1 Introduction

Variação funcional

Variação por estrato

1.1 Registers, dialects, and linguistic variation

Variability is inherent in human language: a single speaker will use different linguistic forms on different occasions, and different speakers of a language will say the same thing in different ways. Most of this variation is highly systematic: speakers of a language make choices in pronunciation, morphology, word choice, and grammar depending on a number of non-linguistic factors. These factors include the speaker's purpose in communication, the relationship between speaker and hearer, the production circumstances, and various demographic affiliations that a speaker can have.[1] Analysis of the systematic patterns of variation associated with these factors has led to the recognition of two main kinds of language varieties: *registers*, referring to situationally defined varieties, and *dialects*, referring to varieties associated with different groups of speakers.

In the present book, *register* is used as a cover term for any variety associated with particular situational contexts or purposes. Although register distinctions are defined in non-linguistic terms, there are usually important linguistic differences among registers as well. In many cases, registers are named varieties within a culture, such as novels, letters, editorials, sermons, and debates. Registers can be defined at any level of generality: for example, academic prose is a very general register, while methodology sections in psychology articles are a much more highly specified register.

Two main kinds of dialects are commonly distinguished in linguistics: *geographic dialects* are varieties associated with speakers living in a particular location, while *social dialects* are varieties associated with speakers belonging to a given demographic group (e.g., women versus men, or different social classes). Most recent dialect studies have used a comparative approach to study social dialects, describing the linguistic patterns of variation across social groups in major urban centers such as New York City, Norwich, Belfast, and Montreal.

Although linguistic differences among geographic and social dialects have been more extensively studied, it turns out that the linguistic differences

among the registers within a language are in many ways more noteworthy. When speakers switch between registers, they are doing different things with language – using language for different purposes and producing language under different circumstances. Many language choices are functionally motivated, related to these differing purposes and production circumstances, and thus there are often extensive linguistic differences among registers. In contrast, dialect differences are largely conventional and therefore less fundamental in nature. Regardless of any dialect differences, speakers using the same register are doing similar communicative tasks; therefore, in most basic respects the characteristic language features used in a given situation are similar across speakers.

To illustrate, text samples 1.1 and 1.2 compare conversations recorded in a working-class (WC) family and a middle-class (MC) family (in the city of Bristol, England).

Text sample 1.1: Working-class conversation

A: I probably yeah. I probably need two cards.
B: Right. ⟨unclear⟩ . . .
C: Nat you gotta have a bath my love it's what . . . twenty-five to nine darling.
B: ⟨unclear⟩ like ⟨unclear⟩. Two different ones.
A: That's a lovely. Any
C: Envelope.
A: Any envelopes. Any envelopes with them Nat?
B: Yep. Get you one now. . . .
C: Nat said her envelopes don't stick very well.
A: Don't it?
B: What?
A: ⟨unclear⟩
C: Yeah. Do you hear her get off the bed then?
A: Mm.
B: What?
C: Lady jumped off the bed. She heard us come in. . . . There's cake in the fridge Nat.
B: I bet he don't eat it.
C: I don't know. He don't eat a lot of cake do he?
B: He eats all my cake though.

Text sample 1.2: Middle-class conversation

A: You should be going to bed now.
B: No I don't wanna go to bed.

A: Come on. . . .
C: [laugh]
A: [laugh] . . . Come on, bed! . . .
C: Don't swear at me my dear.
B: I'm not swearing at you. I just pointed upwards. To go up . . . the stairs . . .
A: We'll be up in a minute. . . .
C: Brush your teeth.
B: I've already done that. I'm coming I'm coming . . . ooh . . . parents!
A: Kids!
B: Parents! . . .
A: Okay.
C: You'd better take her lamp upstairs out of the way as well because that's, remember that's a present. It's gotta be wrapped up.
A: It's got to be wrapped up. Yeah.
C: I know she's picked it out but [laugh]
A: [laugh] . . .
C: I ordered a skirt this morning as well.
A: Right.
C: A navy blue pleated one.
A: Mhm. So you want . . . it'll probably come as black and white spotted
C: Yeah.

The only obvious dialect marker in these two text samples is the use of *do* with a third person singular subject in the WC conversation (*he don't eat it. He don't eat a lot of cake do he?*) Because dialect markers such as these are often highly stigmatized, a relatively rare occurrence of a few features can serve as an important indicator of dialect differences.

Overall, though, the most striking aspect of comparison between text samples 1.1 and 1.2 is the extent of sameness. Situationally, the two conversations are very similar in their production circumstances, primary purposes, and interactiveness. Both conversations are spoken (rather than written), and they are produced on-line, with the words and grammatical organization being assembled on the spot as the conversation unfolds. There is little time to plan ahead, and virtually no opportunity to edit afterwards. In addition, both conversations are personal and directly interactive. Conversational partners express their own personal attitudes, feelings, and concerns, and they interact with one another to build a shared discourse jointly. These

characteristics are tied to conversation as a register and are thus not affected by dialect considerations.

The contextual characteristics shared by these two conversations have important linguistic correlates. For example, the on-line production circumstances of both conversations result in generally short sentences, with many utterances not being structurally complete sentences at all (assuming the traditional concept of a grammatical sentence). These latter include simple responses (*right, what?, okay, yeah*) as well as utterances that build on the shared physical context to supply missing information (e.g., *Any envelopes*; *A navy blue pleated one*). Both conversations also have numerous contracted forms, such as *it's, that's, don't, I'm, we'll*. In addition, many of the referents in both conversations are not explicitly identified, so that hearers must rely on the context for understanding (e.g., *Do you hear her get off the bed then? I bet he don't eat it. I know she's picked it out*). The interactive nature of conversation further results in linguistic similarities between these two texts. For example, both conversations have frequent references to *I* (the speaker) and *you* (the addressee). Similarly, there are frequent direct questions and directives in both (e.g., *Don't it? What? Do you hear her . . .? Nat you gotta have a bath*; *You should be going to bed now*; *Brush your teeth*); these constructions would not be appropriate without a specific addressee (*you*).

The fundamental nature of register variation can be further illustrated by comparing the above two conversational samples to a text representing written, informational prose. Text sample 1.3 below is taken from a science textbook.

Text sample 1.3: Science textbook

> A system of equations that provides an accurate and physically satisfactory representation of an experimental system can be cumbersome, and even complicated and of high order, so solutions may often only be obtained as numerical approximations to solutions. Thus the numerical solutions themselves may be considered to be approaching an equilibrium or periodic solution.

