

## ORALITY

*Merja Kytö*

- Robert Keayne's Notebooks: A Verbatim Record of Spoken  
English in Early Boston? 273

*Brian D. Joseph*

- Textual Authenticity: Evidence from Medieval Greek 309

*Jadranka Gvozdanović*

- Parameters Underlying Punctuation in Older Russian Texts 331

## DIALECT

*Anthony Kroch, Ann Taylor and Donald Ringe*

- The Middle English Verb-Second Constraint: A Case Study in  
Language Contact and Language Change 353

*Frans Gregersen and Inge Lise Pedersen*

- A la Recherche du Word Order Not Quite Perdu 393

- Index of Languages 433

- Index of Subjects 437

## On Textual Parameters and Older Languages

Susan C. Herring, Pieter van Reenen and Lene Schøsler

### 1. Definitions

#### 1.1 Textual Parameters

The term 'textual parameters', as employed in this volume, refers to properties of texts and their contexts that condition variation within individual languages. Examples of textual parameters include text type and genre, poeticality, orality, dialect, writer demographics, scribal influence, cultural status, and whether a text is a translation from another language. The choice of the word 'parameters' is intended to suggest that such factors exercise a partially determining or constraining influence on the choice of linguistic expression within a given text,<sup>1</sup> and moreover that this influence is in principle systematic and predictable, given sufficient understanding of the parameters themselves.

This sense of parameters differs from the specialized use of the term in the Principles and Parameters model of generative syntax, as articulated by Chomsky (1981) and more recently by Kayne (1996). In that model, parameters are defined as dimensions of cross-linguistic syntactic variation that are determined by principles of Universal Grammar (UG). Such parameters are finite in number and binary in value: for example, the 'subject of main finite clauses' parameter has the values null subject and overt subject. A parameter is 'set' when the language learner decides on a value for it based on evidence in the language environment (van Kemenade & Vincent 1997). Like language acquisition, language change in the Principles and Parameters model involves parameter (re)setting in the competence of individual speakers, which is

claimed to have abrupt and widespread consequences for the grammar as a whole (Lightfoot 1991, 1999).

In contrast, textual parameters are defined inductively, on the basis of systematic empirical study of older languages. Rather than having their source in abstract language universals, they arise out of social conventions and practices, their diversity reflecting the diversity of purposes for which humans communicate. As such, their number is in principle unlimited. Nor are textual parameters all of the same type; some can be conceptualized as binary (for example, spoken vs. written production, translated vs. original language), while others have multiple, more or less discrete values (textual genres such as literary narrative, sermon, letter, etc.), and others are scalar dimensions (degree of scribal influence, degree of contact with other languages, etc.). Still others, such as poeticality, comprise both binary (verse vs. prose composition) and scalar (degree of focus on the aesthetics of language) dimensions. Not all of the parameters are necessarily relevant for all texts, but all are potentially relevant in any given language. Children acquire knowledge of textual parameters through the learning of conventions, rather than through innate predisposition. And while textual parameters can and do change as cultural practices evolve over time, language change does not crucially involve parametric change. Rather, the insight into language change offered by the textual parameters approach is of a different order: because language structure can vary according to one or more textual parameters, language itself is revealed to be a heterogeneous affair, and the notion of language evolution as a single process is thereby called into question.

## 1.2 *Older Languages*

While textual parameters are arguably important in the structural analysis of any language (see e.g. Longacre 1979), they are of special importance for the study of older languages. This is so because older languages — whether they have since become extinct or whether they represent earlier stages of contemporary living languages — are preserved primarily in the form of written documents, or texts. ‘Text languages’ (Fleischman, This volume) are limited in comparison with modern spoken languages in several respects. To begin with, the exact place and date of the creation of older texts may be unknown, and their authorship difficult to ascertain. Some texts may effectively have been written by multiple authors, e.g. in cases where older manuscripts were

copied and modified by later hands (Benskin & Laing 1981; Schöslér 1995). The language of the documents may thus reflect a mix of different linguistic systems, the sources of which remain largely unrecoverable to the analyst.

