ŁÓDŹ STUDIES IN LANGUAGE

Edited by
Barbara Lewandowska-Tomaszczyk

22

Stanisław Goźdź-Roszkowski

Patterns of Linguistic Variation in American Legal English

A Corpus-Based Study

PETER LANG
Internationaler Verlag der Wissenschaften

Translators, law students or legal professionals who begin to deal with legal language face a bewildering variety of legal writings. Even though legal language has been examined from a multitude of perspectives, there are virtually no studies explicitly addressing variation in legal English in terms of recurrent linguistic patterns. This book is a first step towards filling this gap. It provides a corpus-based linguistic description of variation among several selected legal genres, including vocabulary distribution and use (keywords), extended lexical expressions (lexical bundles), and lexico-syntactic co-occurrence patterns (multidimensional analysis). The findings are interpreted in functional terms in an attempt to provide an overall characterization of the most commonly encountered types of legal language.

Stanisław Goźdź-Roszkowski received his MA in English Studies and his PhD in English Linguistics from the University of Łódź (Poland), where he now lectures on English Language and Linguistics. His research interests include legal linguistics, corpus linguistics, register variation and genre analysis, specialist translation and terminology.

www.peterlang.de

Patterns of Linguistic Variation in American Legal English

ŁÓDŹ
STUDIES IN LANGUAGE

Edited by
Barbara Lewandowska-Tomaszczyk

Vol. 22

PETER LANG
Frankfurt am Main · Berlin · Bern · Bruxelles · New York · Oxford · Wien

Stanisław Goźdź-Roszkowski

Patterns of Linguistic Variation in American Legal English

A Corpus-Based Study

PETER LANG
Internationaler Verlag der Wissenschaften

Bibliographic Information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at http://dnb.d-nb.de.

ISBN 978-3-653-00659-9 (eBook)
ISSN 1437-5281
ISBN 978-3-631-61581-2

www.peterlang.de

TABLE OF CONTENTS

PREFACE AND ACKNOWLEDGEMENTS

This book was envisaged as an attempt to look from a new angle at some linguistic aspects of what has been commonly referred to as legal English. Even though legal language has been studied from a multitude of perspectives and by a great number of scholars, I still find certain areas related to the linguistic description of this phenomenon inadequately covered. In particular, there are scarcely any studies explicitly addressing linguistic variation in legal English from the genre perspective and in terms of recurrent linguistic patterns. Thus, this work focuses on the description of variation in legal English in terms of preferred arrangement of lexical items into larger strings of units in individual legal genres.

The research presented in this volume was inspired by recent developments in corpus linguistics, especially those exploring the lexis-grammar interface. This book aims to account for some phenomena inherent in legal language in the framework of this relatively recent linguistic approach.

The research that led to this volume was facilitated by the Fulbright Scholarship Program in which I participated in 2007 at the University of Northern Arizona, USA. I am grateful to professor Douglas Biber and professor Randi Reppen for their encouragement and support during my stay in Arizona.

During the preparation of this volume, some findings were presented to a number of colleagues at different conferences and meetings in Poland and abroad. I am thankful to them for their comments and criticisms. I have been particularly fortunate in having professor Barbara Lewandowska-Tomaszczyk as my boss and mentor, who provided me with invaluable advice and guidance making sure I stayed focused on finishing the book. Her help is greatly appreciated. I would also like to thank my colleagues at the University of Łódź for their insightful comments and perspectives on many of the ideas in the book when these were still in the process of development.

I owe a special debt of gratitude to professor Douglas Biber and professor Leszek Berezowski for reading through the draft version and for giving comments on it.

My special thanks are also due to Michał Kornacki for proofreading and editing this volume.

And last but not least, I am grateful to Ania, Kasia and Kuba for being very supportive of my work and putting up with my tedious schedule for so many years.

CHAPTER I: INTRODUCTION

1.1. Variation in legal discourse

Translators, linguists, law students or other individuals who begin to deal with legal language in their university studies or in the course of their professional work face a bewildering variety of legal writings. In fact, they quickly discover that what is routinely referred to simply as "legal language", represents an extremely complex discourse embedded in the highly varied institutional space of a particular legal system and its respective legal culture. The designation "legal language" tends to emphasize the subject matter, the domain in which language is used, i.e. law, at the expense of the linguistic element. Still, it is generally acknowledged that law does not exist without language. Legal rules and regulations are coded in language. Legal concepts and legal processes are accessible only through language. If a legal text is criticised for being abstruse and incomprehensible to the general public, the problem most probably lies with the language. The expression "legal language" hides a multitude of specific classes of texts (genres) employed by various professional groups working in different legal contexts. Legal discourse spans a continuum from legislation enacted at different levels (e.g. state, federal), judicial decisions (judgments, decrees or orders), law reports, briefs, various contractual instruments, wills, power of attorney, etc., academic writing (e.g. journals, textbooks), through oral genres such as, for example, witness examination, jury summation, judge's summing-up, etc. to various statements on law reproduced in the media and any fictional representation of the foregoing. This list is by no means exhaustive. It merely indicates the extraordinary diversity of legal discourse.

