



universität
wien

MASTERARBEIT

Titel der Masterarbeit

„A genre analysis of legal discourse.

Corpus-based analysis of confidentiality agreements“

verfasst von

Milena Lazarevic

angestrebter akademischer Grad

Master of Arts (MA)

Wien, 2013

Studienkennzahl lt. Studienblatt:

A 066 812

Studienrichtung lt. Studienblatt:

MA English Language and Linguistics

Betreut von:

Ao. Univ.- Prof. Mag. Dr. Ute Smit

Acknowledgments

I would like to express my gratitude to my supervisor, Dr. Ute Smit, whose expertise, understanding, and patience added considerably to my studying experience. I would also like to thank my parents and my sister Milica for all the support they provided me. A very special thanks goes to Wolfgang without whose motivation and encouragement I would not have succeeded. Finally, I would like to thank all my friends, in particular Andreas, Marija and Wendy.

Table of contents

1. Introduction.....	1
2. Theoretical and methodological issues of genre analysis	3
2.1. The development of the genre-based approach to written discourse 6	
2.1.1. The Systemic Functional approach.....	8
2.1.2. The New Rhetoric approach	9
2.1.3. The ESP approach.....	10
2.2. A conceptualization of discourse community.....	10
2.3. The methodology of genre analysis	12
2.4. Move analysis	17
3. Legal discourse in professional settings	19
3.1. Research into legal language	19
3.2. The concept of legal language.....	20
3.3. Genres of legal discourse.....	22
3.4. The characteristics of legal language.....	24
4. The study.....	27
4.1. The field of contracts.....	27
4.2. The genre of confidentiality agreements	30
4.2.1. The discourse community	31
4.3. The methodology	32
4.3.1. The data.....	33
4.3.2. The tools used for the corpus analysis	34
4.3.3. Vocabulary of the corpus	35
4.3.4. Concordance-derived analysis	37
4.3.5. Word list-derived analysis	38

4.3.6.	Keywords-derived analysis.....	38
4.3.7.	Formulaic language.....	39
5.	Analysis of lexico-grammatical features of the texts.....	41
5.1.	Lexical features.....	41
5.1.1.	Vocabulary of the corpus of confidentiality agreements	41
5.1.2.	Academic vocabulary of the confidentiality agreements.....	42
5.1.3.	Technical vocabulary of the confidentiality agreements	43
5.1.4.	Common words with uncommon meanings	45
5.1.5.	Wordlist of frequently occurring items.....	47
5.1.6.	Keywords.....	48
5.1.6.1	Concordance of the keyword <i>or</i>	50
5.1.6.2	Concordance of the keyword <i>agreement</i>	51
5.1.6.3	Concordance of the keyword <i>confidential</i>	53
5.1.6.4	Concordance of the keyword <i>any</i>	54
5.1.6.5	Concordance of the keyword <i>party</i>	55
5.1.6.6	Concordance of the keyword <i>shall</i>	56
5.1.6.7	Concordance of the keyword <i>company</i>	58
5.1.6.8	Concordance of the keyword <i>information</i>	59
5.1.6.9	Concordance of the keyword <i>employee</i>	60
5.1.6.10	Concordance of the keyword <i>employment</i>	61
5.1.7.	Antiquated terms.....	62
5.1.8.	Loanwords and expressions	63
5.1.9.	Capitalization	63
5.1.10.	Some other lexical features of the confidentiality agreements.....	64
5.2.	Syntactic features.....	65
5.2.1.	Sentence length and complexity	65
5.2.2.	Binomial and multinomial expressions.....	66
5.2.3.	The use of passive.....	71
5.2.4.	The use of qualifications.....	73

5.2.5. The use of nominalisation.....	74
5.3. Discourse-level features.....	75
5.3.1. Reference as a cohesive device.....	76
5.3.2. Some other cohesive devices	77
5.4. Summary of findings	77
6. Move analysis of the genre	79
6.1. Introduction	80
6.2. Recitals.....	82
6.3. Definitions	85
6.4. Obligation.....	87
6.5. Housekeeping Provisions (Miscellaneous Provisions).....	90
6.5.1. Governing law clause.....	91
6.5.2. Remedies clause.....	92
6.5.3. Entire agreement clause	93
6.5.4. Severability clause	93
6.5.5. Assignment clause	94
6.5.6. Amendment clause.....	95
6.5.7. Term.....	95
6.5.8. Optional steps of the move housekeeping provisions.....	96
6.6. Signatures and dates.....	97
6.7. Summary of findings	98
7. Conclusion	100
Abstract	102
Zusammenfassung.....	104
References:	106
Appendix	114
Curriculum Vitae	120

1. Introduction

According to Bhatia (1993: 101), legal language “has long been criticized for its obscure expressions and circumlocutions, long-winded involved constructions and tortuous syntax, apparently meaningless repetitions and archaisms”. In order to try to explain this complexity, this thesis examines legal language and in particular the language of contracts. Although very often incomprehensible to lay people, “the drafters of this specific genre, construct certain linguistic choices in order to bring in ‘precision, clarity and unambiguity and all-inclusiveness” (Bhatia 1993: 102). “Language is seen as a rich variety of options available for the sender to select the linguistic means which may best further his/her aims and achieve the receiver’s cooperation to that end” (Trosborg 1997: 57). In this study, I will look at what linguistic choices the drafters of legal contracts construct in order to create legal obligations between the parties.

In addition to this and with regard to the language of contracts, another point is worth mentioning. “The lawyer uses what are called in the profession ‘formbooks’ in which documents or paragraphs of document are set out, and it is up to the lawyer to choose the appropriate paragraph or paragraphs for the particular document s/he is drawing up” (Stubbs 1983: 485). Due to the fact that the contracts are often not drafted afresh, but adopted from existing material, I would like to explore how the language reflects this reality and whether the contracts have unique features that are typical across the range of texts. In order to achieve this, apart from looking at the linguistic features, I will also examine how the texts are structured and how they make use of formulaic language.

For the purpose of finding the typical features of the genre of contracts, a corpus-based analysis of thirty confidentiality agreements was conducted. As this genre is considered to be strongly conventionalised, proto-typicality could be established with a smaller corpus (Hüttner 2007: 109). One of the main objectives of this study is to identify what features are specific to legal contracts and in particular, to the genre of confidentiality agreements and how they differ from ‘ordinary language’. These features are identified by using the software which facilitate the corpus analysis, namely the WordSmith Tools and RANGE programmes.

The first part of this thesis is concerned with the theoretical and methodological issues relevant to the research. In the first chapter I will introduce the main concepts of genre analysis and describe its theoretical background. Three different approaches, as well

as the methodology of genre analysis used in this study will be outlined. The second chapter covers different aspects of legal discourse in professional settings. I will look at the research done in the field of legal language, as well as the different legal genres and characteristics of the language. The third chapter deals with the genre of contracts and in particular its sub-genre of confidentiality agreements. Additionally, the methodology employed in this study will be outlined. In the following chapter I will describe the theoretical framework for the analysis, including the significance of different types of vocabulary. The second part of the study describes the findings of the empirical research. Chapter five shows results of the in-depth lexico-grammatical analysis of the corpus, presented as lexical features, syntactic features and discourse-level features. Finally, the structural interpretation of the genre will be examined and the moves and the formulaic language of the confidentiality agreements will be identified.

2. Theoretical and methodological issues of genre analysis

As Dias et al. (1999) state (quoted in Swales 2004: 74), a genre-based approach is „the most developed and comprehensive rhetorical theory to address writing in recent times“. This approach is not only based on mere descriptions of language use, but it offers “an explanation for a specific use of language in conventionalized and institutionalized settings”, trying to answer the question: “why does a particular use of language take the shape it does?” (Bhatia, 2004: 22). Genre analysis as ‘the study of situated linguistic behaviour’ may be seen as:

a reflection of the complex realities of the world of institutionalised communication, or it may be seen as a pedagogically effective and convenient tool for the design of language teaching programmes, often situated within simulated contexts of classroom activities (Bhatia 2002: 3).

The beginning of genre analysis was marked by work of Swales in 1981. In his seminal work on genre analysis (1990: 58), Swales proposes the following definition of a genre:

A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style. Communicative purpose is both a privileged criterion and one that operates to keep the scope of a genre as here conceived narrowly focused on comparable rhetorical action. In addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience.

In these early stages of genre analysis, Swales (1990) and Bhatia (1993) emphasise that the genre is based on its communicative purpose. Swales (1990: 46-47) states that the communicative purpose plays the primary role in determining whether a certain communicative event can be defined as a genre. Bhatia (1993:13-16) additionally notes that although there are many other criteria which influence the genre, such as similarity of form, content, audience, etc., they cannot be the distinguishing factor. The communicative purpose of the genre, which is not always easily identified, shapes the genre and very often there is not only a single purpose but a set of purposes. Bhatia further elaborates that due to the experience or training within the specialist community, the expert members do not only have the knowledge of the specific goals, but they also influence the genre and give it a conventionalized structure (1993:14). The writer of the genre often needs to adhere to the

rules and conventions of the particular genre, and constructs certain lexico-grammatical choices in order to achieve the overall communicative function of the genre in question. This also shows us that certain linguistic features are associated with certain genres but it is not always easy to assign functional values to those features. Additionally, there are often restrictions and constraints which shape and influence the internal structure of the genre.

The concept of communicative purpose has been emphasised as a central point not only by Swales and Bhatia, but also by many others, such as Miller (1984), Martin (1985), Johns (1997), Hyon (1996), among others. However, the focus on communicative purpose as a main or primary criterion has been lately criticized and readdressed. Askehave and Swales (2001: 197) found that “the purposes, goals, or public outcomes are more evasive, multiple, layered, and complex than originally envisaged”. They have referred to Swales (1990: 46) who said that “purpose is a somewhat less overt and demonstrable feature than, say, form and therefore serves less well as a primary criterion. However, the fact that purposes of some genres may be hard to get at is itself of considerable heuristic value”.

Askaheve and Swales (2001: 200) acknowledge that communicative purpose can help the analyst better understand the discourse and show its multi-functionality, as well as that it can help disqualify certain texts as particular genres. However, they say that it cannot by itself help analysts to decide whether a text belongs to a certain genre. Hüttner (2007: 28) refers to Swales (2004: 68-77), who establishes that the communicative purpose is important and should be considered as “part of a discovery process and an outcome of the genre analysis” but not as a main, privileged criterion. In addition to this, she emphasises the importance of the degree of prototypicality of a genre text and how it is ranked. Here, the analyst should consider the information gained from the discourse community with regard to the form and content (Hüttner 2007: 28). She also points out that the information given by the people who use the genre should include the nomenclature they use, and this information all together is to be incorporated in the analysis (Hüttner 2007: 29).

The approach to genre analysis developed by Swales was mainly applied to teaching English for Specific Purposes. Two other approaches were developed at a rather similar time, namely the Australian approach and the one in the field of rhetoric. The three approaches will be discussed in more detail in the text that follows. Bhatia, Flowerdew & Rodney (2008: 10) point out that:

Genre analysis, whether defined in terms of *typification of rhetorical action*, as in Miller (1984), Bazerman (1994) and Berkenkotter and Huckin (1995), *regularities of staged, goal oriented social practices*, as in Martin et al. (1987) and Martin (1993), or *consistency of communicative purposes*, as in Swales (1990) and Bhatia (1993), can be viewed as the study of situated linguistic behaviour in institutionalized academic or professional settings.

These three approaches focus on conventions, as they point out, and we can see the tendency towards more complex analysis of language use and “beyond mere linguistic descriptions to offer explanation for specific uses of language in conventionalized and institutionalized settings” (Bhatia, Flowerdew & Rodney 2008: 10). As Bhatia (2008b: 175) shows, genre theory was until recently concerned more with textual, intertextual and organizational aspects of genres, shown in Swales (1990) and Bhatia (1993), and little on “interdiscursivity and other wider aspects of contextualization”. He says that it was necessary to go beyond the textual analysis of linguistic data and “incorporate a multidimensional and multi-perspective framework for the analysis of academic, professional, institutional and other workplace genres” (Bhatia 2008b: 166). The contextualisation of genre analysis has also contributed to the analysis of more complex professional genres, rather than the convenient, ideal genres (2008b: 166).

Swales (2004: 61-62) emphasizes that genres are frames for social actions, a starting point but no guarantee that a successful rhetoric action will take place. He refers to Beaufort (1998) who distinguishes four areas of knowledge which could shape an expert text, namely the subject matter knowledge, rhetorical knowledge, writing process knowledge and genre knowledge. According to him, genre knowledge is necessary but not a sufficient condition for effectiveness and successful discursal action. Furthermore, Bhatia (2004:25) suggests that when analyzing genres, we should take into account the following, not less significant features of the genre:

1. Although genres are identified on the basis of conventionalized features, they continually develop and change.
2. Genres are associated with typical textualization patterns, yet expert members of professional communities exploit them to create new patterns.
3. Genres serve typical socially recognized communicative purposes; however, they can be exploited or appropriated to convey private or organizational intentions.
4. Although we often identify and conceptualize genres in pure forms, in the real world they are often seen in hybrid, mixed and embedded forms.

5. Genres are given typical names, yet different members of discourse communities have varying perspectives on and interpretations of them.
6. Genres, in general, cut across disciplinary boundaries, yet we often find disciplinary variations in many of them, especially in those used in academic contexts.
7. Genre analysis is typically viewed as a textual investigation, yet comprehensive analyses tend to employ a variety of tools, including textual analyses, ethnographic techniques, cognitive procedures, computational analysis and critical awareness, to name only a few.

In this first part I have looked at some important concepts of genre theory. In the next sections, I will outline the different approaches to genre theory as well as the methodology used in the study.

2.1. The development of the genre-based approach to written discourse

Genre analysis has from the beginning been influenced by many various disciplines and fields, applied and non-applied. Bhatia (2004: 3-26) distinguishes two different views of the development of written discourse. Looking historically at the analysis of written discourse over the past decades, as well as looking at different perspectives, there are certain trends that can be seen in the development of the field. As he points out, in terms of different perspectives, the trends can be defined as: discourse as text, discourse as genre, discourse as professional practice and discourse as social practice. The historical (chronological) development includes the following phases: textualization of lexico-grammar, organization of discourse and contextualization of discourse.

Textualization of lexico-grammar

In the early stages of the development of written discourse, influenced mostly by formalism in linguistics, linguists of the 1960s and early 1970s were very much concerned with studying surface-level features of language use. According to Bhatia (2004: 4-8), this presents the first phase which dealt with textualization of lexico-grammar. Due to the fact that computational procedures were still not developed, the studies were focused mostly on single syntactic features as part of variation studies, as well as on teaching and learning of English for Specific Purposes. Besides, those features were only investigated within clauses and not on a broader scale of discourse. Some of the common syntactic features which were analyzed at that time were, for instance, binominals and multinominals in legal discourse (Gustaffson, 1975), non-verbal combinations (Spencer, 1975), nominalizations

(Bhatia and Swales, 1983), en-participle in chemistry texts (Swales, 1974), to name only a few. This phase presents an important step in the development of discourse analysis as it led to examinations of what is typical in texts which belong to the same genre, as well as to a more multidisciplinary approach and departure from clause-level analysis to larger rhetorical structures.

Organization of discourse

Following this research, linguists turned to exploring organizational patterns of discourse across various academic and professional genres. Those patterns were at first based on structural elements, and only later did Swales (1981, 1990) and Bhatia (1982, 1993) emphasize the relation between these regularities in discourse structure and communicative purposes. The most influential work in the development of genre theory was done by Swales who raised interest in the cognitive structure and Bhatia (1993) who further applied it to other professional genres, considering also the socio-cognitive context (Bhatia, 2004:9-10).

Contextualization of discourse

Further extending the previous research, the focus here was on exploring genres in more detail, particularly regarding:

Purposes: Institutionalized community goals and communicative purposes

Products: Textual artefacts and genres

Practices: Discursive practices, procedures and processes

Players: Discourse and professional community membership (Bhatia, 2004:11).

The analysts acknowledge more the social context and look at the relation of discourse and identities, ideologies, social relations, etc. So, as Bhatia describes it, the development of written discourse analysis started with textual space, then moved into socio-cognitive space, influenced by speech act theory, conversational analysis and teaching ESP and finally into social space (2004: 18). With regard to the four different perspectives of discourse analysis mentioned above, discourse as text, according to Bhatia, deals with formal and functional aspects, such as lexico-grammatical, for instance, rather than with their interpretation or context. The second perspective, discourse as genre additionally involves socio-cognitive analysis and context on a broader scale. Discourse as

professional practice in addition requires professional knowledge and expertise, whereas the final perspective, discourse as social practice requires social and pragmatic knowledge.

In order to provide insights into the development of genre analysis and what different aspects researchers have been taking into account when analysing language use in particular settings, in the next section I will describe the three approaches to genre analysis, namely the Systemic Functional approach, the New Rhetoric and the ESP approach. They are seen as separate approaches, although they do share some similarities. This study mainly focuses on the English for Specific Purposes approach. These three approaches are mainly represented in the works of the following authors:

- American New Rhetoric School (Miller, 1984, 1994; Bazerman, 1994; Berkenkotter and Huckin, 1995)
- Sydney School of Systemic-Functional Approach (Martin, Christy and Rothery, 1997; Martin, 1993)
- British ESP School (Swales, 1981, 1990; Bhatia, 1982, 1993)

2.1.1. The Systemic Functional approach

The Systematic Functional approach or the Sydney School was developed in Australia based on work by Halliday and Martin, and it is “concerned with the relationship between language and its functions in social settings” (Hyon 1996: 693).

The systemic functional approach has designed the theoretical framework which is to describe the “interrelationships between culture, society and language use” (Coffin 2001: 95). According to her, the behaviours, beliefs and values of a social environment “influence and shape both the overall language systems (such as English) and language “instances”, the way people use language in everyday interactions” (2001: 95).

Coffin (2001: 109) states that the concept of genre, developed by Martin together with Rothery (1980, 1981, 1986) reflects Halliday’s theory of the relationship between form, function and context. Martin (1997: 13) says that “as a level of context, genre represents the system of staged goal oriented social processes through which social subjects in a given culture live their lives” (quoted in Coffin 2001: 109). This genre tradition:

focuses on goal orientated social processes and stresses the purposeful, interactive and sequential character of genres, its particular linguistic features, particularly the organization of texts and its links to lexico-grammatical features (Orna-Montesinos 2012: 2).

This approach, as Hyon (1996: 697) further elaborates, compared to the New Rhetoric and the ESP approach, has focused more on primary and secondary schools genres, as well as adult migrant English education, rather than university genres and professional writing. Hyon (1996: 697) names the most distinguished work within this approach by Callaghan, Knapp, & Noble (1993), Christie (1991), Hammond (1987), Joyce, (1992), Martin (1989). As mentioned before, the focus of this approach has been on teaching the social functions and contexts of text, in order to help students succeed (Hyon 1996: 700). Coffin (2001: 109) says that with regard to the educational context, this approach gives emphasis to the texts that are relevant to students' educational and social needs.

Coffin (2001: 109-110) says that this approach proposes that the structure of a particular text of a genre is specified by 'schematic structures' and the texts consist of certain elements or stages which have distinct functions'.

2.1.2. The New Rhetoric approach

The New Rhetoric approach, also known as the North American school has been less educational and more theoretical and research driven (Coffin 2001: 111). She cites Freedman and Medway (1994: 9) who say that the New Rhetoric "places far less emphasis on explicating textual features, its main goal being to unpack complex relations between text and context" (2001: 111). A genre is defined here as follows:

A genre is a socially recognized, repeated strategy for achieving similar goals in situations socially perceived as being similar. A genre provides a writer with a way of formulating responses in certain circumstances and a reader with a way of recognizing the kind of message being transmitted. A genre is a social construct that regularizes communication, interaction and relations. (Bazerman 1988:62).

Hyon (1996: 696) draws upon the work of Bazerman (1988, 1994), Campbell & Jamieson (1978), Coe (1994), Devitt (1993), Miller (1984, 1994), Schryer (1993, 1994), among many others and he says that the focus is more on the "situational contexts in which genres occur than on their forms and have placed special emphasis on the social purposes, or actions, that these genres fulfil within these situations". Coffin (2001: 112) states that

the focus in the New Rhetoric approach is also on understanding the use of genre by particular communities.

This approach has focused less on educational and instructional settings, as mentioned before, but more on “providing descriptions of genres and their contexts and left it up to readers to infer their own teaching applications” Hyon (1996: 703).

2.1.3. The ESP approach

Based on the work of Bhatia (1993), Flowerdew (1993), Gosden (1992), Hopkins & Dudley-Evans (1988), Love (1991), Swales (1990) and Thompson (1994), the ESP school researchers have looked at the genre as a tool for analysing and teaching English in academic and professional settings to non native speakers (referred in Hyon 1996: 695). So, the goal of this approach is mainly educational, and this is where the similarities between the Sydney School and the ESP School lie, although the focus of the ESP school is as we have seen on ESP rather than schools or migrant education.

Hyon (1996: 698) says that the research has shown that, genre-based applications can help non native speakers of English master the functions and linguistic conventions of texts that they need to read and write in their disciplines and professions”. Researchers like Swales (1990), Bhatia (1993) and Dudley-Evans (1994) emphasise the importance of communicative purposes as a determining criteria of genres, as well as the social context of genre. Hyon (1996: 702-703) says that many researchers in the ESP approach have not outlined the methodologies they used for presenting the genre analysis in the classroom, whereas others, like Swales, Bhatia and Flowerdew have given some teaching applications. Swales has for instance analysed some academic genres, Bhatia business and scientific genres while Flowerdew looked at the range of genres in Professional Communication.

2.2. A conceptualization of discourse community

Swales (1990:21) refers to Herzberg who, as early as 1986, wrote that the notion of ‘discourse community’ signifies a number of ideas: “that language use in a group is a form of social behaviour, that discourse is a means of maintaining and extending the group’s knowledge and of initiating new members into the group, and that discourse is epistemic or constructive of the group’s knowledge”. These ideas, although very significant, as Swales points out, do not in fact tell us how we identify those communities, so we need to look further at criteria which help us recognize them as such.

Swales (1990: 24) distinguishes between the concepts of speech community and discourse community, pointing out that the first is concerned with the fact that the determining factors of linguistic behaviour are social, whereas in the case of discourse community they are functional. Members of discourse community pursue certain communicative goals. The example of the speech community can be the speech community of the English-speaking world, where one may be accepted by birth, for instance, whereas to the discourse community by “persuasion, training or relevant qualification”. Swales suggests that a combination of the following set of criteria would lead to the emergence of a discourse community (1990:24-27):

1. A discourse community has a broadly agreed set of common public goals. [...]
2. A discourse community has mechanisms of intercommunication among its members. [...]
3. A discourse community uses its participatory mechanisms primarily to provide information and feedback. [...]
4. A discourse community utilizes and hence possesses one or more genres in the communicative furtherance of its aims. [...]
5. In addition to owning genres, a discourse community has acquired some specific lexis. [...]
6. A discourse community has a threshold level of members with a suitable degree of relevant content and discorsal expertise. [...]

The public goals can be formally written or less explicit but it is essential to note that they are agreed and public. The ways of communication between the members of a discourse community may vary and this communication usually contains specialized terminology, very often unfamiliar to outsiders or novice members. Swales also emphasizes that the existence of a discourse community highly depends on the ratio between novice and expert members and that the conventions of the genre are shaped by the discourse community (1990:24-27).

As an example of a discourse community, Swales introduces a case of a non-academic discourse community, namely the so-called Hong Kong Study Circle - a world-wide organization where most of the members are collectors of Hong Kong stamps. He illustrates how this discourse community fulfils all of the six criteria: the common goal is to promote the knowledge about Hong Kong stamps; communication is achieved by a bi-monthly Journal and Newsletter; there is an information exchange; community specific

genres; very specific lexis not familiar to people out of this circle and a high level of expertise.

Although acknowledging that the concept of discourse community is very important, Hüttner (2007: 38) says that there are certain issues which need to be addressed here. First of all, she says that the status of a researcher has to be defined as either a member of the discourse community or as an observer. Furthermore, the information gained from the expert member of a discourse community is a necessary part of analysing genres. Hüttner (2007: 38) proposes defining clear criteria of implementing such information. The criteria of implementing the information of an expert member will be explained in more detail in the next section, where I look at the methodology of genre analysis, as extended by Hüttner (2007). She addresses the issues of what is considered prototypical by the members of the discourse community and adds that the researcher should not influence the members into providing the prototypical examples of the genre in question. She suggests (2007: 104) that first we need to identify what position the genre has in the lives of the member of the discourse community and identify who the producers and the receivers of the genre are. Throughout the analysis, the analyst is suggested to seek expert advice at several stages, which will be described in the following section.