Moreover, text languages offer an incomplete range of data for linguistic analysis. Only texts from certain genres tend to be preserved from any given historical period, making it difficult for the analyst to separate out the effects of diachrony from the effects of genre (Hock 1997; Gregersen & Pedersen, This volume). Some languages, such as Tamil, are attested in the oldest stages exclusively in poetic verse (Herring, This volume). Others, such as Gothic, are attested primarily in texts translated from another language (Lehman 1973). Depending on the antiquity of the language variety under investigation, the gaps in what kinds of data are available are never small, and sometimes are yawning chasms.

Scholars of older languages have little choice but to work with the texts available. A further problem then arises concerning the status of those texts: how representative are they of the language of the period? For while some historical linguistic studies limit their focus to patterns found only in a specific text or set of texts, many more are interested in describing the features of language X at a given stage of evolution, and for that, generalization is necessary. The question of representativeness is often framed in terms of a text’s ‘authenticity’, with translated texts generally viewed as the least authentic, verse texts as only slightly more authentic, and texts which reflect the spoken language of the times as most authentic and desirable (see e.g. Bakker & Kahane 1997; Jucker, Fritz & Lebsanft 1999). Yet even ‘authentic’ data are necessarily limited: text languages, by definition, are written, and can provide no direct evidence of spoken communication.<sup>2</sup> Nor of course are there any living native speakers of older languages of whom one can ask questions about grammaticality judgments, usage, and the like.

Last but not least, historical texts tend to come down to us through the ages stripped of their social context. Information about the text producers and intended receivers, their social roles and personal relationships, the physical and social setting of text production and reception, and the goals of the text is notoriously sparse for older texts, making pragmatic and sociolinguistic analyses based on such texts difficult.<sup>3</sup> This leads Jacobs and Jucker (1995:7) to conclude that “except for the very immediate past, historical pragmatic hypotheses can never be empirically supported”.

To be sure, historical linguists have long been aware of these properties

of older texts (see e.g. Jeffers & Lehist 1979). However, it seems fair to state that on the whole, they have tended to view them primarily as limitations to be overcome or circumvented, rather than as legitimate foci of inquiry in their own right. In contrast, the approach taken in this volume holds that the properties of older texts are both methodologically important and theoretically interesting — indeed, we ignore or trivialize them at considerable risk. One risk is that of basing linguistic generalizations on data sources that are partial, atypical, or otherwise unrepresentative of a historical period of a language (Hock, *This volume*). When, in addition, language structure and use vary according to parameters that characterize individual texts, as the papers in this volume show, the risk of a partial or misleading analysis is compounded.

The textual parameters approach seeks to minimize such risks by identifying and controlling for situational dimensions that condition variation in older texts. On the one hand, it is a classificatory enterprise that has as its goal to create homogeneous subsets of data out of the heterogeneity of historical records. In so doing, it allows the analyst to make meaningful generalizations within restricted domains. On the other hand, when the conditioning parameters are themselves well understood, the analyst may abstract away from their influence in order to focus on structural (phonological, morphosyntactic, etc.) phenomena, and by comparing data subsets within a language, arrive at more nuanced generalizations about such structures than would otherwise be possible. Alternatively, the analyst may choose to focus on the properties of the parameters themselves, seeking to identify, for example, the genres available in a particular historical period of a language (Herring, *This volume*), or the degree of 'orality' of a body of texts (Kytö, *This volume*). Thus while the textual parameters approach does not in and of itself resolve the problem of limited data, it enables the analyst to extract more value from the data that are available.

## 2. Intellectual Antecedents

A central notion in this volume is the relation between homogeneity and heterogeneity. This relation has been largely neglected in the study of older languages, although it is no less important than in research on contemporary languages. The view that older languages are homogeneous entities through which change spreads uniformly can be traced to the 19th century Neogram-

marians, and is still evident in the work of many historical linguists today. Historical linguistics also inherited from the Neogrammarians a tendency to privilege phonology and morphology in studies of language change, a tendency which has persisted in the comparative method of historical reconstruction, as well as in the more recent variationist approach to language change (see below).

