fact that the corpora contained texts on very specific topics and did not represent 'general' English or -German.

The next step after 'getting a quick overview' was to examine the complete English wordlist with its almost 5,000 words, and get an impression of what could possibly be disregarded to narrow down results. As it was originally assumed that only what is used frequently is important, all word types that appeared only once and twice were considered irrelevant and deleted from the list. After this, only 2,672 word types, just over 50%, were left. Secondly, a method of selection which could be used for choosing relevant terms from the remaining words had to be found, and due to the lack of any other method, intuition was used to highlight all terms on the list that seemed essential, necessary and worth teaching. However, there were still more than 550 terms left and to get a better overview they were then ordered alphabetically; this showed, as mentioned by Flowerdew (1992: 233), that the list included many forms of the same lexeme, e.g.: *authorities, authority, authorization* and *authorized*. After a further selection, which was again not done according to frequency but with intuition and the needs and wants of prospective students in mind, there were still about 350 words left, a mix of 'technical', 'semi-technical' and 'general' vocabulary, which could probably be incorporated into the curriculum of a language course without too much difficulty - depending on the number of lessons planned within a certain time frame.

Another possibility one has to consider is that some words might have the same form but could belong to different categories. *Benefit*, for example, could be a verb or a noun, and it is important to know the frequencies of the two uses. If it

is used as a verb only once out of the 299 times, teaching the verb *benefit* is not as urgent as if it is used 100 times. Hence, it was necessary to investigate such terms closer in regards to word categories.

After having taken the first steps and having worked with AntConc for some time, I found out that the program was 'case sensitive', meaning that it distinguished between words that are written with a capital first letter and the same word starting with a small letter. This did not pose a problem in the German corpus, as in German nouns are always written with a capital first letter, also in the middle of a sentence. However, it does make a difference in English texts and the number of word types in the English corpus shrank considerably after producing a new list that treated all words as lowercase, from 6,216 word types to 4,941, a reduction of about 20%. The list of the 100 most frequently used words did not change dramatically, but there were some slight changes. The 10 most frequently used words in German stayed the same, and in the English version positions 2 and 3 as well as positions 8 and 9 were interchanged. Whereas in the 'case sensitive' wordlist *to* was the second- and *you* the third most frequently used word, this changed in the version were all words were treated as lowercase. The same happened for *for* in position 8 and *or* in position 9 in the 'case sensitive' version; they swopped places in the 'lowercase-version'. It was decided to use the wordlist that does not consider words with capital first letters for the basis of the following analyses.

As the German corpus was only used as a sort of guidance and reference, the AntConc search options were originally also set to 'treat all data as lower case' to provide the same conditions for both corpora. However, it was decided at a later part of the study to repeat the data retrieval for the German wordlist conducting a 'case sensitive' search. Whereas it makes no difference in the English corpus, it is advisable for the German data, as the case sensitive approach has the advantage that one is able to immediately detect nouns and polite forms of address (*Sie*, *Ihr, Ihre*) in the corpus, which makes analysis less difficult and time-consuming. The results of an English corpus, at least in respect to identification of nouns, need to be examined more closely, because, as mentioned above, there

are some words that might be used either as nouns or as verbs, and it is necessary to consult the co-text to find this out.

On closer examination it was found out that the 100 most frequently used words of the English corpus contained many words with 'meaning'. There were 36 nouns, 15 verbs and 3 adjectives, whereas in the German corpus only 25 nouns, 5 verbs and no adjectives were detected. The English 'Top 100' list further contained 5 forms of modal auxiliary verbs, 5 forms of lexical auxiliary verbs, many instances of the conjunction *if* as well as many prepositions.

The fact that more than a third of the words in the 'Top 100' consists of nouns was not very surprising, as the dense use of nouns is one of the characteristics of formal English language. Another indicator for official and rather formal texts is the frequent use of passive forms, which was also expected because of the 1,077 occurrences of *is* and the frequent use of *be* (836 times) (see sub-section 4.5.2.1. below).

Looking at this list, some topics - apart from terminology - that should be taught in an English course based on these texts became immediately apparent and, therefore, were researched more closely. These topics, and also how AntConc tools other than the wordlist-tool were used to get informative results, will be discussed in the next section.

4.5.2 Topics considered worth teaching apart from terminology

4.5.2.1 Passives

As just mentioned, the frequent use of *is* indicated that the passive is a frequent grammatical form used in the texts. Thorough analysis with the help of color coding revealed that out of the 1,077 occurrences, 312 occurrences were actually used in a passive construction, which accounts for almost 29%. To verify the assumption that the passive is frequently used also in other tenses and/or constructions, closer research was conducted on *was, were, been, are* and *be* (see Table 9)

Table 9: Occurrences of the different forms of "be"

Term	Number of tokens	Number of passive usages	Percentage in relation to tokens
are	1,003	292	29
be	836	479	57
been	118	86	72
is	1,077	312	29
was	68	30	44
were	69	27	39
	3,171	**1,226**	**39**

1. […] clinical teams are already encouraged to participate […]
2. […] make a reasonable request to your employer to be accompanied by either a […]
3. […] the way your complaint has been dealt with […]
4. Where no one is appointed as a receiver […]
5. The Constitution was developed as part of […]
6. If you think you were discriminated against […]

This frequent occurrence of passive constructions suggests that the passive in present as well as in past tenses is a grammar topic that should be incorporated into the curriculum, as it can be assumed that it is a form regularly used in this kind of texts.

4.5.2.2 Conditional sentences

The repeated use of *if* (1,631 times) implies that many conditional sentences are produced in this context. Unfortunately, AntCont does not offer a tool to count the total number of sentences in the corpus, but it can be assumed that *if* is used only once in a sentence, so there should be 1,631 *if*-sentences.

To have some reference values, the written texts in the COCA as well as the BNC were consulted for the frequency of *if* (see Table 10) . The COCA with a total word count of 95,565,075 (excluding the spoken texts) had 296,398 instances of

if, the BNC with a total word count of 9,963,663 (excluding the spoken texts) listed 48,319 instances.

Table 10: Percentages of usage of if

Self-made corpus	***COCA[12]***	***BNC[13]***
1.32%	0.31%	0.48%

The frequent appearance of conditionals could be due to the fact that the documents are used to provide explanations about how clients, patients or customers can file applications, claim benefits, or any other support, and what has to be done by the clients, on what conditions they are eligible for certain 'products' or what conditions have to be fulfilled to be entitled to benefits at all.

7. If you were eligible for part A for 2 years but didn't sign up, you'll have to pay the higher premium for 4 years.
8. For SMP purposes, your employment can be taken as continuous if anyone of the following applies:
9. […] your MAP will not change if it has already started.
10. If you lived outside the United States, a current or previous passport […] could be used […]

4.5.2.3 Auxiliary verbs

The many instances of modal- and non-modal auxiliary verbs within the 100 most frequently used words became immediately apparent and – in the case of the non-modal auxiliaries - suggested that certain tenses were used more frequently than others. However, this assumption could only be verified or falsified by finding out how often these verbs were actually used as auxiliary verbs and how often they were used as lexical verbs.

- *Modal auxiliary verbs*

Out of the eight modal auxiliaries, *can*, *may*, *will*, *shall* and *must* were represented in the 'Top 100' list, either in their base- or preterite forms. *Shall* appeared only

[12] data retrieved on January 8th, 2013
[13] data retrieved on January 8th, 2013

once in the corpus, *might* 26 times, the rest was used between 70 and more than 800 times. Altogether, modal verbs were used 2,852 times, which accounts for about 2.3% of the total word count. Below see Table 10 that shows the exact numbers of usage. *Need, ought* and *dare* were not used at all.