The major goal of this book is to demonstrate that the universe of legal texts involves not only different situational characteristics of legal genres, such as different modes (speech, writing) and different production circumstances in which legal genres are created, different participants and the relations among them, or different communicative purposes, but that legal texts differ dramatically in terms of their linguistic characteristics. Important linguistic differences among legal texts can be found even in texts which have been created in the same mode, i.e. written, and which deal with roughly the same topic. For example, the language of Text Excerpt 1.1 from a textbook appears 'normal' in that it is written in standard academic English. It uses several legal terms to clarify the nuances of meaning involved in a legal issue related to contract law. The text employs complex syntax carefully crafted and edited in the form of four sentences cohesively tied and resulting in impersonal and highly nominal style.

Text Excerpt 1.1. *Textbook: Unfair Contract Terms Act*

> Controls were likewise imposed on guarantees and indemnity clauses. In the former case, it was rendered impossible, without any possible recourse to the test of reasonableness, for guarantees to exclude negligence liability where the loss or damage arose from the use of goods by a consumer. As far as indemnity clauses were concerned, they could not impose any liability on one person to indemnify another for that other's negligence or breach of contract, if the term did not first satisfy the test of reasonableness.
> There was a further extension of control in that the Unfair Contract Terms Act also subjected to the test of reasonableness exclusion clauses where they appeared in contracts for the transfer of ownership or possession of goods where the law relating to the sale or hire-purchase of goods did not apply.

The following excerpt from a contract deals with a related concept of indemnification. However, despite the apparent similarity, the contract illustrates strikingly different linguistic characteristics:

Text Excerpt 1.2. *Registration Rights Agreement: "Indemnification"*

> The Company shall indemnify Holder, each of the Holder's officers and directors, and each person controlling such Holder, with respect to such registration or qualification effected pursuant to this Agreement, and each underwriter, if any, and each person who controls any underwriter of the Registrable Securities held by or issuable to Holder, against all claims, losses, damages, and liabilities(or actions in respect thereto) arising out of or based on any untrue statement or alleged untrue statement of a material fact ("Untrue Statement") contained in any prospectus, offering circular or other document (including any related registration statement, notification or the like) incident to any such registrations or qualification, or based on any omission or alleged omission to state therein a material fact required to be stated therein or necessary to make the statements therein not misleading ("Omission"), and shall reimburse Holder, each of the Holder's officers and directors, and each person controlling Holder, each such underwriter and each person who controls any such underwriter, for any legal and any other expenses reasonably incurred in connection with investigating or defending any such claim, loss, damage, liability or action.

Excerpt 1.2 illustrates many of the characteristics typically associated with "legalese", i.e. type of legal language that favours long, convoluted sentences (in fact, this excerpt contains only one sentence which consists of as many as 186 words), impersonal constructions, conjoined phrases and lists of words (usually nouns) resulting in an exceptionally dense use of technical vocabulary (e.g., *claim, loss, damage, liability or action*), multiple negation, the use of *shall,* etc. In comparison with the previous excerpt, the contractual provision is marked by relatively few verb phrases and a heavy reliance on phrasal syntax. Unlike in the textbook, cohesion in this contract is ensured through repetition of lexical items

(e.g. Holder). The sheer length of Text Excerpt 1.2 is attributable to the extremely frequent use of phrasal coordination and past participle forms placed in post-nominal positions (e.g., *registration or qualification effected pursuant to* ...). These two excerpts come from different legal genres representing different levels of specialist communication and different goals. Linguistic characteristics highlighted above reflect those differences. For example, the heavy use of coordination and the determiner *any* is intended to allow for all conceivable contingencies in contractual provisions.

The primary goal of this book is to identify and describe linguistic differences of the type signalled above. The following chapters demonstrate the there are many ways in which legal genres differ. The analyses involve comparing a number of legal genres at different levels, ranging from vocabulary and multi-word expressions to co-occurrence patterns of multiple linguistic features.