2.3. The methodology of genre analysis

Having defined the major concepts of a genre theory and described the different approaches, I will now look at how genres have been approached from a practical view. In order to analyze unfamiliar genres, Bhatia (1993: 22-36) suggests a seven-step analysis:

1. Placing the given genre-text in a situational context
2. Surveying existing literature
3. Refining the situational/contextual analysis
4. Selecting a corpus
5. Studying the institutional context
6. Levels of linguistic analysis

Level 1: Analysis of lexico-grammatical features

Level 2: Analysis of text: patterning or textualization

Level 3: Structural interpretation of the text-genre

7. Specialist information in genre analysis

Although this model has been very influential, it has been further developed and extended by Hüttner (2007). In the text that follows, I will describe the extended genre analysis which will be used in this study. I have chosen to follow this model as I find that it indicates potential gaps in the previous research and that this kind of analysis reveals more about the genre in question. The following text will show which parts of the Bhatia's model have been modified or supplemented. According to Hüttner (2007: 107-113) the extended genre analysis includes the following steps:

1. place genre or genre constituent in situational context

In the first step of analysis, Hüttner (2007: 102) introduces the term *genre-constituent*, in addition to genre, which here relates to the “distinguishable parts of genres, for instance, introductions in academic papers or welcoming statements in business reports, where a clear communicative purpose can be identified”. According to Bhatia (1993:22), the particular genre texts first need to be placed in a situational context. This is based on analyst's background knowledge of the discipline, which is usually obtained from experts who belong to the professional community, that is, the people who use the genre in question. This also includes getting familiar with the communicative conventions of the genre and understanding why the language is used the way it is.

2. survey existing literature on the genre in question

If the analyst does not belong to the professional community, he/she needs to research the existing literature in order to get familiar with the genre. This can be achieved by looking at previous linguistic analysis of the genre or genre variety (if available), relevant methods or theories, guidebooks, manuals or relevant interactions among the members of the professional community (Bhatia, 1993:23). After this Bhatia proposes defining the speaker/writer of the text, the audience, historical and socio-cultural context, linguistic traditions and identifying the topic and the extra-textual reality of the genre texts.

3. select a corpus of genre texts identified as belonging to one genre (or genre-constituent) by the relevant discourse community

When the scene for the analysis has been set, the following step, according to Bhatia (1993:23-24) includes selecting the appropriate texts for the analysis. Hüttner (2007:107) adds here that the genre texts should be “identified as belonging to one genre (or genre-

constituent) by the relevant discourse community”. She emphasises that the members of the discourse community should not be influenced into bringing ‘good’ samples, by which she means the typical genre texts. Bhatia further explains that it is essential to note here that the genre must be well defined and that there are clear design criteria for selecting and compiling the corpus. Well-defined criteria help to distinguish texts from other similar genres as well as help to interpret the research data correctly. Further in the analysis and similarly to step two, Bhatia suggests that the researcher should examine various rules and conventions which can influence (and sometimes constrain) the language use of the genre. The institutional context can be acquired from available literature on the genre and can give some important insights into why the language takes the form it does.

4. establish position of genre (constituent) within discourse community, addressing also question of hierarchical position of producers

Hüttner (2007:109) emphasises two relevant issues here. The first one refers to the importance of the genre given by its users. Some genres are, for example, considered more important than others. It is also important to note, according to her, that the producers of the genre text can have different hierarchical positions. In the business context this can be illustrated by the fact that some texts can be produced only by people in higher positions, while the other only by the lower secretarial positions. This can help to identify the importance for teaching certain genres for instance.

5. levels of linguistic analysis:

a) if genre-constituent, define larger genre it is a constituent of

This involves describing the purpose of the larger genre, as well as the genre-constituent (Hüttner 2007:107-108).

b) define genre structure

Phase one: distinguish between core and optional moves based on quantitative analysis

Phase two: refine quantitative model through incorporation of information from members of discourse community

At this stage of the analysis, the moves are identified and the advice on the communicative intentions and the acceptability of move realisations is sought from the expert members (Hüttner 2007:110). This step of the analysis will be described in more detail in section 2.4.

c) analyse lexico-grammatical features including use of genre-specific formulaic sequences

This step corresponds to Bhatia's first level of linguistic analysis of lexico-grammatical features, but in addition to this, Hüttner (2007:110) proposes the identification of genre-specific formulaic sequences. Formulaic sequences will be described in more detail in section 4.3.8. She refers to Tribble (2001) and Scott and Tribble (2006) who outline the use of concordancing tools, which can facilitate the identification of these sequences. According to Bhatia, this step of analysis may include the analysis of syntactic features such as the use of tenses, different types of clauses, prepositional phrases, use of nominalised forms, binominals/ multinomials, etc. It is based on statistical analysis of the corpus and the findings show the occurrence of these features in the particular genre. Although very significant, at this level of analysis the features are analysed but not assigned the values and what communicative purpose they seek to accomplish (Widdowson, 1979). This type of analysis of surface features was more common in earlier stages of development of text analysis. As Bhatia (1993:26) points out, this surface analysis does not tell us why these syntactical features are used the way they are.

d) analyse text-patterning or textualization

As Bhatia (1993: 26) says this level of analysis called textualization, involves "conventional language use, specifying the way members of a particular speech community assign restricted values to various aspects of language use (they may be features of lexis, syntax or even discourse) when operating in a particular genre". Bhatia takes this term from Widdowson (1979), pointing out that the statistical analysis of lexico-grammatical features, mentioned above, becomes more important when it shows the purpose and function the specific feature performs. Furthermore, some syntactic features can perform more functions in a genre, as well as different functions in different genres (Bhatia, 1993:28).

e) analyse genre-functional formulaic sequences

These sequences can be found within particular moves, and they are “the most obvious genre-functional conventionalised type of linguistic realisation, and as such co-constitutive of the genre as such” (Hüttner 2007:111).

f) establish and refine communicative purpose in bottom-up way

Hüttner (2007:111) says that this is an important part of the analysis, and she argues that defining the communicative purpose of the entire genre is the best done after determining the purposes of the individual moves and after verifying them with the expert members.

6) elicit information on effects, acceptability and appropriacy of linguistic realisations as perceived by members of the discourse community

In this step, the qualitative information gained from the members of the discourse community, can help the researcher to decide on the acceptability of certain atypical examples (Hüttner 2007:111).

7) establish sets of relationship of genre to other genres

a) establish ‘genre set’

b) establish ‘genre-(constituent)- colony’

This step presents an important addition to the previous model, and is concerned with the relationships between genres. Hüttner (2007: 31) takes over the concept of ‘genre sets’, based on Devitt (1991: 340), which presents the genres produced by one discourse community as part of their daily routine. They can be different genres but there is intertextual relationship between them, and more importantly the “relationship to the discourse community which in a sense ‘owns’ the genre set”, which constitutes part of their daily practice” (Hüttner 2007: 31). So, she suggests identifying the producers and receivers of the genre, and finding out what position they have in the discourse community and how they are interrelated (2007:104). The communicative purpose and textuality is here not taken into account. Hüttner (2007: 112) states that in this step, we need to define the genre colony the genre belongs to. Bhatia (2002; 2004: 57) defines ‘genre colonies’ as:

closely related genres, which to a large extent share their individual communicative purposes, although most of them will be different in a number of other respects, such as their disciplinary and professional affiliations,

context of use and exploitations, participant relationships, audience constraints and so on.

In extended genre analysis, the genre colony includes “all member genres sharing at least one of the major communicative intentions” (Hüttner 2007: 112). In order to establish genre relationships, she suggests taking into account the status of the members of discourse community.

8) obtain feedback and further information on analysis by members of discourse community

As Bhatia also suggests, the analyst should not merely describe the genre but also seek a specialist informant who would give further insights and add “psychological reality” to the analysis (1993:34). Although this will be already done throughout the analysis, Hüttner (2007: 112) emphasises that before finalising the analysis, the findings should be verified and refined.

2.4. Move analysis

According to Swales (2004: 228-229), a *move* in genre analysis presents “a discoursal or rhetorical unit that performs a coherent communicative function in a written or spoken discourse”. Bhatia (1993:21) claims that genres show certain regularities, cognitive in nature, which reflect the strategies members use to construct particular communicative purposes and this cognitive structuring “reflects accumulated and conventionalized social knowledge available to a particular discourse or professional community”. The communicative purpose of each move individually further contributes to the communicative purpose of the genre as a whole. It is essential to notice, as Swales (2004) points out, that a move is a functional not a formal unit and can be linguistically realized in different ways: by a sentence, utterance or several sentences. Additionally, they may vary in length. Some move types occur more often in genre texts and they are considered obligatory or core moves; whereas the others do not occur so frequently and they are described as optional moves. This classification is based on the frequency of their occurrence which will be described in the text which follows. Furthermore, they do not necessarily appear in a fixed order.

In chapter 6, I will examine the genre texts according to their structure and identify the moves of the confidentiality agreements. In order to achieve this, I will follow Hüttner’s (2007) methodology for identifying moves. I will first look at the communicative

intentions of individual parts and try to see if they can be grouped into steps which can form a single move. In this process, the expert members will be consulted and the findings verified with them for their acceptability. Then, I will look at how they are textualized and look for the typical formulaic language. Quantitative analysis will then determine if the moves are considered core, obligatory or optional. In my analysis I will follow Hüttner's guidelines based on the frequency of moves (2010: 205) as presented in the table below.

Frequency of occurrence	Status	Comments
90% - 100%	obligatory	genre exemplar usually considered inappropriate or in some way "flawed" <i>without</i> this move
50%-89%	core	typical of the genre, considered part of an appropriate and acceptable genre exemplar
30% - 49%	ambiguous	status can only be decided with further expert information - can be core or optional, acceptable or unacceptable (Phase 2 decisive)
1% - 29%	optional	not considered a typical feature of genre, can be considered an acceptable addition (=truly optional) move or unacceptable (-> Phase 2 decisive)

Table 1. Guidelines for deciding on status of individual moves (Hüttner 2010: 205)

In the present study I will follow the above mentioned methodology proposed by Hüttner and adapt the distinction made by Trosborg (1997: 64). Based on Child (1992: 109-142), she outlines the following structure of a contract: title, introduction, recitals, definitions, body of the document, housekeeping provisions, and signatures and dates. The move structure of the confidentiality agreements will be explained in chapter 6. Having described the methodology which will be used in this study, I will now turn to the description of legal discourse.

3. Legal discourse in professional settings

The main focus of the present study is the analysis of the written legal language, namely the language of legal contracts. The confidentiality agreements, as a subgenre of legal contracts will be discussed in great detail in the next chapters. But before turning to the in-depth analysis of confidentiality agreements, I will give an overview of legal discourse. In this chapter, I will look at how linguistic interest in legal English has developed throughout the time and also at its tradition and change. Furthermore, I will examine what legal language is concerned with and try to outline its main functions. Then, I will look at the classification of legal language into various genres, according to the modes and levels of formality. In the second part of this chapter, the characteristics of legal language will be described, in order to understand what makes this discourse such differentiated a variety – language used for a particular purpose. Finally, the current trends in legal linguistics will be discussed, namely the Plain Language Movement. In this study, legal language will be seen as a specific field of English for Specific Purposes.

3.1. Research into legal language

“Law is necessary bound to language, and in that sense legal language has existed as long as the law”, as Mattilla (2006: 6) points out. Legal language has drawn attention for thousands of years. In the ancient times, the particular interest in legal language was mostly practical and it included legal translation and rhetoric. The first legal document which was translated and is still kept today dates from 1271 BC (Mattilla 2006: 7). The exploration of legal language in the modern sense dates back to the 1960s and among the first ones to carry out linguistic research on legal language were Mellinkoff (1963), Crystal and Davy (1969) and Gustafsson (1975). Mellinkoff focused on the history of legal language, as well as its characteristics and its use. Crystal and Davy analyzed the style, lexical and syntactic features of legal writing, whereas Gustafsson looked closely into the syntactic features. Even at these times legal language was criticized for being beyond the comprehension to lay people.

Later on, discourse analysis and genre analysis emerged, and linguists, like Bhatia in the 1990s, offered new insights into legal language, giving emphasis to lexico-grammatical features and also to rhetorical strategies in legal language (Trosborg 1997). Other linguists who approached legal language from the perspective of genre analysis and discourse analysis were Goodrich (1987), and more recently Shuy (2001) and Stygall (1994, 2002).

In the last two decades, the interest has very much shifted to the issues raised by the Plain Language Movement, which originated in the 1970s. in the Scandinavian countries. Many recent publications on legal language are dealing with the simplification of legal documents and they present guides and manuals for modern legal drafting. Some of the authors (all mentioned in Williams 2005) include Adler (1990), Child (1992), Myers (2001), Asprey (2003), Balmford (2002), to name but a few. This movement advocates that the legal texts should be written in plain language defined as the following.

Plain language is writing that can be understood at first reading by clients, lawyers and judges; it is legally binding but it is also logically organized, concise and unambiguous. It uses normal or standard grammar, punctuation and capitalization. It considers the reader's level of knowledge and state of mind. It uses a tone and style that is professional yet appropriate to the circumstances. It uses non-sexist vocabulary and writing techniques. (Stephens 1990).

Doing the research into legal language includes lawyers as well as linguists, and very often the cooperation between them is necessary. Mattilla (2006: 10) claims that linguists have more distance and are therefore, more objective, whereas lawyers look at the language from the inside of the profession. According to her, they focus more on the diachronic perspective and the history of law, but more recent studies involve legislative drafting giving emphasis on comprehensibility to the wider public. Today's linguistic research often includes quantitative methods exploring the occurrences of various linguistic features.

3.2. The concept of legal language

Mattila (2006: 3) says that legal language is a language for special purposes which is clearly distinguished from ordinary language or from any other language for specific purposes. So, there is a specific legal style, as she calls it, which may not be comprehensible to the general public. This style also varies between different legal genres. She also points out that calling legal language a technical language or “technolect”, that is “a language used by a specialist profession”, should be taken with some reservations (2006: 3). This language is used by a specialist profession, such as lawyers, for instance. But not only by them. The people who use legal language do not have to be professionals and she illustrates this by naming jury-members, lay judges or administrators. Most importantly, Mattila (2006: 3) emphasises that what differentiates this language from other languages for specific purposes is that “the target of messages transmitted in legal language often consists of the whole population, certain layers of population, or a number

of particular citizens”. By saying this, she stresses that legal language is not used only within the legal profession but it “governs all areas of social life” (2006: 4). To confirm the fact that legal language can include language and terminology from any other domain, Mattila (2006: 5) notes that legal terminology can be combined with “non-legal technical terminology: for example, criminal law involves psychiatric terminology, while land law involves surveys, and tax law accountancy”.

As Danet argues, the study of legal language is concerned with “the nature, functions and consequences of language use in the negotiation of social order” (1985: 273). Crystal and Davy (1969: 193) state that all the various activities that the law has implications for, are related to “the imposition of obligations and the conferring of rights”.

Mattila (2006: 31) states that among the several functions legal language has, the most important one is “producing legal effects by speech acts”. He refers to John L. Austin’s (1962) and John Searle’s (1969) theories of speech acts, and argues that language does not only transmit messages, but the acts are realized through this language and this is of great importance. According to him, to be able to express legal relationships, it is necessary to have the language, as the law is “a metaphysical phenomenon that is only “alive” in language” (2006: 31). “The law is a profession of words”, as Mellinkoff (1963: vii) has put it. Based on this, the language of the law has a performative function, which means that the language does not only passively describe the reality but it also changes legal relationships and acts as an “instrument of speech acts” (Mattilla 2006: 3). Considering the functions of the law, Danet (1985: 273-274) asserts, on one hand, the ordering of human relations and on the other hand, the restoration of social order when this does not function any longer. She additionally underscores the regulative and facilitative functions of the law, where the law defines relationships, and what is permitted and what not, and moreover, it creates new relationships.

Here, it is noteworthy to make a distinction between the terms *legal language* and *the language of the law*. Williams (2005: 23) refers to Trosborg (1995: 32) and says that by the expression *the language of the law*, she means the specific area of legal writing, namely the language of the prescriptive texts covered by statute law or common law, such as legislation or contracts and deeds. *Legal language*, on the other hand, has much wider meaning and it covers legal discourse in general, and by this she means “any form of legal documents to the law reports published in newspapers to certain forms of oral language such as legal directions which is read out to jury members” (referred to in Williams 2005:

23). Additionally, with regard to legal discourse, Maley (1994: 13) states that there is a set of discourses where four types can be distinguished. These discourses roughly correspond to the genres identified in the next section. The first one is *judicial discourse*, such as judicial decisions for example, then *courtroom discourse*, the language of judges, witnesses, etc., *the language of legal documents*, such as the language of contracts, deeds, wills, statutes, and *the language of legal consultation*, the language of lawyers and clients for instance.

Danet (1985: 275) observes that O'Barr (1981) as well as Charrow, et al. (1982) believe that legal language is such a differentiated variety of language that it can be called a separate dialect or sublanguage. Even though she acknowledges this differentiation, she prefers calling it a register, as it is, to a great extent, related to the level of formality. In this study legal language will be referred to as a variety as this term covers the language used for specific purposes.

3.3. Genres of legal discourse

Bhatia (1993: 101) expresses that

legal language encompasses several usefully distinguishable genres depending upon the communicative purposes they tend to fulfil, the settings or contexts in which they are used, the communicative events or activities they are associated with, the social or professional relationship between the participants taking part in such activities or events, the background knowledge that such participants bring to the situation in which that particular event is embedded and a number of other factors.

In the view of the above, he names several genres in written form, such as cases, judgements, legislation, contracts and agreements and in lawyer-client consultation and counsel-witness examination in spoken form. This classification into written and spoken genres was further extended in Danet (1985: 275-277), who based on Joos' typology (1961), gives a classification of genres which refer to language use for professional legal purposes in communicative legal settings. She takes in consideration two criteria, namely the modes and the degree of formality. The modes include written, spoken but composed in advance, and spoken and spontaneous. As for the formality, there is a classification into frozen, formal, consultative and casual style. Written documents which use formulaic, fixed language are distinguished as insurance policies, contracts, landlord-tenant leases and wills. The formal types include statutes, briefs and appellate opinions. Marriage ceremonies, indictments, witnesses' oaths, pattern instructions and verdicts are all prepared

in advance, they include fixed expressions but they are spoken aloud. Formal spoken genres which are planned but do not necessarily include fixed formulas, are lawyers' examinations of witnesses in trials and depositions, lawyers' arguments and motions in trials. Unlike expert witness testimony, which are formal and might include prior planning, lay witnesses' testimonies are considered consultative. Other types of consultative but spontaneous genres are lawyer-client interactions and bench conferences. Spoken casual genre includes lobby conferences and lawyer-lawyer conversations. Furthermore, the language in frozen genres has performative function and it shows the most distinctive features. Here is of great importance what is said but more importantly how it is said. This is most evident in witnesses' oaths to tell 'the truth, the whole truth and nothing but the truth', as illustrated in Danet (1985: 277). The classification of legal genres is shown in the table below.

A Typology of Situations in Which Legal English is used, by Style and Mode*

Mode	Style			
	Frozen	Formal	Consultative	Casual
Written	Documents: insurance policies contracts landlord-tenant leases wills	Statutes Briefs Appellate opinions		
Spoken-composed	Marriage ceremonies Indictments Witnesses' oaths Pattern Instructions Verdicts	Lawyers' examinations of witnesses in trials and depositions Lawyers' arguments, motions in trials Expert witnesses' testimony	Lay witnesses' testimony	
Spoken-spontaneous			Lawyer-client interaction Bench conferences	Lobby conferences Lawyer-lawyer conversations

*Danet (1980, p. 471); based on Joos (1961)

Table 2. Classification of legal genres (Danet, 1985: 276)

Looking at the written legal language and its regulatory and informative function, two types of texts can be distinguished: the prescriptive and descriptive. Sarcevic (2000: 9) points out that:

Legal texts whose function is primarily prescriptive include laws and regulations, codes, contracts, treaties and conventions. Such texts are regulatory instruments containing rules of conduct or norms. Accordingly, they are normative texts which prescribe a specific course of action that an individual ought to conform to [...] Today it is generally agreed that normative instruments prescribe how the members of a given society shall act (command), refrain from acting (prohibition), may act (permission) or are explicitly authorized to act (authorization).

According to her, descriptive texts include legal opinion, law textbooks, articles, etc. and they belong to the academic discourse. In addition to these two groups, there are so-called hybrid texts with prescriptive and descriptive characteristics. As examples of this group, Sarcevic (2000: 9) names actions, pleadings, briefs, appeals, requests, petitions, etc. used in judicial proceedings. Bhatia (1983: 2) has the distinction based on the communicative functions, which roughly corresponds to the previously mentioned distinction. He identifies legislative or statutory writing, academic writing, which also includes research journals and textbooks, and juridical writing found in court judgements, case-books and law reports (referred to in Williams 2005: 29).

3.4. The characteristics of legal language

Before outlining the typical characteristics of legal writing, I will focus on what affects and shapes legal writing and try to give some background to why legal language takes the form it does. In order to show the challenges of legal language, Crystal (1995: 374) implies the following:

Legal language is always being pulled in different directions. Its statements have to be so phrased that we can see their general applicability, yet be specific enough to apply to individual circumstances. They have to be stable enough to stand the test of time, so that cases will be treated consistently and fairly, yet flexible enough to adapt to new social situations. Above all, they have to be expressed in such a way that people can be certain about the intention of the law respecting their rights and duties. No other variety of language has to carry such responsibility.

I believe that this should be taken into account when looking at the distinctive features of legal language. Crystal wrote as early as in 1969, that the drafter of legal documents must ensure that the document says exactly what he wants it to say and to leave

no chance for misinterpretation. This obvious demand for precision presents the point of departure for drafting a legal document.

In this section, I will outline the typical features of written legal documents, which are characterized as frozen or formal according to the degree of formality. These features will be in more detail analyzed further in the study where I will closely analyse if the confidentiality agreements are characterised by some of those features.

Legal language is often called *legalese* and this term refers to “the jargon of the legal discourse community” (Williams 2005: 27). But this term, as he notes, mostly has negative connotation, as it indicates the obscure properties and style of legal language. These properties of legal language will be shown in the text which follows. Based on the research of the most prominent linguists who examined the lexico-grammatical features of legal documents, Williams (2005: 31) outlines some characteristics of prescriptive texts and states that legal writing is clearly lexically distinguished from ordinary language by the use of archaic expressions and foreign words, and most of them are only used in legal professions. Some of them include words like *hereinafter*, *heretofore*, *surrejoinder*, *aforesaid*, etc. The foreign words of Latin and French origin came into the language through the long domination of Normans in the fields of law and politics (Williams 2005: 32). They are illustrated by the phrases such as *ex parte*, *ratio legis*, *in situ*, etc. The use of repetition of words or phrases has referential value in the case of legal writing and it serves as a means of avoiding possible ambiguity. The sentences in written legal texts are rather long, with complex inserted qualifications which cause syntactic discontinuities (Bhatia 1993: 111-113). The extensive use of passive forms, gives legal writing an impersonal character as it “brings the object of the action into the foreground” (Mattila 2006: 73). Another device, which contributes to the impersonal style, is the tendency for nominalization. Williams (2005: 37) states that the verbs are often transformed into nouns, illustrated by the example *to amend* transformed into *to make an amendment*.

Another characteristic of legal writing is the information (over)load. The sentences are characterized by a high density of information, and they can be very compact and over-long (Mattila 2006: 72). Due to the fact that the legal terms have multiple meanings, and in order to avoid misunderstandings, legal language often involves many definitions (Mattila 2006: 66). Beside definitions, the language of legal documents includes enumerative lists and they tend to be very inclusive (2006: 72). Legal writing is neutral, “it rejects all that is

affective” and this is also the reason why legal texts hardly ever use exclamations or question marks. (Mattila 2006: 74)

As for the logical structure of legal texts, it can be said that it also plays an important role. The information given in the text is presented in a hierarchical order, and:

a legal text moves from abstract to the concrete, from substantive to the procedural. The structure of the text should be consistent: the principal items are presented before secondary items, and general rules before special conditions and exceptions. (Mattila 2006: 81)

It is worth noting that the structure of legal texts contributes to the formality which additionally includes “fixed formulas at the level of sentences and phrases” (Mattila 2006: 81). These “ready-made sentences and petrified phrases”, as she calls them, will be in more detail discussed further in the study as a part of analysis of formulaic language (Hüttner 2007: 67-96).