Recent decades have seen two major departures from the limitations of the Neogrammarian legacy. First, the notion that a language is homogeneous at any point in time has been challenged by sociolinguists (Labov 1963; Weinreich, Labov & Herzog 1968) who argue that evidence for diachronic change is present in synchronic, socially-conditioned variation.<sup>4</sup> Second, the scope of what counts as legitimate phenomena for historical linguistic study has broadened, such that it now includes grammaticalization (Traugott & Heine 1991), theoretical morphosyntax (van Kemenade & Vincent 1995) and syntax (Lightfoot 1991, 1999), lexicology and semantics (Blank & Koch 1999), discourse markers (Brinton 1996), and pragmatics (Jucker 1995). However, these two new research directions have been pursued largely independently of one another: with few exceptions, the variationist paradigm tends not to focus on older stages of language,<sup>5</sup> dependent as it is on the social context provided by modern spoken languages, and the expanded historical approach, while focusing squarely on older languages, tends not to address patterns of variation.<sup>6</sup> As a result, there is as yet no single coherent approach committed to accounting for heterogeneity in historical texts.

The textual parameters approach is an attempt to bridge this gap. Narrowly conceived, its subject matter is variation conditioned by properties of texts themselves — their manner of production, their notional type, their structural organization, and the like — properties which are directly reflected in the textual artifacts that come down to us through time, and hence are accessible for empirical analysis. In this sense, the approach has antecedents in text linguistics (cf. Longacre 1983/1996). A text-focused approach is represented in the present volume in the chapters by Bolkestein, Brinton, and Gvozdanović, among others. However, our definition also admits of a broader conceptualization, which includes within its scope socio-historical factors — writer and reader demographics, their geographical and social dialect(s), contact with other languages, the effects of language standardization, and other features of the (not always directly accessible) external context in which a text was produced. The chapters by Gregersen and Pedersen, and by Kroch,

Taylor and Ringe, are examples of sociolinguistic approaches to textual parameters.

Our broad conceptualization of textual parameters overlaps partially with Jacobs & Jucker's (1995) equally broad definition of historical pragmatics, which encompasses, in addition to "a pragmatized semantics, speech act theory, the research into function words, [and] the analysis of maxims of conversation" in older languages, historical applications of "text analysis (text types, communication forms, text pragmatics)" and "language norms and varieties" (p.10). As examples of this latter type of analysis, several papers in Jucker (1995) — e.g. Tuija Virtanen's on locative and temporal markers in Early Modern English travelogues, and Heinz Bergner's on the "openness" of medieval texts — could almost as easily have appeared in the present volume.

More generally, the questions addressed by the textual parameters approach have been raised and considered to a greater or lesser extent by individual scholars, too numerous to name here, scattered across different historical sub-disciplines. Many scholars of older languages have recognized the need to take account of textual parameters in order to make sense of the heterogeneity of their data, albeit without necessarily focusing their analysis on those textual properties *per se*.<sup>7</sup> This work constitutes the broader context out of which the impetus for a textual parameters approach arose. The present volume provides a venue for individual analyses of older languages which address heterogeneity, in the hopes that by bringing them together in one place, their insights might inform one another at a more abstract level, leading ultimately to the systematic identification of parameters of variation that characterize 'text languages' more generally. Such an approach holds the promise of advancing historical linguistic inquiry by contributing to it a principled understanding of heterogeneity, in the process allowing homogeneous patterns of structure and use more clearly to be seen.

### 3. The Evidence

#### 3.1 Overview of the Chapters

The 12 chapters in this volume report on original linguistic research involving nine older languages, two of them non-Indo-European (Japanese and Tamil),<sup>8</sup> the remaining languages representing the Germanic, Hellenic, Indic, Ro-

mance, and Slavic branches of the Indo-European family. The language varieties examined date from ca. 1500 B.C. (Vedic Sanskrit)<sup>9</sup> to the 20th century A.D. (Modern Danish); two (Latin and Sanskrit) have since become extinct.<sup>10</sup> The linguistic phenomena investigated also cover a broad spectrum, from morphology to syntax to discourse particles to punctuation. The absence of papers dealing with phonological and lexical variation we view as accidental gaps; in principle, there is no reason why the study of sound change or lexical change should be incompatible with the textual parameters approach.

Diachrony has no privileged status in this collection. Although all of the authors analyze phenomena that have undergone change over time, only a subset explicitly compares different historical stages of the same language, and only two (Brinton and Gregersen & Pedersen) attempt to track a linguistic structure over a continuous period of evolution. For other authors, diachrony is one of a number of possible parameters that explain patterns of variation, albeit not necessarily the most revealing one (e.g. Hock; van Reenen & Schøsler).