The contextual characteristics of this science text are strikingly different from those of the conversational texts. The science text is written, carefully planned, edited, and revised. It is produced by an author who does not overtly refer to himself in the text. The production is not interactive; the text is addressed to a large audience of scientists, but these addressees are never directly referred to. The primary purpose of the text is to present information about non-linear science, as opposed to the (inter)personal purposes of conversational participants from both dialects in text samples 1.1 and 1.2.

Due to the influence of these contextual factors, the linguistic characteristics of the science text are dramatically different from those of the conversational texts. The sentences of the science text are all grammatically complete, and many of them are quite long and grammatically complex. None of the reduced or interactive linguistic characteristics common in conversation occurs in this text. However, science texts do contain numerous linguistic characteristics rarely found in conversation. In text sample 1.3, these characteristics include technical vocabulary (e.g., *numerical approximations, equilibrium, periodic solution*), complex noun phrase constructions (e.g., *A system of equations that provides an accurate and physically satisfactory representation of an experimental system*), and passive constructions (e.g., *solutions may often only be obtained, solutions . . . may be considered*).

The extensive linguistic differences between the conversational texts (from both dialects) and this science text reflect the fundamental influence of register, associated with contextual differences in production circumstances, purpose, interactiveness, etc. Because these contextual factors operate across dialects, a full understanding of the associated patterns of register variation is essential to any comprehensive analysis of language use in a speech community.

The present book takes these considerations one step further, investigating the extent to which register factors operate in similar ways across languages and cultures. Register variation is widely considered to be intrinsic to all cultures. For example:

> each language community has its own system of registers . . . corresponding to the range of activities in which its members normally engage (Ure 1982: 5)

> register variation, in which language structure varies in accordance with the occasions of use, is all-pervasive in human language (Ferguson 1983: 154)

> no human being talks the same way all the time . . . At the very least, a variety of registers and styles is used and encountered (Hymes 1984: 44)

Given the ubiquity of register variation, analysis of the linguistic patterns across registers is of central importance for both the linguistic description of particular languages and the development of cross-linguistic theories of language use. Hymes (1984: 44) argues that the analysis of register variation – 'verbal repertoire' in his terms – should become the major focus of research within linguistics:

> [the] sociolinguistic perspective . . . has the possibility of taking the lead in transforming the study of language, through developing and consolidating the systematic study of verbal repertoire.
>
> The abilities of individuals and the composite abilities of communities cannot be understood except by making 'verbal repertoire', not 'language', the central scientific notion. To do so requires a mode of description in linguistics which can address

the organization of linguistic features in styles, that is, in ways which cut across the standard levels of linguistic structure.

The present study, by comparing the patterns of register variation across four quite different languages and cultural settings, provides a major step towards these goals.

1.2 Introduction to previous research on register

Interest in register analysis can be traced back to the merging of situational, social, and linguistic descriptions by anthropological linguists such as Boas, Sapir, Malinowski, Whorf, Bloomfield, and Firth during the first half of the twentieth century. These observers typically focused on non-western languages and cultures, and occasionally examined language in different social contexts. Two important early studies focusing on situated language are Malinowski's (1923) discussion of the 'context of situation' and Firth's (1935) elaboration of that concept. Early in the century, Boas established the value of collecting and analyzing natural texts, and that became accepted practice for linguists such as Sapir, Kroeber, and Bloomfield. Predating the early studies of social dialect variation in the mid-1960s, researchers such as Ferguson, Gumperz, Halliday, and Hymes examined linguistic variation across social situations and communicative uses, as well as across speaker groups.

In the late 1950s and the early 1960s, a number of papers and books appeared describing particular registers of various languages and the ways in which linguistic form is influenced by communicative purpose and the context of situation. These include studies by Ferguson on 'high' and 'low' diglossic varieties and on baby talk, by Gumperz on the range of 'argots' used in rural South Asian villages, by Brown and Gilman on the role of second person pronouns in relationships of 'power' and 'solidarity', by Brown and Ford on American English address terms, by Leech on press advertising and on the language of poetry, and by Crystal and Davy on conversation, unscripted commentary, newspaper reporting, legal documents, and the language of religion. Hymes (1964), Pride and Holmes (1972), and Giglioli (1972) have collected many valuable papers from this period.

Over the last thirty years, there have been an increasing number of register studies undertaken by scholars in linguistics, communication, rhetoric, education, anthropology, and other related disciplines. Atkinson and Biber (1994) survey these studies, grouping them into four major categories: synchronic descriptions of a single register; diachronic descriptions tracing the evolution of a single register; synchronic descriptions of the patterns of variation among multiple registers; diachronic descriptions tracing changes

in the patterns of variation among multiple registers. Register studies also differ in the linguistic characteristics chosen for analysis, the use of quantitative and/or qualitative methodologies, and the language studied.

Section 1.2.1 surveys the terminological differences found in previous register studies, while section 1.2.2 briefly surveys previous studies of a single register (both synchronic and diachronic). Then in sections 1.3 and 1.4, I turn to previous research on register variation, which provides the immediate background to the present study.

1.2.1 Terminology in register studies: *register, genre, style, sublanguage*, and *text type*

The term *register* is used in the present book as a general cover term for situationally defined varieties. This usage is similar to that of sociolinguists such as Ure (1982), Ferguson (1983, 1994), and Hymes (1984). For example:

> the register range of a language comprises the range of social situations recognized and controlled by its speakers – situations for which appropriate patterns are available (Ure 1982: 5)

> register variation 'is the linguistic difference that correlates with different occasions of use'. (Ferguson 1994: 16)

Register distinctions are defined in non-linguistic terms, by differences in purpose, interactiveness, production circumstances, relations among participants, etc. For this reason, particular registers can be more-or-less constrained in their linguistic characteristics: for example, experimental psychology articles are highly constrained in their linguistic characteristics, while novels employ a wide range of differing linguistic configurations. Thus, register analyses must describe the extent of linguistic variability within a register, in addition to the typical linguistic characteristics of the register. (*Text types* differ from *registers* in that they are defined in linguistic rather than situational terms – see below.)

Registers can be defined by situational characteristics at any level of generality. Halliday (1988: 162) also notes this fact, stating that registers 'can be identified at any delicacy of focus'. There can be major differences among registers specified at different levels of generality. At one extreme, very general registers such as speech and writing are defined only by reference to their physical mode; at the other extreme, highly specified registers such as methodology sections in psychology articles are defined in terms of their physical mode, production circumstances, intended audience, micro-purpose, topic, etc. In the present study, register differences are

assumed to define a continuous space of variation;[2] thus, no attempt is made here to identify discrete register levels or to propose separate terms for varieties at different levels of generality (cf. Biber 1994).