Most of the typical features mentioned above will be discussed in greater detail in chapter 6. The following section deals with the current trends in the exploration of legal language.

4. The study

In the present study, I will follow the methodology proposed by Hüttner, namely the extended genre analysis which was outlined in the second chapter. The texts will be analysed using the corpus-based approach. This quantitative approach includes the statistical analysis of the lexical and grammatical features of the specialised corpus which consists of a collection of legal contracts.

Many linguists have pointed out that the understanding of professional discourse is to a great extent connected to the understanding of the culture of the discipline (Bazerman, Littl, & Chavkin 2003; Bhatia 2004, 2008; Hyland 2000, 2003; Hüttner 2007). In order to get the insights into the genre, and following the aforementioned researchers, in this chapter I will cover the first four levels of extended genre analysis, as suggested by Hüttner (2007).

In the first part of the chapter, I will begin by identifying the genre of contracts, of which the confidentiality agreements are a sub-genre. Secondly, I will describe the communicative purposes of the confidentiality agreements and identify the discourse community which uses the genre in question. In the second part of the chapter, the methodology used in the current study will be described. Here, I will also define the data and the corpus which present the main material of the study. Finally, I will describe the tools used in the analysis of the agreements.

4.1. The field of contracts

As mentioned previously in section 3.3. contracts fall into the category of written legal genres with a great formality in style and the use of some formulaic, fixed expressions. Bhatia, Langton and Lung (2004: 206) say that contracts are legal genres which are very conservative, with a fixed relation between the form and function. The language of contracts is very much restricted by the fact that it should be as precise as possible and in order to avoid any misinterpretation, “every attempt is made to restrict the number of interpretations that a particular legislative statements can attract” (Bhatia, Langton & Lung 2004:206). Redmond (1979: 62) defines contracts as the following.

A contract is a legally binding agreement, that is, an agreement imposing rights and obligations on the parties which will be enforced by the courts (cited in Trosborg 1997: 62).

Another complementing definition found in the *Black's Law Dictionary* states that:

A contract or agreement is either where a promise is made on one side and assented to on the other; or where two or more persons enter into engagement with each other by a promise on either side. A contract is an agreement by which one person obligates himself to another to give, to do or permit, or not to do something expressed or implied by such agreement (<http://thelawdictionary.org/>, 10.10.2012).

When it comes to contracts and the effect they have, it is noteworthy pointing out that they do not affect the society as a whole, as some other legal genres, but only a group of people or a few members (Trosborg 1997: 57)

With regard to the communicative purposes, it can be said that the language of contracts has a regulative function in a sense that contracts regulate legal behaviour, create rights and obligations of the parties involved (Trosborg 1997: 22). She refers to Searl's theory of speech acts which analyses the effects an utterance has on the behaviour of speakers and hearer. Within the category of regulative acts, he makes a classification into directives and commissives (1997: 62-63). The directive acts involve the rules of the contract which order human relations, that is, the obligation one party incurs over the other. On the other hand, commissive acts, according to Austin (1962), "commit the speaker to a certain course of action". It is clear that the communicative purpose of the certain clauses or provisions in the contracts is achieved through the use of various regulative acts. The following table shows the classification of regulative acts for the purpose of analyzing the contracts, presented through the speech act categories on one hand, and some of their linguistic representation on the other hand.

Regulative strategies for simple contracts (presented at levels of increasing directness)		
I.	Constitutives	unmarked
II.	Permission	allow [-] · can · could · grant [-] · may · might · need not · [VP] no obligation · offer [-]
	Rights	(a) shall/will {“benefit”} [$\pm$ negation] (b) shall/will {“limitation of liability”} [negation]
III.	Obligation	are to · has to · has to · have to · is to · must · obligate · obliged to · shall · should · will
	Prohibition	can [negation] · may · [negations] · shall [negation] · will [negation]
	Performatives	demand [-] · order [-] · request [-]
IV.	Promises	accept [-] · acknowledge [-] · agree [-] · certify [-] · commit [-] · confirm [-] · covenant [-] · declare [-] · guarantee [-] · promise [-] · represent [-] · understand [-] · undertake [-] · warrant [-]
[] : symbol for syntactic item { } : symbol for semantic item denoting stated value or equivalent - : symbol for morphological item		

Table 3. Domain-specific hierarchy of strategies for issuing regulatives (Trosborg 1997: 69)

As we can see, promises use performative verbs such as *accept*, *promise*, *covenant*, *commit*, *agree*, etc. which are more direct because the parties of the contract commit themselves by a promise to do something, they take the legal obligation to do the promised act.

4.2. The genre of confidentiality agreements

The confidentiality agreements are also known as non-disclosure agreements or secrecy agreements (Radack 1994: 68) and they can be defined as the following.

agreements that are used when the owner of confidential information wishes to disclose that information to another party (either an individual or a company) usually in the course of business negotiations, and wishes the information to remain confidential. By signing a confidentiality agreement, the recipient undertakes the obligation not to disclose the confidential information as defined in the agreement. (Hanson, Thompson & Dahlgren 2004)

The language of contracts, as Trosborg (1997: 62) states, “refers to mutual rights and obligations in relation to promise and considerations”. As she states, the drafting of the legal documents, such as legislation, contracts and wills, is considered “preventive law” (1997: 19). By doing this effectively, it is possible to avoid litigation. In this particular case of the confidentiality agreements, we could say that the purpose the agreements want to communicate is to protect the information which is considered confidential. The two parties agree what information will be disclosed and what information will be kept confidential. In this way, as we have seen before, the parties avoid engaging in legal proceedings.

The confidentiality agreements can deal with different types of confidential information, depending on what kind of work the parties are involved with. When the information is already known to the public, it is not considered confidential. The lists of definitions of the confidential information tend to be inclusive as shown below.

The type of information that can be included under the umbrella of confidential information is virtually unlimited. Any information that flows between the parties can be considered confidential—data, know-how, prototypes, engineering drawings, computer software, test results, tools, systems, and specifications. This list is certainly not exhaustive but does illustrate the breadth of items that can be deemed confidential. (Radack 1994: 68).

According to the distinction between bilateral and unilateral contracts made by Atiyah (1989), most confidentiality agreements are unilateral, which means that one party of the contract agrees not to disclose particular information. If both parties supply the information which needs to be protected, then the agreements are bilateral or mutual. This type of agreements is less common than the former ones.

The linguistic features, employed to achieve the communicative purpose of protecting confidential information, are very distinctive, as will be shown in the next chapters, and are very much shaped by the main features of the confidentiality agreements, such as “precision, clarity, all-inclusiveness and unambiguity” (Bhatia 1993: 102).

4.2.1. The discourse community

The concept and the criteria which would lead to the emergence of a discourse community have been described in section 2.2. In order to define the discourse community of the genre of confidentiality agreements, I will first of all consider two kinds of discourse communities - the legal specialist and non-specialist discourse communities. The specialist discourse community of the legal contracts would include lawyers, legal draftsmen, judges, and other legal professionals, whereas the non-specialists would involve the parties to the contract which most probably do not have any formal law education. As Bhatia (1993: 103) says, the lawyers and judges have the role of interpreters who interpret the legal message to the clients, the ordinary citizens. With regard to the specialist and non-specialist communities, it is important to mention that the lay people or these ordinary citizens are never the producers of the genre text, and the text is produced only by a legal specialist or a legal drafter. The parties to the contracts mostly belong to the non-specialist discourse community and they are defined in the introductory part..

Trosborg (1997: 59) emphasises that the parties to the contract are at the same time senders and receivers.

The parties are senders, as before the contract is drawn up by the lawyer, they have agreed on the subject-matter of the contract, its contents, i.e. the specific “promise” and “consideration” and the particular circumstances and conditions involved, and they are receivers, as they verify and witness the agreement by their signature, and thus approve the contents of the contract.

According to Trosborg (1997: 58-59) it is important to mention the sender/ receiver relations. These relations can be symmetrical or asymmetrical, depending on the “authoritative status of the interlocutors” as well as the distinction between them as the insiders or outsiders of the domain (Trosborg 1997: 58). The relation between the parties in the contract can be said to be symmetrical as they both “hold something of great value and interest to the other - promise and consideration”, and both parties witness the document by signing it (1997: 59). Both parties have a designated ‘title’ according to the role they play (Trosborg 1997: 59), and in the particular genre of confidentiality agreements these

role can be, for example, *receiving party/disclosing party, company/employee, employer/employee, company/executive*, etc.

4.3. The methodology

“Corpus linguistics is a relatively new approach to language studies that has the opportunity to revolutionize the teaching and learning of discourse for specific purposes” (Connor & Upton 2004: 1). Baker (2006: 1) defines corpora as “large bodies of naturally occurring language data stored on computers”. The purpose of the corpora is defined by Sinclair (1991:17) who says that one of the main uses of a corpus is identifying what is central and typical in a language. In the early days of the corpus linguistics, the corpora were used mainly in the field of lexicography and grammar, and only lately has the corpus approach been applied to discourse studies (Flowerdew 2008: 7). In the initial stages of corpus linguistics, the large general corpora were designed and they gave insights into how language is used in a variety of contexts. Furthermore, they gave information on lexicogrammatical features as well as functional variations (Flowerdew 2008: 7). In recent years, the tendency towards compiling smaller, specialized corpora can be seen. Connor and Upton (2004: 2) emphasise that specialized corpora can give more contextual information about the communicative situation, as they are often collected by the researcher.

One of the most obvious advantages of corpus-based analysis is that the large quantities of texts can be analysed with minimal effort, and in this way fighting the constraints of manual analysis used until then (Flowerdew 2008: 8). Lee (2008: 92) outlines some of the major contributions of the corpus-based approach as the following. The use of corpora has not only facilitated the access to a huge amount of data, but also made the analysis of different text types more available. This means that researchers do not only have to focus on the ‘popular’ genres, but can additionally examine the genres that are less studied. The corpus-based approach to text analysis in ESP also gives “quantitative, empirically-based information about frequency/typicality/idiomaticity” (Lee 2008: 93). With the help of software programs such as WordSmith Tools or AntConc, the textual features can be easily identified. The coding and analysis of the data is also facilitated by the development of new mark-up programs which add extra information to a corpus and can improve data interpretation. This information added can be general or descriptive, or texts can also be tagged for parts-of-speech, phrases, clauses, etc. Another important contribution of the corpus-based analysis is that the researcher can identify the trends that would be hard to notice otherwise (2008: 92).

Biber (1998: 4) emphasises that corpus-based analysis involves qualitative as well as quantitative techniques. Furthermore, the most important aspect of corpus research is how the researcher interprets the findings (Baker 2006: 148).

4.3.1. The data

The data of the present analysis has been drawn from the corpus of legal documents, namely confidentiality agreements, which I have retrieved from the online public domain securities filings (<http://contracts.onecle.com/> 23.09.2012). The types of contracts that were used as my main material are concerned with the disclosure of information, or more precisely, with keeping such information confidential. The corpus consists of thirty confidentiality agreements and their titles are presented in the appendix of this study. As contracts belong to strongly conventionalised genres, a smaller corpus is enough to find proto-typicality, in contrast to the genres which are more ‘creative’ (Hüttner 2007: 109).

The idea of compiling the confidentiality agreements corpus initially came from a personal interest in legal language, which is a variety of language - language use for professional purposes, that has characteristics very different from other varieties. The choice of texts was primarily based upon availability, but the main criterion on which the decision was made was the representativeness and the authenticity of the texts. The contracts used in the study are genuine in nature and they include parties from different countries. Most of the confidentiality agreements used in the study are unilateral agreements. The texts reflect the contemporary legal language and are drafted by native English speakers with legal expertise. As I wanted to present the contemporary picture of the confidentiality agreements, relatively recent contracts were chosen, ranging from 2001 to 2010. Only one text dates back to 1997.

For the purpose of this study, the texts were considered in their entirety, containing the original content, which could give us the insights into all the features of this genre. Nevertheless, for the sake of confidentiality, the texts were anonymised which means that the data which contains the names of people, companies, etc., is not presented in the study. The examples retrieved from the texts and presented in this study to illustrate certain features of the genre, will be marked with their title and the number by which they appear in the appendix. Throughout the study, the confidentiality agreement corpus will be referred to as the CAC.

During the course of this research, the relevant information was gained not only by the corpus-based analysis of the genre, but also by consulting the literature on legal language, and in particular the language of contracts. This literature is referred to throughout the study. In order to familiarise myself with the important legal concepts found in the texts of confidentiality agreements, particularly helpful were the textbooks of international legal English (Day, Krois-Lindner & Translegal 2006), the legal dictionaries (Black's Law Dictionary) as well as the online information on confidentiality agreements published in journals or by law firms (Radack 1994), (Hanson, Thompson & Dahlgren 2004). In addition to this, I sought opinion of a lecturer of legal English, as well as discussed certain issues, mostly about the concepts and the structure of the confidentiality agreements, with a practicing lawyer who works in contract law, and a graduate law student. Having described the corpus used in the current study, I will now turn to the tools used to facilitate the quantitative part of the analysis.

4.3.2. The tools used for the corpus analysis

Nation (2001: 32) observes that there are three important points to consider when doing a corpus research, namely a set of good research questions, the corpus as a source of data and the computer programs to facilitate this. Having defined the first two, I will in this section describe the programs used in the study. The two computer programs that will be used in this study are WordSmith Tools and RANGE. These corpus linguistics tools help investigate the type of vocabulary in different texts.

WordSmith Tools

“Oxford WordSmith Tools is an integrated suite of programs for looking at how words behave in texts” (Scott 2007: 2). It is one of the most widely used software which analyses corpora and it includes three main tools which can perform different jobs. These tools are Concord, Wordlist and Keyword and their significance will be explained in the next chapter.

RANGE

RANGE is a freeware which contains the ready made word lists, such as the general service word list of the first 1000 most frequent words, the second 1000 most frequent words, and the academic word list (<http://www.victoria.ac.nz/lals/about/staff/paul-nation>). The specialized corpus can then be compared with these word lists, which can help identify the different types of vocabulary in the corpus, namely general service and academic

words. Additionally, the program can show a list of words that do not belong to these two lists and in this way the researcher can identify the technical and low frequency words. The significance of this will be described in the next chapter.

The reference corpus

As early as in 1991, Sinclair (1991: 14) wrote that the importance of a general corpus “as a place of reference” is extremely great and it opens new horizons. A reference corpus “consists of a large corpus (usually consisting of millions of words from a wide range of texts) which is representative of a particular language variety” (Baker 2006: 30). The reference corpus used in this study is the British National Corpus (BNC). This corpus consists of around 100 million words and it includes written and spoken language. Baker (2006: 43) outlines that the importance of the reference corpus lies in the fact that it is representative enough as to allow for different kind of research; it also represents “what is ‘normal’ in language’ and by comparing our data to it, certain characteristics can be revealed.

4.3.3. Vocabulary of the corpus

In this section, I will emphasise the importance of the lexical analysis of the texts. Nation (2001a: 205) states that the use of vocabulary in a text is the result of the communicative purpose of that text. According to him, there are two important points:

First, vocabulary use signals and contributes to the uniqueness of the text, that is, what makes this text different from all other texts. Second, vocabulary use carries general discourse messages which are shared with other texts of similar types (Nation 2001a: 205).

In view of the above, we can conclude that the words which occur with high frequency in the texts, can show some important features of the discourse in question. The function words present the significant proportion of all words, and Nation (2001a: 206) refers to the findings of Francis and Kucera (1967) where the functional words cover 43-44% of the words in the text. It is also noteworthy pointing out that the words which appear frequently are very often the topic related words. This will also be proven in my collection of texts by the high occurrence numbers of the words like *agreement*, *company*, *party*, *confidential*, etc.

Nation (2001b: 32) emphasises that not all words in a language are equal, and in order to determine the importance of a word, we need to look at its “frequency and range

of occurrence”. This is particularly true when it comes to the specialized areas of a language. Four types of vocabulary can be distinguished in the text, namely high-frequency words, academic words, technical and low-frequency words.

According to him, the words more useful for a language learner are the ones which appear frequently in a wide range of uses. These words are called high-frequency words or general service words, based on West (1953). It is generally accepted that the high frequency words cover about 2000 word families. The program RANGE identifies two base word lists, each consisting of 1000 words.

The RANGE program can also identify the academic words in a corpus by analysing the corpus in relation to the academic word list developed by Coxhead (1998, 2000). This word list includes 570 words that are not in the first 2000 most frequent words. These words are also called sub-technical words, and they are rather formal but there are not technical words (Nation 2001:17)

The words that do not appear in either of the lists present the technical words and low-frequency words. Nations (2001a: 198) defines a technical word as a word that ‘is recognisably specific to a particular topic, field or discipline’ and one that cannot be usually found in other subject areas. Finding technical words is rather problematic as they can be found also in high frequency word lists and academic word lists. The list of technical words can be developed through the use of technical dictionaries and the use of corpus based frequency counts, comparing a general corpus with a specialized corpus (Nation 2001a: 201). Intuitive judgement plays also an important role here. Regarding the low-frequency words, Nation (2001a: 19) says that they can be proper names, the words with moderate frequency which did not get into the high-frequency list, or they are just in infrequent use in the language. A words which appears only once in the corpus is called *hapax legomena* and there is usually a great number of these words (Scott 200: 54). *Hapax legomena* can be non-English words, proper nouns, words without accents – symbols which can indicate stress, vowel quality, etc., or they can also be typos or mistakes (Scott and Tribble 2006: 26).

There is a distinction of what words are counted in the analysis. The RANGE programs gives results according to the three categories: tokens, types and word families, further examined by Nation (2001a: 7). Tokens or ‘running words’ present every occurrence of a word in the texts, that is, all forms of a word are counted. Types involve

only one form of the same word. The tokens can answer the question ‘How many words are there?’, whereas types can tell us how large the vocabulary of the text is. Word families include a headword and its inflected and derived forms although it is not always so transparent what should be included in a word family and what not (Nation 2001a: 8).

4.3.4. Concordance-derived analysis

The Concord tool identifies “all references to any given word or phrase within our corpus” (Scott 2001: 47). This tool is extremely useful because it gives the researcher examples of words in context (Scott 2001: 50). The word we want to look up is called a *node*, and *collocates* are words which co-occur with that word. Baker defines the term as the following.

A concordance is simply a list of all of the occurrences of a particular search term in a corpus, presented within the context that they occur in; usually a few words to the left and right of the search term (Baker 2006: 71).

The significance of the concordance identification lies in the “rich contextual information which illuminates meaning and characteristic use within the text(s) studied” (Scott & Tribble 2006: 33). The concordance lines show us how the words behave, and Hunston (2002: 42) distinguishes four types of what can be observed by concordance lines, namely “observing the ‘central and typical’, observing meaning distinctions, observing meaning and pattern, and observing detail”.

Apart from the concordance lines, the Concord tool can also identify collocations. Collocation is defined as “the tendency of words to be biased in the way they co-occur” (Hunston 2002: 68) and it “illustrates the idiom principle” (Sinclair 1991: 115). Collocation can include a relation of lexical items, termed ‘colligation’, or a lexical item and its grammatical environment (2002: 12). Here, it is important to mention the MI- score or Mutual Information score. This concept tells the strength of a collocation. Hunston (2001: 71) points out that any MI-score higher than three, can be considered significant. Simple repeated strings of words found within the concordance lines are called clusters, but they do not always have the psychological reality for speakers (Scott and Tribble 2006: 19) as very often the words which appear in clusters are high frequency words and they do not reveal some important features (Baker 2006:100).

4.3.5. Word list-derived analysis

The WordSmith tool Wordlist is an important tool which creates word lists, which “help identify the common words in a corpus” (Scott 2001: 54). They may be ordered in alphabetical order (or reverse alphabetical order), or according to frequency. The former one can be more convenient when looking for specific items, whereas the latter gives us a straightforward idea of which words appear most frequently in the corpus.

Considering the word-lists, we can talk about high-, medium- and low-frequency words. The most common high-frequency words are prepositions, determiners, pronouns, conjunctions and these words supply the grammatical information to nouns, verbs, adjectives or adverbs (Scott and Tribble 2006:24). They do not carry great lexical content. The medium-frequency words are mostly lexical items. Scott and Tribble looked at BNC word-lists and they noted that 40% of the words are *hapex legomena* and have frequency of one (2006: 26).

As Scott and Tribble (2006: 31) point out, the word-list can be a great starting point for understanding the lexis of the texts in question and it can show certain characteristics of the genre. While the word list only provides frequency, the keyword list can provide saliency, as Baker states (2006: 125). This will be explained in the following section.

4.3.6. Keywords-derived analysis

The Keyword tool gives a list of unusually frequent words compared with a reference corpus and those words are characterized as ‘outstanding’ (Scott 2001: 6). The important concept here is the keyness, which is defined by Scott and Tribble (2006: 55) as “a quality words may have in a given text or set of texts, suggesting that they are important, they reflect what the text is really about, avoiding trivia and insignificant detail”. The keyness can be positive, which means the words appear unusually frequently, that is, more often than expected compared to the reference corpus, or negative, which appear unusually infrequently. This means that they show characteristics in comparison with the norm, which is in this case British National Corpus.

The importance of keywords lies in the opportunity “to characterise a text or a genre” (Scott 2007: 115). By comparing the specialized corpus with the large reference corpus, we can, for instance, analyze the words which make this corpus unique (Baker 2006: 147). We can in this way identify the words which appear more frequently compared to the reference

corpus, termed positive keywords, and the ones which appear less frequently, negative keywords.

4.3.7. Formulaic language

In the next chapter, following Hüttner's fifth level of genre analysis, I will also examine the use of genre specific formulaic sequences in the genre of confidentiality agreements. This area of analysis has not received too much attention from linguists and this is Hüttner's immense contribution and addition to the methodology of the genre analysis proposed by Bhatia. Hüttner (2007: 98) states that the formulaic language use can be found across different genres, oral or written, native or non-native, such as for instance course descriptions, research articles or textbooks (Biber et al. 1999; Biber 2006) and many others. Wray (1999: 214) defines the term formulaic sequence as the following:

[a] formulaic sequence [is] a sequence, continuous or discontinuous, of words or other meaning elements, which is [...] prefabricated: that is, stores and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar (cited in Hüttner 2007:73).

Hüttner (2007: 75) makes an addition to this definition and states that "formulaic sequences can serve to pragmatically further the communication of discourse meaning within the communities that use them". She outlines two important functions of formulaic language (2007: 70-72). The first involves formulaic sequences as production strategies, where the speaker uses this kind of sequences instead of creating "all strings of language from scratch", so there is less emphasis on "the processing capacities of the short-term memory" which can then deal with novel ideas created from rules. The second function of formulaic sequences is to aid the communication in a sense that it can organize the discourse and help hearers identify what is coming and ease their understanding of the content. Formulaic language occurs in both spoken and written language, and is stored in the speakers' minds as whole chunks of information, as she notes.

With regard to formulaic sequences and genres, she distinguishes two types: genre-specific formulaic sequences and genre-functional formulaic sequences. The first type involves "multi-word chunks that are quantitatively typical of specific genres and thus constitute parts of their specific 'idiomaticity', whereas the second type "furthers the communicative purposes of a particular genre move" (Hüttner 2007: 98).

In this study I will look at the structure of confidentiality agreements and distinguish between core and optional moves. Within these particular moves I will then identify genre-

specific formulaic sequences. Having presented the major theoretical concepts of a genre theory in the first chapter, as well as the concept of legal discourse, the methodology which will be used in the study and a larger genre of contracts, I will now turn to the empirical investigation of the corpus of confidentiality agreements. In the next chapter, the lexicogrammatical characteristics of the confidentiality agreements will be analyzed.

5. Analysis of lexico-grammatical features of the texts

Following Hüttner's fifth step of analyzing genres (2007), in the next two chapters I will look at the levels of linguistic analysis, namely analysis of lexico-grammatical features, text-patterning or textualization and the structural interpretation. In this chapter I will focus on distinctive linguistic features of legal language, in particular the language of written legal contracts, and try to explain how these features are assigned certain values in the corpus of confidentiality agreements. The in-depth lexico-grammatical analysis of the thirty confidentiality agreements was conducted and the results will be presented as lexical features, syntactic features and discourse-level features.