Conversely, a question that informs all of the contributions to some degree is that of authenticity, framed by some authors as whether and how the textual data reflect the spoken language of the time (see especially the chapters by Joseph and Kytö). Other authors address the authenticity issue at a theoretical level, questioning what it means for a particular textual variety to be 'authentic' (Gergersen & Pedersen; Herring; Hock).

Most importantly for the purposes of the present volume, all of the authors invoke textual parameters to explain variation in their data. Four parameters recur as particularly significant in the data examined in this collection: *text type*, *poeticity*, *orality*, and *dialect*. In the chapter summaries that follow, after a critical introduction that sets the stage for the textual parameters approach, the chapters are grouped according to these four parameters. This is largely a presentational convenience; most authors invoke more than one parameter, including those not discussed by any other author. It may serve, however, to illustrate the kinds of evidence that a textual parameters approach brings to bear in the analysis of older languages.

#### 3.2 Methodology and Ideology

The chapter by **Suzanne Fleischman**, originally delivered as a plenary lecture at the 11th International Conference on Historical Linguistics in Los Angeles, was a direct inspiration for the present volume. Fleischman's critical method-



ological survey of historical linguistic and philological practice sets the agenda for the textual parameters approach (although she does not refer to it as such) by identifying variation as "the essence of medieval textuality", and by calling for a new approach to the linguistic description of "text languages" that makes variation its foundation.

Text languages, according to Fleischman, raise unique methodological challenges because they lack native speakers and have finite corpora; moreover, the texts that are available tend to be heavily "processed" — that is, copies at one or more removes from their manuscript sources. Drawing on the example of Old French, she characterizes authentic medieval texts as heterogeneous at multiple levels: inconsistent orthography, variable morphosyntax, genre-specific features, and dialectal heterogeneity, across different copies of the same manuscript and within a single manuscript. This heterogeneity has been the "nemesis of philology", which (along with the enterprise of so-called descriptive grammar) has sought to uncover the underlying structural homogeneity of Old French, in the process regularizing variation, "correcting" scribal "errors", supplying "missing" suffixes, and constructing a largely fictional medieval dialect (Francien) from which modern standard French is supposed to have developed.

Fleischman analyzes these practices as reflective of Old French scholars' *ideologies of language*, including a belief in the incompetence of scribes, the normative status of the standard language, language change as decay, and "the myth of monoglossia" — the belief that a language is a unified, homogeneous entity. The "unarticulated — and no doubt unconscious — desire for the stability and regularity that institutionalized written language offers" also leads scholars into the methodological traps she terms, following Cerquiglini et al. (1976), the *Historicist Reflex* — the tendency to derive the grammar of Old French from Latin — and *Conceptual Inertia* — the anachronistic application of grammatical categories of modern French to Old French, or the failure to recognize a category or distinction operative in Old French because it is no longer operative in the modern language. Her critique sounds a warning to scholars of older languages in general, not just those interested in Old French: methodological assumptions influence the nature of historical linguistic findings. If our methodologies are ideologically-based, as in the case of Old French, rather than truly descriptive, the result will likely be "descriptions" at odds with the reality of the textual data.

In light of these problems, Fleischman acknowledges the "normative and

provisional status of our generalizations about (...) text languages generally", suggesting that historical linguists might usefully adopt practices of variationist sociolinguistics as a methodological corrective. However, variationist sociolinguistics does not consider all of the types of heterogeneity Fleischman identifies as being present in medieval manuscripts, nor does it allow for critical analysis of the sort that she so insightfully engages in in her essay. This difference aside, we see Fleischman's essay as raising a set of challenges that the textual parameters approach — indeed, any responsible empirical approach to the study of text languages — must ultimately seek to address. The remaining chapters in this volume all attempt to meet this challenge in one or another respect, by reporting on case studies of variable linguistic phenomena in text languages.