Readers should be aware, however, that there is no general consensus within sociolinguistics concerning the use of *register* and related terms such as *genre* and *style*. A few writers (e.g., Trudgill 1974; Wardaugh 1986) restrict the term *register* to refer only to occupational varieties, such as computer programmer talk or auto mechanic talk. Several other researchers attempt to make a theoretical distinction between *register* and *genre*. Ventola (1984) and Martin (1985) refer to register and genre as different 'semiotic planes': genre is the 'content-plane' of register, and register is the 'expression-plane' of genre; register is in turn the 'content-plane' of language. Martin (1985) and Couture (1986) both describe registers as comprising particular configurations of the contextual categories of field, tenor, and mode (proposed by Firth and Halliday). Martin (1985: 250) further states that 'genres are how things get done', listing poems, narratives, expositions, lectures, recipes, manuals, appointment making, service encounters, and news broadcasts as examples of genres.

Gregory and Carroll's (1978: 64) and Couture's (1986:80) characterization of register – 'language in action' – is similar to Martin's characterization of genre. In contrast, Couture (p. 80) characterizes genre as 'conventional instances of organized text'. As examples of registers, Couture lists the following: the language used by preachers in sermons; the language used by sports reporters in giving a play-by-play description of a football game; and the language used by scientists reporting experimental research results. Genres include both literary and non-literary text varieties; for example, short stories, novels, sonnets, informational reports, proposals, and technical manuals.

Ferguson (1994) also attempts a theoretical distinction between *registers* and *genres*. *Register* variation is associated with 'a communication situation that recurs regularly in a society', while *genre* variation is associated with 'a message type that recurs regularly in a society' (pp. 20–21). Examples of registers cited by Ferguson include sports announcer talk and cookbook recipes, while chat, conversation, directions in a strange town, love-letters, newspaper articles, and obituaries are given as examples of genres.

The IPrA Survey of Research in Progress (Nuyts 1988) uses a number of different terms. The survey exemplifies terms in a hierarchical presentation rather than explicitly defining the terms. At a high level of generality, the survey distinguishes between *language varieties* and *discourse types*. *Register* is considered to be a language variety, along with *dialect, argot, slang,* and *jargon*; examples of registers in this framework include aviation language, journalese, legalese, literary language, religious language, scientific language, technical

language, and mythical language. The survey identifies two major discourse types: *conversation types* and *text types*. Conversation types include adult–child, classroom, interview, dinner, meeting, narrative, and courtroom. Text types include advertisement, comic strip, essay, joke, legal text, letter, literature, message, monologue, narrative, obituary, report, and summary.

In all these discussions, the distinction between register and genre tends to be quite abstract and vague. In practice, it is difficult to decipher the empirical correlates of these supposedly discrete categories, or to determine whether a particular variety should be classified as a register, a genre, or some other text category. For this reason, the present study uses only the term *register* as the general cover term associated with all aspects of variation in use.

The term *sublanguage* (Kittredge and Lehrberger 1982; Grishman and Kittredge 1986) has been used within computational linguistics to refer to 'a subsystem of language that . . . [is] limited in reference to a specific subject domain'; each sublanguage is claimed to have its own 'distinctive grammar' (Grishman and Kittredge 1986: ix). Sublanguages are quite restricted; for example, analyses of sublanguages have focused on medical articles about lipoprotein kinetics (Sager 1986) and Navy telegraphic messages (Fitzpatrick, Bachenko, and Hindle 1986).[3]

Finally, the term *style* has perhaps been used for a wider range of concepts than any of the other terms. Crystal and Davy (1969) use the term in a similar way to the use of register here; Joos (1961) similarly uses the term style to refer to registers at a high level of generality. Labov (1972) uses style to refer to language production under circumstances that require differing amounts of attention to speech (e.g., reading word lists versus an interview). More commonly, style has been treated as a characteristic way of using language. This usage often has an evaluative sense, as when writing handbooks discuss writing with style (which carries the implication that many people write without style). With regard to literary language, style in this sense has been studied as a characteristic of particular genres, particular periods, particular authors, and even particular texts (see discussion in Leech and Short 1981). A similar notion of style has been used to study conversational interactions, where each culture can be described as having a distinctive *communicative style* (e.g., Tannen 1984a, b).

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 (e.g., Biber 1988; Biber and Finegan 1989a). 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* in the present book corresponds closely to *genre* in these earlier studies.

In contrast, the term *text type* has been used in my own previous analyses to refer to text categories defined in strictly linguistic terms (Biber 1989). That is, regardless of purpose, topic, interactiveness, or any other non-linguistic factors, text types are defined such that the texts within each type are maximally similar with respect to their linguistic characteristics (lexical, morphological, and syntactic), while the types are maximally distinct with respect to their linguistic characteristics. After the text types are identified on formal grounds, they can be interpreted functionally in terms of the purposes, production circumstances, and other situational characteristics shared by the texts in each type. A comparison of text types in English and Somali is presented in chapter 9.

1.2.2 Previous studies of a single register

Studies of a single register have three major components: description of the situation in which the register is used; description of the linguistic characteristics of the register; and analysis of the functional or conventional associations between the situational and linguistic features. These relationships are schematized below:

SITUATIONAL FEATURES ⟷ FUNCTIONS and CONVENTIONS ⟷ LINGUISTIC FORMS

These relationships are bidirectional, with situational characteristics influencing the choice of linguistic form, while the choice of linguistic features in turn helps to create the situation. Positing a functional association does not entail a one-to-one mapping between form and function. Rather, the mapping across form–function–situation often comprises complex many-to-many kinds of relations.

Associations between form and situation can be motivated either by functional communicative requirements or by simple conventions. Functionally motivated patterns can be related to a number of situational characteristics, such as the physical setting, the extent of shared context or background knowledge, the degree of interactiveness, the production circumstances, the primary purposes or communicative goals of participants, and the social relations among participants. Finegan (1982) illustrates form-function associations in his analysis of common linguistic features (e.g., pronouns,

imperatives, passives) in last wills and letters of transmittal; Zwicky and Zwicky (1980) cite the use of past participle modifiers and 'tasty adjectives' as examples of conventional features of American restaurant menus. Ferguson (1983) describes both types of association in his analysis of baseball game broadcasts.