Table 11: instances of modal auxiliary verbs / percentages are rounded

Modal auxiliary	Number of tokens	Percentage
can	840	0.68
may[14]	634	0.51
will	577	0.46
must	336	0.27
should	294	0.24
would	74	0.06
could	70	0.05
might	26	0.02
shall	1	0.0008
	2,852	**2.3**

- will and would

Because *will* and *would* are often used in conditional sentences, the joint appearance of these two modal verbs and *if* was searched for. This was done by means of the 'collocates' tool, which had to be adjusted accordingly. As *if* and the modal verb *will* never appear next to each other, adjustments were changed to 'from 10L to 10R, which means that all words are shown that appear up to 10 words to the left and right of the search term. In the case of *if*, this produced the result that it appears 170 times with *will*. To verify this result, the same search was conducted for *will*, and it showed that the result was the same. Of course, this does not prove that there are necessarily 170 conditional sentences with the joint use of *if* and *will*, because punctuation is disregarded in the computer search. Therefore, it can easily be possible that AntConc includes a case where *will* is used

[14] It should be noted that eleven instances of *may* had to be disregarded, as they referred to the month of *May*.

towards the end of one sentence and the next sentence starts with *if* (11.), provided there are not more than nine words in between.

11. […] the employee helpline will advise you on what you should do next. If your contract ends […]

With the help of the concordancing tool different uses of *will* could be detected and it is used about one third in a conditional clause to express a future consequence (12.) and in the rest of the instances it is used as a part of the future simple (13.).

12. If it does not, the state agency will decide […]
13. We will ask for evidence to show if you do not have a Social Security number.

The different usages of *would* were more difficult to detect. The most obvious use of the auxiliary was in hypothetical conditional sentences in combination with *if* or *unless* (14. and 15.), but also in past conditionals sentences (16.):

14. Your employer is someone who is liable to pay the employer's share […] or would do if you earned enough […]
15. […] unless to do so would put anyone at risk.
16. […] you are entitled to the same MA you would have got if your baby had been born alive.

However, it was also used to express hypothetical volition in sentences such as:

17. [this means] she would get 90 % of £ 150 again for the first 6 weeks […]

- Can and could

Can was used in all three senses stated in Leech (2004: 74-75), namely for expressing possibility, ability and permission. Leech lists the use of *can* for possibility as "very common", its use for expressing ability as "common" and its use for expressing permission as "less common". However, this seems not to apply for the texts in the corpus. *Can*, which was used 840 times in total, was only used 34 times or 4% to express ability (18.), which does not reflect the "common" usage mentioned by Leech:

18. You may get the lower […] rate if you can walk, but need guidance […] from another person.

The rest of the instances could be allocated almost equally between their use for expressing permission and possibility. Whereas meaning was very clear for some sentences (19., 20.), in some cases boundaries were fuzzy (21.):

19. We can accept only certain documents as proof of identity.
20. How can I protect myself from identity theft?
21. In some cases you can file a fast appeal.

Whereas (19.) clearly expresses permission and (20.) possibility, the meaning of (21.) seems not so clear. It could mean that in some cases you are allowed to file a fast appeal or that you have the possibility to file one. As the possibilities in this context are often the result of a legal condition, I would argue that it is almost impossible to distinguish between the two meanings, and that Leech's contention that *can* to express permission is "less common", cannot be confirmed for the texts in the current corpus.

In comparison, *could* was used only 70 times, once to express the past of *can* (22.), four times to express 'tentative suggestion' (23.), once to express 'hypothetical permission' (24.), and the remaining 54 times to express 'hypothetical possibility' (25.)[15]:

22. […] if your baby is born prematurely, before the maternity certificate could be issued […]
23. If you don't have a textphone, you could check if your local library has […]
24. You could do this kind of course for up to two weeks […]
25. You and your partner could be liable to a tax charge called […]

The fact that *could*, which is normally used to make an expression of possibility tentative, is not used very often in the texts might be due to the fact that the authorities who issued the documents want to ensure that their formulations are very clear and do not leave much space for speculations.

- May and might

[15] for more information see: Leech, 2004

May was used 634 times in the texts to express either possibility or permission. The two meanings of may are very close to each other and, therefore, distinction is blurred in some cases. Leech (2004: 76) states that the use to express possibility (26.) (with almost 500 instances in the current corpus) is by far more common than the 'permission' use (27.) (about 50 instances). Even though some sentences could not clearly be allocated to the one or the other use (28.) the findings correspond to Leech's contention.

26. There may be several private companies that offer […]
27. Your employer's […] procedures may allow the accompanying person to take a fuller […] role […]
28. We may use the information […]

Like in the case of *can* and *could*, *might* is normally used as a polite alternative to *may* in the 'hypothetical permission' or it is used as a substitute for *may* in expressing factual possibility. In the corpus it is used only 26 times to express 'hypothetical possibility' (29.):

29. Your child might not go back into education or training […]

Another observation that was made in the connection with *may* is that it was used in only seven of the 15 texts and in only two of the seven American texts. This might verify Leech's (2004: 129) assumption that *may* is used less in American English.

- Must and have to

According to Leech (2004: 78-81), both modals express obligation or requirement. Whereas they have the same meaning when they express requirement, there is a difference when used to express obligation. *Must* is used when "the speaker is the person who exerts authority over the person(s) mentioned in the clause" (ibid: 78), and *have to* is used when the authority of the speaker is not involved and obligation is expressed

> without specifying the person exercising power or influences. The constraining power may be some authority figure such as a doctor or an employer, the government or simply the power of 'circumstances' (ibid: 80).

In the case of the investigated texts, the distinction is very hard to make because it can be argued that the speaker and the authority figure are one and the same 'person'.

Must was used 336 times (30. and 31.), and it is very difficult to determine whether it is used to express requirement or obligation, as a legal requirement can be seen as an obligation.

30. Workers must also meet the eligibility and qualifying requirement […]
31. The information provided on the claim form must be correct at this time.

Have to was used only 131 times, and also in these instances it is not completely clear whether they express an obligation or a legal requirement, which – in my understanding – is an obligation (32., 33.):

32. Di I have to get a referral to see a specialist?
33. You don't have to sign a private contract.

Leech (2004: 78) also mentions that (especially in spoken American English) the 'obligation' use of *must* is being replaced more and more by the use of *have to*. This contention cannot be confirmed for the written documents of the corpus, as *must* was used in all 15 documents and almost three times as often as *have to*, whereas *have to* was not used in three of the documents, and all three were American texts.

- Shall and should

Shall, which is hardly in use anymore, occurred only once in the corpus (34.) to express obligation and could be replaced by *must*. Leech (2004: 89) states that this usage is found only in legal or quasi-legal documents.

34. […] no individual who […] has registered for work shall be considered ineligible […]

Should was used 294 times and in none of the senses mentioned in Leech (2004), who ascribes *should* the same features as *must* with the one difference that *should* lacks the confidence of the other modal. However, in the texts of the corpus *should* was used to express a recommendation or a piece of advice:

35. You should claim Child Benefit as soon as …

36. If you are seeking hospital care you should contact your local commissioner…

- *Non-modal auxiliary verbs*

The frequency of the non-modal auxiliaries (see Table 12) in the texts was much higher than the frequency of the modal auxiliaries, which is why they were examined a little closer than the modals. The verbs *be, have* and *do* were used very often in the corpus, but in different forms. However, as mentioned above, it had to be found out whether they really functioned as auxiliaries or if they were used in a lexical sense.

For the forms *is, have* and *are* the percentage of occurrences in the COCA and the BNC is given to provide reference values.

Table 12: instances of non-modal auxiliary verbs / percentages are rounded

non-modal auxiliary	Number of tokens	%	COCA	BNC
is	1,077	0.87	1.25	1.03
have	1,017	0.82	0.69	0.77
are	1,003	0.81	0.60	0.48
be	836	0.67	-	-
do	311	0.25	-	-
has	189	0.15	-	-
were	69	0.05	-	-
was	68	0.05	-	-
had	51	0.04	-	-
did	31	0.03	-	-
	4,652	3.8	-	-

- Is

To find out how *is* was used, it was first tried to separate the different forms with the help of AntConc, by going to the concordance tab and typing in *is *ing* to get all the instances for the present progressive use. AntConc listed 43 occurrences, three of which had to be deducted as they did not show a verb in the progressive

form, but the term *something*. However, the rest had to be separated manually and the concordance list with all the 1,077 instances was examined in detail and color coded according to its use in a passive construction, in the present progressive or as a lexical verb and it was found out that *is* was most often used as a lexical verb, with 725 instances.