1.2. Previous research on legal language and linguistic variation

It is no exaggeration to say that the last two decades have seen a veritable explosion of interest in studying the relationships between law and language. The complexity of this area is reflected in the multitude of perspectives from which research has been carried out. These range from genre and discourse analysis (Bhatia, 1983, 1993; Conley and O'Barr, 1998; Danet, 1980; Goodrich, 1987; Klinck, 1992; Kurzon, 1994, 1997; Maley, 1985, 1994; Shuy, 2001; Stygall, 1994, 2002; Tiersma, 1999; Trosborg, 1995, 1997), to semiotics (Jackson, 1994, 1995; Kevelson, 1988, 1989; van Schooten, 1999), modality (Bennett, 1989, 1990; Gotti, 2001; Kimble, 1992; Lauridsen, 1992), and forensic linguistics (Berk-Seligson, 2002; Coulthard and Johnson, 2009; Cotterill, 2002; Gibbons, 2003; Hollien, 2001; McMenamin, 1993; Olsson, 2004). The rapidly growing interest in the converging fields of law and language has prompted some scholars to postulate the emergence of a new interdisciplinary field of legal linguistics (cf. French *jurilingustique,* or German *Rechtslinguistik*) covering a range of different, albeit related, areas such as legal terminology and lexicography, legal translation and interpreting, analysis of legal discourse, courtroom discourse, linguistic human rights, language policy and planning, etc. (see, more recently, Šarčević, 2009; Sočanac, Goddard and Kremer, 2009).

However, relatively few major linguistic publications raise the issue of variation in legal language. Early research on the language and law interface dating back to the 1960s and 1970s (e.g. Crystal and Davy, 1969; Gustaffsson, 1975; Spencer, 1975) focused on statistically significant features of lexico-grammar used in a particular type of legal texts. Such analyses belong to the early studies of language variation as "register" (Halliday, McIntosh and Strevens, 1964). In the absence of advanced computational and corpus procedures, they

were severely limited to only few significant features. It was simply not feasible to provide a comprehensive description of register variation by considering the full range of texts and linguistic characteristics. The advent of corpora and computational analytical tools has made such analyses possible.

Tiersma in his book *Legal Language* acknowledges the existence of variation in legal language by noting the following: "It should be evident by now that there is great variation in legal language, depending on geographical location, degree of formality, speaking versus writing, and related factors. The language and style of lawyers also differs substantially from one genre of writing to another" (1999: 139). Veda Charrow (1982: 84) notes that the designation "legal language" spans a continuum "from almost "normal" formal usage to highly complex varieties that differ substantially from normal formal usage". Klinck (1992: 134) asks a very pertinent question: "If on the one hand we justify our use of the term "legal language" by saying that it is a distinctive sublanguage of English, and on the other, we recognize further diversity within "legal language" itself, should we not be talking about legal languages?" A similar view is found in Maley (1994) who asserts that "There is not one legal discourse but a set of related legal discourses. Each has a characteristic flavour but each differs according to the situation in which it is used" (Maley, 1994: 13).

This perception of legal language has led to the emergence of different taxonomies and typologies. For example, Danet (1980: 471) proposes a well-known classification based on stylistic categories of degree of formality related to different modes of text production. Thus, in terms of mode, Danet distinguishes between written and oral texts, with the latter subdivided into spoken-composed and spoken-spontaneous. The degree of formality is marked by means of four distinct categories (frozen, formal, consultative and casual). Various legal genres can be thus distinguished and compared according to this set of contextual factors. This important sociolinguistic scheme suggests that legal genres vary depending on the degree of codification, standardisation and predictability of lexico-grammatical structures. For example, genres such as insurance policies, contracts, leases and wills are subsumed under the "frozen" and "written" categories, reserved for texts with highly formal traits and with features typical of the written mode. According to Danet's scheme, other written genres such as statutes, briefs and appellate opinions exhibit a lower degree of formality in comparison with that of witness examinations and motions, which belong to the oral mode (Danet, 1980: 471).

After conceding that legal discourse is expressed in various legal situations, Maley (1994: 13) lists several categories such as judicial discourse, courtroom discourse, the language of legal documents (contracts, regulations, deeds, wills,

statutes), the discourse of legal consultation[1]. Tiersma (1999) provides a tripartite division of legal texts into three major categories of *operative* legal documents (those that create or modify legal relations such petitions, statutes, contracts, wills, etc.), *expository* documents (e.g. judicial opinions which analyse objectively legal points and *persuasive* documents (e.g. briefs or memoranda). He concludes by asserting "Clearly, legal language is not monolithic. Even if we limit ourselves to the written variety, there is substantial variation among different genres of documents. Generally speaking, operative documents have by far the most legalese, as compared to persuasive and expository documents" (Tiersma 1999: 141). More recently, Gibbons (2003: 15) reiterates some fundamental and classic distinctions of legal language. One basic dividing line runs along the written vs. spoken mode. Thus, there are written, largely monologic texts of legislation and other legal documents and the spoken more interactive and dynamic texts found in a variety of law-related processes, such as, for instance, courtroom interaction, police investigations, prisons and consultations between lawyers and clients. Another well-known distinction proposed by Trosborg (1995: 32) concerns "language as realized specifically in legal documents, i.e. texts covered by the scope of statute law and common law, namely (i) legislation, and (ii) simple contracts and deeds", referred to as "the language of the law", which should be distinguished from other uses of "legal language".