Most studies of a single register focus on a specialized kind of language. Early studies of this type include the following: Mellinkoff (1963), which describes the characteristic lexical and grammatical features of legal language; Ferguson (1964), which describes the typical linguistic characteristics of baby talk; and Leech (1966), which provides a comprehensive register description of advertising language. More recently, Ferguson (1983) has investigated the distinctive situational and linguistic characteristics of radio sportscasting; Heath and Langman (1994) have undertaken a related description of the sports coaching register; Bruthiaux (1994) presents a register analysis of personal advertisements in the *LA Weekly*; and Jucker (1992) provides a book-length treatment of syntactic variation in British newspapers.

There have been fewer diachronic studies of single registers, and these usually focus on the evolution of some professional variety (science or legal prose). For example, Hiltunen (1990) traces the evolution of legal prose in English from 635 AD to the present. Halliday (1988) analyzes syntactic changes in English scientific prose (in physics), from the fourteenth century onward. Bazerman (1984, 1988) uses several different methodological approaches to describe experimental science articles and presents two diachronic analyses: a rhetorical analysis of the *Philosophical Transactions of the Royal Society* from the years 1665–1800, and a register analysis of historical change in spectroscopic articles from the journal *Physical Review* from the years 1893–1980. Similar to synchronic studies of single registers, diachronic studies include both detailed linguistic analyses and descriptions of the salient situational characteristics. These diachronic studies show how historical change in the characteristic linguistic features of a register is closely tied to changes in the perceived purposes or primary audience of a register.

1.3 Previous research on register variation

The study of register can also be approached from a variationist perspective. Register variation studies compare the linguistic characteristics of two or more registers, usually using quantitative techniques to isolate linguistic similarities and differences among registers. Most previous studies of register variation have compared registers varying along a single situational parameter, such as formality, attention paid to speech, extent of planning,

or physical mode. The following subsections survey previous synchronic and diachronic studies of register variation, as well as register studies of non-western languages and computational analyses of sublanguages. Section 1.4, then, is devoted to studies adopting the Multi-Dimensional approach to register variation.

1.3.1 Previous synchronic studies of register variation along a single situational parameter

Comparisons of spoken and written registers have been the most common type of variation study. These studies have generally relied on quantitative methods to analyze differences in the relative distributions of surface linguistic features, such as adjectives, nominalizations, passives, and various clause types.

An early study comparing spoken and written registers is Blankenship (1962), which analyzed the lectures and published writing of public figures, focusing on linguistic features marking complexity (such as sentence length and passives). In a more elaborate study, Blankenship (1974) compared six spoken and written registers with respect to a wide variety of linguistic features, including word length, sentence length, type-token ratio, adjectives, and prepositions.

In an influential series of articles, Chafe (1982; Chafe and Danielewicz 1986) compared four spoken and written registers: dinner-table conversations, personal letters, lectures, and academic papers. These registers were analyzed with respect to linguistic characteristics from two functional/linguistic parameters: integration/fragmentation and involvement/detachment. The parameter of integration/fragmentation was realized by linguistic features such as nominalizations, participles, and attributive adjectives versus clause co-ordination; the parameter of involvement/detachment was realized by linguistic features such as first-person pronouns versus passives.

Tannen (1982a) compared spontaneous conversational narratives and elicited written versions of the same story, with respect to discourse phenomena such as the use of repetition, parallelism, and structural elaboration. Chafe and Tannen (1987) survey the extensive literature comparing spoken and written registers.

Another closely related parameter of variation is the distinction between planned and unplanned discourse described by Ochs (1979). This study compares unplanned, spoken narratives versus planned, written narratives on the same topic. Linguistic features analyzed include referent deletion, differing uses of demonstratives and definite articles, and active and passive voice.

Finally, a number of register variation studies have been directed towards

analysis of the differences between restricted and elaborated codes (building on the theoretical framework developed by Bernstein). These studies typically involve comparisons across social groups as well as comparisons across registers. For example, Rushton and Young (1975) compared three written registers across working-class and middle-class groups of students: imaginative descriptive writing, opinionative discursive writing, and technical explanatory writing. This study focused primarily on syntactic characteristics, such as complex nominal modifiers, passives, WH subordinators, and deeply embedded clauses. Poole and Field (1976) use a factor analysis to compare interviews and written personal essays by working-class and middle-class undergraduates; linguistic features analyzed include mean sentence length, subordinate clauses, adjectives, adverbs, passives, and pronouns.

1.3.2 Previous diachronic studies of register variation

There have been far fewer studies dealing with diachronic register variation, that is, the analysis of the changing relations among registers across time. Such studies are essential for a complete understanding of the processes of language standardization, modernization, and adaptation. In most previous linguistic studies, language change is treated as a mechanical process influenced only by language-internal (rather than social/situational) factors. Further, languages are usually treated as if they were homogeneous constructs uniformly affected by diachronic developments. However, recent studies show that a register perspective is crucial to a complete understanding of the processes of language development and change: that linguistic change interacts in complex ways with changing patterns of register variation.

Romaine's (1980, 1982) 'sociohistorical' approach shows how structural changes enter a language in particular registers and subsequently evolve at different rates in different registers. This approach is based on analysis of the relative frequency of forms across registers from different historical periods. Romaine (1980) traces the development of WH relative clause markers in Scots English, analyzing the alternation among relative clause markers in registers such as verse play, narrative prose, epistolary prose, and record prose; this study shows how WH relative clause markers first entered the language in the most complex literate registers.

Kytö and Rissanen (1983), Rissanen (1986), Kytö (1986, 1991), Nevalainen (1986), and Nevalainen and Raumolin-Brunberg (1989) adopt Romaine's sociohistorical approach to study the patterns of historical syntax in early British and American English. For example, Rissanen (1986) describes the differing development of periphrastic *do* in formal written registers (e.g., chronicles) versus speech-based registers (e.g., sermons and records of meetings). Devitt (1989a, b) also uses this approach to analyze the

influence of register variation on historical processes of standardization in Scots English and American English. She considers five registers (private records, private correspondence, official correspondence, religious treatises, and public records) and shows how linguistic features such as present participle inflections and relative pronouns were standardized at different rates in different registers.

While these empirical studies are important for their analysis of the interplay between language change and register variation, they do not provide comprehensive analyses of the changing patterns of register variation in a language. Other recent programmatic articles, however, have noted the importance of overall changes in the register system of a language and have called for comprehensive historical analysis of these developments. For example:

> The register range of a language is one of the most immediate ways in which it responds to social change. The difference between developed and undeveloped languages (Ferguson, 1968) is fundamentally one of register range, and language contact, which contributes to language development . . . is mediated by particular registers . . . This issue is concerned with both the pressures that make for change and the way in which these changes are realized linguistically. (Ure 1982: 7)

> [one of the two main tasks requiring attention within sociolinguistics at present is] the description and analysis of the organization and change of verbal repertoires in relation to the main processes of societal evolution of our time (Hymes 1984: 44–45)

Section 1.4.2 below surveys studies that extend the Multi-Dimensional approach in ways that enable such comprehensive analyses of the changing relations among registers in a language.