37. […] the evidence suggests that there is a clear connection between […]
38. Medicaid is a joint federal and state program […]

As already mentioned above, a passive construction with the auxiliary *is* occurred 312 times:

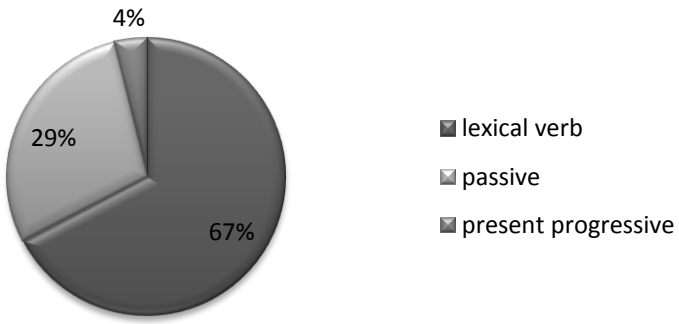
39. [...] make sure nobody is excluded or left behind.
40. The money you get is based on your income […]

Is as a non-modal auxiliary in the present progressive could be detected 40 times.

41. Each PCT is also opening a new health care center.
42. Women's State Pension Age is rising […]

Figure 2 shows the percentages of the different usages.

Figure 2: Distribution of the different usages of is (1,077 occurences)



- Have

The same procedure was used for *have*, and it was found out that the term was used 554 times as a lexical verb:

43. […] a person must have a personal, direct interest […]
44. […] care trusts have a duty to make arrangements […]

228 times it was used as a non-modal auxiliary verb in perfect constructions:

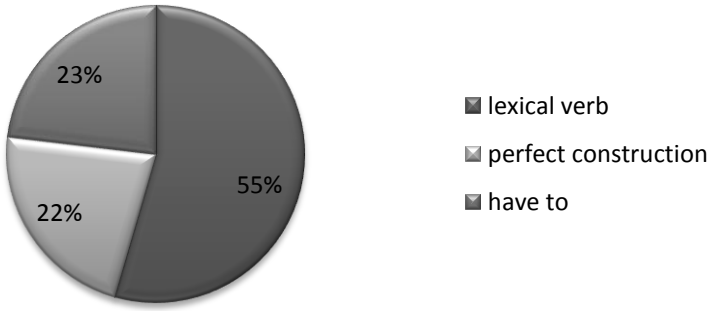
45. Your child needs to have started […]
46. […] decisions by the courts have made it clear[…]

Additionally, it was used 235 times in combination with *to,* to express obligation, hence, actually belonging to the category discussed above.

47. You do not have to pay SMP back […]
48. […] so you do not have to worry about payment […]

See Figure 3 for exact percentages of distribution.

Figure 3: Distribution of the different usages of have (1,017 occurrences)



- Are

The last word that was examined this closely was *are*, because there were also more than 1,000 instances in the corpus. *Are* was also used most frequently as a lexical verb in combination with either a noun or an adjective with 621 occurrences:

49. There are a number of national bodies […]
50. But there are circumstances when your employer […]

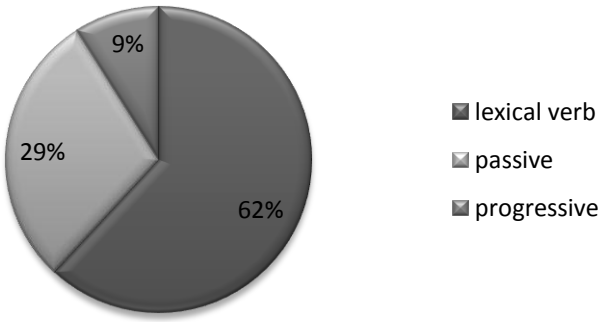
It is followed by the use as an auxiliary in a passive construction with 292 occurrences:

51. If you are paid weekly […]
52. ..these bills are sent out every 13 weeks […]

Finally, it appeared 92 times as an auxiliary verb in a progressive form:

53. …because you are already getting another kind of […]
54. If you are getting another government benefit […]

Figure 4: Distribution of the different usages of are (1003 occurrences)



The rest of the non-modal auxiliaries were not examined as thoroughly: however, there are some facts that are noteworthy.

Be in its base form was used 836 times in this corpus, and mostly together with modal auxiliary verbs (580 times including negations as well):

55. There may be a limit on the amount…
56. […] you should be able to find the information […]
57. […] weekly earnings must be at least equal to […]
58. There is a risk that evidence could be destroyed […]
59. […] treatment that you would be entitled to receive […]
60. […] that you can be given treatment…

Second most frequently it was used in the form of *to be* (127 times), whereof 93 occurrences were passive constructions.

61. […] it is intended to be a living document […]
62. […] priorities that need to be decided […]

4.5.3 Terminology

The 20 most frequently used nouns in the English corpus (excluding names such as medicare or NHS) were: *service(s)* (786), *care* (779), *health* (774), *plan(s)* (755), *benefit(s)* (618), *right(s)* (582), *part(s)* (567), *information* (515), *drug(s)* (437), *employer(s)* (429), *coverage* (394), *week(s)* (368), *allowance* (344), *pay* (334), *state(s)* (306), *child*(ren) (305), *security* (271), *period(s)* (259), *insurance* (253), *claim(s)* (206).

This list is not entirely in accordance with the 'Top 100' list, as in the cases of *care*, *pay, right, need* and *state*, the instances where the words were used as verbs or adjectives (in the case of *right*), respectively, had to be deducted. *Help* and *need*, even though they would have been within the 20 most frequently used nouns according to the 'Top 100' list, where they were in positions 61 and 64, respectively, did not make it on the 'Top 20' list, because of their frequent use as a verb. However, some positions changed, as plural usage was included in the 'Top 20' list as well, which was not the case in the AntConc wordlist. All the English nouns were found within the 100 most frequently used words, the German nouns were found within the first 106 words. As the German language has four cases and nouns have different forms, depending which case they are used in, these forms were also searched for and included in the count as well as plurals. However, just as in the search for the English nouns, names such as WGKK were ignored. The 20 most frequently used nouns in the German corpus were: *Kind(er)* (261), *Anspruch(-sprüche)* (198), *Jahr(e)* (142), *Leistung(en)* (130),*Versicherte(n)* (119), *Information(en)* (111), *Seite(n*) (110), *Familienbeihilfe(n)* (104), *Telefon* (92), *Antrag(-träge)* (84), *Person(en)* (75), *Kinderbetreuungsgeld* (70), *Kosten* (65), *Wochengeld* (65), *Höhe* (56), *Mail* (54), *, Frau(en)* (54), *Krankenversicherung(en)* (50), *Website* (49), *Fax* (48*).*

What is interesting is that the nouns in the two corpora do not really correspond to each other. There are no translations or equivalents in the German list for *care*, *plan, right, part, drug, employer, coverage, week, allowance, pay, state, security* or *period*. *Service* in some cases corresponds to *Leistung*, *health* in the English corpus is often used in connection with *insurance* which would then correspond to *Krankenversicherung*, *benefits* is in some cases used like *Leistung*, *allowance* is *Beihilfe*, *children* is the English term for *Kinder* and *claim* for *Anspruch*. There is also a slight difference in how much the first 20 nouns account for in the respective corpus. Whereas they account for more than 7.5% in the English corpus, they only make up 5.2% of the German corpus.

Obviously the emphasis of each corpus is on different 'products' in the two languages or importance of certain factors is different in the two (or three if we

distinguish between American and British) cultures. This difference might be explained by cultural factors, but it might also be a result of a not entirely balanced selection of texts for the corpora.