### 3.3 *Text Type*

One of the problems of heterogeneity identified by Fleischman is the existence of "genre-specific" features in older texts. However, the term 'genre' as it is used in the field of linguistics is rather imprecise (Swales 1990), in that it can refer to categories of discourse at multiple levels of generality. Thus for some contributors to this volume, genre is broadly construed to include such categories as 'prose' and 'verse', while for others, a genre is a specific category whose members share a common purpose and set of linguistic conventions (e.g. mystery fiction, romantic fiction, academic prose, official documents; see Biber 1988). In this introduction, we subsume the former distinction under 'poeticity' (see below), reserving the term 'genre' for the latter, more restrictive notion. In addition, linguists sometimes find it desirable and useful to classify texts at an intermediate level; for example, mystery fiction and romantic fiction (along with folk tales, epics, and personal experience stories) are all subtypes of 'narrative', and academic prose and official documents are subtypes of 'exposition'. The categories narrative and exposition can themselves be contrasted in terms of their notional purpose (narrating vs. explaining) and characteristic surface features (e.g. temporally-sequenced events vs. logically-related assertions; Longacre 1983/1996). Confusingly, however, the term 'genre' is sometimes employed for this intermediate classificatory level as well, including by the authors in this volume. To distinguish this latter notion from the others, we refer to it using the text linguistic term 'text type'.

Studies of text typology vary in the number and nature of types they

identify, although most include narrative as a basic monologue type, along with some form of expository and some form of argumentative or hortatory discourse (Longacre 1983/1996; Virtanen 1992). These, in turn, contrast with interactional types such as dialogue, drama, and conversation. Little work has yet been done on the identification of text types in older languages; however, there is reason to believe that these may differ from those posited for contemporary languages, and may vary across languages and even within a single language according to ideological and cultural change over time (as, for example, the status of religious texts in many European nations; Gregersen and Pedersen discuss this point further). Accordingly, the chapters in this volume take the perspective that text types and their associated genres must be allowed to emerge inductively from the available linguistic and socio-historical evidence, rather than imposed from above. However, while the possible values for 'genre' might be considered to constitute an open set, to which the investigation of older languages would lead us to add members such as 'epics', 'charters', and the like, the 'text type' parameter presumably has a finite number of possible realizations within any given language — most typologies posit between four and six basic types (Virtanen 1992).

Several chapters in this volume identify text type as a parameter that contributes significantly to variation in older languages. **Van Reenen and Schøsler** propose that a prerequisite for a correct understanding of seven Old French particles is an examination of the pragmatic structure of Old French (12th to 14th century), in particular with respect to the distinction between 'topic continuity' and 'topic switch' across adjacent clauses. The pragmatic structure of Old French, in turn, depends on "textual genres", three of which are considered in van Reenen and Schøsler's investigation: (1) direct discourse (dialogue), (2) charters (exposition), and (3) narration in fictional texts. Direct discourse is characterized by frequent occurrences of 'topic switch' that are often unpredictable and implicit. By contrast, charters are characterized by 'topic continuity', and consequently by predictable pragmatic structures that are explicit. Finally, narration is characterized by a mixture of 'topic continuity' and 'topic switch' that is not always predictable and not always explicit. These textual characteristics influence the frequency and syntactic patterns of use of the particles in question. As van Reenen and Schøsler put it, "[t]he fundamental pragmatic choice with respect to the linguistic means of expressing the message is that of the appropriate textual genre: monologue or dialogue, narrative or not".

The second level that determines particle choice concerns thematic structure. The authors undertake a statistical analysis of a large corpus of Old French, balanced according to text type, which reveals four of the seven particles to be explicit markers of thematic structure, according to two dimensions: +/-opposition of the Theme, and +/-opposition of the Rheme. Once these pragmatic choices are made, syntactic restrictions concerning word order and presence or absence of the subject follow more or less automatically. When the particle APRES is chosen, chronological variation in the word order can also be observed, with the original word order *Particle Verb (Subject)* attested in early Old French, and the "modern" word order *Particle Subject Verb* found predominantly in the later Old French corpus. Other particles exhibiting chronological variation are AINZ, which is receding during the 14th century, and LORS, the use of which increases in the later texts. However, the authors find no clear tendencies regarding dialectal variation.

By taking into account these various parameters, van Reenen and Schøsler identify properties of Old French particles which previously had gone unnoticed. Their investigation further suggests that for any given analysis, the influence of different textual parameters can be ranked: in the case of Old French, text type is more important for an understanding of the particles than are chronological or dialectal variation.