1.3.3 The study of register variation in non-western languages

Most studies of register variation have dealt with Indo-European languages, with English receiving by far the most attention. There have, however, been a few studies of register variation in non-western languages. For example, Duff (1973) compares oral and written stories in the Arawakan language Amuesha, finding differences in discourse organization, explicitness (e.g., the use or omission of temporal markers), and noun phrases. Deibler (1976) compares letters and conversations in Gahuku, a Papua New Guinean language, with respect to such features as contractions, shortened verbs, imperatives, and sentence length. Hurd (1979) presents a comparative description of a traditional legend, a personal narrative, and a procedural text in the Papua New Guinean language Nasioi, based on analysis of repetitions, grammatical reductions, conjunctions, and vocabulary selection.

Clancy (1982) and Tannen (1984a) compare spoken and written narratives

about the 'pear film'. Clancy presents a comparison of Japanese spoken and written narratives, focusing on referential choice (full NP, pronoun, or zero), word order, and relative clause types. Tannen compares spoken and written narratives in Greek and English, showing how Greeks in both modes and speakers in both languages tended to be more interpretive. Li and Thompson (1982) compare the characteristics of written classical Chinese with spoken and written modern Mandarin, focusing on zero anaphora, clause length, and the presence of grammatical morphemes. Chafe (1982) includes a comparison of conversational and ritual Seneca texts, briefly describing ways in which ritualistic language is more integrated and detached (more structural elaboration, and frequent impersonal constructions), while conversational language is more fragmented and involved (e.g., frequent evidentials related to emphasis or hedging).

Even fewer studies undertake diachronic register analyses of non-western languages. Romaine (1994) provides one of the few existing historical descriptions of register development in a non-western language. This study documents the early development of sports reportage in Tok Pisin, as it emerged as a distinct register in the newspaper *Wantok*. Hared (1992) adopts Romaine's earlier sociohistorical approach to analyze the historical changes in three Somali press registers over the first seventeen years of their history (1972–89). This study shows how the complementary processes of standardization and modernization interact with register factors in influencing the course of linguistic change.

Most of the studies cited above are restricted in scope, analyzing relatively few registers with respect to relatively few linguistic characteristics. In large part, these restrictions are due to the difficulties of fieldwork in many non-western languages. That is, the lack of previous grammatical and functional descriptions makes it difficult to decide which linguistic features to study, and it is often a time-consuming and labor-intensive job to compile a large corpus of texts from a non-western language. There have, however, been three comprehensive studies of register variation in non-western languages: analyses of Korean, Nukulaelae Tuvaluan, and Somali using the Multi-Dimensional approach. These studies, which provide the basis for the cross-linguistic comparisons in the present book, are introduced in section 1.4.3 below.

1.3.4 Computational analyses of sublanguages

Researchers in computational linguistics concerned with issues of linguistic variation have focused on the computational analysis of *sublanguages*, highly specified registers of a language that operate within specific domains of use with restricted subject matter. Many sublanguage studies have compared

varieties across languages, making them especially relevant to the purposes of the present book.

Sublanguages are usually taken from science and technology domains and are restricted to a particular topic. Examples of sublanguages analyzed to date include scientific journal articles on lipoprotein kinetics (Sager 1986), Navy telegraphic messages (Fitzpatrick, Bachenko, and Hindle 1986), weather reports (Lehrberger 1982), aviation maintenance manuals, and stock market reports (Kittredge 1982). Because they are so restricted in purpose and topic, sublanguages are much more systematic in structure and meaning than the language as a whole; thus computational systems for applied natural-language processing can achieve greater success when they are designed specifically for texts from a single sublanguage.

As noted above, research on sublanguages is particularly relevant to the present investigation because some of these studies are explicitly cross-linguistic: for example, Lehrberger (1982) and Kittredge (1982) discuss the use of sublanguage grammars as the basis for information retrieval and automatic translation between English and French. The kinds of issues raised in these studies are directly related to the issues investigated here. For example, Lehrberger (1982: 105) raises the following questions:

1 To what extent do corresponding sublanguages in different languages have similar characteristics?
2 Are there groups of sublanguages that have grammatical characteristics in common?
3 Are there systematic patterns underlying the evolution of sublanguages, associated with scientific developments and cultural changes?

Kittredge (1982) differs from most sublanguage studies in that it explicitly adopts a variationist perspective, comparing the extent and kinds of variability within and across sublanguages. Further, Kittredge compares the extent of sublanguage variability cross-linguistically, presenting highly interesting and provocative conclusions such as:

'the written style of English and French tended to be more similar in specialized technical texts than in general language texts' (1982: 108).

'parallel sublanguages of English and French are much more similar structurally than are dissimilar sublanguages of the same language. Parallel sublanguages seem to correspond more closely when the domain of reference is a technical one' (1982: 108).

'purpose of text and semantic domain have a powerful influence on text and sentence structure' (1982: 135).

With respect to structural characteristics, shared purpose is much more important than shared semantic domain, and in fact, a 'common semantic domain has little to do with structure at the sentence or text level' (1982: 135). Thus, recipes, aviation manuals, and other

kinds of manuals or assembly instructions show strong structural similarities, while English weather bulletins and weather synopses show important structural differences even though they share the same semantic domain.

'One is therefore drawn to conclude that English and French technical texts show the strongest parallels because the text purpose is more similar here than in descriptive texts . . . Most of the unexpected structures one finds in a sublanguage text can be associated not so much with a shift in semantic domain as with a shift (usually quite temporary) in the attitude which the text producer takes towards his domain of discourse' (1982: 135).

Questions needing additional research identified by Kittredge include:

1 What are the parameters of sublanguage complexity?
2 How does the 'professionalization' of a sublanguage affect the rigidity of its style? In less professionalized (e.g., colloquial) sublanguages, which lack recognized norms, is there considerably less consistency of structure and any greater 'distance' between English and French style?
3 How do the constraints of sublanguage semantics and pragmatics influence sentence structure and text structure? Can structural resemblances between semantically different sublanguages be related to similarities of text purpose?
4 How are the boundaries of a sublanguage determined? To date, most sublanguages chosen for analysis have tended to be quite restricted. To what extent do the same generalizations hold as more general kinds of registers are analyzed?