Despite the recognition that legal language is indeed heterogeneous, most linguistically-oriented studies have so far either treated legal language[2] as a largely monolithic phenomenon (e.g. Mellinkoff, 1963, Crystal and Davy, 1969, Tiersma, 1999[3], Alcaraz and Hughes, 2002) defined largely in terms of several distinctive lexico-grammatical features such as, for example, the excessive use of the passive voice, conditionals, archaic adverbs and prepositional phrases, the use of *shall,* etc.) which should apparently hold true for all types and categories of legal texts, or there have been a number of studies that focus on a single textual category and a limited range of linguistic features (e.g. Gustaffsson, 1975) focused on binomials and multinomials, Finegan (1982) looked at form and function in testament language; more recently Williams (2005) focuses on verbal constructions used in prescriptive legal texts). Either way, research efforts were essentially directed at identifying classes of factors that make legal language distinctive *relative to* general, non-specialized language. However, all the statements cited above as related to the issue of variation in legal language have

1 Maley (1994:16) also provides a classification of legal discourse reflecting the consecutive stages of a "conflict resolution scheme".

2 It should be noted that the appropriateness of using the designation "legal language" has been the subject of much debate. Klinck 1992 contains a detailed discussion of various perspectives adopted with respect to this issue. See also Mattila (2006).

3 While acknowledging the existence of variation in legal English, the book focuses on providing a general description of legal language disregarding the genre/register perspective.

not been followed up by empirically-grounded research practice which would document with reasonable accuracy the many different ways in which legal texts are different or similar. A case in point is Mattila (2006). This study acknowledges that legal language varies according to a particular genre or discourse community. As a result, one can talk about "the language of *legal authors, legislators* (laws and regulations), *judges* and *administrators,* as well as *advocates*" (2006: 4). Then, the description of a given "language" is worded in extremely broad terms:

> "The language of legal authors is characterised by greater freedom than the other sub-genres of legal language. At the same time, legal authors employ a good deal of scholarly vocabulary, notably Latin words and sayings. Courtroom language is especially formal, often archaic. It often has a categorical character in that judges use unreserved declarations and peremptory orders" (2006: 4).

It appears that there are virtually no studies that would provide an explicit description of linguistic variation *within* legal language or a description of variation between legal language and other specialised languages. To paraphrase Halliday's introductory comment on the term "scientific English", from his seminal 1988 article "On the language of physical science", the term "legal language" has been all too often used as a convenient label for a generalized functional variety, or register, of the modern English. Unfortunately, such labelling often implied that it was either stationary or homogeneous ignoring a great degree of variability of legal language and its constant evolution.

1.3 Register and genre perspectives on legal language

Previous linguistic investigations of legal language have relied on the concepts of register and genre in the construction of their analytical frameworks. The way these concepts have been understood and used is therefore of fundamental importance in further discussion of variation in legal language. Thus, the next section provides a brief overview of how register, genre and style have been used in previous studies of language variation.

1.3.1 Different perspectives on text varieties: register, genre, style

There is a long and rich tradition of studying variation in language use. Linguistic variation can be approached from at least four major perspectives (see Biber and Conrad, 2009). Each perspective attempts to account for the patterns of linguistic variation by focusing on a different factor. In the first approach, there is a focus

on investigating the informational properties of text elements, such as "given" or "new" informational status, "focus", "topic", and the preference to place "heavy" constituents at the end of a clause in English. Such investigations conducted within the field of linguistic pragmatics or functional linguistics study how these factors influence the choice of one linguistic variant over another. In the second, historical linguistics approach, variation in language is examined across time periods. In the third approach, demographic characteristics of a speaker are explored within the realm of dialectal and sociolinguistics studies. Finally, the fourth perspective, adopted in this book, examines the situational context and communicative purpose. It should be pointed out that these approaches differ in terms of how they define linguistic variation.