5.1. Lexical features

In this section I will examine the typical lexical features of the corpus of confidentiality agreements including the use of formulaic sequences characteristic of this genre. First of all, I will look at the overall size of the corpus as well as the distribution of the lexical items throughout the corpus. This involves comparing the corpus with the larger general English corpus, the British National Corpus, and also identifying the general service words, academic words, technical and semi-technical words. This will be done by using the Word Smith Tools and the Range program. With regard to their frequency, I will examine the keywords in their context, as this can give us considerable insights into this specific genre. Furthermore, other lexical features will be investigated. These include archaic expressions and foreign words, lexical repetition or redundancy, tendency towards formality and unusual prepositional phrases.

5.1.1. Vocabulary of the corpus of confidentiality agreements

As Crystal and Davy (1969: 207) point out, the range of vocabulary that may be found in legal language is immensely wide, as, in one way or another, everything can become its subject. Nevertheless, the lawyers or the legal draftsmen tend to develop "preferences in their choice of words" (Crystal & Davy 1969: 207), so it is easy to notice certain characteristics.

According to Paul Nation's program RANGE, the corpus of confidentiality agreements consists of a total of 60.107 tokens, 3.925 types and 1.090 families. In this section I will explore the distribution of lexical items analyzed by the RANGE program. I will discuss the results obtained by comparing the confidentiality agreement corpus with

the base wordlists of the first most frequent 1000 words, the second 1000 words and the Academic Word List. The results are indicated in the following table.

Wordlist	Tokens/%	Types/%	Families
1 st 1000 most frequent words	44868/74.65	1221/31.11	596
2 nd 1000 most frequent words	2737/4.55	296/7.54	157
AWL	4969/8.27	736/18.75	337

Table 4. General service vocabulary and academic word list of CAC

According to the analysis, the words that appear in the first base General Service List present about 31% of the corpus, whereas the words found in the second list present around 7.5 %. This amounts to a total of 39% of GSL types and 753 families. The percentage of running words found in GSL is about 79 %, which is very close to the 80% of the typical coverage of the text (Nation 2001a: 11). The most frequent words in my corpus, also found in the GSL list are, as assumed, the functional words. These include words like *the, of, or, and, to, any, in*, etc. Some of the common lexical words are *agreement, company, party, shall, employee*, etc. They will be dealt with in more detail in section 5.1.5.

5.1.2. Academic vocabulary of the confidentiality agreements

The words in the CAC, which are also found in the Academic Word List, cover about 19% of the types and 337 word families. The percentage of running words found in the academic word list is 8.27 % which also corresponds to the typical coverage of 9%, indicated by Nation (2001a: 12). Table 5 shows some of the most common academic words found in the corpus.

period, termination, prior, legal, data, entity, documents, section, corporation, required, consultant, acknowledges, consent, available, successors, defined, benefit, restrictions, contract, purchase, computer, paragraph, financial, research, processes, equipment, technical, access, etc.

Table 5. Some common academic words of CAC

Some of the examples of academic words are presented in the following sentences in bold and italicised print.

1. Capella shall ***compensate*** employees for ***assigning*** their rights in inventions that Capella ***seeks*** to protect under patent laws in an amount not to ***exceed*** \$100 per invention (evenly ***allocated*** among all inventors). (Confidentiality, Non-Competition and inventions Agreement, 12.)
2. For this purpose, "***successor***" means any person, company, organization or other ***entity*** which at any time, whether by ***purchase***, merger or otherwise, directly or indirectly ***acquires*** all or substantially all of the assets or business of the Company. (Confidentiality and Non-Competition Agreement, 29.)

5.1.3. Technical vocabulary of the confidentiality agreements

According to the program RANGE, the words that do not belong to any of the previously mentioned lists present the technical vocabulary and the low-frequency words. In my corpus this number amounts to almost 12.53 %. However, not all of them are technical words. After the analysis of the given words, it is concluded that around 6% of the corpus cover those technical words which are characteristic of the genre of confidentiality agreements. This number is rather high and as a consequence, the language of contracts (and generally legal language) is usually very dense in nature. The ten most frequent technical words in my corpus identified by the program Range are presented in the table below.

Technical words	Frequency
confidential	469
proprietary	120
breach	119
confidentiality	108
disclose	93
affiliates	90
hereby	72
recipient	68

Table 6. The ten most frequent technical words identified by RANGE

Other very significant legal concepts are presented in the table below. Some of them account for distinctive features of my corpus and they will be further explained starting from the next section to section 5.2.

confidential, proprietary, breach, forfeit, confidentiality, affiliates, disclose, disclosure, hereby, recipient, executive, hereto, obligations, jurisdiction, non-competition, herein, covenants, deemed, foregoing, thereof, intellectual, transaction, solicit, construed, client, patent, waiver, copyright, unenforceable, arbitration, injunctive, tangible, severability, stipulate, etc.

Table 7. Some technical vocabulary of CAC identified by RANGE

The use of some of the frequent technical words is illustrated by following examples taken from the corpus.

3. In the event of a *breach* or *threatened breach* by Seller of the provisions of this Agreement, the parties *hereto* agree that Buyers shall be *entitled to equitable relief* to protect its interests *therein*, including, but not limited to, preliminary and permanent *injunctive relief*. (Noncompetition and Confidentiality Agreement, 22.)
4. Notwithstanding the *foregoing*, each party (or its permitted successive *assignees* or *transferees hereunder*) may *assign* or transfer this Agreement as a whole without consent to a Person that succeeds to all or substantially all of the business or *assets* of such party as long as such Person agrees to accept all the *terms and conditions set forth herein*. (Master Confidential Disclosure Agreement, 26.)

After having dealt with the distribution of the words typical of the language of legal confidentiality agreements or so-called genre specific terms, in the next section, I will examine the most common semi-technical words that appear in my corpus.

5.1.4. Common words with uncommon meanings

Mellinkoff (1963:11-12) says that the legal language is characterized by the “frequent use of common words with uncommon meanings”. What he means by this is that some words that have a certain meaning to the non-lawyer may have a very different meaning to the lawyer. They are highly idiosyncratic as they have a very specific meaning within legal discourse and are incomprehensible to non-experts. Some of such common polysemous words found in the corpus of confidentiality agreements and very much related to the language of legal contracts include the following: *party*, which in contrast to its general meaning, in the practice of law means a ‘person contracting or litigating’, *action* a ‘lawful pursuit for justice or decision under the law’, *executed* ‘signed and delivered’, *covenant* ‘sealed contract’, *instrument* ‘legal document in writing’, *said* ‘mentioned before’, *without prejudice* ‘without loss of any rights’, *counterpart* ‘duplicate of the document or a copy’, *virtue* ‘force or authority’ and other commonly found words. All these words have a different meaning in their general use. The legal definitions are adapted from *Black’s Law Dictionary* (<http://thelawdictionary.org/>, 10.10.2012). The table below indicates some of the commonly found semi-technical words.

Semi-technical words	Frequency
Party	818
Provision	71
Action	33
Damages	28
Assign	27
Remedy	25
Covenant	20
Execute	14

Table 8. The frequency of some semi-technical words

These are illustrated in the following examples taken from the corpus:

5. *Without prejudice* to the rights and *remedies* otherwise available to either *party* hereto, each *party* hereto shall be entitled to equitable relief by the way of injunction or otherwise if the other *party* hereto or any of its Representatives breach or threatened to breach any of the *provisions* of this letter agreement. (Confidentiality Agreement, 2.)
6. This Agreement may be *executed* in separate *counterparts*, each of which is deemed to be an original and all of which taken together constitute one and the same agreement. (Confidentiality and Non-Competition Agreement, 24.)
7. Thus, should the Employee breach and term of this Agreement, he/she shall be required to pay the Company or its affiliates as liquidated *damages* 100% of the value of all cash payments...(Agreement Concerning Exclusive Services, Confidential Information, Business Opportunities, Non-Competition, Non-Solicitation and Work Product, 19.)

In the previous sections, I have focused on identifying different types of vocabulary, namely general, academic, technical and common words with uncommon meanings. The next section deals with the frequent lexical items and the less frequent ones.

5.1.5. Wordlist of frequently occurring items

In this section, I will look at the most frequent words that appear in the corpus of confidentiality agreements. Table 9 indicates the ten most frequent words in my collection of texts compared to those in the British National Corpus. As assumed, the functional (or grammar) words occur most frequently in both corpora. The words *the*, *of*, *or*, *and*, *to*, *in*, and *a* match those in the British National Corpus but only appear in different order. The words *agreement* and *party* are unusually frequent, which is explained by the fact that this specialized corpus deals with the issues related to agreements and the parties involved in those agreements. The unusual occurrence of the word *any* indicates another linguistic feature typical of the language of contracts which will be addressed later in the study.

RANK	CAC	BNC
1	THE	THE
2	OF	OF
3	OR	AND
4	AND	TO
5	TO	A
6	ANY	IN
7	IN	THAT
8	AGREEMENT	IS
9	A	IT
10	PARTY	FOR

Table 9. The most frequent items in the CAC and BNC

A significant discrepancy can be seen when examining the content (or lexical) words of the two corpora. The most frequent lexical words in the BNC corpus are *time*, *know*, *people*, *see*, *get*, *way*, *work*, *years*, *think*. Very differently, the CAC corpus shows words like *agreement*, *company*, *information*, *shall*, *party*, *other*, *confidential*, *employee*, *employment and business*. This very evidently indicates the difference between the general

English language and the language used for specific purposes. These words show only a glimpse of the specialized lexicon used in my corpus of confidentiality agreements and the use of some of them will be described in this chapter.

The analysis shows that a great number of words appears very infrequently. It is indicated that 1109 words appear only once in the whole corpus, which amounts to 33.26%, 512 words appear twice and 281 words occur three times in the corpus. As assumed, a big number of words does not appear so frequently, but what is even more distinctive about the confidentiality agreements is that over 75 words appear with the frequency higher than 100, and their frequency ranges from 100 to 3671 occurrences in the corpus. The feature of lexical repetition will be described in section 5.1.10. The hash symbol # also appears frequently in the corpus.¹ In the next section, I will look at the keywords that were identified by the program Word Smith Tools.

Words that appear:	CAC	Percentage
1 time	1109	33,26 %
2 times	512	15,36 %
3 times	281	8,43 %

Table 10. The occurrence of infrequent words

5.1.6. Keywords

In order to identify the keywords in the texts, I compared the list of the words in my corpus with the reference set of words in BNC. Table 11 shows the ten words in my corpus that are outstanding in their frequency which means that they are unusually high in comparison with the norm which is in this case the BNC.

¹ The hash symbol # was ranked number six in the word list of the CAC and number seven in the BNC, but as this symbol represents numbers or strings including numbers (Scott & Tribble 2006: 16), it was not included in my analysis, as it was not considered very significant.

RANK	KEYWORD	FREQUENCY
1	Or	2432
2	Agreement	853
3	Confidential	469
4	Any	1212
5	Party	818
6	Shall	675
7	Company	708
8	Information	720
9	Employee	339
10	Employment	346

Table 11. List of keywords in the corpus of confidentiality agreements

As shown above, the keywords are of great importance for the analysis as they can show linguistic features of the texts which are not likely to be noticed otherwise. Even a glance at the list of keywords in my corpus shows that one of the most remarkable characteristics of the language of confidentiality contracts is the tendency towards lexical repetition. The frequent occurrence of the word *or* reveals a tendency towards binominal expressions. The keywords presented here appear strikingly frequently compared to the BNC norm and this repetition of words carries a certain value which will be discussed in more detail later in this chapter. Apart from the lexical words which are undoubtedly related to the specialized topic of confidentiality agreements, such as *agreement*, *confidential*, *company*, *employee*, *information*, *party* and *employment*, another evident feature is the high frequency of the word *or*, *any* and the verb *shall*, the most common verb identified in my collection of texts.

Some of the words with negative keyness which appear relatively infrequently in the corpus are *was*, *he*, *it*, *had*, *were*, *they*, *she*, *we*, *there*, *a*, *up*, *on*, *you*, etc. This shows us that the notion of past, for instance, is not so important for the confidentiality agreements as the conveying of information is more related to the present and the future. The use of pronouns is also rather infrequent because their referential value, in the case of legal documents where the precision is very important, would bring ambiguity.

Due to the fact that the keywords are very important for the analysis of lexicogrammatical features of the genre in question, in the text that follows I will examine the concordance of the top ten keywords found in my corpus. These keywords show lexicon that clearly distinguishes this specific genre from other legal genres as they identify the important topics, such as confidentiality of information, employment, agreements, etc. as well as other features, like lexical repetition, tendency towards inclusiveness, etc.

5.1.6.1 Concordance of the keyword *or*

The purpose of a concordance is to show us how certain words are used in the text and what meaning they convey. As shown in Table 12, the keyword *or* appears striking 2527 times in the corpus of confidentiality agreements. The most frequent collocates are listed below:

Collocates	Frequency
Other	110
Any	105
Company	97
Its	97
In	88
Otherwise	66
To	66
Directly	62
Indirectly	61

Table 12. The collocates of the keyword *or*

Considering the Mutual Information score which tells the strength of the collocation, however in the case of *or* shows that the collocates worth mentioning are *unenforceability*

with the extremely high MI score of 8655, followed by *affiliates* 6704, *patentable* 6539, *solicit* 6292, etc. The following example shows the typical use of *or* which in this instance appears even seven times in a single sentence.

8. The Employee hereby agrees that, from time to time upon the reasonable request of the Company, the Employee shall assist the Company in connection with any pending *or* future dispute, litigation, arbitration *or* similar proceeding *or* investigation *or* any regulatory requests *or* filings involving the Company, any of its employees *or* directors *or* the employees and directors of any subsidiary. (Confidential Settlement Agreement and General Release, 1.)

As example 8 illustrates, the word *or* functions as one of the connecting words (together with the word *and*) which form binomial or multinomial expressions. As we can see, these expressions are highly frequent in the language of contracts and they serve as a means of expressing all-inclusiveness and precision, among other things. The binomial and multinomial phrases will be discussed in more detail in section 5.2.2.

Looking at the significant clusters of the word *or*, some very frequent ones worth mentioning would include *the company or*, which appears 111 times in the corpus, *directly or indirectly*, 86 times, *or its affiliates*, 85 times, *whether or not*, 52 times, *person or entity*, 37 times, *in whole or in part*, 19, and many other.

5.1.6.2 Concordance of the keyword *agreement*

Collocates	Frequency
This	577
Shall	139
And	80
The	69
Or	40
Is	37
Confidential	28

Purchase	27
Entire	25

The word *agreement* is the second most frequent word in the corpus with its remarkable appearance of 853 times in all 30 texts. Table 13 shows the most frequent collocates of this word.

Table 13. The collocates of the keyword *agreement*

According to the analysis, the collocate *this agreement* seems to be the most frequent in this collection of texts as it appears 577 times. That can be explained by the fact that the confidentiality agreements, and contracts in general, strive for precision and prevention of misinterpretation, so by specifying *this agreement*, the legal draftsmen of the contract emphasize that only this agreement shall be taken into account and not any other (unless otherwise specified). The definite article *the*, which appears 69 times in the corpus together with *agreement*, and *this* are used to refer to that particular agreement and in this case they bring about the precision, one of the critical features for the language of contracts. Trosborg classifies this as “legal concretes” as these items make up ‘the specialized apparatus of the contract’. (Trosborg 1997: 119) This is illustrated in the following example:

9. ***This Agreement*** shall be binding on and inure to the benefit of the parties hereto and their permitted and respective heirs, successors and assigns. (Noncompetition and Confidentiality Agreement, 22.).

According to the MI score, the most important collocates include *agreement supersedes* with the score of 9068, then *confidential agreement*, 8329, *disclosure agreement*, *agreement shall*, *definitive agreement*, *purchase agreement*, *entire agreement*, *separation agreement*, *agreement constitutes*, *transaction agreement*, etc.

In addition to this, the following clusters, which also appear very frequently, are worth mentioning:

Clusters	Frequency
Of this agreement	233
This agreement shall	122
This agreement shall be	63
Provision of this agreement	37
Terms of this agreement	26

Table 14. The clusters of the keyword *agreement*

5.1.6.3 Concordance of the keyword *confidential*

As expected, another keyword with a high frequency very specifically related to the topic of the texts in question is the word *confidential*. This word occurs 469 times in the corpus and just as the word *agreement*, it appears in all texts. Some of the most frequent collocates are shown in Table 15 below.

Collocates	Frequency
Information	363
The	80
Any	41
Of	37
Business	33
Such	32

Table 15. The collocates of the keyword *confidential*

The word combination with the most extraordinary but not surprising number of appearance is *confidential information*. It also has the highest MI score of 9736. This collocate appears 363 times throughout the CA corpus and presents one of the essential concepts of the agreements. This is illustrated in the following example:

10. If the Discloser fails to mark or identify disclosed ***Confidential Information*** as “**Confidential**”, the Recipient shall be responsible for protecting such disclosures in accordance with this Agreement from the date of receipt of written notice by the Discloser identifying the disclosure as “**Confidential**” and requesting that such disclosure be treated as ***Confidential Information*** under this Agreement. (Confidentiality Agreement (Mutual), 9.)

Other collocates with very high MI score are *disclose, non, secrets, disclosure, company's, strictly, business, master, highly*, among others. Furthermore, some relevant clusters are identified in the corpus. Some of them include *the confidential information*, which occurs 64 times in the texts, *confidential information and*, 41, *any confidential information* 35, *confidential business information*, 31, *of confidential information*, 31, *such confidential information*, 24, *all confidential information*, 18, *confidential or proprietary*, 13, etc. As we can see, the most common clusters include *confidential information* which is not surprising as this is the most important concept of this genre.

5.1.6.4 Concordance of the keyword *any*

The language of legal contracts is characterized by the frequent use of the word *any*, as Danet points out (1985). This has also been proven in the analysis of my texts which has shown that the word *any* appears outstanding 1212 times.

Collocates	Frequency
Of	267
In	139
Other	131
Or	105
To	99

Table 16. The collocates of the keyword *any*

As the study suggests, *any* is a distinctive word in the language of contracts. It refers to categories in general, but most importantly, here it also carries the meaning of inclusiveness, one of the very important features of the confidentiality agreements, as indicated in the following example:

11. (...) between the Company and ***any*** other person including, without limitation, ***any*** consultant, contractor, customer, potential customer, or supplier of the Company; or (h) engage in or participate in ***any*** business conducted under any name that shall be the same as or similar to the name of the Company or ***any*** trade name used by the Company. (Employee Confidentiality, Invention Assignment and Non-Compete Agreement, 13.)

According to the relation computed by MI score, the most important collocate is *any copyrightable*, with the score of 11254, *waives any*, 7254, *supersedes any*, then *solicit any*, *any confidential*, *any unauthorized*, etc. Additionally, the following clusters were identified as rather frequent: *at any time*, which occurs 45 times, *any of the*, 43 times, *any confidential information*, 39 times, *or any other*, *any and all*, *in any way*, *any and all*, *for any reason*, etc.

5.1.6.5 Concordance of the keyword *party*

The keyword *party* is another lexical item extremely frequently found in my corpus with the occurrence number of 648. The most common collocates are *party A*, *party B*,

disclosing party, *receiving party*, *third party*, with the frequency numbers 226, 192, 96, 88, 62, respectively. The collocates with high MI score are *disclosing party*, with the score of 10894, *party hereto*, *party A's*, *party B's*, and *receiving party*. The example below shows the use of some of the collocates.

12. Title to the Confidential Information will remain solely in the ***Disclosing Party***. All use of Confidential Information by the ***Receiving Party*** shall be for the benefit of the *Disclosing Party* and any modifications and improvements thereof by the ***Receiving Party*** shall be the sole property of the ***Disclosing Party***. (Non-disclosure Agreement, 17.)

The combinations of words commonly found in the CAC are shown in the table below.

Clusters	Frequency
The disclosing party	89
The receiving party	87
Party B shall	52
By party A	49
Party A and	48

Table 17. The clusters of the keyword *party*

Some other common clusters are *party A and party B*, *the other party*, *between party A and party B*, and *a third party*.

5.1.6.6 Concordance of the keyword *shall*

The use of modal *shall* is widely common in the language of legal contracts which has been confirmed by the number of 675 appearances in the CAC. It is also specific as it has a rather unusual sense in the language of legal texts. Apart from its general use of implying futurity, *shall* in legal texts serves as a means of expressing an obligation or a duty. Williams (2005: 116) states that it also expresses deontic modality which “is intrinsically projected towards regulating behaviour and situations located in the future”. Butt and

Castle (2006: 200) also mention command (mandatory) and directory provisions which here mean “provisions requiring a certain course to be taken”. Tiersma (1999) defines the use of *shall* also as a proposition of the performative phrase. For instance, the parties perform the act of promising by signing the contract, as he points out.

The use of *shall* is clearly shown in the two examples that follow.

13. Party B ***shall*** have the obligation to disclose to Party A all the intellectual property rights applied or obtained by Party B during the Service Term of Party B and within one year after Party B’s Separation. (Confidentiality and Non-Competition Agreement, 6)
14. Upon Separation from Party A due to any reason, without the prior written consent of Party A, Party B ***shall not*** hold any position in any Competing Unit within the period to which the economic compensation fee for non-competition paid by Party A is applicable. (Confidentiality and Non-Competition Agreement, 6.)

The most frequent collocates are shown in the table below.

Collocates	Frequency
Be	241
Not	145
Agreement	139
And	70
B	42
Have	33
Employee	33

Table 18. The collocates of the keyword *shall*

The collocate *shall be* is the most common and it occurs in all texts. Nevertheless, the collocates with the high MI score are *shall inure*, *hereunder shall*, *hereto shall*, *shall*

supersede, and herein shall. Beside the mentioned collocates, the following clusters could be of relevance: *this agreement shall be*, occurring 62 times in the texts, *shall not be*, 43 times, *party B shall*, 42, *shall be deemed*, 27, *shall be entitled to*, 24, and many others. The use of *shall* clearly shows the tendency towards impersonalization and formality.

5.1.6.7 Concordance of the keyword *company*

The lexical item *company* is another extremely frequent item and it appears 796 times within the CAC. This is not surprising as all the texts deal with the topic of agreements between companies or companies and private persons. The significant collocates of *company* are shown in the table below.

Collocates	Frequency
The	623
And	99
Or	97
To	28
May	19
In	18
shall	17

Table 19. The collocates of the word *company*

The collocates with the highest MI score are *affiliate company* with 6762 MI score and *company agrees* with 3146.

As for the typical clusters, the following ones are noteworthy: *of the company*, occurring 169 times throughout the corpus, *by the company*, 94, *the company or*, 93, *the company and*, 84, *with the company*, 81, etc. Some of the clusters are presented in the following sentence taken from the contract.

15. Except as may be required and authorized in the course of his/her employment *with the Company*, the Employee shall not at any time during his/her employment *with the Company* or after the termination thereof for any reason disclose or use, directly or indirectly, any confidential or proprietary information *of the Company* or its affiliates. (Agreement concerning exclusive services, confidential information, business opportunities, non-competition, non-solicitation and work product, 19.)

5.1.6.8 Concordance of the keyword *information*

The word *information* presents a very important concept for the topic of confidentiality agreements and it appears 720 times in the corpus. The most common collocates of *information* are *and* which appears 91 times, then *proprietary*, 73 times, *of*, 58, *such*, 45, etc. As assumed, the collocate *confidential information* has the highest MI score. The two other collocates with high scores are *proprietary information* and *disclosed information*. The important clusters and their frequency are indicated in the table below.

Clusters	Frequency
The confidential information	63
Confidential information and	41
Any confidential information	35
Of confidential information	33
Confidential information of	32
Confidential business information	31

Table 20. The clusters of the keyword *information*

Other important clusters include *such confidential information*, *confidential information means*, *all confidential information*, *among others*. This is illustrated in the example below.

16. “**Confidential Information**” means *business information*, technical data, know-how and other information that is not otherwise in the public domain and of which the

owner actively undertakes to restrict or control the disclosure to Third Parties...(Master Confidential Disclosure Agreement, 26.)

5.1.6.9 Concordance of the keyword *employee*

The keyword *employee* appears 339 times in the corpus of confidentiality agreements and in 20 texts. The most frequent collocates are listed in the table below.