The present study investigates similar research questions, but it adopts a quite different analytical approach. First, the registers investigated here are defined at a higher level of generality than typical sublanguages (e.g., academic prose as a general register versus scientific journal articles on lipoprotein kinetics as a restricted sublanguage). Further, the present study distinguishes between registers, defined on the basis of situational characteristics, and text types, defined on the basis of linguistic characteristics. Both of these points are discussed further in chapter 9.

More importantly, the Multi-Dimensional approach used here attempts to be comprehensive in coverage, with respect to both linguistic and non-linguistic characteristics. From a linguistic perspective, most sublanguage studies (and most register studies generally) attempt to identify the few linguistic characteristics that are 'distinctive' for a register. In contrast, the cross-linguistic comparisons here are comprehensive in that they are based on a representative range of linguistic features in each language, with the analyses based on the systematic co-occurrence patterns among features (see chapter 2).

With respect to non-linguistic characteristics, most sublanguage studies have focused only on the influence of topic, although Kittredge (1982) also emphasizes the importance of purpose. In contrast, the present study investigates the importance of a large number of non-linguistic characteristics, including interactiveness, production circumstances, and personal stance, in addition to communicative purpose.[4] In sum, the cross-linguistic analyses presented here are more comprehensive, as well as more macroscopic, than the kinds of analyses typically included in sublanguage studies. In chapters 7 and 10, I discuss the extent to which these Multi-Dimensional analyses can be used to provide a theoretical framework for the more microscopic investigations required for sublanguage research.

1.4 Introductory overview of the Multi-Dimensional approach to register variation

Although there have been a large number of previous register studies adopting a comparative perspective, there have been few attempts to provide a comprehensive analysis of register variation in a language. Rather, most previous studies have compared only a restricted range of registers varying along a single situational parameter. The Multi-Dimensional (MD) approach to register variation – the analytical approach adopted in the present book – was developed to fill this gap.

MD analyses describe the relationships among the full range of registers in a language, with respect to multiple linguistic parameters of variation. As shown in later chapters, the MD approach also enables motivated register comparisons across languages. The following subsections briefly introduce previous MD studies, while fuller theoretical and methodological descriptions of the MD approach are given in chapters 2 and 5 respectively.

1.4.1 Multi-Dimensional studies of synchronic register variation in English

The MD approach to register variation was originally developed by Biber (1984c, 1985, 1986, 1988) for comparative analyses of spoken and written registers in English. Methodologically, the approach uses computer-based text corpora, computational tools to identify linguistic features in texts, and multivariate statistical techniques to analyze the co-occurrence relations among linguistic features, thereby identifying the underlying *dimensions* of variation in a language. The analytical goal of the MD approach is to provide comprehensive descriptions of the patterns of register variation having two components: (1) identification of the underlying linguistic parameters, or

dimensions

dimensions, of variation; and (2) specification of the linguistic similarities and differences among registers with respect to those dimensions.

Dimensions are defined by distinct groupings of linguistic features that co-occur frequently in texts. Dimensions are identified statistically by a factor analysis, and they are subsequently interpreted in terms of the communicative functions shared by the co-occurring features. Interpretive labels are posited for each dimension, such as 'Involved versus Informational Production', 'Narrative versus Non-narrative Concerns', and 'Explicit versus Situation-Dependent Reference'.

Two primary motivations for the MD approach are the assumptions that: (1) generalizations concerning register variation in a language must be based on analysis of the full range of spoken and written registers; and (2) no single linguistic parameter is adequate in itself to capture the range of similarities and differences among spoken and written registers. The approach thus requires analysis of numerous spoken and written registers with respect to numerous linguistic features. In earlier synchronic MD analyses of English (e.g., Biber 1986, 1988), approximately 500 texts from twenty-three registers were analyzed, including face-to-face conversations, interviews, public speeches, broadcasts, letters, press reportage, official documents, academic prose, and fiction. Linguistic features analyzed in these studies include lexical features (e.g., type–token ratio, word length, hedges, emphatics), grammatical features (e.g., nouns, prepositional phrases, adjectives), and syntactic features (e.g., relative clauses, adverbial clauses).

multi-dimensional

Registers can be compared along each dimension. Two registers are similar along a dimension to the extent that they use the co-occurring features of the dimension in similar ways. MD analyses show that registers are often similar along one dimension but quite different along other dimensions.

The early MD studies of English have been extended in several ways. Biber (1988) gives the fullest account of the methodology and a synchronic analysis of the relations among spoken and written registers; Biber and Finegan (1986) and Biber (1989) analyze the linguistically well-defined *text types* of English (i.e., text varieties which are maximally similar in terms of their linguistic characteristics; see chapter 9); Biber (1990, 1993a, b, c, 1994) and Biber and Finegan (1991) elaborate on methodological considerations; Biber (1987, 1991) considers specialized registers of English (British versus American writing, primary-school reading materials); Biber and Finegan (1994a) analyze the patterns of variation across subsections within medical research articles (viz., Introduction, Methodology, Results, and Discussion); and Biber and Finegan (1988, 1989b) and Biber (1992a, b) consider specialized linguistic domains (stance, discourse complexity, and referential strategies).

Two of the major conclusions of these synchronic MD studies of English are that: (1) no single dimension of variation is adequate in itself to account for the range of similarities and differences among registers – rather, multi-dimensional analyses are required; and (2) there is no absolute difference between spoken and written language – rather, particular types of speech and writing are more or less similar with respect to different dimensions.

1.4.2 Multi-Dimensional studies of diachronic register variation in English

The MD approach has also been used to study diachronic patterns of register variation in English. These studies use the dimensions identified and interpreted in Biber (1988) to trace the development of registers across time periods. Biber and Finegan (1989a, 1992, 1994b) study the development of English written registers from 1650 to the present with respect to three linguistic dimensions. The 1989a study traces the development of fiction, essays, and letters; this study interprets the observed patterns of change relative to the changing purposes and readership of written texts. The 1992 study adds an analysis of dialogue in drama and dialogue in fiction to the 1989 description. Finally, the 1994b study uses the framework developed in these previous studies to compare the written styles of particular eighteenth-century authors (Swift, Defoe, Addison, and Johnson) across different registers.