Atkinson and Biber (1994) and Biber (1995) contain an extensive survey of previous synchronic and diachronic empirical studies of register variation. Most recently, a useful overview of how the concepts of register, genre and style have been used in previous research on language variation can be found in Biber and Conrad (2009). Interestingly, there does not seem to be a general consensus on the exact meaning of such seemingly obvious and related terms. In many studies, one concept is adopted and used exclusively while the others are ignored. For example, in his seminal book *Variation across Speech and Writing,* Biber exclusively uses the concept of genre. The term genre is also employed in the now classic studies done by Bhatia (2002) and Swales (1990). In other studies, however, register is the preferred concept. The following excerpt from Biber's 1995 book *Dimensions of Register Variation* is a good illustration of this perspective (1995: 9-10):

> In my own previous studies, I have used the term *genre* as a general cover term, similar to my use of register in the present book. In Biber (1988: 68), I describe *genres* as "text categorizations made on the basis of external criteria relating to author/speaker purpose" and "the text categories readily distinguished by mature speakers of a language; for example … novels, newspaper articles, editorials, academic articles, public speeches, radio broadcasts, and everyday conversations. These categories are defined primarily on the basis of external format" (Biber 1989: 5-6). In practical terms, these categories are adopted because of their widespread use in computerized language corpora. The use of the term *register* corresponds closely to *genre* in these earlier studies.

In this approach, register distinctions are defined in non-linguistic terms such as the speaker's purpose in communication, the topic, the relationship between the addressor of a message and its recipient, and the production circumstances. In studying variety across registers, both situational characteristics as well as corresponding important linguistic features are taken into account. Biber uses the term *register* to refer to varieties that other scholars would name genres. Thus,

novels, biographies, letters, book reviews, sermons, lectures etc. are all registers. The use of this term is extended to cover language variety at different levels of generality. This applies both to very specialized varieties such as "methodology sections in experimental psychology articles" and fairly general ones as, for instance, "academic prose".

Register can also be discussed in conjunction with another term widely used in sociolinguistic studies, i.e. *style.* To some extent, the two terms overlap. These two words can be used as cover terms for all kinds of situational variation. As a result, Crystal and Davy (1969) use *style* while Biber (1995) and other scholars, particularly those working in the field of language learning and teaching, prefer the term *register* to what is essentially the same type of situational variation. Many sociolinguists, following Trudgill (1983) tend to employ *style* as the more general term, reserving *register* for the specialized language that occurs when certain topics are discussed by people with shared background knowledge and shared assumptions about those topics, particularly when related to their occupation or profession. For example, American lawyers might employ legal register by using the expression "Miranda warning" where laypeople would need to resort to a lengthy explanation to the effect that it is a police warning that is given to people held in police custody or in a custodial situation before they are asked questions relating to the commission of a crime. For some scholars, the term *register* in this sense is primarily a matter of vocabulary, either because of the use of special words or because ordinary words are used in a special sense. Many words associated with a particular register are technical or semi-technical words. Legal register has many words that are used in a technical sense (e.g. *trust, consideration, party*).

There are, however, research studies which distinguish between register and genre (e.g. Ventola, 1984; Martin, 1985; Ferguson, 1994). Such studies are based on Systemic Functional Linguistics which has an extensive theoretical framework associated with these concepts (see Halliday, 1985; Martin, 1985; 1997, 2001; Matthiesen, 1993). Martin argues that register and genre are on different "semiotic planes" (Martin, 1985). Genre is viewed as a social process in which participants belonging to a certain culture use language in predictable sequential structures to fulfil certain communicative purposes. Genres have been also perceived as "conventional instances of organized text" (Couture, 1986: 80). The dynamic and interactive nature of genres as closely involved in human and more specifically professional communication is acknowledged in Swales' definition of genre (1990: 58):

> 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 (1990: 58).

On the other hand, registers have been referred to as the "expression-plane" of genre (Martin, 1985) and they tend to be associated with typical linguistic choices within different genres. The choices are influenced by three contextual variables called field, tenor and mode in Systemic Functional Linguistics.

1.4. Register, discipline, genre and legal language

Despite the absence of a general consensus concerning the use of register, genre and style, research on legal language has been more consistent in their use, probably due to the influence of the conceptualization proposed and popularized by Bhatia for over a decade (see, for example, Bhatia, 1993; 1994; 1997; 2004). Thus, by and large, the term *register* has been used to refer to a general kind of language associated with a domain of use, i.e. law. More recently, Bhatia (2004: 30) insists on distinguishing between register and *discipline* noting that the latter "represents the content", while the former the "language associated with it". On the other hand, the term *genre* has been understood as referring to more specialised varieties, such as brief, statute, contract, judgment, textbook or academic essay. At the same time, genres cut across disciplines. For example, textbooks from different disciplines can all share certain typical generic characteristics attributed to the common communicative purpose of providing students with established, state-of-the-art disciplinary knowledge. Thus, Bhatia contends, textbooks tend to rely on certain shared generic resources, such as typical lexico-grammatical features, rhetorical organization and "the use of multiple modalities to make disciplinary knowledge accessible to the uninitiated readership" (Bhatia, 2004:31). In addition, studies of a legal genre tend to involve considering its communicative purpose(s), the setting(s) or context(s), the social or professional relationship between the participants, the background knowledge of the participants, etc. (Bhatia, 1993: 101).