Collocates	Frequency
The	127
And	36
Shall	33
Of	32
Consultant	26
by	17

Table 21. The collocates of the keyword *employee*

In view of the above, here are some examples that include and show the use of the given keyword.

17. ***Employee*** hereby acknowledges and agrees that any breach by ***Employee*** of the provisions of this Agreement may cause (name of the company) irreparable harm for which there is no adequate remedy at law. (Confidentiality, Non-Competition and Invention Agreement, 12.)
18. ***Employee*** acknowledges that, during the period of ***Employee's*** employment with the Company, ***Employee*** has had or will have access to Confidential Information of the Company. (Employee Confidentiality, Invention Assignment and Non- Compete Agreement, 10.)

As shown in the examples above, the collocate *employee hereby* has extremely high MI score of 11075, followed by *employee acknowledges*, *employee consultant*, *employee agrees and employee understands*. The following clusters are identified: *to the employee*, which occurs 18 times, *by the employee*, 14, *employee shall not*, 12 times, *employee and the*, 11, *as an employee*, 11 and *the employee shall*, 11 times.

5.1.6.10 Concordance of the keyword *employment*

The last word among the top ten keywords is the word *employment* which appears 346 times throughout the corpus. The most frequent collocates are *my*, *with*, *the*, *of*, *or*, *by*, *your employee's*, and they appear with the following frequencies 96, 92, 48, 36, 34, 32, 31 and 24 respectively. The collocates with high MI scores are *associate's employment*, *employee's employment*, *executive's employment*, etc. This is shown in the following sentence.

19. I agree that during the period of ***my employment*** I shall devote my full time, skill, energy and efforts to (name of the company) and shall not participate, directly or indirectly, in any capacity, in any business or activity that is in competition with (name of the company). (Employee Non-competition, confidential information and work product agreement, 8.)

The table below indicates some important clusters of the keyword *employment*.

Clusters	Frequency
Of my employment	65
Employment with the	49
Employment with the	49
Employment with the company	48
My employment with	33
Termination of my employment	22

Table 22. The clusters of the keyword *employment*

In this section, I have scrutinized the top ten keywords found in the CAC with regard to their frequency and significant collocates and clusters. The findings have shown that the words *or* and *any* appear extremely frequently and they point to the critical feature of confidentiality agreements. This feature is to try to include all cases and possibilities and avoid any ambiguity and possible misinterpretation. It is also very clear that the words like *party*, *company*, *information*, *confidential*, *employee* and *employment* appear very frequently as they present the main concepts of this genre. In the following sections, I will examine some other distinctive lexical features of the confidentiality agreements.

5.1.7. Antiquated terms

The language of prescriptive legal texts is often characterized by a high frequency of archaic or antiquated expressions, the terms which are no longer in use in ordinary language and came from Old and Middle English. Tiersma (1999:95) states that “legal language often strives toward great formality, it naturally gravitates towards archaic language”. These archaic expressions may include adverbial expressions such as: *hereto*, *herewith*, *hereinafter*, *thereby*, etc., archaic morphology: the ending ‘-eth’ instead of third person singular ‘-es’, as in *witnesseth* for example, the legal use of *such*, *aforesaid*, etc., the use of subjunctive as in ‘*be it known*’ and many others (Tiersma, 1999).

Most archaic words that appear in the language of contracts have referential values. The word *hereinafter*, for example, refers to the following part of the document, *hereto* means ‘to this document’, *aforesaid* relates to the part mentioned earlier, etc. The examples of archaic terms found in the corpus are the following:

20. 1. NOW THEREFORE, in consideration of the mutual promises and agreements contained ***herein***, the parties ***hereto***, with the advice of counsel and intending to be legally bound, ***hereby*** agree to the foregoing and as follows...
21. ...the Employee shall not at any time during his/her employment with the Company or after the termination ***thereof*** for any reason disclose or use, directly or indirectly, any confidential or proprietary information of the Company or its affiliates.

Some of the most frequent antiquated legal adverbials found in the CAC are the following: *herein*, which appears 57 times, *deemed* 45 times, *thereof*, 48 times, *whereas*, 16, *hereinafter*, 16, *whereof*, 13, *therein*, 12, *thereby*, 6, etc. Other examples include *hereon*, *therewith*, *thereunder*, *whereby*, *witnesseth*, *duly*, *such*, etc.

Except for bringing formality as Tiersma (1999:96) says, these expressions are often used by legal draftsmen because they are safe and convenient. As they have been in use for a very long time, they have a certain degree of authoritative interpretation and therefore he states that the drafters try to avoid changes because what has proven to function before, does not need to be changed.

5.1.8. Loanwords and expressions

A great number of words and phrases in legal English was borrowed from Latin or indirectly through French. It is considered that the Latin legal terms in their original forms such as *ex parte*, *bona fide*, *corpus delicti*, *ab initio*, *per capita*, *res judicata*, etc., occur relatively frequently in legal writing. Generally speaking, although still present as technical legal terms, they are much less common than they once were but still show a distinctive feature of certain legal genres (Crystal and Davy, 1996: 208-209). Even though they are accepted as part of English language, as Mellinkoff (1963: 14) points out, many are not in active use. In my collection of texts, I failed to find any instance of these terms and it has been proven that the Latin expressions do not characterize this specific genre and their use is most probably decreasing.

On the other hand, a number of French terms which are anyway mostly derived from Latin, commonly occur in the language of legal contracts. Some of the terms of French origin, frequently found in the confidentiality agreement corpus, are the following: *agreement*, occurring 853 times, *party*, occurring 818 times, *property*, 123 times, *agree*, 142, *condition*, *subject*, *damages*, *possession*, *purchase*, *court*, *action*, *terms and respect*, to name just a few. These words also occur frequently in everyday English.

5.1.9. Capitalization

Capitalised terms are terms which are written with a capital letter. The initial capitalization appears frequently in legal contracts and it plays an important role. The capitalised terms are used in legal documents to identify the defined terms, while the uncapitalised terms refer to a general meaning. For instance, the *Shareholder* can only refer to the specific shareholder defined in the contract, while the uncapitalised *shareholder* can include any other shareholder. Although, capitalization is a textual feature rather than lexical, it is embedded in this section as it is more related to lexical than syntactic features.

To illustrate this, I give examples taken from one of the agreements, referred throughout the text as the “Agreement”. Other capitalised terms in the text include the

Company, the *Employee*, the *Parties*, the *Effective Date* being for example August 15, 2004, then the *Payment* \$1,540,000, the *Confidential Information*, the *Expiration Date*, etc. The terms are once defined and they only refer to that defined term. In this way, they are used like proper nouns (instead of repeating for instance the names of people, companies, etc.) and they convey exactness and precision which is in this context of great significance.

In some instances, certain words or whole chunks of text are printed in capitals, as a means of emphasizing their importance or the importance of what follows.

22.I UNDERSTAND THAT THIS AGREEMENT AFFECTS MY RIGHTS TO INVENTIONS I MAKE DURING MY EMPLOYMENT, AND RESTRICTS MY RIGHTS TO DISCLOSE OR USE THE COMPANY'S CONFIDENTIAL INFORMATION OR TO COMPETE WITH THE COMPANY DURING OR SUBSEQUENT TO MY EMPLOYMENT.

I HAVE READ THIS AGREEMENT CAREFULLY AND UNDERSTAND ITS TERMS. I HAVE COMPLETELY FILLED OUT EXHIBIT A TO THIS AGREEMENT. (Employee Confidentiality Agreement, 25.)

5.1.10. Some other lexical features of the confidentiality agreements

As shown in section 5.1.6. some lexical items are extremely frequently used in the corpus of confidentiality agreements. This is demonstrated by the great number of occurrence of words such as *agreement*, *company*, *information*, *party*, *confidential*, *employee* and many *others* and it brings about another very striking characteristic of legal language and that is lexical repetition or redundancy. The repetition is also illustrated by the high frequency of the words *any* and *such*, found 1212 and 470 times respectively. As a consequence, a higher precision is achieved. As pronouns can sometimes be ambiguous, the repetition of nouns can enhance precision and show the exact meaning. This is why the pronouns are rather scarce in the sample of confidentiality agreements and apart from the pronouns *I* and *you*, all other pronouns have negative keyness. Two other words with negative keyness are *could* and *might*, which is not surprising as they carry a degree of imprecision. The repetition is also shown by the use of words with the same or almost the same meaning. The use of synonyms is discussed later in the text as a part of binomial or multinomial expressions.

Additionally, the language of legal contracts shows a tendency towards high formality (Bhatia, 1993, Danet, 1985, Mellinkoff, 1963, etc.). This is, as already mentioned before, proven in the CAC by the frequent use of *shall*, antiquated expressions and Old and Middle English words, the use of French words, passive forms, nominalization, etc. The use of complex prepositional phrases also contributes to this feature. (Bhatia, 1993: 107). The pattern P-N-P (Preposition+ Noun + Preposition) is commonly found in the CAC. The most frequent is *in accordance with*, identified 34 times in the corpus and in 23 texts, then *for the purpose of*, *in the event of*, *in respect of*, *by virtue of*, etc.

With regard to the lexical features, as we have seen, the language of the confidentiality agreements analyzed in this study shows rather distinctive features. Some of them, as we have seen, include lexical repetition, tendency towards all-inclusiveness, binominal and multinominal expressions joined by the words *or* and *and*, frequent use of *shall* for regulating behaviour, etc. Even more distinctive are syntactic features, which will be analyzed in the following section.

5.2. Syntactic features

The second step in the analysis includes the analysis of syntactic features. On a syntactic level, the language of the confidentiality agreements has shown a tendency towards long and complex sentences, the frequent use of passive forms, nominalization, and binomial and multinomial expressions. The use of qualifications is also found to be important for the language of the contracts. It is also evident that the legal writing tends to be impersonal and very formal. These are some of the features I will discuss in the next sections.

5.2.1. Sentence length and complexity

One of the most remarkable syntactic features of legal contracts is extremely lengthy and complex sentences. Barber (1962) identified the typical sentence in written scientific English as consisting of 27.6 words (referred to in Bhatia 1993). Gustaffson (1975) found that the average legal sentence in the corpus he analyzed contained about 55 words. The legal register sentences are not only longer than the typical sentences in written scientific English for example, but their structure is often more complex as it typically consists of more subordinate clauses. Gustaffson (1975) identified an average of 2.86 clauses per sentence. The average sentence in my corpus of confidentiality agreements contains 39.07 words per sentence, which is substantially higher compared to 27.6 words in written scientific English, for instance.

Crystal and Davy (1969) observed that other types of discourse tend to convey the information by series of short sentences linked together in a continuity, while on the other hand, legal prescriptive writing is characterized by long, complex sentences, which could stand alone. They imply that the distinctive characteristic of legal writing is to “conflate, by means of an array of subordinating devices, sections of language which would elsewhere be much more likely to appear as separate sentences” (Crystal and Davy, 1963: 201). Legal sentences, particularly in written context, very often present “self-contained units which convey all the sense that has to be conveyed at any particular point” (1963: 201). This is why legal prescriptive writing is characterized by the lack of words which function as sentence connectors and they are rather low in cohesive devices. The texts found in my corpus also show a tendency to state a whole idea in a single paragraph as a whole sentence. To illustrate the complexity of the sentences in my corpus, I enclose a whole section which consists of only one sentence with an above-average length of 175 words and different types of clausal embedding:

23. Subject to the immediately succeeding paragraph, unless otherwise agreed to in writing by the Disclosing Party, the Receiving Party and its Representatives (i) shall not use Proprietary Information for any purpose other than evaluating the Transaction and consummating the Transaction in a manner approved by the Company, and (ii) except as required by applicable law, regulation (including, without limitation, any rule, regulation or policy statement of any national securities exchange, market or automated quotation system on which any of the Receiving Party's securities are listed or quoted) or legal process, shall keep all Proprietary Information confidential, shall use at least the same degree of care to protect the Proprietary Information as it uses with its own confidential information and shall not disclose or reveal any Proprietary Information to any person other than those persons who are employed or engaged by the Receiving Party or its Representatives and actively and directly participating in its evaluation of the Transaction or who otherwise need to know the Proprietary Information for the sole purpose of evaluating the Transaction. (Confidentiality Agreement, 2.)

5.2.2. Binomial and multinomial expressions

Other characteristics prevalent in legislative texts include binomial and multinomial expressions (Bhatia, 1984, 1993; Gustafsson, 1975). The term *binomial* was first introduced by Yakov Malkiel (1959) and defined as “a sequence of two words pertaining to the same form-class, placed on an identical level of syntactic hierarchy, and ordinarily connected by some kind of lexical link” (quoted in Gustafsson, 1975:9). Furthermore, Gustafsson (1975:16) refers to Malkiel’s distinction of the *irreversible* binomials where the mutual order is fixed, and *reversible* where their positions may be changed. According to him, they can also be differentiated as *formulaic* or permanent and *unformulaic* or temporary language combinations.

Binomials and multinomials as linguistic tools are not only common in colloquial language spoken and written, but are also particularly frequent in scientific and professional language such as technical and legal for instance, where it is essential to have technical accuracy and precision. The importance of binomials and multinomials in legislative writing is extremely high. Their significance for conveying the information lies primarily in achieving precision and all-inclusiveness and they are highly effective (Bhatia, 1993:108). To illustrate this, Bhatia gives the following examples taken from various legislative texts: *sign and delivered, in whole or in part, by or on behalf of, under or in accordance with, wholly and exclusively*, but also the more complex ones: *from such activities or from the use of any such machinery or plant, by the Government or by any government, public or local authority... or by any person other than the person claiming relief*. The binominal and in particular multinomial expressions (identified manually when examining the binominal expressions) are extremely frequently found in the corpus of confidentiality agreements and the most common binomial expressions are indicated in the table below.

Binomials	Frequency
directly or indirectly	61
whether or not	40
person or entity	27
any and all	29
company or its affiliates	28
successors and assigns	20
in whole or in part	19
terms and conditions	15

Table 23. Binomial expressions in the corpus of confidentiality agreements

The classification of semantic relationship which will be used in the present study is based on Gustafsson's (1975:85-115) classification of relationship into semantic opposition, semantic homoeosemy which means similarity of meaning, semantic complementation and semantic hyponymy. As she stresses, the distinction between the semantic relation the members of binomials imply is not always particularly easy to define. According to her, the semantic opposition stands for the relation where the members express the contrast and the polarization of the information. This opposition may be expressed in different ways. The following examples found in my corpus explicitly show the opposition by a negative morpheme, a prefix or a suffix, such as *un-*, *in-*, *dis-*, *non-*, etc. Additionally, this relation includes time relation, chronological sense, reciprocity and result, among many others. This will be exemplified in the following binomials found in my corpus. The examples presented in this section include also some multinomials. The expressions which belong to the semantic relation in question are marked in bold whereas all the other binomials or multinomials are written in italic.

24. The Employee shall not, without prior written permission of the Company, ***directly or indirectly*** either as an *officer, director, employee, agent, advisor, consultant, principal, stockholder, partner, owner (...)* in any way *engage in, represent, be connected with or have* a financial interest in, any business...(Agreement concerning exclusive services, confidential information, business opportunities, non-competition, non-solicitation and work product, 19.)
25. ...whether such requests occur ***prior to or after*** termination of Executive's employment with the Company. (Confidentiality and Non-competition Agreement, 24.)
26. ...communication of such information to any other person when it is reasonably foreseeable that such other person is likely to ***purchase or sell*** such securities in reliance upon such information.(Confidentiality Agreement, 2.)
27. This is the entire agreement and supersedes all prior *agreements and understandings*, whether ***oral or written***, between the parties with respect to the subject matter hereof. (Associate Confidentiality Agreement, 27.)
28. ...to waive compliance, ***in whole or in part***, with the terms of this latter agreement. (Confidentiality Agreement, 2.)

Other antonym pairs of binomials found in the corpus include: *foreign and domestic*, *current and former*, *vested and unvested*, *patentable and unpatentable*, *with or without cause*, *his or her*, *registered and not (registered)*, and many others.

Gustafsson (1975: 95-99) defines the semantic homoeosemy as a relation where there is a notational similarity and the same set of features between the members of binomials, which could roughly correspond to synonymy. In addition, he asserts that at a high level, many binomials could fall into this category. This relation could be expressed by repetition showing formal similarities, although not many cases of repetitive binomials were found in my collection of texts. Another example of homoeosemy would be foreign and familiar words which occurred in my corpus, as well as words, which to ordinary speaker appear as if they have the same meaning but they in fact belong to legal jargon and require a certain level of expertise. These examples perfectly show the importance of binomials in legislative writing as a linguistic device to show the accuracy and precision:

29. ...and shall be primarily (and not as a guarantor) ***responsible and liable*** for any ***breach or violation*** by any of its Representatives of the terms of this letter agreement that applies to Representatives. (Confidentiality Agreement, 2.)
30. ... to hold such information ***in confidence and trust...*** (Agreement regarding confidential information and proprietary developments, 14.)
31. In any administrative ***proceeding or litigation*** in which SOHU may be involved relating to any Work Product...(Employee non-competition, confidential information and work product agreement, 8.)
32. ...the Company must protect its good will, its vase of members and prospective members, its employees, its ***confidential and proprietary*** information, and the work product of its employees. ... (Agreement concerning exclusive services, confidential information, business opportunities, non-competition, non-solicitation and work product, 19.)
33. ...or the taking of actions which impair the Employee's ability to perform such ***duties and responsibilities***; or any material violation of Company ***rules and regulations***. (Agreement concerning exclusive services, confidential information, business opportunities, non-competition, non-solicitation and work product, 19.)

34. Executive is not a party to or bound by **any employment agreement, noncompete agreement or confidentiality agreement** with any other person or entity and upon the **execution and delivery** of this Agreement by the Company... (Confidentiality and Non-competition Agreement, 24.)

As Gustafsson (1975:99) points out semantic complementation cannot be strictly defined. The members of binomials are here not synonyms or antonyms but they are semantically related. Very often one of the members has one feature which the other one does not have, be it time, logical order or some other feature. She refers to Koskeniemi (1968) who considers the ‘contiguity of meaning’ a determining aspect (referred to in Gustafsson, 1975:86). In view of the above, I display the following examples which express different relations such as sequence, complement, representativeness and specificity among others:

35. ... means any **neglect of, or refusal or inability to perform**, the Employee’s *duties or responsibilities* with respect to the Company with the same level of contribution as in part periods of employment; or any **insubordination, dishonesty, negligence or malfeasance** in the performance of such *duties and responsibilities*. (Agreement concerning exclusive services, confidential information, business opportunities, non-competition, non-solicitation and work product, 19.)
36. ... developed and maintained relationships with its **clients, talent, producers, designers, programmers, distributors, merchandisers, advertisers, and employees**, and participated in and was responsible for ...(Confidentiality, Non-competition, Non-solicitation, and Non-recruitment Agreement, 21.)
37. ...under the terms of this Agreement shall immediately be **forfeited and terminated**, and any amount already paid (but not including any **federal, state and local** taxes that were withheld at payment) by the Company to the Employee... (Confidentiality Agreement (Mutual), 9.)

The distinction between the third and the fourth type of relation, the semantic complementation and hyponymy, according to Gustafsson (1975:86), depends on interpretation and is not easily defined. However, she bases the distinction on Bierwisch’s (1971: 170) definition of hyponyms where one member ‘contains all the components occurring in the meaning of the item, but not vice versa’ (Gustafsson, 1975:86). Croft and

Cruse (2004: 142) define this relation as “If X is a hyponym of Y, then the semantic content of Y is a proper subpart of a semantic content of X”. A further instance of this is illustrated in the following examples:

38. ... ***the press or the media*** or any other administrative agency, in either case that would portray the Company... (Confidentiality Agreement (Mutual), 9.)
39. ...to provide that all Proprietary Information that is so disclosed will be accorded confidential treatment to fullest extent available under applicable ***laws and regulations***. (Confidentiality Agreement, 2.)
40. ***ACCEPTED AND AGREED TO*** (Employee Confidentiality Agreement, 25.)

As this section has shown, the binomials and multinomials are extremely frequent in legislative texts and they are of great significance when it comes to expressing precision, accuracy, emphasizing, intensification and all-inclusiveness.

5.2.3. The use of passive

Legal language is highly characterized by the extensive use of passive constructions (Danet, 1985; Williams, 2004; 2005). Williams argues (2004:231) that nearly a quarter of all verbal constructions in legal prescriptive texts are passive constructions. As Sarcevic (2000:177) points out, the use of passive brings the impersonalization to the language of contracts. In contrary to the generally accepted view that because the agent is not specified, the legal sentences can be ambiguous and confusing, Williams (2005:159) justifies the use of passive forms in legal documents in certain cases and states that “it may not be possible to specify the agent without drawing up a long (and probably incomplete) list of bodies that might be responsible for communicating ratification”. This is when passive construction might come handy also in order to avoid further wordiness of the document.

The fact that passive voice is prominent in the language of legal English has also been proven in my collection of texts. In order to identify passive forms, I have looked at the forms of *be* and searched for their combinations with the past participle. The following table indicates the frequency of all the forms of the verb *to be* with the right column presenting only passive forms. The past perfect passive has not been found.

The verb to be	Frequency	Used as passive
Be	610	442
Is	375	223
Are	167	87
Was	33	26
Were	6	3
Have been	18	16
Has been	21	24

Table 24. Passive forms in CAC

The results show that passive forms have been found in 821 instances throughout the corpus, which presents quite a significant number. Example 41 shows that the agents in passive constructions can also be specified, as shown in example 41.

41. The Receiving Party acknowledges that the Confidential Information is proprietary to the Disclosing Party, *has been developed* and *obtained* through great efforts by the Disclosing Party and that Disclosing Party regards all of its Confidential Information as trade secrets...(Non-Disclosure Agreement, 17.)

42. These Covenants *will be interpreted* and *enforced* in accordance with the laws of the State of Washington as applied to agreements made and performed in Washington, without regard to the State's conflict of laws provisions. (Development, Confidentiality and Noncompetition Covenants, 23.)

43. This Agreement *may be executed* by the parties in separate counterparts, each of which when so executed and delivered shall be an original, but all such counterparts shall together constitute but one and the same instrument. (Noncompetition and Confidentiality Agreement, 22.)

44. This Agreement *shall be governed* and *construed* and *enforced* in accordance with, the laws of the State of Delaware without giving effect to its conflicts of laws rules, and *shall be deemed* to be effective as of the first day of my employment with SOHU. (Employee Non-Competition, Confidential Information and Work Product Agreement, 8.)

5.2.4. The use of qualifications

Bhatia (1993:110-111) identifies the use of qualifications as the most important feature of legal writing. The qualifications present the conditions that have to be met in order for something else to happen. He also emphasises that what is more interesting is the way these qualifications are inserted within a single sentence. According to him, the language of legal documents is “extremely rich in qualificational insertions within their syntactic boundaries [...] The qualifications seem to provide the essential flesh to the main proposition without which the provision will be nothing more than a mere skeleton, of very little legal significance” (1993: 111). These long qualifications, like in the following example from the corpus, considerably add to the awkwardness and complexity of the legal sentence. Very often, as he states, they are inserted next to the word they qualify.

45. In addition, if, in connection with the business or affairs of any member of the Group, Mr. Fuller shall have obtained Confidential Information belonging to any third party under an agreement purporting to bind any member of the Group which contained restrictions on disclosure, he will not, without the previous written consent of the Purchasers (as defined in the Stock Purchase Agreement) and, if appropriate, the relevant third party at any time infringe such restrictions.

Qualifications are in a syntactical sense presented by conditional sentences which can be, in legislative writing, very often rather complex. Crystal and Davy (1969:203) state that a great number of legal sentences have a logical structure which is here simplified such as the following:

“If X, then Z shall be Y”

or,

“If X, then Z shall do Y”,

where ‘X’ presents the description of case, ‘Y’ is the legal subject and ‘Z’ the legal action. They point out that “every action or requirement, from a legal point of view, is hedged around with, and even depends upon, a set of conditions which must be satisfied before anything at all can happen” (Crystal & Davy, 1969:203). They also emphasize that the logical structure of the legal writing certainly has effects on its syntax used by people who have to conform to this logical or semantic framework. Such propositions may be formulated in various ways using the words *if*, *in the event of/that*, *unless*, *provided that*, *shall*, etc. Here are some examples taken from the CA corpus:

46. **If** a Court determines that at the time this Agreement is presented for enforcement any provisions are overly broad or unenforceable, the parties agree that the Court shall reform the Agreement to make it enforceable to the maximum extent possible and shall enforce the other terms as written. (Confidentiality, Non-Competition, Non-Solicitation, and Non-Recruitment Agreement, 21.)
47. **In the event that** any action or proceeding is commenced by any party hereto for the purpose of enforcing any provision of this Agreement, the parties to such action or proceeding may receive as part of any award, judgement, decision, or other resolution of such action or proceeding, their costs, and reasonable attorneys' fees. (Noncompetition and Confidentiality Agreement. 22.)