Below find a table of the 20 most frequently used nouns in both corpora for comparison

Table 13: 20 most frequently used nouns in the English and German corpus

English				**German**			
		occurr.	%			occurr.	%
1	service(s)	786	0.63	1	Kind(er)	261	0.70
2	care	779	0.63	2	Anspruch(-sprüche)	198	0.53
3	health	774	0.63	3	Jahr(e)	142	0.38
4	plan(s)	755	0.61	4	Leistung(en)	130	0.35
5	benefit(s)	618	0.50	5	Versicherte(n)	119	0.32
6	right(s)	582	0.47	6	Information(en)	111	0.30
7	part(s)	567	0.45	7	Seite(n)	110	0.30
8	information	515	0.42	8	Familienbeihilfe(n)	104	0.28
9	drug(s)	437	0.35	9	Telefon	92	0.25
10	employer(s)	429	0.35	10	Antrag(-träge)	84	0.23
11	coverage	394	0.32	11	Person(en)	75	0.20
12	week(s)	368	0.30	12	Kinderbetreuungsgeld	70	0.19
13	allowance(s)	344	0.28	13	Kosten	65	0.17
14	pay	334	0.27	14	Wochengeld	65	0.17
15	state(s)	306	0.25	15	Höhe	56	0.15
16	child(ren)	305	0.25	16	Mail	54	0.14
17	security	271	0.22	17	Frau(en)	54	0.14
18	period(s)	259	0.21	18	Krankenversicherung(en)	50	0.13
19	insurance	253	0.21	19	Website	49	0.13
20	claim(s)	206	0.17	20	Fax	48	0.13
		9,282	7.52			1,937	5.20

The 20 most frequently used verbs (not including modal and auxiliary verbs) were: *is* (725), *are* (621), *have* (554), *work* (453), *see* (246), *go* (175), *provide* (173), *make* (171), *use* (144), *start* (133), *join* (130), *like* (121), *take* (119), *use* (118), *means* (117), *apply* (107), *tell* (107), *give* (106), *click* (105), *covers* (105).

In the case of the verbs, they were counted only in this one specific form. However, some of them were also used in other forms or tenses. What is noteworthy here is that *is* and *are* are the most frequently used verbs (their use as auxiliary verbs has already been considered and deducted). The regular use of *is* indicates that many 'things' are described, the frequent use of *are* can be explained with the fact that the texts are written for clients, who are directly addressed (*you are*).

In the German corpus there also was a high number of modal verbs within the "Top 100" and *ist* (*is), sind (are)* and *haben (have)* were used frequently. As the German corpus was not examined in as much detail as the English one, it can only be assumed that they were also partly used as auxiliary verbs and partly as lexical verbs. The following most frequently used verbs were (in this specific form): *besteht* (47), *erhalten* (42), *beträgt* (41), *gibt* (36), *gewährt* (35), *beziehen* (29), *gebührt* (26), *gelten* (24), *bezogen* (22), *nehmen* (20), *beantragt* (159, *bestimmen* (15), *erfolgt* (15), *erworben* (15), *liegt* (15), *übersteigt* (15), *bezahlen* (14), *übernehmen* (14), *ausbezahlt* (13) and *beachten* (13).

Again, there is a big gap between the two corpora. As it was observed with the nouns, the verbs used in the English corpus and the ones used in the German corpus are to a great extent different. Apart from *is, are* and *have*, there are only two other verbs that are also used in the German corpus, which are: *take (nehmen)* and *covers* could sometimes be used in the sense of *übernehmen*. *Give*, here, is a 'false friend' and cannot be translated with *gibt*. *Gibt* in the German corpus is used in the construction *es gibt* or *noun+gibt* which can be translated as *there is*.

The 20 most frequently used adjectives in the English corpus were: *social* (306), *medical* (202), *local* (200), *covered* (161), *able* (150), *national* (127), *paid* (126), *certain* (119), *available* (115), *public* (112), *primary* (108), *due* (105),

original (101), *approved* (100), *legal* (99), *long* (96), *based* (89), *entitled* (88), *employed* (84) and *eligible* (77).

The German corpus produced the following results regarding adjectives: *versichert(e/r),* (52)*medizinische(n/r)* (48), *erforderlich(en)* (45), *möglich* (41), *folgende(n/r)* (36), *ärztliche(n/r)* (33), *gesetzliche(n/r)* (30), *grundsätzlich* (27), *notwendig(e/er/n)* (26), *zuständige(n/r)* (25), *allgemeine(n/r)* (24), *eingetragen(e/r)* (21), *freiwillig(e/er/en)* (21), *privat(e/er/en)* (23), *geboren* (14), *erheblich* (12), *kostenlos(e/r/en)* (16), *ausländischen* (10), and *geldleistungsberechtigt* (10).

As adjectives in German change their form depending on gender and case of the noun, it is extremely difficult to provide 100% reliable information, especially because the German corpus is not the main object of research here, but is used only as a reference. However, even after having searched for different forms of the adjectives, the 'Top 20' list did not change regarding the contained words, but only positions were changed.

Regarding these six lists, it can be said that terminology differs widely in the two corpora, and that with the help of frequency lists alone, no obvious tendencies regarding terminology can be detected. As the corpus was originally designed to find out more about terminology in the Social Security sector, to get information about which terms should not be left out of the curriculum under any circumstances, and possibly also to learn new words myself, a decision had to be made about which terms should be included in a syllabus.

It was predicted that many words of the self-made specialist corpus would be technical and needed to be incorporated into the curriculum. However, as can be seen from the lists above, the results of the wordlist showed a different picture. In the English as well as in the German corpus, there were a large number of 'semi-technical' and general words students most probably already know, and it almost seemed that there are only a handful of terms that might be new and useful to the learners. The only 'semi-technical' terms in the most frequently used words of the English texts were: *benefits, coverage, allowance, claim, apply, cover, covered, due, approved, entitled and eligible*. The German texts contained

some 'semi-technical' but also some 'technical' terms: *Anspruch, Leistung, Familienbeihilfe, Antrag, Kinderbetreuungsgeld, Wochengeld, Krankenversicherung* and *geldleistungsberechtigt.*

The higher rate of technical terms in the German corpus can be explained by the fact that English compound nouns, such as *health insurance*, are very often not closed or hyphenated, whereas in German they form a single, longer word like *Krankenversicherung*. *Health insurance* would then be listed as two separate words that do not belong into the category 'technical' or 'semi-technical'. To detect compound nouns in the English corpus and give information on the influence this has on the wordlist, a special analysis would have to be performed.

4.5.4 Recommendations for the curriculum

Regarding modal auxiliary verbs, some topics could be identified that should not be missing in the curriculum. It is suggested that units which cover these topics should not teacher-centered and pure grammar lessons, but that they should be embedded and presented in context.

4.5.4.1 Modal auxiliaries 'in general'

The epistemic, deontic and dynamic uses and meanings of the modals should be presented and explained. Especially, because potential students need to be able to give information on different social security topics, it might be a good idea to present the auxiliaries for expressing obligation, permission, requirements, possibility of different degrees, suggestion or advice. These terms could easily be presented in the context of giving advice to customers and telling them what rights and obligations they have.

In this context, the teacher should also review the differences in grammatical behavior from lexical verbs, such as the subject-auxiliary inversion, the behavior in negations or the lack of inflectional forms.

4.5.4.2 Conditional clauses / will / would

The review of conditional sentences using *will* is an absolute necessity. It is suggested, though, that also the hypothetical conditionals and the past conditionals are given consideration.

In connection with the non-modal auxiliary verbs some more topics that seem worth teaching could be detected. As with the modal auxiliaries, the topics should be presented using vocabulary that appears in the corpora and also choosing examples from the corpus texts.

4.5.4.3 Passive

As already emphasized the passive voice as a topic should not be left out and its frequent use in formal language should be stressed.