While this perception of genre is shared in the present study, the concept of legal register is rejected for two reasons. First, it somehow erroneously implies a hierarchical relationship between these two concepts, whereby "legal register" represents some vague superordinate term covering various types of legal texts (genres) and being practically synonymous with the notoriously imprecise notion of 'legal language'. This concern is echoed in Hatim and Mason (1990: 54) who rightly observe:

> […] We need to beware of positing such unrestricted registers as "commerce" and "journalism". To attempt to quantify the frequency of items of vocabulary and grammar in such wide domains cannot lead to any meaningful characterization of a register.

Second, the use of "genre" and "register" may signal different methodological approaches. The genre perspective usually focuses on issues related to discourse communities, ideology and power, while the register-oriented study tends to deal with characteristic lexico-grammatical linguistic features.

1.5. The concepts of register and genre as used in the present study

After Biber and Conrad (2009), this study uses the concepts of genre and register as different *approaches* or *perspectives* for analyzing legal texts rather than different kinds of *texts* or different *varieties*. Table 1.1 below (adapted from Biber and Conrad 2009:16) summarizes the most important methodological characteristics of registers and genres[4].

Table 1.1. *Methodological characteristics of registers and genres*

Defining characteristic	Register	Genre
Textual focus	sample of text excerpts	complete texts
Linguistic characteristics	any lexico-grammatical feature	specialized expressions, rhetorical organization, formatting
Distribution of linguistic characteristics	frequent and pervasive in texts from the variety	usually once-occurring in the text, in a particular place in the text
Interpretation	features serve important communicative functions in the register	features are conventionally associated with the genre: the expected format, but often not functional

The two perspectives appear to differ in terms of textual focus, linguistic characteristics relied upon in the analysis, distribution of linguistic characteristics and the interpretation. In the register perspective, the analysis is usually based on

4 Biber and Conrad 2009 also include *style* as yet another perspective but since the concept of style is not referred to in this study, it has been dropped from further discussion.

a sample of text excerpts representative of a particular variety and the focus is on lexical and grammatical features which are frequent and which are pervasive (widely distributed) across this variety. Such typical features are then examined according to their function(s) in the situational context of the variety. In contrast, the genre perspective focuses on language characteristics which may occur only once in a text and which are usually located at a specific place in the text. These linguistic characteristics can be specialised, formulaic expressions crucial to the construction of a particular genre. As a result, an analysis is based on complete texts. The language features are conventionally associated with the genre. They conform to the cultural expectations of how a particular genre should be constructed. It should be pointed out that the same texts can be analyzed from both register and genre perspectives. For example, a genre study of a legal document known as *power of attorney* would analyze the expected textual conventions for complete texts of this type of legal documents. These conventions specify that a power of attorney should begin with the date when the document was made, the person who made it (*This Power of Attorney is made this xx day of xx 20xx by me xxx)* and the designation of a person appointed to be the attorney (I/*We hereby appoint Mr/s xxx to be my/our Attorney*)[5]. This type of analysis would result in proposing a macrostructure, i.e. format or organizational outline. A macrostructure in most such documents consists of two major types: *commencement and performative act* and the *operative parts* (see Alcaraz and Hughes 2002: 144-146). A typical power of attorney closes with the so-called testimonium clause (*In witness whereof I have hereunto set my hand this ....[date]*). All these conventional expressions arranged in accordance with a specific organizational format contribute to the creation of what a legal culture recognizes as the genre of power of attorney. From a register perspective, this specialized variety contains a frequent occurrence of *to-infinitive* clauses, prepositional phrases (e.g. *on my behalf, in the name of*), nominal chains (e.g. *any other act matter or thing*), nominalizations, etc. For example, the dense use of nominalization can be attributed to the highly specialist and informational nature of this text variety.

It should be noted that whether a particular text belongs to the category of *legal* texts is not determined solely on the basis of lexico-syntactic or conventional features. Such text must be accepted by the professional community in the course of actual legal proceedings (cf. also Gizbert-Studnicki, 1986; Jopek-Bosiacka, 2006).

The present book combines these two perspectives by treating them as complementary whenever possible. As a result, only complete texts are considered in terms of both important lexico-grammatical features and specialised

5 These examples come from authentic texts published in *Selection of English Documents* by Tepis Publishing House (1998).

expressions. These are identified by means of frequency counts and then analyzed in their immediate co-texts as well as against the overall document macrostructure. In other words, the following chapters show that certain linguistic patterns with a clear functional focus tend to cluster in particular places in the legal texts. Thus, both register and genre perspectives have been integrated into the analytical approach adopted in this book. The object of study are seven written varieties formally recognized in the US legal culture i.e. briefs, academic journals, contracts, opinions, professional articles, legislation, and textbooks. However, instead of focusing on extra-linguistic issues of ideology, power and discourse communities, this book takes a narrower approach by examining the ways in which lexico-grammatical linguistic features are patterned across different legal texts.