5.2.5. The use of nominalisation

Another common feature of legal language is the use of nominalisation, where noun phrases are used instead of verb phrases (Bhatia 1993, 1997; Maley 1994; Tiersma 1999; Crystal and Davy 1969). Crystal and Davy (1969: 206) emphasise that most of these nouns are either abstract or they do not primarily refer to physical objects. Nominalisation contributes to formality and the impersonal tone of the documents.

In order to show how the confidentiality agreements are highly nominal in nature, I present the findings of the analysis, where the nouns which end with *-tion* appear 2148 times, the number of nouns which end in *-ment* is 1500, the once with *-ance* is 161, and the ones with *-ence* is 89. Some of the nominal words found in the corpus are:

agreement, termination, consideration, violation, performance, maintenance, assignment, payment, regulations, amendment, enforcement, development, restriction, invention, transaction, modification, compensation, injunction, etc.

Table 25. Some nominal words in the CAC

They are illustrated by the following examples.

48. The *provisions* of this *Agreement* shall survive the **termination** of my *employment* and the **assignment** of this *Agreement* by the Company to any successor in interest or other assignee. (Employee Confidentiality Agreement, 25.)

49. The Employee hereby agrees and covenants, that: (a) he shall not divulge to any person or entity other than the Company, without express written **authorization** of the Company's Chief Executive Officer, any proprietary or confidential information, whether written or oral, ... (Confidential Settlement Agreement and General Release, 1.)

In section 6.2. I have examined some of the most common syntactic features of the language of confidentiality agreements. The data obtained in the analysis of the corpus have shown some very distinctive features, such as the extensive use of binomial and multinomial expressions, complex sentences, the frequent use of passive, nominalisation and qualifications. In the next section, I will briefly look at the discourse level features.

5.3. Discourse-level features

Not too much emphasis has been put on the analysis of discourse-level features of legal discourse so far, but as Danet (1985: 285-286) states, these features are also contributing to its distinctiveness. Swales and Bhatia (1983) claim that syntactic and discoursal characteristics of legal writing are interconnected and that some syntactic features help us understand 'regularities of organization' of the genre in question (referred to in Bhatia 1993: 105-106). For example, the extensive use of nominalization and the number of qualifications bring in the syntactic discontinuities, which then make the discourse structure of the sentence complex and compound. (Bhatia 1993: 105)

In this section, I will focus primarily on looking at cohesive devices as discoursal features. Cohesion is defined in the following way:

Cohesion is the network of lexical, grammatical, and other relations which provide links between various parts of a text. These relations or ties organize and, to some extent create text, for instance by requiring the reader to interpret words and expressions by reference to other words and expressions in the surrounding sentence and paragraphs. (Baker 1992: 180)

Considering the discourse level, as I have mentioned before, the sentences in my corpus tend to be rather long and they function as self-contained units. Therefore, it is noticeable that they do not need to be linked to what follows or proceeds. This is one of the features which distinguishes the language of legal contracts from other varieties.

5.3.1. Reference as a cohesive device

It is generally assumed that anaphoric devices are rather scarce in legal language. This is proven in my collection of texts, where the pronouns are avoided and repetition increased to enhance precision and clarity. The presence of pronouns as referential devices is considered to bring ambiguity in legal language, so when we encounter a pronoun, we usually look at a surrounding context to find a referent. Except for *I* and *you*, which have more exophoric than anaphoric function, meaning that they do not refer back to the previous of the following expressions but rather to a reference outside the immediate discourse, the other pronouns appear rather infrequently. The pronoun *it* appears 125 times in the texts, but mostly as a filler of the subject position than a substitute for something else, like in the phrase *It is further understood and agreed*. However, the examples of referential *said* and *such* are very typical of the language of legal texts. They are presented in the following example.

50. In the event that any ***such*** geographic, activity or time limitation is deemed to be unreasonable by a court, Employee shall submit to the reduction of either ***said*** activity or time limitation to ***such*** activity or period as the court shall deem reasonable. (Employee Confidentiality, Invention assignment and Non-Compete Agreement, 13.)

Demonstrative reference is achieved most frequently by determiners such as *this* or *these*, occurring 690 and 36 times, respectively. Expressions like *this agreement*, *this contract*, *this clause*, etc. appear very frequently in the corpus of confidentiality agreements and they are termed ‘legal concretes’ (Trosborg 1997: 119). They also have referential value and bring about precision and unambiguity, as they specify the terms like in the following example. The word *this* collocated in most cases with the word *agreement* in the corpus.

51. ***This Agreement*** shall be binding on and inure to the benefit of the parties hereto and their permitted and respective heirs, successors and assigns. (Noncompetition and Confidentiality Agreement, 22.).

Some of the terms mentioned in section 6.1.7. such as *hereinafter*, *hereto*, *hereby*, *aforesaid*, *foregoing*, etc. also contribute to the cohesion of the texts.

5.3.2. Some other cohesive devices

Looking at the distribution and the frequency of the cohesive devices, what is most remarkable is lexical repetition. This feature has been discussed previously in the study and its occurrence is not surprising as it is one of the safest ways of achieving precision. As for the conjunctions which appear in the texts, the most common are *and* and *or*. *And* appears 2120 times and *or* 2432 times. Ellipsis as a cohesive device is not common in legal language due to the fact that it carries ambiguity. The devices like synonyms, hyponyms or antonyms were in great detail discussed in the section about binomial and multinomial expressions.

5.4. Summary of findings

In the present analysis, I have first of all focused on identifying and describing the major lexical features of the corpus of confidentiality agreements. I have looked at the vocabulary, namely the technical vocabulary, common words with uncommon meanings, have identified general service words and academic words in the corpus, and the frequent and less frequent lexical items. Additionally, I examined in great detail the words that appear with unusual frequency compared to a large general English corpus. The top ten words were then scrutinized considering their frequency and the typical collocates and clusters. It was identified that the language of contracts has a tendency towards lexical repetition, proven by the outstanding number of occurrences of some lexical items. Moreover, it has been shown that the language of the confidentiality agreement includes a great number of common words with uncommon meanings, that is, the words that have a very specific meaning within the legal discourse in contrast to the general meaning. The archaic words, such as *hereto*, *herein*, *thereof*, etc. are also very common in the corpus. All these features contribute to the great formality in style. The common prepositional phrases, as well as words borrowed from other languages occur also very frequently in the texts.

In the second part of this chapter, I have analysed the syntactic feature of the texts and identified certain common characteristics. First, I looked at the sentence length and noticed that most sentences in the corpus are rather lengthy and with complex structure. This is proven by an unusual number of 39.20 words per sentence. The sentences are also characterised by the use of qualifications which add to their complexity. Other features which contribute to the formality of the genre are the extensive use of passive constructions, the legal use of *shall*, as well as nominalisations. It was found that binomial and multinomial expressions are extremely frequent in the language of the contracts and

they convey the meaning of inclusiveness and precision. Conditional sentences found in the texts tend to be extremely complex with a number of different types of clausal embeddings.

Finally, I examined the discourse-level features such as the cohesive devices like anaphora, conjunctions, ellipsis, lexical repetition, synonymy, etc. The given texts also show some distinctive discursive features typical of legal genres.

The data analysed indicated that apart from sharing the same lexical features with other genres of contracts, this genre of confidentiality agreements differentiates itself by a rather specific lexis. By this I mean that the topic of confidentiality includes lexis which is related to information and its protection which will most probably not be included in other genres of contracts. Other important findings contribute to the fact that this genre uses different lexical and syntactical strategies in order to achieve another critical feature of contracts, and that is precision. What is essential here is that there should be no place left for misinterpretation and ambiguity. The parties of the contract are committed to act according to what is prescribed in the contract and in this way avoid any litigation. According to this, the purpose of the contracts can also be defined as a preventive purpose. Different lexical strategies are also employed to achieve all-inclusiveness which in a way serves the purpose of avoiding ‘finding the gaps’. This is accomplished by the use of binominals and multinominals, long lists and item-by-item enumeration, frequent use of *any*, etc.

Having analysed the salient lexical, syntactical and discourse-level features of the confidentiality agreements, I will now look at the structural interpretation of the texts.

6. Move analysis of the genre

The theoretical and methodological concepts of the move analysis have been described in the previous chapters. In this chapter I will scrutinise the genre texts in order to identify the communicative intentions of the individual parts. In order to accomplish these communicative purposes, the authors use different ‘rhetorical strategies’ which will be referred as steps throughout the text. During this part of the analysis, the literature on the structure of the contracts as well as the specialist members of discourse community were consulted in order to substantiate the findings.

Considering the organization of the contracts and based on Child (1992: 109-142), Trosborg (1997: 64) outlines the following structure of contracts: *title, introduction, recitals, definitions, body of the document, housekeeping provisions, and signatures and dates*. In my opinion, the part of a contract above defined as a *body of the document* is a rather broad term, and I suggest that, in the particular genre of the confidentiality agreements analysed in this study, this move should be termed *obligation*. The reason for this is that the term obligation is a more specific term and it presents one of the most important concepts of the contracts. In the case of confidentiality agreements, this section carries one of the main communicative purpose, which is inferring the rights and obligations on the parties to keep certain information confidential.

The analysis of the corpus revealed that there are six moves typical of the confidentiality agreements and they are presented in the following table. Within these moves different rhetorical strategies or steps will be identified. In the genre of confidentiality agreements, these rhetorical strategies are presented through legal clauses as building blocks of a contract.

Moves	Percentage	Status of the moves
Introduction	93,33%	Obligatory
Recitals	53,33%	Core
Definitions	100%	Obligatory

Obligation	100%	Obligatory
Housekeeping provisions (Miscellaneous)	100%	Obligatory
Signatures and dates	96,67%	Obligatory

Table 26. The moves of the confidentiality agreements

6.1. Introduction

According to Trosborg (1997: 64-65), the introduction declares ‘the nature of the contract’. In this move, the parties of the agreement are identified, that is, the persons or entities involved in the agreement. Additionally, the parties acknowledge and agree to what proceeds in the document but the purpose of the document or the promises or obligations, are not specified in the introduction. This section often repeats the title of the agreement and it very often includes the date the agreement is entered into and starts being effective, as well as the addresses of the parties of the agreement. Trosborg (1997:65) points out that repeating the title of the document is important because as title is technically not part of the text of the document, repeating it brings more value by specifying which document is in force.

The introduction appears in 28 out of 30 contracts, which amounts to approximately 93 % and makes this move obligatory. The two examples below show the introductory part of the agreements.

52. This Confidentiality and Non-Competition Agreement (the "Agreement") is made as of this 10th day of September 2003 ("Effective Date") by and between (...) and (...) (the "Director"). (Confidentiality and Non-Competition Agreement, 29.)

53. This Mutual Confidentiality Agreement (the “Agreement”) is entered into and is effective as of July 8, 2002 (the “Effective Date”) by and between (...), with places of business at (..) and (...), with a place of business at (...). (Confidentiality Agreement (Mutual), 9.)

In view of the above, one can identify the following four steps:

1) specifying the document in force

2) identifying the parties of the agreement

3) stating the date of the effectiveness

4) stating the address

Specifying the document in force is, as mentioned above, achieved mostly by repeating the title of the contract. The language commonly used for this purpose is the referential *this* as in the following examples:

54. This Confidentiality and Non-Disclosure Agreement... (Confidentiality and Non-Disclosure Agreement, 28.)

55. THIS EMPLOYEE CONFIDENTIALITY, INVENTION ASSIGNMENT AND NON-COMPETE AGREEMENT...(Employee Confidentiality, Invention Assignment and Non-Compete Agreement, 10.)

With regard to the parties of the agreement, the most relevant clusters typical of the first step are identified and presented in table 27 below, whereas the illustration of it is given below the table.

Between (...) and (...)

By and between (...) and (...)

By and among (...) and (...)

Table 27. The language typical of step 2 move 1

56. THIS AGREEMENT (“Agreement”) is made by and between (...) (“Employer”) and (...) (“Employee”). (Employee Patent and Confidential Information Agreement, 15.)

The clusters which frequently occur in the third step and are related to the date the contract is entered into are the following:

As of
On the date of
Dated as of
As of this (date) day of (month), (year)
Effective as of

Table 28. The language typical of step 3, move 1

Stating the address in the introductory part seems to be optional and the language which shapes this step is shown in the table below.

Residing at
Having an office at
Whose registered office is at
With places of business at

Table 29. The language typical of step 4 move 1

6.2. Recitals

This move of the contract ‘gives information that forms the foundation or background for the document’ (Trosborg 1997: 65). The function of this section of the document is defined in the European Union’s Practical Guide to European Law:

The purpose of the recitals is to set out concise reasons for the chief provisions of the enacting terms, without reproducing or paraphrasing them. They shall not contain normative provisions.... (<http://europa.eu/>, 22.10.2012)

The recitals present the basis and state the purpose and reasons for the document. The information given in this section does not include the obligations of the parties but

only states the background and the information known before the agreement has come into force.

The recitals appear in 16 out of 30 texts in the corpus, which is around 53 %, and according to the distinction mentioned above, they can be considered core moves. The typical recitals are introduced with the formal word *WHEREAS* and recitals which begin with *WHEREAS* occur 8 times throughout the texts (see example 56).

57. *WHEREAS*, the parties may disclose to each other certain confidential information defined below and InterTrust and/or Company desires to keep such information confidential;

WHEREAS, in consideration of the disclosure of such information to InterTrust and/or Company, InterTrust and/or Company is willing to keep such information confidential in accordance with the terms and conditions set forth in this Agreement;

NOW, THEREFORE, InterTrust and Company hereby agree as follows:
(Confidentiality Agreement (Mutual), 9.)

In order to achieve the communicative purpose recitals have, there are several steps which may or may not be employed. These steps have the aim to:

- 1) introduce the line of work; the research and the services the company provides;
- 2) state that the receiving party has access to certain confidential information; the disclosing party wants to share or disclose certain information with the receiving party;
- 3) state that the disclosing party wants to keep certain information confidential;
- 4) make an introduction to the next part of the agreement.

As shown above, in a broad sense the recitals state the subject matter of the agreement. The first step, which is rather optional, gives the background information about the company, which is usually the disclosing party. This can be illustrated by the following phrases:

(The company) is engaged in the business of providing...
(The company) is a (...) enterprise engaging in ...
(The company) carries on the business of ...
(The company) provides, develops, sells, and markets...products and services.

Table 30. The language typical of step 1

The second step indicates that during the employment, the parties may share the information which is considered confidential or proprietary. More importantly, as indicated in the third step, one of the parties wishes to keep this information confidential and wishes, further in the text, to define the exclusive circumstances when this information may be disclosed. The second and the third step, in fact, state the main purpose or the reason for the whole agreement. So, this kind of background information of the document can be articulated by the following phrases:

(The party) may have access to or become aware of ...
(The party) may share certain proprietary information with the ...
(The party) will develop (...) and will become aware of ...
(The parties) may disclose to each other ...

Table 31. The language typical of step 2

(The party) desires to keep such information confidential.
(The party) must protect its ...
(The parties) wish to protect ...

Table 32. The language typical of step 3

The final step indicates the introductory line where the parties acknowledge or agree to everything that follows and is mentioned in the entirety of the agreement. It is often introduced by the phrase *NOW, THEREFORE*. The following formulaic phrases indicate this.

NOW, THEREFORE, in consideration of the mutual promises and agreements contained herein, (...) (the parties) hereby agree as follows:

NOW, THEREFORE, in consideration of (...) the parties hereby covenant and agree as follows:

In consideration of (...), we agree that:

Table 33. The language typical of step 4

As Trosborg (1997: 65) points out, the recitals precede definitions, unless the term is used in the recitals in which case the order is reverse. In my collection of texts, the recitals are always used before the definitions. As mentioned above, the recitals do not include the provisions which state the obligations of the parties. These are included in the body of the document.

6.3. Definitions

Trosborg (1997: 65) notes that the purpose of this part of the contract can be twofold. According to her, the definitions are used to:

- 1) define technical words or terms of art that the user of a document is likely not to know, and
- 2) to stipulate for the purposes of the document definitions that defer from the usual definitions of words in common use.

The importance of this part of the contract is shown by the fact that definitions appear in all of the texts in the corpus, making them obligatory. The two most frequent definitions in the agreements are:

- 1) the definition of confidential information, and
- 2) the exclusion from confidential information.

It is essential to define the secrets, property, knowledge, information, etc., which parties want to share with each other and which need to be protected and held confidential. These usually include long lists, item-by-item enumeration and they are often presented as a single sentence in a long paragraph. They have a tendency to all-inclusiveness. The definition of the confidential information appears in all texts, accounting for the most important definition in these genre texts. The definition of the confidential information can be seen in the example that follows.

58. CONFIDENTIAL INFORMATION DEFINED. "Confidential Information" means trade secrets, proprietary information, and confidential knowledge and information which includes, but is not limited to, matters of a technical nature (such as discoveries, ideas, concepts, designs, drawings, specifications, techniques, models, diagrams, test data, scientific methods and know-how), and matters of a business nature (such as the identity of customers and prospective customers, the nature of work being done for or discussed with customers or prospective customers, suppliers, marketing techniques and materials, marketing and development plans, pricing or pricing policies, financial information, plans for further development, and any other information of a similar nature not available to the public). (Employee Confidentiality. Invention assignment and Non-Compete Agreement, 10.

The language typical of this part is presented in the table below and it shows clearly the tendency for inclusiveness.

"Confidential Information" means ...which includes, but is not limited to, ...including, without limitation, "Confidential Information" means any and all information ...

Table 34. The language typical of the definitions of confidential information

The exclusion from confidential information involves items which do not need to be kept confidential. They usually include the information which was either discovered prior to the involvement, was gained through a different source, is developed independently or is

known to the public. This kind of definition appears 11 times in the corpus and it is optional. It is illustrated below.

59. ... however, that "Information" shall not include information that (a) is or becomes generally available to the public other than as a result of a disclosure by us or any Representative, (b) is or becomes available to us on a non-confidential basis from a source other than the Company or Hallmark Capital which is not bound by a duty of confidentiality to you or the Company, (c) is independently developed by us solely from publicly available information, or (d) is disclosed pursuant to an order or requirement of a court, government administrative agency or other governmental body. (Confidentiality and Non-Disclosure Agreement, 1.)

“Confidential Information” shall not include...

... are excluded from

... shall not apply to

Table 35. The language typical of the definition of exclusion from confidential information

Other definitions found in the corpus include the definitions of the following terms: disclosing party, receiving party, external communications, the third party, work product, agreement, highly confidential information, affiliates, subsidiary, and some others.

6.4. Obligation

An obligation is defined by the *Black's Law Dictionary* as ‘a legal duty, by which a person is bound to do or not to do a certain thing’ (<http://thelawdictionary.org/>, 10.10.2012). The obligations present the main body of the document and carry one of the major communicative purposes of the agreement. In this case this is to make a promise or obligate oneself to respect the confidentiality of certain information defined in the agreement. As assumed, this move appears in all the texts. In my collection of texts, the obligation is mostly expressed by the use of *shall*, and it makes use of regulative speech acts.

The most common language which textualizes the communicative purpose of this move is presented in the table below.

(the party) shall / shall not
I agree that
I understand that
(the party) acknowledges that
(the party) acknowledges and agrees that

Table 36. The typical language of the move *obligation*

In order to express the obligation of the parties, the following main steps will be identified and presented in the next sections. The more frequent steps include *obligation of confidentiality, return of property, ownership of the confidential information, covenant not to compete and covenant not to solicit*. Some of the other steps identified in the corpus, but with a rather small number of occurrences include *the payment of the compensation fee, no warranty regarding the accuracy, third party information, etc.*

In the step obligation of confidentiality, the party to the contract agrees to keep the specified information confidential. This step appears in all the agreements. During the course of work, the employees often have to disclose certain information. The confidentiality agreements can also regulate how certain information can be disclosed in a proper way. This involves using the information only for the purposes specified in the agreement. This step is illustrated by the following sentence.

60. I agree at all times during the term of my employment and thereafter to hold in strictest confidence, and not to use, except for the benefit of the Company, or to disclose to any person, firm or corporation without the prior written authorization of a duly authorized officer of the Company, any Confidential Information of the Company. (Confidentiality & Non-compete Agreement, 5.)

The return of property clause, as the name suggests, obligates the employee to deliver the property or any confidential information, upon the termination of the agreement or the employment. This clause usually consists of long lists which tend to include all the

property which is in the ownership of the company. The following section is a good example of this.

61. I agree that, at the time of leaving the employ of the Company, I will deliver to the Company (and will not keep in my possession, recreate or deliver to anyone else) any and all devices, records, data, notes, reports, proposals, lists, correspondence, specifications, drawings blueprints, sketches, materials, equipment, other documents or property, or reproductions of any aforementioned items developed by me pursuant to my employment with the Company or otherwise belonging to the Company, its successors or assigns. (Employment, confidential information, and arbitration agreement, 16.)

This clause is rather frequent and it appears in 20 of the texts. The phrases which depict the language of the return of property clause are presented below.

(the party) shall/will return
I will deliver to the company
...will be returned to
...destroy or deliver

Table 37. The language typical of the return of property clause

The ownership of confidential information step specifies what rights the recipient has with relation to the information. In other words, here, it is acknowledged that the information is the property of the company or the disclosing party. This is shown in the example below.

62. Employee acknowledges and agrees that all Confidential Information of the Company and all reports, drawings, blueprints, (...) concerning the Company's Confidential Information are and shall remain the Company's property. (Confidentiality, invention assignment and non-compete agreement, 10.)

Covenant not to compete obligates the party not to provide similar services to another person or company that is in direct or indirect competition with the company. This is shown in the example below.

63. The Director agrees that he shall not engage in any business directly competitive with that carried on by the Company, provided that nothing in this clause shall preclude the Director from holding or being otherwise interested in any shares or other securities of any company... (Confidentiality and Non-Competition Agreement, 29.)

In the clause covenant not to solicit the employee needs to agree that he/she will not solicit other employees, that is he/she shall not encourage other employees to leave the company and offer them employment in another company. The typical example is given below.

64. You agree that for a period of one (1) year after your employment ends, whether voluntarily or not, you will not induce or attempt to influence, directly or indirectly, any employee of RN to terminate his/her employment with RN or to work for you or any other entity. You agree that this means you will not identify to a third party RN employees as potential candidates for employment. You further agree not to, directly or indirectly, solicit or assist in soliciting orders from any current or known prospective customers or to encourage negotiations with RN. (Development, confidentiality and noncompetition covenants, 23.)

6.5. Housekeeping Provisions (Miscellaneous Provisions)

As Trosborg (1997: 66) emphasises the main function of the housekeeping provisions is to “attend to the business of managing the document itself”. This means that this section of the document is not related to the subject matters of the contract, in this case the disclosure and keeping the information confidential, but in fact it is concerned with diverse aspects of executing the contract. In other words, it carries out the functioning and managing of the agreement. The clauses or provisions which deal with these issues are sometimes called *miscellaneous provisions* and they appear towards the end of the contract.

The housekeeping provisions are also of great significance because they deal with important issues such as choosing what law will govern the agreement, how the disputes will be resolved in case there are any, the assignment of the rights and obligations, the

termination of the agreement, the severability, etc. Only the clauses which appear more than 50% in the corpus will be described, whereas the clauses which do not appear so often will be only briefly outlined in section 6.5.8. The most common clauses which constitute housekeeping provisions are presented in the table below.

Clauses	Frequency of occurrence
Governing Law	86,67 %
Remedies	86,67 %
Entire Agreement	73,33 %
Severability	70 %
Assignment	63,33 %
Amendment	60 %
Term	53,33%

Table 38. The clause that constitute housekeeping provision

6.5.1. Governing law clause

This clause states what laws and jurisdictions govern the agreement. This becomes important in the case of disputes. In my collection of texts, it appears to be one of the most frequent housekeeping clause. The governing law clause is illustrated by example 64.