4.5.4.4 Present progressive tense

Is as well as *are* were used repeatedly as auxiliary verbs in the present progressive tense. Hence, this verb form and its usage should be taught or reviewed.

4.5.4.5 Progressive aspect of tenses

In the course of a review of the present progressive, it might be a good idea to explain the differences between the simple and progressive tense forms once more. Many students are not aware that progressive tenses refer to temporary situations or situations that are not complete and subject to change. In German there is no tense form that expresses these notions, which makes it extremely difficult for German speakers to understand the concept.

4.5.4.6 Present perfect tense

Another aspect that is missing in the German language is the perfect aspect to express possible continuation, experience or results. The perfect tenses are the most difficult tenses to understand for German speakers, which is why they should be included in any curriculum of any ESP course. The fact that the present perfect tense was used very often in the corpus only confirms this.

4.5.4.7 Terminology

Due to the unexpected results, it was quite difficult to make a decision on which terms would be worth investigating further and need to be taught and which would not. Hence, for the decision my intuition and experience had to complement results of the corpus research. Based on the research results, I suggest teaching the German words *Anspruch, Leistung, Versicherter, Familienbeihilfe, Antrag, Kinderbetreuungsgeld, Wochengeld, Krankenversicherung, besteht, erhalten, beziehen, gelten, beantragt, bestimmen, erworben, bezahlen, übernehmen, grundsätzlich, zuständig, gesetzlich, kostenlos, erforderlich, freiwillig, entsprechend, geldleistungsberechtigt* and *notwendig.* Following English terms might be new to prospective students and should be taught: *benefit, drug, coverage, allowance, pay, insurance, claim, provide, join, cover, available, due, approved, entitled* and *eligible*.

Additionally, I would include words that are used often in different kinds of documents, words that - even if not on the same form – appear as different word categories such as nouns and verbs and where I thought that presumably interesting exercises could be created as soon as the results are available.

To ensure that all forms of a word were displayed, only as many letters as necessary to get all forms were typed in followed by an asterisk. For the first round the following terms were entered: accord*, allow*, appl*, approv*, authori*, benefi*, claim*, contribut*, counsel*, cover*, entitle*, insur*, oblig*, period*, provid*, requir*, right*, support*

Below you will find Tables 14-17 with the search results for some of the terms mentioned above. The table has been designed in such a way that whoever has to plan a course or design exercises can use it for their purposes. It shows the different forms of the words that appeared in the corpora, the forms that might be interesting to teach, clusters or collocations which appear frequently and seem helpful for the learners as well as example sentences containing the terms taken from the corpora. The tables also include definitions according to the Oxford English Dictionary online edition[16] as a help for preparing exercises where a

[16] www.oed.com

mixture of new and familiar words have to be matched to their definition(s). For some terms, also example sentences retrieved from the COCA were provided to show the usage of the respective terms in a more general context if available.

Table 14: 'accord*'

'accord*' (16)	
Forms found	according (3), accordance (13)
Forms that should be taught	according, accordance
Clusters/collocations	in accordance with, according to
Example sentences from corpus	You have the right to be treated with dignity and respect, in accordance with your human rights. (=You have the right to be treated with dignity and respect, in agreement with the human rights laws.) However, the law changes all the time, and rights and duties will also vary according to individual contracts.(= However, the law changes all the time, and rights and duties will also vary as formulated by individual contracts)
Usage in more general contexts	According to Spector (1997), job satisfaction is important for two reasons:[17] According to data from the U.S. Commerce Department, corporate profits were 23.8 percent of total corporate income in 2010[18] We want to help Mr Chen achieve his goal, which is to come here, and we want to do it in accordance with our values[19]
Definitions for terms	according to: as stated or formulated by; in accordance with (also to) : in agreement or harmony with; in conformity to

[17] from: http://corpus.byu.edu/coca/ (220) retrieved on December 30, 2012
[18] from: http://corpus.byu.edu/coca/ (237) retrieved on December 30, 2012
[19] from: http://corpus.byu.edu/coca/ (2) retrieved on December 30, 2012

Table 15: 'allow*'

'allow*' (393)	
Forms found	allowance(s) (344), allowing (3), allow (8), allows (11) allowed (17)
Forms that should be taught	allowance(s)
Clusters/collocations	Jobseeker's Allowance, Disability Living Allowance, Attendance Allowance, Maternity Allowance, Carer's Allowances, Severe Disablement Allowance, Employment and Support Allowance, allowance period, allowance claim form, allowance on behalf of, allowance is stopped, allowance regulations, claim allowance, get allowance
Example sentences from corpus	You can get Attendance Allowance even if no-one is helping you to look after yourself. Disability Living Allowance is money to help you with extra costs you may have because you are disabled
Usage in more general contexts	Taylor spent his allowance money on a radioactive dining set[20]
Definitions for terms	A fixed sum of money provided for expenses; *spec.* a regular fixed sum given by a parent, guardian, etc., to a person who has no other significant income.

Table 16: 'approv*'

'approv*' (109)	
Forms found	approve (1), approval (7), approved (101)
Forms that should be taught	approve, approval, approved
Clusters/collocations	approved by, approved amount, approved treatment, approved training, approved application, getting prior approval
Example sentences from corpus	It is important that you follow the plan's rules, like getting prior approval for a certain service when needed. A course provided by an employer as part of a job contract doesn't count as approved. If your application is approved, the letter will show the amount of your benefit and when your payments start.
Definitions for terms	approve: To confirm authoritatively; to sanction. approved: Pronounced good; justified, sanctioned, commended, esteemed. approval: The action of approving; sanctioning approbation.

[20] from: http://corpus.byu.edu/coca/ (20) retrieved on December 30, 2012

Table 17: 'appl*'

'appl*' (262)	
Forms found	application(s) (49), apply (107), applies (71), applied (10), applicable (3), applying (19), appliances (1)
Forms that should be taught	application, apply, applies.
Clusters/collocations	application for, application is approved, application is successful, application is not approved, application is unsuccessful, application on behalf of, earlier application, process your application, sign the application, review the application, submit the application, complete the application, you can/may/must/should/could apply, rules may apply, how to apply for, apply for, apply to treatment, apply as soon as possible, apply by phone, apply online, applies if, applies to
Example sentences from corpus	We will review your application to make sure you meet some of the basic requirements for disability benefits. If you or a member of your family live or work in another European Economic Area Country or Switzerland, different conditions may apply. There are circumstances when, by law, you must apply for an NI number. For SMP purposes, your employment can be taken as continuous if any one of the following applies: …
Usage in more general contexts	The meeting was designed to determine why community colleges are not able to apply successfully for federal funding for computing programs[21].
Definitions for terms	application: A formal request to be considered for a position, licence, etc., submitted to an authority or institution apply: to have valid or suitable reference *to*; to make a (formal) request *for*.

Based on the insights into terminology gained from the current corpus analysis I suggest teaching the 20 most frequently used nouns, verbs and adjectives in an ESP course as they seem to be essential in the relevant texts. For this it would be advisable to research the terms the way that has been described above, prepare tables for each of them and create different kinds of exercises to clarify meaning and different forms.

As teaching vocabulary – as any other topic - out of any context is always problematic, there should be a clearly defined aim. For teaching terminology I find it always quite inspiring and challenging at the same time to prepare a text on a

[21] from: http://corpus.byu.edu/coca/ (1) retrieved on December 30, 2012

specific topic. After having raised the awareness for the topic by asking students what they know about it or if they have any experience with it, there are various kinds of exercises that help in the acquisition of new words. The teacher has to decide beforehand which terms might be new to the students and which terms they already know.

- Matching exercises

  To clarify meaning of new terms, exercises where definitions have to be matched to the words are very useful. It is advisable to prepare exercises where familiar words are mixed with new terminology so that students can work using the method of elimination, and that they do not feel de-motivated.

- Example sentences

  The words should then be presented in context to show students with what words the newly learned terms usually co-occur and collocate, and to consolidate meaning.