In addition, two studies by Atkinson use the MD approach to trace the evolution of professional registers in English. Atkinson (1992) combines a multi-dimensional approach with a detailed analysis of rhetorical patterns to study the development of five subregisters of medical academic prose from 1735 to 1985, focusing on the *Edinburgh Medical Journal*. Atkinson (1993) employs a similar integration of multi-dimensional and rhetorical methodologies to analyze the evolution of scientific research writing, as represented in the *Philosophical Transactions of the Royal Society of London* from 1675 to 1975. Chapter 8 of the present book integrates these previous MD studies (by Biber, Finegan, and Atkinson) to provide a more comprehensive picture of diachronic register variation in English.

1.4.3 Multi-Dimensional studies of register variation in non-western languages

Three MD studies of register variation in non-western languages have been completed to date: Besnier's (1986, 1988) analysis of Nukulaelae Tuvaluan; Kim's (1990; Kim and Biber 1994) analysis of Korean; and

Biber and Hared's (1992a, b, 1994) analysis of Somali. Taken together, these studies provide the first comprehensive investigations of register variation in non-western languages.

The first MD study of a non-western language was undertaken by Besnier (1986, 1988) on Nukulaelae Tuvaluan. This study analyzes the characteristics of seven spoken and written registers (e.g., conversations, political meetings, private-setting speeches, sermons, letters) with respect to three underlying linguistic dimensions of variation ('Attitudinal versus Authoritative Discourse', 'Informational versus Interactional Focus', 'Rhetorical Manipulation versus Structural Complexity'). As in the MD analyses of English registers, each dimension represents a distinct grouping of linguistic features that co-occur frequently in texts, reflecting shared communicative functions.

The second MD analysis of a non-western language, carried out by Kim (1990; Kim and Biber 1994), was even more ambitious, analyzing the relations among twenty-two spoken and written registers in Korean. This study identified five major dimensions of variation: 'Informal Interaction versus Explicit Elaboration', 'Discourse Chaining versus Discourse Fragmentation', 'Stance', 'Narrative Concern', and 'Honorification'.

Finally, Biber and Hared (1992a) study the relations among twenty-six spoken and written registers in Somali with respect to five major dimensions of variation: 'Structural Elaboration: Involvement versus Exposition', 'Lexical Elaboration: On-Line versus Planned/Integrated Production', 'Argumentative versus Reported Presentation of Information', 'Narrative versus Non-Narrative Discourse Organization', and 'Distanced, Directive Interaction'. Biber and Hared (1992b, 1994) extend this MD analysis of Somali to study historical change following the introduction of native-language literacy in 1973. The 1992b study compares the range of register variation found before 1973 (when only spoken registers existed) with that found immediately after 1973 (among spoken *and* written registers). The 1994 study traces the evolution of seven press registers from 1973 to 1989, analyzing the historical evolution of written registers in their initial periods of development.

These studies of Korean, Nukulaelae Tuvaluan, and Somali, together with the earlier MD analyses of English, provide the basis for the cross-linguistic investigations of register variation in the present book.[5]

1.5 The importance of cross-linguistic studies of register variation

The present study investigates patterns of register variation from a cross-linguistic perspective, addressing the extent to which the underlying dimensions of variation and the relations among registers are configured in

similar ways across languages. Given the research gaps identified above – with relatively few comprehensive studies of register variation in English and very few studies of register variation in non-western languages – it will come as no surprise that there have been essentially no previous cross-linguistic investigations of register variation. Similarly, since there have been few diachronic analyses of register variation in any language (western or not), it is not surprising that there have not been previous cross-linguistic diachronic comparisons of register variation. However, there are at least four reasons why such cross-linguistic studies (both synchronic and diachronic) are needed at the present time.

First, as the following chapters show, there are highly systematic similarities in the patterns of register variation across languages, suggesting the operation of underlying form–function associations tied to basic aspects of human communication. The four languages analyzed here show striking similarities in their basic patterns of register variation, as reflected by:

the co-occurring linguistic features that define the dimensions of variation in each language;
the functional domains represented by those dimensions; and
the linguistic/functional relations among analogous registers.

The cultural and linguistic diversity among these four languages raises the possibility that some of these shared patterns will turn out to reflect universals of register variation. Assuming that all languages comprise a range of registers, the analysis of these shared patterns of register variation is central to any comprehensive theory of cross-linguistic typology and universals.

Second, diachronic comparisons of register variation across languages are crucial to a broad range of theoretical issues, including the general mechanisms of historical change, the processes of language standardization and modernization, and the influence of literacy on language change. In particular, the following chapters show that languages as diverse as English and Somali have undergone similar patterns of evolution following the introduction of written registers. These similarities again raise the possibility of universals of register variation, in this case relating to the historical development of written registers in response to pressures of modernization and language adaptation.

Third, cross-linguistic analyses of register variation are needed as a basis for on-going research in computational linguistics (e.g., analyzing bilingual text corpora to develop computational systems for machine translation). As noted above, related research in this area has focused on the analysis of sublanguages. Studies such as Kittredge (1982) have found that technical sublanguages can be more similar cross-linguistically than disparate sublanguages within the same language. However, there are

several unresolved issues in sublanguage research, many of which relate to the overall linguistic characterization of sublanguages and the question of how the range of sublanguages relate to one another within and across languages. The following chapters show that such issues can be addressed in the MD approach.

Finally, cross-linguistic analyses of register variation are essential for the development of comprehensive text typologies. As chapter 9 shows, the text categories that are well defined linguistically (i.e., the *text types*) are remarkably similar across languages, when compared within the multi-dimensional space of each language.

The analyses in chapters 4–7 show that the MD approach is particularly well suited to cross-linguistic comparisons because it is based on the co-occurrence patterns that are well represented in each language, reflecting the communicative functions that are well represented in the corresponding culture. Comparisons based on individual linguistic features cannot address these issues. As Hymes predicted in 1974, 'it is essential to isolate the dimensions and features underlying taxonomic categories. These features and dimensions, more than particular constellations of them, will be found to be universal, and hence elementary to descriptive and comparative frames of reference' (p. 41). The analyses in chapters 7–9 show that there are indeed strong similarities in the underlying dimensions of English, Nukulaelae Tuvaluan, Korean, and Somali. Future research is required to determine which of these shared patterns reflect underlying universals of register variation.[6]

1.6 Overview of the present study

The present study synthesizes earlier MD analyses of register variation in four quite different languages: Biber's (1986, 1988; Biber and Finegan 1989a) analysis of English; Besnier's (1986, 1988) analysis of Nukulaelae Tuvaluan; Kim's (1990; Kim and Biber 1994) analysis of Korean; and Biber and Hared's (1992a,b, 1994) analysis of Somali.[7] The study explicitly compares the patterns of register variation across languages, addressing the extent to which there are cross-linguistic similarities with respect to:

1 the co-occurrence patterns among linguistic features, and the ways in which features function together as underlying dimensions;
2 the synchronic relations among registers;
3 the diachronic patterns of change within and among registers;
4 the text types that are well defined linguistically.