The major claim made in this book is that rather than use general, umbrella terms such as "legal language" or "legal register", the complexity and interrelatedness within legal language appear to be best viewed in terms of some sort of grouping of genres yoked together within the same domain. There have been a few interesting concepts proposed which are relevant to legal genres.

Meritorious claims usually cite legal authorities, such as statutory laws or case law, to support their arguments. Undoubtedly, law essentially depends on the two most conventionally standardized disciplinary genres, i.e. legislation and judgments to realize its disciplinary goals (Bhatia, 2004:55). This centrality is reflected in the *intertextual* and *interdiscursive patterning* that these mutually dependent generic constructs display in various forms of legal discourse, including academic journals and textbooks. In fact, all legal genres are intertextually linked to a varying extent. The degree of such interrelatedness depends on how narrowly or broadly a given legal action or professional activity is defined. Using the case of tax accountants, Devitt (1991) proposes the concept of *genre set* to denote a range of written genres that this particular professional group creates in the course of their work. She discusses several distinct but intertextually linked letters, such as an opinion letter to the client, a response letter to the client, a letter to tax authorities, etc. Similarly, Candlin and Bhatia (1998; and more recently Bhatia 2004) demonstrate how this concept could be employed in the case of a solicitor's professional activity. The interaction between a solicitor and a client usually involves completing a set of tasks leading to the production of various documents, such as a client file, a legal brief and a letter of advice. This set of written products can be referred to as a genre set. A genre set is thus a grouping of distinct albeit related generic constructs which are employed within the same, usually narrowly defined professional activity with a clearly designated agent (a professional) who performs this activity.

The concept of genre set has been extended by Bazerman (1994: 97) who proposes the concept of *systems of genres.* While the concept of genre set is somehow one-sided in that it represents the work of one participant to a

professional activity, e.g. the solicitor, systems of genres attempt to account for the full range of genres. Bazerman (1994: 97) asserts that:

> The system of genres would be the full set of genres that instantiate the participation of all the parties – that is the full file of letters from and to the client, from and to the government, from and to the accountant. This would be the full interaction, the full event, the set of social relations as it has been enacted

Thus, the concept of systems of genres is broader and more comprehensive than genre set. However, it might still prove to be inadequate in order to capture fully the complexity of legal language. The reason for this is that law involves a range of strictly defined professional activities as well as larger sets of domain-specific generic constructs. Systems of genres as proposed by Bazerman are limited to the description of a particular professional activity.

This book favours a recent concept of domain-specific *disciplinary genres* proposed in Bhatia (2004). It represents the universe of particular systems of genres connected with different professional activities and combined with larger legal genres such as legislation, contracts, legal briefs, etc. In the course of professional activity, its participants need to refer to, interpret and exploit such larger generic constructs in order to achieve their professional objectives. Taken together, all such genres combined form a set of domain-specific disciplinary genres. Worth stressing is that such a set can be found in different configurations depending on the nature of a particular professional legal activity and the extent to which other genres are referred to or relied upon. The perception of legal language as consisting of a multitude of different, albeit related, domain-specific disciplinary genres seems amenable to corpus work in general and this study in particular. While, it is probably not feasible to build an exhaustive collection of all genre sets or genre systems existing in the domain of law, the contemporary corpus resources and tools allow the analyst to compile a dataset encompassing a set of selected domain-specific genres arranged in a particular configuration. Seen against such background, the ALC (The American Law Corpus) contains a set of interrelated disciplinary genres consisting of primary genres (Bhatia, 2006) or constitutive text type (Kjaer, 2000), which ─ through intertextual links ─ determine legal practice and target genres (Bhatia, 2006) or reproductive text types (Kjaer, 2000), such as, for instance, contracts.

1.6. Overview of the book

The approach adopted in this book is closely related to the commonly accepted usage-based, context-of-use or inductive stance in linguistics (see e.g. Kemmer and Barlow, 2000; Langacker, 2008, or Tomasello, 2003). Recent contributions in