65. This Agreement is executed under seal and will be governed by and construed according to the laws of the Commonwealth of Massachusetts. (Employee Confidentiality Agreement, 25.)

The common language used to accomplish this is presented below.

(This Agreement) shall be governed by and construed in accordance with the laws of (...)

(...) shall be construed in accordance with the laws of (...)

(...) will be interpreted and enforced in accordance with the laws of (...)

Table 39. The language typical of the governing law clause

6.5.2. Remedies clause

Remedy is in the *Black's Law Dictionary* defined as ‘the means by which the violation of a right is prevented, redressed, or compensated’ (<http://thelawdictionary.org/>, 20.10.2012). It can be seen as a threat of legal action in case of the breach of contract, that is, if one of the parties does not fulfil its obligations stated in the agreement (Day & Krois-Lindner 2006: 97). Hence, it can be said that the communicative intention is to prevent the breach of contract or if that happens to “restore some form of justice” (Day & Krois-Lindner 2006: 97).

This clause appears 26 times in the corpus, which presents 86,67 % of all texts. The remedies can range from a few sentence- to a page-long clause, and the common patterns recognized in the corpus are presented by the following examples:

be liable for compensation
in the event of a breach or threatened breach
be entitled to injunctive (or other equitable) relief
suffer irreparable injury

Table 40. The language typical of the remedies clause

The following example illustrates this clause.

66. Employee acknowledges that money damages alone would not adequately compensate the Company in the event of a breach or threatened breach by Employee of this Agreement, and that, in addition to all other remedies available to the Company at law or in equity, the Company shall be entitled to injunctive relief for the enforcement of its rights and to an accounting of profits made during the period of such breach. (Employee Confidentiality, Invention Assignment and Non-compete Agreement, 10.)

6.5.3. Entire agreement clause

This clause is defined as “a contract clause overtly stating that the contract contains all aspects of agreement between an insurer and the insured“ (Black’s Law Dictionary, <http://thelawdictionary.org/>, 20.10.2012). This means that the communicative purpose of this clause is to state that this agreement presents the complete and final agreement and that there are no other agreements between the two parties that should be taken into account.

The entire agreement clause occurs 22 times within the corpus, which amounts to 73,33 % of all texts. This clause usually contains only one sentence and the language typical of this clause is presented in the table below.

This Agreement contains the entire agreement ...
This Agreement supersedes all prior agreements ...
This agreement constitutes the entire agreement ...

Table 41. The language typical of the entire agreement clause

The example of the entire agreement clause is shown below.

67. This Agreement constitutes the entire agreement with respect to the Confidential Information disclosed herein and in connection herewith and supersedes all prior or contemporaneous oral or written agreements concerning such Confidential Information. (Confidentiality Agreement (Mutual), 9.)

6.5.4. Severability clause

The communicative purpose of the severability clause is defined by the *Black’s Law Dictionary* as the following:

In the event a fraud is committed per a poorly written term of an insurance policy this clause protects the insurer from having to make a settlement payment. (<http://thelawdictionary.org/>, 12.10.2012)

That is to say, if any part of the contract is by a court determined to be illegal or invalid, the remainder of the document should still be in effect. In addition, it is often

stated that the objectionable provision can still be, if possible, modified or rewritten. The example of a severability clause is shown below.

68. In the event that any portion of this Agreement is held to be invalid or unenforceable for any reason, that invalidity or unenforceability shall not affect the other portions of this Agreement and the remaining terms and conditions, or portions thereof, shall remain in full force and effect. A court of competent jurisdiction may so modify the objectionable provision as to make it valid, reasonable, and enforceable. It is the intention of the parties that the restrictions imposed by this Agreement be enforced to the maximum permissible extent. (Confidentiality, Non-Competition and Inventions Agreement, 12.)

Severability clause appears 21 times in the texts, which is 70 % of all texts and is considered a core clause. The common speech acts that constitute the severability clause are obligation, prohibition and permission.

6.5.5. Assignment clause

Black's Law Dictionary defines the assignment in the contract as a transfer of rights or obligations from one person to another (<http://thelawdictionary.org/>, 12.10.2012). This usually happens when one party sells the business so the rights and duties are transferred to the buyer (Day & Krois-Lindner 2006: 110). This clause may state that the rights and obligations are not assignable, or that they may be transferred to the respective successors and assigns. The case of the former is illustrated in the following example:

69. Associate may not assign this Agreement or any of Associate's rights or obligations hereunder. (Associate Confidentiality Agreement, 27.)

This clause occurs 19 times within the corpus and is considered a core clause. The prohibition and permissive acts can be employed to achieve the communicative purpose of this clause. The language commonly found in the clause is presented in the table.

the rights and obligations of (...) may not be assigned...
(...) shall not assign
(...) may assign

This Agreement shall / will be binding upon

Table 42. The language typical of the assignment clause

6.5.6. Amendment clause

The amendment clause occurs 18 times in the corpus, which amounts to 60 %. This clause of the agreement states that the agreement may not be changed without the prior written agreement signed by both parties. The following sentence taken from the CAC illustrates this point.

70. This Agreement may not be changed, modified, released, discharged, abandoned or otherwise terminated in whole or in part except by an instrument in writing, agreed to and signed by the Employee and a duly authorized officer of the Company. (Employee confidentiality, invention assignment and non-compete agreement, 10.)

The verbs that commonly indicate the change in this sense are the following: *change, amend, modify, waive, terminate and abandon*. The phrases that follow can be typically found illustrating this clause.

This Agreement may not be changed
This Agreement may not be amended
No amendment to

Table 43. The language typical of the amendment clause

6.5.7. Term

The term clause in the confidentiality agreement specifies ‘the time period during which confidential information will be disclosed and the time period during which the confidentiality of the information is to be maintained’ (Hanson, Thompson & Dahlgren 2004). This clause occurs 16 times within the corpus. The typical language that can be found in this clause is the following:

This Agreement shall remain in effect
This Agreement shall take effect
This Agreement shall be effective
This Agreement shall terminate

Table 44. The language typical of the term clause

The following example from the CAC illustrates the typical term clause.

71. This Agreement shall remain in effect for a two-year term (subject to a one year extension if the parties are still discussing and considering the Transaction at the end of the second year). Notwithstanding the foregoing, the parties' duty to hold in confidence Confidential Information that was disclosed during term shall remain in effect indefinitely. (Non-disclosure Agreement, 17.)

6.5.8. Optional steps of the move housekeeping provisions

In the previous sections, I have examined the clauses which constitute the move of housekeeping provisions. I have analysed in more detail the ones that appear more than 50% in the corpus and have looked at the communicative purposes and how they are textualized in particular clauses. I have tried to illustrate the clause by presenting the examples from the corpus and by showing the language typical of the clause.

The clauses which do not appear so frequently in the corpus are the following:

Clause	Frequency	Communicative purpose
Counterparts	8 texts	Defining the number of copies or duplicates of the agreement which are considered to be an original
Headings	7 texts	Defining that the headings of the agreement are only descriptive and do not affect the meaning and interpretation of the agreement

Notices	7 texts	Stating that the notices, as a means of communication, are only to be given in written
No employment rights	4 texts	Stating that the agreement does not confer any rights related to continuation or termination of the employment
Attorney's fees	2 texts	Defining the costs and expenses that may arise in the event of a dispute

Table 45. The less frequent clauses of the housekeeping provisions

6.6. Signatures and dates

Signatures and dates can be found at the end of the contract and they make the contract valid and enforceable, unless stated otherwise. By signing the contract, the parties accept and agree to all what is stated in the contract. The dates are of great importance as they state the time when the contract comes into force and very often the termination period of the obligations of the parties is written in relation to the date the contract is entered into. Only one text in the corpus does not have a designated place for a signature and date, but it is assumed that they would still be included in the final validation of the contract.

The most common formulaic phrases found in the corpus of confidentiality agreements include the sentences starting with the phrase *IN WITNESS WHEREOF*, written in capitalized form in order to show the importance of what follows, making the agreement a valid document. This phrase occurs 13 times throughout the corpus. The following examples illustrate the typical ending of the contract and show the language commonly used in this part.

72. IN WITNESS WHEREOF, the parties hereto have executed this Agreement as of the date first above written.

Disclosing Party

Receiving Party

Name:

Name:

Title:

Title:

(Non-Disclosure Agreement, 17.)

73. This Agreement has been signed on the date first stated on page 1, above.

Signed for and on behalf of the Client. (Confidentiality Agreement, 4.)

74. Understood and Agreed (Confidentiality Agreement (Mutual), 9.)

75. ACCEPTED AND AGREED as of the date first written above: (Confidentiality Agreement, 2.)

6.7. Summary of findings

The analysis of steps and the moves of the confidentiality agreements in the corpus has shown that the texts have a rather conventional and fixed structure. This is proven by the fact that five out of six moves are obligatory. Some of the moves, such as the definitions, are also less fixed in the order of their occurrence. For example, the move *definitions* does not necessarily appear immediately after *introduction*, or *recitals* in the event there are any, but in few cases, they were incorporated into the move *obligation*.

The high percentage of the moves occurrence clearly shows the conventionalisation of the genre structure. The fact that there were no optional moves found in the texts, contributes to the notion that the structure of this genre is clearly defined and fixed.

According to the communicative purposes, six moves of the confidentiality agreements were identified and the structure of the confidentiality agreements is presented in the following table.

Moves	Steps
Introduction	1) specifying the document in force 2) identifying the parties of the agreement 3) stating the date of the effectiveness 4) stating the address
Recitals	1) introducing the line of work; the research and the services the company provides; 2) stating that the receiving party has access to certain confidential information; the disclosing party wants to share

	or disclose certain information with the receiving party; 3) stating that the disclosing party wants to keep the certain information confidential; 4) making an introduction to the next part of the agreement
Definitions	1) the definition of confidential information 2) the exclusion from confidential information
Obligation	1) obligation of confidentiality, 2) return of property, 3) ownership of the confidential information, 4) covenant not to compete and 5) covenant not to solicit
Housekeeping Provisions (Miscellaneous)	1) governing law 2) remedies 3) entire agreement 4) severability 5) assignment 6) amendment 7) term
Signatures and dates	

Table 46. The moves and steps of the genre texts

7. Conclusion

The genre analysis of professional discourse carried out in this study has tried to describe the typical features of the confidentiality agreements and reflect on the complexity of the language employed. In this study, I have examined the most common lexico-grammatical features of the genre in question, and have looked at the structural organization of the genre.

In the first part of the analysis by using Mike Scott's program WordSmith Tools and Paul Nation's RANGE, I have conducted the analysis with which I identified the lexical, syntactic and discourse-level features. Firstly, the percentages of vocabulary which falls into the category of general service words, academic words, and technical and low-frequency words was calculated. It has been shown that the most common words in the corpus are the general service words, then academic words and finally a small amount of words belongs to the category of technical and low-frequency words. The list of the most frequent words in the corpus was compared to such a list of British National Corpus, which resulted in the list of keywords indicating some distinctive features of the genre. The top ten keywords were scrutinised and I looked at their collocates and clusters. This analysis has shown some relevant genre-specific formulaic sequences. The analysis of the corpus has further established other genre specific lexical features, such as a number of common words with uncommon meanings. These words have a very specific meaning within the legal discourse. Other features include antiquated terms, foreign words and expressions, a tendency towards high formality and an extraordinary lexical repetition or redundancy. Syntactic features have shown even more distinctive features, such as an outstanding use of binomial and even more of multinomial expressions whose major purpose is all-inclusiveness, among others. The use of passive forms and nominalizations contributes to the great formality in style. The sentences in the confidentiality agreement corpus tend to be very long and complex, with the average number of 39 words per sentence. The most common cohesive device found in the corpus is the lexical repetition.

The structural interpretation conducted in the second part of the analysis has shown that the moves occur with a great frequency in the texts, which tells about the fixed structure of the contracts. Optional moves were not found in the corpus, which shows that the drafters of the contracts stuck to certain conventions. The moves of the

confidentiality agreements include introduction, recitals, definitions, obligations, housekeeping provisions and signatures and dates.

With the reflection on the research question set at the beginning of the study, I can conclude that all the analysed features have shown that the language of legal contracts has a great tendency towards precision, all-inclusiveness and avoiding ambiguity, and this is shown in the linguistic choices employed. The characteristics of the language identified in this study, have shown a great typicality across the corpus and we can say that the samples analysed are the prototypes of the genre. It can be concluded that the genre texts analysed in the study belong to a very conventionalised genre and the common features are found across the genre. This analysis has proven that a corpus-based genre analysis is a great tool for analysing various genres including the more complex, not convenient 'ideal' genres.

Abstract

Genre analysis of English for Specific Purposes has in the recent years attracted a great deal of interest, and in this study I have used the corpus-based approach, the quantitative approach which involves the statistical analysis of the lexical and grammatical features of a specialised corpus. Researchers concerned with genre analysis have been mostly interested in academic genres such as research articles, dissertations, textbooks, etc. The recent focus in the field of legal language has been more on simplifying complex legal writing and making it more comprehensible to ordinary people. Little research has been done on analyzing the language of legal contracts, in particular the genre of confidentiality agreements, and investigating whether they have unique features that are typical across the range of texts. One of the main objectives of this study was to identify what features are specific to the language of confidentiality agreements and how they differ from ‘ordinary language’. I will look at what linguistic choices the drafters of legal contracts construct in order to create legal obligations between the parties.

For the purpose of finding the typical features of the genre of contracts, I conducted a corpus-based analysis of thirty confidentiality agreements. Contracts have prescriptive or regulative function as they regulate legal behavior and create obligations of the parties. In this particular case of confidentiality agreements, we could say that the purpose the agreements want to communicate is to protect the information which is considered confidential. In order to identify lexical, syntactical and discourse level features, I have used the software which facilitate the analysis, namely the WordSmith Tools and RANGE programs and the results were compared to the reference corpus, the British National Corpus.

The findings of the analysis have shown that this specific genre of contracts is strongly conventionalised and certain features were found throughout the texts. The words which appear most frequently in the corpus are, as assumed, general service words, then academic words and technical words. Keywords and important collocations and clusters were also presented in the study. Other lexical features which frequently occurred in the texts were common words with uncommon meanings, antiquated terms, loan words and expressions, capitalisation, lexical repetition or redundancy, scarce use of pronouns, etc. Some syntactical feature included lengthy and complex sentences, frequent use of binomial and multinomial expressions, passive forms, nominalisation, the use of qualification. The

move analysis has also shown that there is a rather conventional and fixed structure of the texts as five out of six moves were identified as obligatory. According to the communicative purposes, the following moves were identified: introduction, recitals, definitions, obligation, housekeeping provisions, and signatures and dates.

The analyzed features have shown that the language of legal contacts has a great tendency towards precision, all-inclusiveness and avoiding ambiguity, and this is shown in the linguistic choices employed. The research has proven that the genre of confidentiality agreements is a rather important genre in today's business-driven world where information plays a vital role.

Zusammenfassung

Genre-Analyse für Englisch für spezielle Zwecke hat in den letzten Jahren hohe Aufmerksamkeit erlangt. Für meine Studie habe ich den Textkörper basierten Ansatz herangezogen sowie die quantitative Herangehensweise, welche die statistische Analyse lexikalischer und grammatikalischer Merkmale eines speziellen Korpus beinhaltet. Forscher, die sich mit der Genre-Analyse beschäftigen, waren zumeist an akademischen Genres interessiert, wie zum Beispiel Forschungsartikeln, Dissertationen, Textbüchern etc. Im Feld der Rechtssprache lag das Forschungsinteresse vor allem auf der Vereinfachung komplexen rechtlichen Schreibens und auf dem Versuch, dieses für Laien verständlicher zu gestalten. Nur wenig Forschung hat sich mit der Sprachanalyse rechtsgültiger Verträge auseinandergesetzt, im Speziellen mit dem Genre der Vertraulichkeitserklärung, und dabei herauszufinden versucht, ob diese Textkörper einzigartige Eigenschaften besitzen, die typisch sind für diesen Textbereich. Eines der wichtigen Ziele dieser Arbeit war es herauszufinden, welche Eigenschaften spezifisch für die Sprache von Vertraulichkeitserklärungen sind und wie sich diese von „normaler Sprache“ unterscheiden. Ich werde dabei meinen Fokus darauf legen, welche linguistischen Entscheidungen die Verfasser von Rechtsverträgen fällen, um legale Verpflichtungen zwischen den Vertragspartnern herzustellen.

Um die typischen Eigenschaften des Textgenres „Vertrag“ herauszufinden, führte ich eine korpusbasierte Analyse von dreißig Vertraulichkeitserklärungen durch. Verträge haben präskriptive oder regulative Funktion, da sie rechtliches Verhalten regulieren und Verpflichtungen zwischen den Vertragsparteien schaffen. Im speziellen Fall der Vertraulichkeitserklärungen kann gesagt werden, dass die Vereinbarungen den Zweck vermitteln wollen, Information zu schützen, die als vertraulich erachtet wird. Um lexikalische und syntaktische Eigenschaften sowie solche der Diskurs-Ebene zu identifizieren, habe ich Software verwendet, die die Analyse unterstützt – namentlich WordSmith Tools und RANGE – und die Ergebnisse wurden mit dem Referenzkorpus des British National Corpus verglichen.

Die Ergebnisse der Analyse haben gezeigt, dass dieses spezielle Vertragsgenre in hohem Maße konventionalisiert ist und sich gewisse Eigenschaften durchwegs in den Texten wiederfanden. Wie vermutet, waren die Worte, die mit der häufigsten Frequenz im Textkörper aufschienen, allgemeine Begriffe, sodann akademische Begriffe und technische

Begriffe. Schlüsselbegriffe und wichtige Wortzusammenstellungen und Wortgruppen werden in der Arbeit ebenfalls präsentiert. Andere lexikalische Merkmale, die in den Texten häufig aufschienen waren gewöhnliche Begriffe mit einer ungewöhnlichen Bedeutung, antiquierte Terme, Lehnwörter und –ausdrücke, Großschreibungen, lexikalische Wiederholungen und Redundanzen, seltener Gebrauch von Pronomen usw. Syntaktische Merkmale beinhalteten Ausführlichkeit und komplexe Sätze, den häufigen Gebrauch von binominalen und multinominalen Ausdrücken, Passivformen, Nominalisierungen sowie den Gebrauch näherer Bestimmungen. Die „move analysis“ hat ebenfalls gezeigt, dass hier eine eher konventionelle und fixierte Textstruktur vorherrscht, da fünf von sechs „moves“ als obligatorisch identifiziert wurden. Übereinstimmend mit den kommunikativen Absichten konnten folgende „moves“ identifiziert werden: Einleitung, Präambeln, Definitionen, Verpflichtungen, organisatorische Klauseln, Unterschriften und Datumsangaben.

Die analysierten Merkmale haben gezeigt, dass die Sprache rechtlicher Verträge eine starke Tendenz zu Präzision, Allumfassendheit und Vermeidung von Mehrdeutigkeit aufweist, und sich dies in den linguistischen Entscheidungen zeigt, die zur Anwendung gelangen. Die Arbeit hat nachgewiesen, dass das Genre der Vertraulichkeitserklärung in der heutigen Geschäftswelt, in der Information eine unerlässliche Rolle zukommt, ein sehr wichtiges Genre darstellt.

References:

- Adler, Mark 1990. *Clarity for Lawyers: The use of Plain English in Legal Writing*. London: The Law Society
- AntConc software, <http://www.antlab.sci.waseda.ac.jp/software.html> (24.01.2012)
- Askehave, Inger & Swales, M. John. 2001. *Genre Identification and Communicative Purpose: A Problem and a Possible Solution*. *Applied Linguistics* 22/ 2: 195-212
- Asprey, Michele 2003 (1991). *Plain Language for Lawyers*. Annandale, NSW: The Federation Press.
- Atiyah P. S. 1989. *An Introduction to the Law of Contract*. Clarendon Law Series, 4th edition, Clarendon Press, Oxford.
- Austin, John L. 1962. *How to do things with words*. Oxford: Oxford University Press.
- Balmford, Christofer 2002. *Plain Language: 'Beyond a Movement'*. Repositioning Clear Communication in the Minds of Decisionmakers, Fourth Biennial Conference Proceedings of the Plain Language Association International (PLAIN), Toronto, Canada, 26-29 September 2002 1-32.
www.nald.ca/PROVINCE/ONT/PLAIN/movement/movement.pdf
- Barber, Charles L. 1962. 'Some measurable characteristics of modern scientific prose' in *Contributions to English Syntax and Phonology*, Stockholm
- Baker, Mona 1992. *In other words: a coursebook on translation*. London and New York: Routledge
- Bazerman, C. (1988). *Shaping written knowledge: The genre and actiuity of the experimental article in science*. Madison: University of Wisconsin Press.
- Bazerman, C. 1994. *System of genres and the enhansment of social intentions*. In Aviva Freedman and Peter Medway (eds) *Genre and New Rhetorics*, London: Tylor & Francis
- Bazerman, C. 1997. *The life of genre, the life in the classroom*. In W. Bishop and H. Ostrum (Eds.), *Genre and writing*. Portsmouth, Nh: Boynton/Cook. 19-26.
- Bazerman, C., Little, J.. & Chavkin, T. 2003. *The productions of information for genred activity spce: Informational motives and consequences of the Environmental Impact Statement*. *Written Communication*, 20 (40), 455-477
- Beaufort, A.1998. *Writing in the real world: Making the transition from school to work*. New York: Teachers College Press.