**Bolkestein** examines factors that determine the distribution of the pronominal variants available for anaphora in classical Latin in two types of texts, Caesar's historiographical narrative and Cicero's informal correspondence. Anaphora in Latin is indicated by zero anaphora, the pronoun *is*, the demonstrative pronouns *hic*, *iste*, *ille* and the independent relative pronoun *qui*. In the case of the demonstrative pronouns, referential function is traditionally derived from their spatial deixis. However, as Bolkestein points out, a simple extension of the spatial parameter is not satisfactory and cannot predict which anaphoric pronoun will be selected under which circumstances.

Bolkestein therefore investigates the possibility of applying Givón's (1983) model of anaphoric reference to classical Latin. According to that model, there is a direct correspondence between choice of formal coding and degree of accessibility of the referent, with more accessible referents being coded by weaker, and less accessible referents by stronger, means of expression, accessibility being assumed to operate linearly, on the basis of frequency of mention and distance in the text since the last mention. Bolkestein shows that in classical Latin texts, however, non-linear factors must be taken into

account. Specifically, the acceptability of different forms of reference depends on a heterogeneous set of factors including text type, position, information status of the referent, switches in syntactic and semantic function of the referent, and hierarchical discourse structure.

The selection of anaphoric pronouns is not evenly distributed across different text types, but is characteristic of a specific type, as evidenced by a comparison of the private correspondence and the historical narrative. The thematic structure of the letters differs from that of Caesar's narrative in that each recognizable sub-unit of the letters tends to treat a subject matter that differs from the preceding one, whereas the narrative shows a "tail-head linking", or thematic continuity, between sub-units. The high frequency of *ille* in the letters and its almost total absence in the narrative is shown to be dependent on the difference in thematic structure in the two text types. Bolkestein relates the different positional tendencies of *is* and *hic* in the two text types to information structure, especially to the distinctions of focus and topic.

With respect to anaphoric distance and discourse hierarchical distinctions, all — even the two "weakest" forms of anaphora coding, zero anaphora and independent relative pronouns — allow their antecedents to be more than one clause away, provided certain conditions are satisfied with respect to the nature of the intervening material and the contextual status of the host clause of the anaphor. Bolkestein devotes particular attention to less obvious "breaks", such as a change in the speaker's "stance" (is the speaker narrating? is the speaker commenting or explaining?) which may affect the selection of referring expression. In this way, Bolkestein shows that the variation in means of anaphoric expression in Latin is functionally motivated, and provides clues which allow the analyst to predict which anaphor will most probably appear in which type of context, as determined by discourse factors including, but not restricted to, text type.

Text type figures importantly in Brinton's diachronic analysis of the English word *anon*, which from its Old English source meaning 'in one X' evolved to an adverb meaning 'at once' in Middle English, to the meaning 'soon, coming' in Early Modern English. Brinton analyzes this evolution as a case of grammaticalization, questioning whether it follows the directionality of change proposed by Traugott's (1982) three-step model which posits a linear evolution from *propositional* to *textual* to *interpersonal* meaning.

*Anon* has many characteristics of a verbal filler or a metrical expedient, providing extra syllables for a line or increased flexibility in rhyming. How-

ever, *anon* also occurs quite often in prose work: a non-metrical description is, therefore, needed. Brinton studies the use of Middle English *anon* in two narratives, one verse and one prose, and finds that it shows significant variation according to what she calls "discourse type": narrative or dialogue. In narration — verse as well as prose — *anon* is a "peak marker", signaling "discourse-semantic prominence" or foreground, the latter being defined by the notions of temporal sequence, causal importance, human importance, thematic importance, evaluative marking, cognitive accessibility, and figurative importance. *Anon* thus shifts from a lexical form with propositional meaning to a pragmatic marker with textual meaning, in keeping with the first two stages of Traugott's model.