these areas have led to the observation that "human beings construe non-linguistic reality the way they think about themselves or speak and write to others, becoming manifest and accessible predominantly in large computer-readable collections of spoken and written natural texts" (Schulze and Römer, 2009: 2). The present study links up with the extensive evidence on the interrelatedness of vocabulary and syntax provided by corpus linguistics, which has been emerging by means of various theoretical and methodological assumptions, some of which include (and are, in fact directly or indirectly drawn upon in the present analysis) idiom principle (Sinclair, 1991), collocation, colligation, semantic preference, and semantic prosody, the pattern grammar work (Hunston and Francis, 1999), the theory of lexical priming (Hoey, 2005). As previously mentioned, this book makes use of a corpus of seven different disciplinary legal genres, which, until now, appears to be the first to utilise a large-scale collection of annotated data of legal discourse. I explore legal discourse in various types of situations and contexts, e.g. academic, judicial, commercial, providing legal advice, etc. My interest lies with identifying and accounting for lexical, lexico-syntactic and phraseological patterning present in various textual manifestations of contemporary American legal discourse. I follow an empirical research design that relies on a number of analytical approaches including corpus linguistics and genre and register analysis in the sense described in Section 1.4 above. Both quantitative and qualitative examinations are combined in respect of data obtained from corpus and computational methodologies in my investigation of a range of lexico-syntactic features of individual legal genres relative to one another and to other specialist genres.

The central goal pursued in the book is to provide a corpus-based, relatively comprehensive linguistic description of variation among several selected legal genres by surveying the distinctive linguistic characteristics and their co-occurrence within each genre. These linguistic descriptions include vocabulary distribution and use (*keywords*), extended lexical expressions (*lexical bundles*), and lexico-syntactic co-occurrence patterns (*Multi-Dimensional Analysis*). The findings are then interpreted in functional terms in an attempt to provide an overall characterization of the most commonly encountered types of legal language and to describe the extent to which selected legal genres are different or similar linguistically. The claim made in this study is that what is commonly referred to as "legal English" should be more accurately described as a system of related domain-specific genres, which vary widely in terms of patterning, understood here as recurring lexical and lexico-grammatical combinations discernible in large collections of authentic texts by means of quantitative and qualitative analytical techniques. Chapter 2 focuses first on the design and construction of the American Law Corpus and then it moves on to introduce a range of methodologies employed in the analyses provided in this book. Chapters 3-6 explore linguistic patterning at three different levels. Chapter 3 starts with

overall patterns of vocabulary use by examining the proportion of the total number of words (tokens) and types (different words) and the distribution of high-frequency words across the legal genres. Thus, this chapter attempts to offer insights into the diversity of lexical choice in legal language. The latter part of Chapter 3 uses the Keyword Analysis to investigate which words are unusually frequent in a particular legal genre relative to the other genres. This quantitative analytical part is followed by an in-depth functional analysis of these words in their relevant contexts culminating in proposing functional categories for the previously selected keywords. Chapter 4 takes us beyond the level of individual words to examine the distribution and functions of multi-word expressions across the different text genres. It starts from the assumption that collocations differ depending on the text type and they thus represent an extremely useful means to discriminate between different legal genres. Chapters 5 and 6 both use Multi-dimensional Analysis, a research technique for studying co-occurrence patterns in a given language variety. The former relies on the 1988 MD model, which enables one to examine variation in legal genres relative to a number of other written genres in English. The latter discusses findings based on a new Multi-dimensional Analysis after carrying out a new factor analysis with a view to identifying co-occurrence patterns unique to the domain of legal texts.

Finally, Chapter 7 will summarize the results of the analysis, offer pedagogical implications and emphasize directions for future research.

CHAPTER II: THE METHODS AND THE CORPUS

2.1. Introduction

The descriptions of legal language provided in this book emerged from different research methodologies applied to examine a multi-genre corpus of legal texts. Much of the published work on legal discourse[6] has focused on the functions of a specific feature (or features) in a particular legal genre or on the analysis of legal discourse in different courtroom situations or other legal contexts, especially in the field of forensic linguistics. The research goal in this book is to provide a comprehensive linguistic description of a range of legal genres based on large authentic data. The data comes from what is probably the largest and most representative corpus of legal discourse. Thus, this chapter first describes the design and construction of the corpus and it demonstrates how individual genres fit in within the larger framework of legal institutions or what could be collectively described as the world of law. The latter part of this chapter introduces different methodologies employed throughout the book.

2.2. Design and construction of the American Law Corpus (ALC)

The collection of texts (hereinafter called the *American Law Corpus* or the ALC) contains over 5,500,000 words and represents seven major genres which are part of the American legal culture and education. Table 2.1 below shows the overall composition of the ALC by genre category.

Table 2.1. *Composition of the American Law corpus*

Genre	# of texts	# of words
Academic journals	71	552,487
Briefs	64	763,222
Contracts	177	1,178,616
Legislation	60	1,178,516
Opinions	114	1,182,246
Professional Articles	100	201,404
Textbooks	104	519,116
Total	**687**	**5,578,393**

6 The term "discourse" is used in this book in the sense of "the study of language use". (See Schiffrin, Tannen, and Hamilton, 2003:1) for a recent survey of the range of definitions given to this term).