The four languages analyzed here complement one another in several respects. From a strictly linguistic point of view, these languages are from four quite different language families: English from Indo-European,

Tuvaluan from Austronesian, Korean from Altaic, and Somali from the Cushitic subfamily of Afroasiatic. They are notably different in their geographic locations, as well as in their cultural and religious associations. Geographically, the specific variety of English used for the present study is spoken in England, and it thus represents a major European language with millions of speakers. Tuvaluan represents the opposite extreme: the variety studied here is spoken on the central Pacific atoll of Nukulaelae, which has only about 310 inhabitants. Korean represents one of the major Asian languages, with approximately 65 million speakers. And Somali represents one of the major languages of Africa, spoken by approximately 5 million inhabitants of the countries of Somalia, Djibouti, Ethiopia, and Kenya.

Culturally, these four groups range from the primarily urbanized speakers of English and Korean, to the mixed urbanized and rural speakers of Somali (ranging from city-dwellers to nomadic camel-herders), to the traditional fishing/gathering/farming culture of Nukulaelae Tuvaluan. Religion plays a central role in the daily lives of speakers from two of these cultural groups: Christianity in Nukulaelae Tuvaluan and Islam in Somali; thus, speech events in these cultures almost always contain some reference to God or other religious associations. Religion plays a less central role in the English and Korean cultures.

These languages also differ with respect to their status: English is spoken in many countries around the world, it has a long history of literacy and standardization, and it has an extremely broad range of spoken and written registers; Nukulaelae Tuvaluan represents the other extreme in that it has very few speakers, restricted primarily to a single atoll in the Pacific, a relatively short history of literacy, only two written registers, and a generally restricted range of spoken registers. Korean is spoken primarily in Korea, but it has a long history of literacy and presently has a wide range of spoken and written registers with well-established uses. Finally, Somali is spoken in four countries of East Africa (with some official status in two of those countries), but it is not well known outside of those countries. It has a very short history of literacy, although at present it has a wide range of spoken and written registers.

Overall, the languages considered in the present investigation represent four quite different language types and social situations. Obviously any attempt to identify cross-linguistic universals of register variation must be based on a larger sample of languages. The goal here is more modest: to investigate the possibility of such universals, to identify and interpret the generalizations concerning register variation that hold across these four languages, and to explore the utility of the MD approach for such analyses. Although the sample of languages is small, the investigation here can be regarded as a relatively strong test of these research questions. That is,

given the extreme linguistic, sociocultural, and situational differences across the four languages considered here, any marked cross-linguistic similarities in the patterns of register variation can be interpreted as reflections of basic communicative functions shared cross-culturally, indicating the potential for cross-linguistic universals (or at least universal tendencies) governing the patterns of register variation.

1.7 Outline of the remainder of the book

Chapter 2 further develops two of the themes introduced in the present chapter: it describes the analytical requirements of comprehensive studies of register variation, and it presents a more detailed theoretical introduction to the MD approach. In chapter 3, I outline the important cultural and social aspects of the four language situations included in the study. This chapter also includes a brief description of the social history of literacy in each case, as well as situational descriptions of the registers represented in each language. Chapter 4 illustrates the difficulties involved in cross-linguistic register comparisons based on individual linguistic features. The chapter compares three registers in English and Somali with respect to selected individual linguistic features, showing that conclusions differ considerably depending on the structural level used for comparison. Chapter 4 concludes with a brief discussion of how the MD approach provides a solution to the methodological indeterminacy of cross-linguistic comparisons based on individual features.

Chapter 5 summarizes the methodology used for MD studies and describes the text corpora, tagging programs, linguistic features, and statistical procedures used in the analysis of each of the four languages. Chapter 6 then presents and interprets the four MD analyses of register variation. For each dimension in each language, this chapter describes the set of co-occurring linguistic features grouped on the dimension, the similarities and differences among registers with respect to the dimension, and the functional underpinnings.

Chapter 7 presents the heart of the book: a cross-linguistic comparison of the patterns of register variation from several perspectives. First, the dimensions themselves are compared across the four languages, with respect to both the sets of co-occurring linguistic features and their underlying communicative functions. The strongest similarities exist for those dimensions that serve similar functions; in many cases these dimensions are represented cross-linguistically by structurally similar sets of linguistic features, and they define similar relations among spoken and written registers.

The analysis in chapter 7 further shows that it is possible for communicative functions to be given different prominence in two languages; for

example, two dimensions might be clearly distinguished in one language but conflated in a second language. Several dimensions are represented in all four languages and are thus potential candidates for universal status. Other dimensions are shown to reflect the particular communicative demands of a given culture and language. An overall comparison of the multi-dimensional space for each language shows that dimensions are differently salient depending on the importance of particular functional domains in the four cultures.

In addition, chapter 7 compares the dimensions in each language from a structural perspective, showing that many communicative functions have quite similar structural bases in these four languages. Finally, this chapter compares selected registers across the four languages, showing that there are similar multi-dimensional characterizations associated with characteristics of the physical situation, while there are linguistic differences associated with different particular communicative purposes. The conclusion to chapter 7 discusses the theoretical implications of these observed cross-linguistic patterns.

Chapters 8 and 9 are more specialized, comparing only English and Somali. Chapter 8 focuses on diachronic register variation, comparing the evolution of written registers in English and Somali. Despite important differences in their social histories – English with a long history of written registers, in contrast to the extremely compressed history of written registers in Somali – there are notable patterns of similarity in the development of written registers in the two languages. Chapter 9 then compares the text typologies of English and Somali. For this analysis, the register distinctions are set aside, and cluster analysis (a multivariate statistical technique) is used to identify the text types that are linguistically well defined in each language. In this chapter, it is shown that certain text types are linguistically well defined in both languages, while others appear to be distinctive. As in chapter 8, these patterns are discussed relative to the differing functional priorities of English and Somali.

In conclusion (chapter 10), I briefly summarize the main points of the study, describe related on-going research, and identify a number of research issues that require further investigation.

The book also includes two appendices, presenting brief grammatical descriptions of the linguistic features used in the analyses of Korean and Somali. Appendix I on Korean is written by Yong-Jin Kim, adapted from his 1990 dissertation; appendix II on Somali is written by Mohamed Hared and myself.