After having done these exercises, the text could be presented to the course participants with the aim to fulfill specific tasks. Here it may be a good idea to let them read through the text quickly and answer one very general question about the text, having done that they should read through the text once more and answer 3 or 4 very specific questions.

- Introduction of different forms

  When the text has been read and the questions have been answered, meaning of the new words should once more be clarified, e.g. by asking 'concept questions'. If the students feel confident, other forms can be introduced. They can for example be elicited from students or be provided by the teacher.

- Gap fill exercises

  When learners are familiar with the different forms a term can have, gap-fill exercises where learners have to decide which form of the word can be used in the respective sentences could be provided.

In a follow-up lesson, a similar text with gaps could be presented to the students.

4.5.4.8 General suggestions

The suggestion by Robinson (see 2.1. page 8) that specialist terminology is not a necessary part of an ESP course can to some extent be confirmed. Research has shown that there are many other topics that need to be introduced to course participants. As mentioned in 2.4, teachers need to design curricula that cover aspects of the field of expertise, terminology and communication patterns that are typically used within the learners' discourse community. However, a combination of 'specialist terminology', 'general terminology' and any other topics is probably what makes a course efficient and varied. This would also correspond to Basturkmen's contention (see 2.4.4, page 19) that students' motivation can be maintained by teaching not only very restricted expert language that cannot be used outside the classroom.

A balanced and interesting lesson could then revolve around a specific topic such as 'Maternity Benefit'. After having raised their awareness for the topic and having gone through the stages of vocabulary work as described above, students could be asked to produce sentences for informing women who are inquiring about Maternity Benefit about their rights and obligations. In this way the different modal verbs could be reviewed or introduced and their meaning could be determined.

To connect the vocabulary and grammar work, a role play could be initiated in which one half of the students represent claimants and the other half themselves, and the claimants need to find out in a talk what they need to do and what requirements they have to fulfill to be eligible for Maternity Benefit.

To bridge the "gap" between special and general language mentioned in 2.4.4 it is advisable to introduce the newly acquired vocabulary and grammar also in more general contexts.

5 CONCLUSION

When it is said that Corpus Research is a very useful and quick tool to create teaching material, I can agree only partly. In my experience the process from collecting the texts until teaching material is actually created is very long and asks for meticulous precision in every respect.

First, one has to think about what texts would be useful to include in the corpus to get a wide range of relevant language for the specific topic. This might be not as easy to answer as it seems, because relevance cannot always be determined beforehand, and factors such as balance and representativeness have also be taken into consideration. Even if one has a clear notion of what texts to search for, it is not certain that these kinds of texts can be easily accessed and prepared for use on the computer or in the concordancing program. In the current research, fortunately, all texts were available on the internet and could easily be converted into .txt files to be processed by AntConc.

As soon as the choice of texts has been made, it is advisable to find a suitable system for storing the text files and results on the computer. The idea behind that is that the researchers should be able to quickly find, access and retrieve the documents that are used for the corpus study, and this is only possible with a very clear and simple system. I made the mistake to store the text files under very long names which made it difficult to find documents when I needed them, simply because I could not remember the names and always had to check twice or three times if the file was the one I actually wanted to access. Long file names also complicated the work with AntConc, as the documents always had to be counted to guarantee that all of them are included in the search. So after some time, file names were changed into shorter versions, each file preceded by a serial number, until towards the end of the study the names were replaced by serial numbers. This way, it was visible at one glance whether all documents were opened in AntConc. Another advantage of this system was that now not the whole names of the documents had to be remembered but only the numbers, which was quite convenient and saved much time. To keep track of the texts, an extra

document was created, where the documents were listed by number, with a short explanation what the document contained and where it had been retrieved from. Another, originally unconsidered, advantage of shortening the filenames was that it saved paper, as in concordance lists, file names are always listed at the end of the lines.

The next big step is to get acquainted with the concordancing software, which is obsolete if one regularly works with it. However, for this study it was necessary to read through the AntConc handbook carefully to find out what kind of information could be obtained with the program, what different options there were and how the program needed to be handled. I had worked with AntConc before, but with an older version that was a little different to the one used for the current study. Even though the software is very easy to handle, it took some time to "warm up" and find the way to the things I wanted to find out.

The first form of analysis was to look at the wordlist, as this is the easiest function available, and results not in concordance with what I had expected. As results could not be changed, it was important to leave expectations behind, change the point of view and see what the list had actually to tell. This was an interesting expedition and it also needed some time, but finally revealed insights into what grammar topics need to be taught in an ESP course for employees in the Social Security sector, which was not necessarily intended to be found out.

Is Corpus Analysis a suitable tool or method to create syllabi or teaching materials? It could be – to some extent. No matter how manifold options of analysis are, a corpus is, on the one hand, only as useful as its user is accurate and creative, and, on the other hand, a corpus alone is not sufficient for creating a syllabus or material. Aston (2001: 5) emphasizes that

> [c]orpora can provide teachers and learners with information about the language and the culture which can complement and integrate that available from other sources, such as textbooks, teachers and reference materials. This can effectively increase their competence and awareness, as well as developing their ability to evaluate information from other sources more critically.

Corpora are also obviously not the appropriate place of finding the definitions of terms. If one needs to know what a word means, dictionaries and encyclopedias

are a better source. And as Gavioli suggested (3.5.2, pg 41) also grammars have to be consulted to optimize the preparation- and teaching process. However, corpora provide us with examples and give us contextual clues other sources lack. If teachers consult grammars and dictionaries they are provided with some indications on certain matters, "but concordances can provide much more information, since they supply more examples and can highlight relatively subtle tendencies" Partington (2001: 82). Partington also highlights the fact that there are apparent limits to the information that can be obtained from corpora like the ones I have used for this study. One limit has to do with the size and he states that in small corpora "rare words occur too infrequently in a small corpus to be studied with any confidence" (2001: 84). A corpus, then, should be considered as a complimentary tool rather than the only source.

If one goes through all steps carefully and works accurately, Corpus Analysis is very time intensive and lesson planners have to be exact observers. As the results might deviate from what is expected, some extent of flexibility is also necessary, and one might not find the teaching content that is actually expected or needed. For this study it was expected to get real obvious and clear results regarding terminology, which it did not. This strongly supports Hutchinson and Water's observation that ESP teaching is not just a matter of teaching specialized words (see 2.1, pg 7).

Corpus Research is definitely very revealing and inspiring in the sense that it shows regularities that would not have been found without the concordancing program, and it can be recommended if it is used to design material for parallel courses or for long-term training. A corpus shows quite well what needs to taught without spending endless hours reading the texts one by one and it is a useful source for finding example sentences, researching the different forms of terms that are used in this expert field, and it can be used over and over again for different purposes. Grammatical peculiarities can be detected and considered for implementation into the curriculum. However, due to the work involved, it does only seem a sensible method to use, if material can be used numerous times. I would strongly advise against performing Corpus Research with a DIY corpus if

materials have to be designed for only one or two courses. Even if one only wants to make use of the wordlist tool to give a hint to what terms should be taught, searching for the texts, balancing them, converting them, listing them, etc. seems much too elaborate. If a teacher also considers looking for grammatical constructions and finding out about differences in usage, research would take three times or four times longer than the actual teaching, and this does not seem to be worth the effort.

Should one consider performing Corpus Research for short-time courses, it would be much easier to use annotated and marked-up corpora, which are probably not available for many domains.

Another aspect is that Corpus Research definitely asks for a certain amount of linguistic expertise to be able to interpret the results correctly, and there is also the danger of getting "carried away", and as Kennedy (1991: 367) puts it, it is important to "distinguish between what is scientifically interesting and what is pedagogically useful." Aston, Bernardini and Stewart (2004: 13) also addressed this fact when they stated: "[…] corpora are great sources of serendipitous findings, as we all know." This, in my experience, is on the one hand, the greatest advantage of performing Corpus Research and, on the other hand, it is at the same time the greatest disadvantage. If one starts this 'expedition' without knowing where it will lead to, distractions are preprogrammed. While working on a specific topic and checking lists, another peculiarity or interesting fact appears and withdraws the attention from what one was working on. It is crucial to avoid investigating every interesting aspect immediately and to take notes of the observations made and return at a later point of time.