- Berkenkotter, C. and Huchkin, T. N. 1995. *Genre Knowledge in Disciplinary Communication – Cognition/Culture/Power*. Hillsdale, NJ: Lawrence Erlbaum Associates
- Bhatia, Vijay K. 1993. *Analysing Genre: Language Use in Professional Settings*. New York: Longman
- Bhatia, Vijay K. 2002. *Applied Genre Analysis: A multi-perspective model*. *Iberica* 4. 3-19
- Bhatia, Vijay K. 2004. *Worlds of written discourse. A genre-based view*. London: Continuum
- Bhatia, Vijay K. 2008a. *Genre Analysis, ESP and Professional Practice*. *English for Specific Purposes* 27 (2) 161-174
- Bhatia, Vijay K. 2008b. *Towards critical genre analysis*. In Bhatia, Vijay K., Flowerdew, John & Jones H. Rodney. *Advances in discourse analysis*. Oxon: Routledge
- Bhatia, Vijay K., Flowerdew, John & Jones H. Rodney. 2008. *Approaches to discourse analysis*. In Bhatia, Vijay K., Flowerdew, John & Jones H. Rodney. *Advances in discourse analysis*. Oxon: Routledge
- Bhatia, Vijay K., Langton, M. Nicola & Lung, Jane. 2004. *Legal discourse: Opportunities and threats for corpus linguistics*. In *Discourse in the professions: Perspectives from corpus linguistics*. Connor, Ulla & Upton, A. Thomas (eds.). Amsterdam, Philadelphia: John Benjamins
- Bhatia, Vijay K. & Swales, J. M. 1983. *An Approach to the linguistic study of Legal Documents*. *Fachsprache* 5, 3, 98-108
- Biber, Douglas. 2006. *University language: a corpus-based study of spoken and written registers*. Amsterdam & Philadelphia: John Benjamins.
- Biber, Douglas., Conrad, S. & Reppen, R. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: CUP
- Biber, Douglas, Connor, U. & Upton, T. A. 2007. *Discourse on the Move. Using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins
- Biber, Douglas ; Johansson, Stig; Leech, Geoffrey; Conrad, Susan; Finegan, Edward. 1999. *Longman grammar of spoken and written English*. Harlow: Longman/Pearson Education.
- Bierwisch , M. 1971. *Semantics*, in *New Horizons in Linguistics*, ed, by Lyons, J.
- Black's Law Dictionary, Free Online Legal Dictionary 2nd Ed. <http://thelawdictionary.org/> (10.10.2012)
- “Business English Contracts” <http://contracts.onecle.com/> (23.09.2012)
- Butt, Peter, Castle, Richard 2006. *Modern Legal Drafting. A guide to using clearer language*. Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo: CUP

- Callaghan, M., Knapp, P., & Noble, G. 1993. Genre in practice. In B. Cope & M. Kalantzis (Eds.), *The powers of literacy: A genre approach to teaching writing* (pp. 179–202). Bristol, PA Falmer Press.
- Charrow, Robert P. & Charrow, Veda R. 1979. *Making Legal Language Understandable: A psycholinguistic study of jury instructions*. Columbia Law Review, 79,
- Charrow, Veda P., Crandal, J. A., & Charrow, Robert P. 1982. *Characteristics and functions of the legal language*. In R. Kittredge & J. Lehrberger (Eds.), *Sublanguage*. Berlin: Walter de Gruyter
- Child, Barbara 1992. *Drafting Legal Documents: Materials and Problems*. St. Paul: West Publishing Co.
- Christie, F. 1991. *Genres as social processes*. In *Working with genre: Papers from the 1989 LERN conference* (pp. 73–88). Leichhardt, Australia: Common Ground.
- Coe, R. M. 1994. *An arousing and fulfillment of desires”: The rhetoric of genre in the process era—and beyond*. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 181-190). London: Taylor & Francis.
- Coffin, Caroline 2001. *Theoretical approaches to written language: A TESOL perspective*. In: Burns, Anne and Coffin, Caroline eds. *Analysing English in a Global Context: A Reader. Teaching English Language Worldwide*. London: Routledge,
- Connor, U. and T. A. Upton (Eds.). 2004. *Discourse in the Professions. Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins.
- “Covenants Not to Compete in Intellectual Property Transactions“. Florida Intellectual Property Law Firm. Lott & Friedland
http://patentfla.com/articles/covenants_not_to_compete_in_ip_transactions.htm
 (12.10.2012)
- Coxhead, A. 1998. *An academic word list*, Occasional Publication Number 18, LALS, Victoria University of Wellington, New Zealand
- Coxhead, A. 2000. *A new academic word list*. TESOL Quarterly 34 (2), 213-238
- Croft, William & Cruse, Alan 2004. *Cognitive linguistics*. Cambridge University Press: Cambridge
- Crystal, David & Davy, Derek 1969. *Investigating English style*. London: Longman
- Danet, Brenda 1985. *Legal Discourse*. In: Van Dijk, T.A. *Handbook of Discourse Analysis*. London: Academia Press, Volume 1.
- Day, Jeremy, Krois-Lindner, Amy & Translegal. 2006. *International Legal English. A course for classroom or self-study use*. Teacher’s book. Cambridge: Cambridge University Press
- Devitt, Amy, J. 1991. *Intertextuality in tax accounting: generic, referential and functional*. In Bazerman, Charles; Paradis, Jean (eds). *Textual dynamics of the*

- professions: historical and contemporary studies of writing in professional communities*. Madison, WI: University of Wisconsin Press, 336-357
- Devitt, Amy, J. 1993. *Generalizing about genre: New conceptions of an old concept*. *College Composition and Communication*, 44, 573-586.
- Dias, P., Freedman, A., Medway, P., & Pare, A. 1999. *Words apart: Acting and writing in academic and workplace contexts*. Mahwah, NJ: Lawrence Erlbaum
- Flowerdew, J. 1993. *An educational, or process, approach to the teaching of professional genres*. *ELT Journal*, 47, 305-316.
- Goodrich, Peter 1987. *Legal Discourse: Studies in Linguistics, Rhetoric and Legal Analysis*. Basingstoke: Macmillan
- Gosden, H. 1992. *Discourse functions of marked theme in scientific research articles*. *English for Specific Purposes*, 11, 207-224.
- Gustafsson, Marita. 1975a. *Binomial Expressions in Present-day English: a Syntactic and Semantic Study*. Turku: Turun Yliopisto
- Gustafsson, Marita 1975b. *Some Syntactic Properties of English Law Language*. Turku: Department of English, University of Turku
- Hammond, J. 1987. *An overview of the genre-based approach to the teaching of writing in Australia*. *Australian Review of Applied Linguistics*, 10, 163-181.
- Hanson, J. Mark, Thompson, R. Joe, Dahlgren, J. Joel. 2004. "Overview of Confidentiality Agreements", Department of Economics University Extension, Iowa State University <http://www.extension.iastate.edu/agdm/wholefarm/html/c5-80.html> (12.10.2012)
- Herzberg, B. 1986. *The Politics of Discourse Communities*. New Orleans: Conference on College Composition and Communication
- Hopkins, A., & Dudley-Evans, T. (1988). *A genre-based investigation of the discussion sections in articles and dissertations*. *English for Specific Purposes*, 7, 113–121
- Hüttner, Julia 2007. *Academic Writing in a Foreign Language: An Extended Genre Analysis of Student Texts*. (Sprache Im Kontext). Frankfurt am Main: Peter Lang.
- Hüttner, Julia. 2010. *Purpose-Built Corpora and Student Writing*. *Journal of Writing Research*, 2 (2), 197-218.
- Hyland, Ken 2000. *Disciplinary discourses: Social interactions in academic writing*. Harlow: Pearson Education Limited.
- Hyland, Ken 2003. *Genre-based pedagogies: A social response to process*. *Journal of Second Language Writing*. 12 (1), 17-29
- Hyon, Sunny. 1996. *Genre in Three Traditions: Implications for ESL*. *TESOL Quarterly* Vol. 30, No. 4

- Johns, A. M. 1997. *Text, Role and Context: Developing Academic Literacies*. Cambridge: Cambridge University Press.
- Joos, M. 1961. *The Five Clocks*, New York: Harcourt, Brace and World.
- Joyce, H. 1992. *Workplace texts in the language classroom*. Sydney, Australia: New South Wales Adult Migrant English Service.
- Kanoksilapatham, B. 2005. *Rhetorical structure of biochemistry research articles*. English for Specific Purposes, 24, 269-292
- Koskenniemi, I. 1968. *Repetitive Word Pairs in Old and Early Middle English Prose*, Annales Universitatis Turkuensis, B:107, Turku
- Kucera, Henry & Francis W. Nelson. 1967. *A computational analysis of present-day Americal English*, Providence, Rhode Island: Brown University Press.
- Kwan, B. 2006. The schematic structure of literature reviews in doctoral theses of applied linguistics. English for Specific Purposes, 25(1), 30-55
- Love, A. M. (1991). *Process and product in geology: An investigation of some discourse features of two introductory textbooks*. English for Specific Purposes, 10, 89-109.
- Maley, Yon. 1994. *The language of the Law*. In John Gibbons (ed.) *Language and the Law*. London: Longman
- Martin, J. R. 1984. *Language, Register and Genre*. In F. Christie 9ed). *Children writing: Reader*, Geelong, Australia: Deakin University Press.
- Martin, J. R. 1989. *Factual writing: Exploring and challenging social reality*. Oxford: Oxford University Press.
- Martin, J. R. 1993. *A Contextual Theory of Language*, in *The Powers of Literacy- A Genre Approach to Teaching Writing*. Pittsburgh: University of Pittsburgh Press
- Martin, J. R. Christy, F. and Rothery, J. 1987. *Social Processes in Education: a reply to Sawyer and Watson (and others)* in I. Reid (ed.), *The place of Genre Learning: Current Debates*. Geelong, Australia: Deakin University Press
- Martin, J. R. and Rothery, J. (1980, 1981, 1986). *Writing Project Report, Nos. 1, 2, and 4*. Sydney: Department of Linguistics, University of Sydney
- Mattila E. S. Heikki. 2006. *Comparative Legal Linguistics*. Aldershot: Ashgate Publishing Company
- Maley, Yon 1994. *The Language of the Law* in John Gibbons (ed.) *The Language and the Law*. London: Longman
- Mellinkoff, David 1963. *The language of the law*. Boston: Little, Brown
- Miller, C. R. 1984. *Genre as Social Action*. Quarterly Journal of Speech, 70, 157-78

- Miller, C. R. 1994. *Rhetorical Community: the cultural basis of genre*, in Freedman, A. and Medway, P. (eds) *Genre and the New Rhetoric*. London: Taylor and Francis
- Myers, Judith Gillespie 2001. *Banishing Bureaucratese: Using Plain Language in Government Writing*. Vienna, VA: Managements Concepts Inc.
- Nation, Paul. 2001a. *Learning vocabulary in another language*. Cambridge University Press
- Nation, Paul. 2001b. *Using small corpora to investigate learner needs. Two vocabulary research tools*. In Ghadessy, Mohsen, Henry Alex & Roseberry, Robert L. *Small corpus studies and ELT. Theory and practice*. Amsterdam, Philadelphia: John Benjamins
- Nation, Paul. RANGE program. <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- “Non-disclosure Agreements”, Intellectual Property Office, UK Government
<http://www.patent.gov.uk/patent/info/cda.pdf> (12.10.2012)
- O’ Barr, W. M. 1981. *The language of the law*. Boston: Little, Brown.
- Orna-Montesinos, Concepcion. 2012. *Constructing Professional Discourse: A Multiperspective Approach to Domain-Specific Discourses*. Newcastle upon Tyne: Cambridge Scholars Publishing
- Posteguillo, S. 1999. *The schematic structure of computer science research articles*. English for Specific Purposes, 18(2), 139-158
- Radack, V. David. “Understanding Confidentiality Agreements”. Technical Journal JOM, 45 (5), 1994, p. 68
<http://www.tms.org/pubs/journals/JOM/matters/matters-9405.html> (12.10.2012)
- Redmond, P. W. D. 1979. *General Principles of English Law*. Plymouth: Macdonald and Evans.
- Sarcevic, Susan 2000. *New Approach to Legal Translation*. The Hague: Kluwer Law International.
- Schryer, C. F. (1993). *Records as genre*. Written Communication, 10, 200-234.
- Schryer, C. F. (1994). *The lab vs. the clinic: Sites of competing genres*. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 105-124). London: Taylor & Francis.
- Scott, Mike 2001. *Comparing corpora and identifying key words, collocations, frequency distribution through the WordSmith Tools suite of computer programs*. in Ghadessy, Mohsen, Henry, Alex and Roseberry, Robert L. *Small Corpus Studies and ELT*. Amsterdam: John Benjamins
- Scott, Mike. 2007. *Oxford WordSmith Tools. Version 4.0*. Oxford University Press.
- Scott, Mike. 2011. *WordSmith Tools version 6*, Liverpool: Lexical Analysis Software

- Scott, Mike and Tribble, Christopher 2006 *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins
- Scott, Mike. WordSmith Tools, Version 6.0 <http://www.lexically.net/wordsmith/> (24.01.2012)
- Searle, John R. 1969. *Speech Acts*. Cambridge: Cambridge University Press.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Shuy, Roger W. 2001. *Discourse Analysis in the Legal Context*. In Schiffrin D. / Tannen D. / Hamilton H. (eds.), *Handbook of Discourse Analysis*. Oxford : Blackwell
- Spencer, A. 1975. *Noun-Verb Combination in Law*. Birmingham: LSU, University of Aston in Birmingham
- Stephens, Cheryl M. 1990. *What is really wrong with Legal Language?* Remarks Delievered to the Wills and Estates Section, Canadian Bar Association, Vancouver, British Columbia, 27 November 1990.
www.plainlanguagenetwork.org/Legal/wills.html.
- Stubbs, Michael 1983. *Can I have that in writing, please? Some neglected topics in speech act theory*. Journal of Pragmatics 7: 479-494.
- Stygall Gail 1994. *Trial Language: Differential Discourse Processing and Discursive Formation*. Amsterdam: Benjamin
- Stygall Gail 2002. *Legal Texts and Narrative Discourse Analysis*. In Barton E. / Stygall G. (eds.), *Discourse Studies and Composition*. Cresskill, NJ: Hampton Press
- Swales, John M. 1981. *Aspects of Article Introduction*. Birmingham AL: University of Aston
- Swales, John M. 1990. *Genre Analysis: English for Academic and Research Settings*. Cambridge: CUP
- Swales, John M. 2004. *Research Genres: Exploration and Applications*. Cambridge: Cambridge University Press
- Swales, John M. and Bhatia, Vijay K. 1983. *An approach to the linguistic study of legal documents*, Fachsprache, 5 (3), 98-108
- Tiersma, Peter 1999. *Legal Language*. Chicago: University of Chicago Press.
- “The European Union. Guide of the European Parliament, the Council and the Commition. Joint Practical Guide” <http://europa.eu/> (22.10.2012)
- Thompson, S. 1994. *Frameworks and contexts: A genre-based approach to analysinglecture introductions*. English for Specific Purposes, 13, 171–186.
- Tribble, Chris. 2001. *Small corpora and teaching writing: towards a corpus-informed pedagogy of writing*. In Mohsen, Ghadessy; Henry, Alex; Roseberry, Robert I. *Small corpus studies and ELT*. Amsterdam & Philadelphia: John Benjamins

- Trosborg, Anna. 1997. *Rethorical Strategies in Legal Language. Discourse Analysis of Statutes and Contracts*. Tübingen: Narr
- West, M. 1953. *A general service list of English words*. London: Longman, Green & Co.
- Williams, Christopher 2004. *Pragmatic and Cross –cultural Considerations In Translating Verbal Constructions in Prescriptive Legal Texts in English and Italian*, in Candlin, Christopher N./Gotti Maurizio (eds.) *Intercultural Discourse in Domain-Specific English*, special issue of Textus 17/1
- Williams, Christopher 2005. *Tradition and change in legal English: Verbal constructions in prescriptive texts*. Bern: Peter Lang
- Williams, R. 1999. *Results section of medical research articles: An analysis of rhetorical categories for pedagogical purpose*. English for Specific Purposes, 18(4), 347-366
- Wray, Alison. 1999. *Formulaic language in learners and native speakers*. Language Teaching 32, 213-231

Appendix

The list of Confidentiality Agreements used in the empirical study

1. Confidentiality and Non-disclosure Agreement, August 3, 2005
2. Confidentiality Agreement, August 24, 2010
3. Confidentiality Agreement,
4. Confidentiality Agreement, August 26, 2006
5. Confidentiality & Non-Compete Agreement November 1, 2000
6. Confidentiality and Non-Competition Agreement
7. Confidentiality and Non-Competition
8. Employee Non-Competition, Confidential Information and Work Product Agreement
9. Confidentiality Agreement (Mutual), July 8, 2002
10. Employee Confidentiality, Invention Assignment and Non-Compete Agreement, January 1, 2001
11. Proprietary Rights and Confidentiality Agreement, 2008
12. Confidentiality, Non-Competition and Inventions Agreement
13. Intellectual Property Rights, Confidentiality and Non-Competition Agreement,
14. HP Agreement Regarding Confidential Information and Proprietary Developments (With Protective Covenants Relating to Post-Employment Activity)
15. Employee Patent and Confidential Information Agreement
16. Employment, Confidential Information, and Arbitration Agreement, December January 2, 2005
17. Non-Disclosure Agreement
18. Employee Agreement Concerning Protection of Company Property and the Arbitration of Legal Disputes,
19. Agreement concerning exclusive service, confidential information, business opportunities, non competition, non –solicitation and work product, October 30, 1997

20. Executive confidentiality, non-solicitation, non-competition, intellectual property rights, and code of conduct agreement, October 15, 2004
21. Confidentiality, non-competition, non-solicitation, and non-recruitment agreement, March 17, 2005
22. Noncompetition and confidentiality agreement,
23. Development, confidentiality and noncompetition covenants
24. Confidentiality and non-competition agreement, December 30, 2005
25. Employee confidentiality agreement
26. Master confidential disclosure agreement, December 3, 2001
27. Associate confidentiality agreement
28. Confidentiality and non-disclosure agreement, 2002
29. Confidentiality and non-competition agreement, September 10, 2003
30. Non-Competition and Confidentiality Agreement

Wordlist

N	Word	Freq.
1,00	THE	3671,00
2,00	OF	2649,00
3,00	OR	2432,00
4,00	AND	2120,00
5,00	TO	2009,00
6,00	#	1391,00
7,00	ANY	1212,00
8,00	IN	1201,00
9,00	AGREEMENT	853,00
10,00	A	850,00
11,00	PARTY	818,00
12,00	BY	774,00
13,00	INFORMATION	720,00
14,00	COMPANY	708,00
15,00	THIS	690,00
16,00	SHALL	675,00
17,00	THAT	613,00
18,00	BE	610,00
19,00	NOT	536,00
20,00	WITH	535,00
21,00	FOR	485,00
22,00	OTHER	487,00
23,00	CONFIDENTIAL	469,00
24,00	AS	472,00
25,00	SUCH	471,00
26,00	IS	377,00
27,00	EMPLOYMENT	346,00
28,00	EMPLOYEE	339,00
29,00	ALL	335,00
30,00	BUSINESS	314,00
31,00	ITS	310,00
32,00	I	301,00
33,00	WHICH	291,00
34,00	B	250,00
35,00	WILL	242,00
36,00	MAY	205,00
37,00	PARTIES	195,00
38,00	ON	186,00
39,00	FROM	177,00
40,00	MY	171,00
41,00	AT	169,00
42,00	ARE	168,00
43,00	TIME	166,00
44,00	WITHOUT	157,00
45,00	RIGHTS	151,00
46,00	AN	150,00
47,00	HAVE	149,00
48,00	INCLUDING	143,00

49,00	AGREE	142,00
50,00	IF	142,00
51,00	UNDER	142,00
52,00	DURING	138,00
53,00	YOU	135,00
54,00	NON	130,00
55,00	TERMINATION	126,00
56,00	IT	125,00
57,00	PERIOD	123,00
58,00	PROPERTY	123,00
59,00	USE	123,00
60,00	PROPRIETARY	120,00
61,00	HAS	121,00
62,00	BREACH	119,00
63,00	BETWEEN	118,00
64,00	DATE	116,00
65,00	WRITTEN	114,00
66,00	CONFIDENTIALITY	110,00
67,00	PERSON	109,00
68,00	DISCLOSING	108,00
69,00	WHETHER	107,00
70,00	PRIOR	106,00
71,00	WORK	106,00
72,00	OTHERWISE	105,00
73,00	DISCLOSURE	98,00
74,00	NO	98,00
75,00	RECEIVING	98,00
76,00	SERVICES	97,00
77,00	HIS	95,00
78,00	PROVISIONS	94,00
79,00	DISCLOSE	93,00
80,00	ADSI	92,00
81,00	RIGHT	92,00
82,00	LAW	91,00
83,00	AFFILIATES	90,00
84,00	TRADE	90,00
85,00	UPON	89,00
86,00	AFTER	88,00
87,00	AGREES	87,00
88,00	INVENTIONS	87,00
89,00	BUT	84,00
90,00	TERMS	84,00
91,00	MADE	83,00
92,00	LIMITED	82,00
93,00	SOHU	82,00
94,00	THIRD	82,00
95,00	CAPELLA	80,00
96,00	TERM	80,00
97,00	HP	78,00
98,00	PROVIDED	77,00
99,00	COMPETITION	76,00
100,00	RN	76,00

Keyword list

N	Keyword	Freq.
1,00	OR	2432,00
2,00	AGREEMENT	853,00
3,00	CONFIDENTIAL	469,00
4,00	ANY	1212,00
5,00	PARTY	818,00
6,00	SHALL	675,00
7,00	COMPANY	708,00
8,00	INFORMATION	720,00
9,00	EMPLOYEE	339,00
10,00	EMPLOYMENT	346,00
11,00	ADSI	92,00
12,00	DISCLOSING	108,00
13,00	TERMINATION	126,00
14,00	SOHU	82,00
15,00	PROPRIETARY	122,00
16,00	AFFILIATES	90,00
17,00	BUSINESS	314,00
18,00	NON	130,00
19,00	SUCH	471,00
20,00	CAPELLA	80,00
21,00	CONFIDENTIALITY	110,00
22,00	B	250,00
23,00	INVENTIONS	87,00
24,00	PARTIES	195,00
25,00	RN	76,00
26,00	OTHER	487,00
27,00	DISCLOSE	93,00
28,00	DISCLOSURE	98,00
29,00	HERETO	62,00
30,00	BREACH	119,00
31,00	HEREBY	72,00
32,00	AGREES	87,00
33,00	AGREE	142,00
34,00	HEREIN	57,00
35,00	EMPLOYEE'S	65,00
36,00	RECIPIENT	73,00
37,00	RIGHTS	151,00
38,00	HP	78,00
39,00	RECEIVING	98,00
40,00	PRIOR	106,00
41,00	ASSOCIATE	72,00
42,00	CKX	34,00
43,00	PROVISIONS	94,00
44,00	ENTRUST	42,00
45,00	BY	773,00
46,00	HEREUNDER	39,00
47,00	PROPERTY	123,00
48,00	ASSIGNS	42,00

49,00	HEREOF	37,00
50,00	SECRETS	66,00
51,00	INDIRECTLY	62,00
52,00	ENTITY	63,00
53,00	THIS	698,00
54,00	SOLICIT	39,00
55,00	OTHERWISE	105,00
56,00	FOREGOING	46,00
57,00	OBLIGATIONS	67,00
58,00	NEPHROS	27,00
59,00	COVENANTS	50,00
60,00	THEREOF	48,00
61,00	INCLUDING	143,00
62,00	JURISDICTION	64,00
63,00	BRITESMILE	26,00
64,00	WRITTEN	114,00
65,00	ACKNOWLEDGES	47,00
66,00	WAIVER	38,00
67,00	AGREEMENTS	67,00
68,00	COMPANY'S	72,00
69,00	DATE	116,00
70,00	UNENFORCEABLE	32,00
71,00	LIMITATION	46,00
72,00	PALMSOURCE	21,00
73,00	PERIOD	123,00
74,00	PARTYS	21,00
75,00	SUCCESSORS	39,00
76,00	CONSTRUED	39,00
77,00	ACKNOWLEDGE	48,00
78,00	WITHOUT	157,00
79,00	ITS	310,00
80,00	LAWS	66,00
81,00	RELATING	61,00
82,00	OF	2653,00
83,00	FULLER	39,00
84,00	OBLIGATION	51,00
85,00	PURSUANT	35,00
86,00	LIMITED	82,00
87,00	INJUNCTIVE	22,00
88,00	DEEMED	45,00
89,00	INTELLECTUAL	54,00
90,00	DISCLOSED	40,00
91,00	CONSULTANT	45,00
92,00	PERSON	109,00
93,00	COMPETITION	76,00
94,00	ASSOCIATE'S	18,00
95,00	FORTH	48,00
96,00	REPRESENTATIVES	58,00
97,00	VIOLATION	33,00
98,00	DISCLOSER	17,00
99,00	REMEDIES	39,00
100,00	DIRECTLY	72,00

Curriculum Vitae

Personal Data

Name: Milena Lazarevic

Education

since 2009	M.A. English Language and Linguistics, University of Vienna, Austria
2008	MBA (MIT Sloan School of Management, USA) Sun Yat-sen University, Guangzhou, China (winter semester)
2006	CELTA, Cambridge Certificate in English Language Teaching to Adults, INSEARCH, UTS, Sydney, Australia
2003-2004	Peace Studies, NGO Group MOST, Belgrade
1998-2004	B.A. General Linguistics, Faculty of Philology, University of Belgrade, Serbia

Work History

since 2012	English teacher at WIFI, Vienna, Austria
since 2012	English teacher (ESP teacher) SPIDI, Vienna, Austria
since 2012	English teacher (International Legal English Certificate (ILEC) preparation) at Delphin Sprachservice, Vienna, Austria
since 2009	English teacher at VHS Alsergrund, Vienna, Austria
2008	English teacher at Nanhai Experimental Middle School, Foshan, China
2007- 2008	English teacher at KangDa Middle School affiliated to the University of Southern Queensland Australia in Guangzhou, China
2007	English teacher at EF International Language Schools, Sydney, Australia

2005-2006	English teacher at Crows Nest Community Centre, Sydney, Australia
since 2005	Translating and interpreting services, freelance
2005	Youth Services at Crows Nest Community Center, Sydney, Australia
2005-2007	Serbian language tuition, Sydney, Australia
2001-2004	English teacher at private language school “Kontext”, Belgrade, Serbia
2002-2004	English teacher at Community College Bozidar Adzija, Belgrade, Serbia
2003	English teacher at private language school “Content”, Belgrade, Serbia

Further Qualifications

2003	Scholarship ESSLLI-Summer School in Logic, Language & Information, Vienna, Austria
2002	EGG – European Summer School in Generative Grammar, Novi Sad, Serbia
2001	EGG – European Summer School in Generative Grammar, Nis, Serbia

Language skills:

Serbian:	Native language
English:	Professional knowledge
German:	Working knowledge
French:	Working knowledge
Mandarin:	Limited knowledge