The original idea to use and exploit the corpus mainly for retrieving terminology was discarded quite soon, as so many other relevant topics appeared and literally asked for examination.

For me, personally, the research was very helpful and I will be able to use the results in my courses in the AMS and the Hauptverband. Grammar topics that were never really considered, such as the use of modal verbs or the passive voice, will definitely be incorporated into my curricula from now on.

Especially because of this 'oversupply' of potential topics it is extremely important to never lose focus. It is interesting from the researcher's point of view to follow a different track from time to time if it seems promising. However, one has to be careful not to go astray and always recall the original intention of conducting the analysis.

In the course of this study it has become very clear that creating teaching material for an ESP course with the help of specialized corpora does not make the preparation process easier if teachers are confronted with a domain they are completely unfamiliar with. Even the set-up process of a study is much easier if one has some expertise in the respective industry.

My advice to ESP teachers would be to find a domain that interests them and specialize, because then the many hours invested in Corpus Analysis are more than worth it.

6 BIBLIOGRAPHY

Aston, Guy. 2001. Learning with Corpora. Bologna: CLUEB

Aston, Guy; Bernardini, Silvia; Stewart, Dominic. 2004. *Corpora and Language Learners*. Amsterdam: John Benjamins Publishing Company.

Barnard, Roger and Zemach, Dorothy. 2003. "Materials for Specific Purposes". In: Tomlinson, Brian (ed) 2003. *Developing Materials for Language Teaching*. London: Continuum, 306-323.

Basturkmen, Helen. 2010. *Developing Courses in English for Specific Purposes*. New York: Palgrave Macmillan.

Cook, G. 1998. "The uses of reality: a reply to Ronald Carter". *ELT Journal 52/1*, 57-64.

Dudley-Evans, Tony. 1998. *Developments in ESP: A multi-disciplinary approach*. Cambridge: Cambridge Univ. Press.

Flowerdew, John. 1993. "Concordancing as a Tool in Course Design". *System* (21); Vol. 2, 231-244.

Garcia Mayo, Maria del Pilar. 2000. *English for Specific Purposes: Discourse Analysis and Course Design.* Bilbao: Universidad del País Vasco, Servicio Editorial.

Gavioli, Laura. 2005. *Exploring Corpora for ESP Learning*. Studies in Corpus linguistics (21). Amsterdam: John Benjamins Publishing Company.

Holme, R. 1996. *ESP Ideas*. Harlow: Longman.

Howatt, A.P.R. 1984. A *History of English Language Teaching*. Oxford: Oxford University Press.

Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hutchinson, Tom; Waters, Alan. 1987. *English for Specific Purposes. A learning-centred approach*. Cambridge: Cambridge University Press.

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

Kennedy, Graeme. 1991."Preferred ways of putting things with implications for language teaching". In Svartvik, Jan (ed), 1992. *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82. Stockholm, 4-8 August 1991*. New York: Mouton de Gruyter, 335-373.

Kim, Dan. 2008. *English for Occupational Purposes. One Language?* London: Continuum**.**

Leech, Geoffrey. 2004. *Meaning and the English verb.* (3rd edition). Harlow: Pearson Longman.

McEnery, Tony; Xiao, Richard; Tono, Yukio. 2006. *Corpus-based language studies. An advanced resource book*. London: Routledge.

Partington, Alan. 2001. "Corpus-based Description in Teaching and Learning". In: Aston, Guy (ed). *Learning with Corpora*. Bologna : CLUEB, 63-84.

Rogne, Leah; Estes, Caroll L.; Grossmann, Brian R.; Hollister, Brooke A. and Solway, Erica (ed). 2009. *Social Insurance and Social Justice. Social Security, Medicare and the Campaign Against Entitlements*. New York: Springer Publishing Company.

Robinson, Pauline. 1991. ESP Today: A Practitioner's Guide. New York: Prentice Hall.

Sinclair, John et.al. 1990. Collins Cobuild English Dictionary – New Edition. London: Harper Collins.

Sinclair, John M. 1991."The automatic analysis of corpora". In Svartvik, Jan (ed), 1992. *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82. Stockholm, 4-8 August 1991*. New York: Mouton de Gruyter, 379-397.

Steiner, Günther. 2011. *Johann Böhm in der österreichischen Sozialversicherung*. Wien: Verlag des ÖGB Gmbh.

Tudor, I. 1997. "LSP or Language Education?". In: Howard, R. and Brown G. (eds). *Teacher Education for LSP*. Clevedon: Multilingual Maters, 90-102.

The Oxford English Dictionary. Online edition. Oxford University Press. www.oed.com (29 December, 2012)

Widdowson, Henry G. 1978. *Teaching Language as Communication*. Oxford: Oxford University Press.

Widdowson, Henry. G. 1998. "Communication and community: The pragmatics of ESP". *English for Specific Purposes* 17(1), 3-14. (Extended version of Widdowson 2003)

Yates, F. 1965. *Sampling Methods for Censuses and Surveys*. 3rd edition. London: Carles Griffin & Company Limited.

7 APPENDIX

Texts used for the English corpus

Name of document	**Address**	**No of words**
UK attendance-allowance-leaflet.pdf	http://www.nidirect.gov.uk/attendance-allowance-leaflet-aa5dcs.pdf	2931
UK disability-living-allowance-leaflet.pdf	http://www.nidirect.gov.uk/disability-living-allowance-leaflet.pdf	3836
UK Health	http://www.nhs.uk/choiceintheNHS/Rightsandpledges/NHSConstitution/Documents/nhs-constitution-interactive-version-march-2012.pdf	34367
UK Info Appl NI Number	http://www.nidirect.gov.uk/applying-for-a-national-insurance-number-nic.pdf	1074
UK Jobseeker's Allowance	http://www.direct.gov.uk/prod_consum_dg/groups/dg_digitalassets/@dg/@en/@benefits/documents/digitalasset/dg_199994.pdf	685
UK state-pension-claim-form	http://www.nidirect.gov.uk/state-pension-claim-form-br1	2950
US Appl Soc Sec Card	http://www.socialsecurity.gov/ssnumber/ss5.htm	483
US Disability benefit	http://www.ssa.gov/pubs/10029.pdf	3434
US Government pension questionnaire	http://www.socialsecurity.gov/online/ssa-3885.pdf	1190
US Info Appl Sosec no	http://www.socialsecurity.gov/ssnumber/ss5doc.htm	3299
US unemployment benefits_ Eligibility for Benefits_.docx	http://www.socialsecurity.gov/policy/docs/progdesc/sspus/unemploy.pdf	2099
USA Medicare	http://www.socialsecurity.gov/pubs/10043.pdf	3484
USA Medicare long	http://www.medicare.gov/pubs/pdf/10050.pdf	38272
UK Child Benefit.docx	http://www.nidirect.gov.uk/index/information-and-services/money-tax-and-benefits/child-benefit-and-child-trust-fund/keeping-your-child-benefit-up-to-date/child-benefit-when-your-child-is-aged-16-or-over/child-benefit-if-your-child-is-in-further-education-or-training.htm	3220

UK Maternity benefits	http://www.nidirect.gov.uk/nil-17a.pdf	19480

Texts used for the German corpus

Name of document	**Address**	**No of words**
Pflegegeld	http://www.bva.at/mediaDB/MMDB116686_30%20-%20Informationsblatt%20Bundespflegegeld.pdf	1334
Pflegegeld	http://www.pensionsversicherung.at/mediaDB/838349_Pflegegeldinformation%202012.pdf	755
Krankenversicherung	http://www.wgkk.at/mediaDB/648107_Ratgeber.pdf	17983
Pensionsversicherung	http://esv-sva.sozvers.at/mediaDB/655795_B%206.pdf	3185
Arbeitslosengeld	http://www.ams.at/sfa/14080_18650.html	1128
Familienbeihilfe	http://bmwa.cms.apa.at/cms/content/attachments/8/9/2/CH0567/CMS1058739520786/familienbeihilfe_mehrkindzuschlag_2009.pdf	3708
Pensionsantrag	http://www.pensionsversicherung.at/mediaDB/667229_EigpenKEZ.pdf	2295
Arbeitslosenvers SVA	http://esv-sva.sozvers.at/mediaDB/671463_Arbeitslosenversicherung.pdf	813
Krankenvers SVA	http://esv-sva.sozvers.at/mediaDB/641573_B%204.pdf	6768
Kinderbetreuungsgeld	http://www.arbeiterkammer.at/online/kinderbetreuungsgeld-varianten-53313.html	940
Wochengeld	https://www.help.gv.at/Portal.Node/hlpd/public/content/8/Seite.082100.html	741

7.1 Zusammenfassung

Die vorliegende Masterarbeit versucht die Frage, ob Corpus Analysis / Corpus Research eine geeignete Methode für die Erstellung von Unterrichtsmaterial für "English for Specific Purposes" (ESP) Kurse im Sozialversicherungsbereich ist, zu beantworten.

Die Autorin bietet im ersten Teil eine Definition des Begriffes ESP, gibt einen Überblick über die Entwicklung dieser Richtung des "Teaching English to Speakers of Other Languages" (TESOL) und geht dann besonders auf die Herausforderungen der ESP Trainer ein, die sehr oft auch "Non-Native Speaker" sind. Im Besonderen werden die Schwierigkeiten behandelt die in Zusammenhang mit der Vorbereitung auf ständig neue Spezialgebiete entstehen, aber auch die Verantwortung die ESP Trainer gegenüber ihren Schülern wahrnehmen sollten. In weiterer Folge, bietet die Autorin einen Überblick über die verschiedenen Stadien und Aspekte, die beim Kursdesign und der Entwicklung von Unterrichtsmaterialien zu beachten sind.

Im Hinblick auf den empirischen Teil der vorliegenden Arbeit, wird im zweiten Teil ein theoretischer Überblick über den Bereich von Corpus Linguistics geboten. Er bietet Begriffserklärungen, stellt verschiedene Arten von Corpora vor, erklärt kurz wie Corpora erstellt werden, was man mit ihrer Hilfe herausfinden kann und diskutiert zum Schluss die Vor- und Nachteile der Methode.

Der empirische Teil der Thesis beinhaltet die Beschreibung einer Corpus Analyse, die von der Autorin durchgeführt wurde. Anfangs werden die Beweggründe für die Durchführung der Studie beschrieben, danach der Aufbau derselben erklärt und schließlich wird auch die verwendete Software und deren Funktionen vorgestellt. Anschließend werden die Resultate präsentiert und eine Interpretation derselben geboten. Anhand der Ergebnisse schlägt die Autorin einige Unterrichtsthemen vor, die in einem ESP Kurs im Sozialversicherungsbereich behandelt werden sollten und gibt abschließend einige Tipps zur Präsentation von neuer Terminologie.

7.2 Lebenslauf

Name:	Anches Michaela **Holzweber**
akad. Titel:	Mag. phil.
Geburtsdatum:	01-08-1968
Staatsbürgerschaft:	österreichisch
Kinder:	eine Tochter, Rebecca, geboren 1995

Bildungswerdegang:

seit 10/2010	Master Studium an der Universität Wien "English Language and Linguistics" mit Spezialisierung auf ESP, Abschluss voraussichtlich Februar 2013
11/1999	Abschluss des Studiums an der Uni Wien, Studienrichtungen Geschichte / Anglistik und Amerikanistik
1989:	Abschluss des "Lehrgang für Werbung und Verkauf" an der Wirtschaftsuniversität Wien
1987:	Matura im musischen Zweig des BORG 1010 Wien, Hegelgasse 12

Zusatzqualifikationen:

seit 09/2006	Diverse interne Workshops der FH Burgenland zum Thema Cambridge Exam Preparation, Invigilation, Assessment, etc.
12/2004	Abschluss des CELTA Kurses (Cambridge University; Certificate for Teaching English to Speakers of Other Languages) am bfi Wien (grade „B")

Berufliche Erfahrung

	Tätigkeit als selbständige Sprachtrainerin
2004 - 2013	Leitung von Fachenglischkursen für MitarbeiterInnen des Hauptverbandes der österreichischen Sozialversicherungsträger, des AMS Wien (Rechtsabteilung, Infozone, Servicezone, Beratungszone, Geschäftsleitung) und AMS Niederösterreich (Infozone, Servicezone und Beratungszone)
2009 - 2013	Leitung von Businessenglish Kursen für MitarbeiterInnen von Adfors, Liebherr, Kirchdorfer Industries, MABA Track, DELTA BLOC, MABA Fertigteilindustrie undHPN (Netzwerkspezialisten)
2007 - 2011	Leitung von Fachenglischkursen für MitarbeiterInnen der Abteilung Technische Produktion bei PALMERS
2006 - 2012	Leitung von Fachenglischkursen für die Mitarbeiterinnen der Personal- und Rechtsabteilung sowie Englischtraining für die Lehrlinge aller drei Lehrjahre im Bereich Büro bei METRO Cash& Carry
2004	Unterricht der 2. und 3. Schulstufe AHS am „INFINUM" (Institut für individuelle Unterrichtsmethoden)
2003 - 2006	regelmäßige Leitung von Business English Kursen für „Biz Talk" und „Venetia"
2003	Gründung der AustrianLanguageCompany
1998 - 2000	freiberufliche Tätigkeit als Englischlehrerin für (Klein)kinder (Helen Doron Early English)

	Tätigkeit als unselbständige Sprachtrainerin
2010 - 2012	Lektorin am ITM, International College of Tourism and Mangement
2006 - 2013	Lektorin an der Fachhochschule Eisenstadt, Lehrgang für Internationale Wirtschaftsbeziehungen und Lehrgang für Internationales Weinmanagement
2005 - 2011	Leitung von Englisch Kursen am bfi Wien,
	Leitung eines „Political English" Kurses an der Sozialakademie in Mödling (organisiert vom bfi) im Rahmen der Ausbildung von Betriebsräten
2005	Sprachtrainerin bei IKEA Vösendorf
	Tätigkeit als selbständige Übersetzerin und Lektorin
2006 - 2013	für verschiedenste Firmen in unterschiedlichsten Branchen, z.B.: Volvo Austria, Bentley Austria, Via Donau, VCE, FCP, MCC, Back Office Service für Banken, Die Hochzeitshummel, Bundesjugendvertretung, Minigolf Company, RSB, Gulliver's Reisen, und etliche Privatpersonen
	Tätigkeit als unselbständige Übersetzerin
1998 - 1999	Mitarbeiterin bei *Translex*, Büro für juristische Fachübersetzungen;
	andere unselbständige Tätigkeiten
2001 - 2003	Marketing Managerin und Übersetzerin bei den Schreibprofis, einem Bürodienstleistungs-anbieter mit Spezialisierung auf Rechtsanwaltskanzleien
06-09/2001	Vorstandssekretärin bei Marwah & Partner, internationale IT Firma, als einzig Deutschsprachige unter englischsprachigen Mitarbeitern verantwortlich für Übersetzung sämtlicher Verträge, Korrespondenz etc.
05/1999-06/2001	Assistentin bei Baxter AG, internationaler Pharmakonzern, im Bereich Virussicherheit/Präklinische Dokumentation (Übersetzungen E-D, D-E und allgemeine administrative Aufgaben)
1995-1997	Karenzurlaub
1995	März bis August: Mitarbeiterin im Sales- und Marketingbereich bei EF (Europäische Ferienschule)