*Anon* also goes on to develop interpersonal meanings in Early Modern English — notably, those of 'coming, presently' as spoken by a servant to his or her master, and the (now lost) meaning of 'what did you mean/say?' However, Brinton argues convincingly that this process cannot be explained as a simple linear continuation of Traugott's model, for two reasons. First, the interpersonal meaning of 'coming' must be derived directly from the propositional meaning 'soon', rather than from the textual meaning 'foreground'. Second, the interpersonal meanings are evident in, and could only have arisen in, dialogue — as, indeed, the textual meaning could only have arisen in narrative. This leads Brinton to propose a refinement to Traugott's model of grammaticalization which allows textual and interpersonal meanings to develop independently in "grammaticalization chains", rather than as necessary components of a single, linear evolution.

The case of *anon* illustrates two important points for the textual parameters approach. The first is that differences in text type condition synchronic variation, as shown also in the chapters by van Reenen & Schöslér and Bolkestein. Second, it provides empirical evidence that diachronic change can operate independently in different text types, thereby problematizing any study of language change that does not control for text type.

### 3.4 Poeticity

Many older texts are in whole or in part poetic. In the strictest sense, 'poeticity' refers to the degree to which a text follows formal poetic conventions, as opposed to being produced in prose without regard for such conventions. Formal poetic conventions include rules of versification involving metre,



rhyme, alliteration, repetition and/or structural parallelism, as well as the setting off of lines of text as stanzas or other sub-units. A text may also be considered 'poetic' if it makes use of suggestive or imaginative language, or if it generally foregrounds the aesthetics of language use, rather than the referential functions of language (Jakobsen 1960). The poeticality parameter may be conceptualized as a binary distinction (poetry vs. prose) or as a continuum, e.g. if the foregrounding of language aesthetics is considered to be a relative, scalar notion. The continuum approach is especially useful in analyzing non-prototypical phenomena such as 'prose poems' (cf. German *Kunstprosa*) and referential works composed in verse (see e.g. Herring, This volume).

Although these definitions are straightforward, in practice, it is difficult to separate out the notion of 'poeticality' from that of 'text type' and 'genre', since poetic and prose styles can express a variety of different text types and genres. Especially confusing for purposes of classification is the situation in which one or the other style becomes culturally associated with a single genre, thereby creating a *de facto* identity between the two phenomena, while the other style asymmetrically expresses a range of genres. Thus in most contemporary occidental languages (such as English), poetry is largely restricted to a single aesthetic, expressive genre and prose is the default for all other communicative purposes, while in other cultures and other historical periods, the reverse situation may obtain: in the classical Tamil period, for instance, poetry was the default and prose was restricted in its communicative functions to recordings of land grants. For this reason, the chapters in this volume that address the poeticality parameter also consider distinctions of genre.

Hock raises the issue of poeticality as a methodological prerequisite for the analysis of classical Indo-European languages. Many early IE texts were composed in verse and obey metrical constraints. However, poetic genres have been claimed to be unsuitable for syntactic research in that they tend to be characterized by 'poetic license', resulting in highly marked structures 'metri causa'. This problem has potentially serious methodological consequences for the study of historical syntax.

Hock argues that poetic genres should be included as data for analysis in order to gain the widest possible picture of the range of grammatical constructions available. 'Poetic license' is not a significant problem for syntactic analysis, he claims, because most poetic effects are limited to phonology and morphology. Moreover, phonological variation in poetry can contribute insights into the relationship between prosody and syntax, in that poetry im-

poses clearly visible prosodies of its own (such as line breaks and caesuras), unlike ancient prose texts which normally provide no clues to prosodic structure. This latter claim is illustrated with examples of prosodically-sensitive syntactic constructions in Old English and Vedic Sanskrit.

Having established the desirability of analyzing poetic as well as prose texts, Hock then goes on to consider the role of genre in early IE syntax. Rather than elevating one genre above another as more 'natural', he advocates the examination of a variety of genres and sensitivity to genre differences. Nowhere is the need for such sensitivity more apparent than in cases where different chronological stages of a language are attested in different genres. One example of this is Sanskrit word order, where what has previously been claimed to be a diachronic shift from more to less flexible OV order can be seen in large part as an artifact of genre differences, e.g. in Vedic hymns (more flexible) vs. didactic prose (less flexible). The danger of confusing diachronic with genre effects is further illustrated with a discussion of Sanskrit appositive relative clauses.

In the final section of his chapter, Hock puts into practice the genre-sensitive mode of analysis he advocates. Revisiting the claim that V1 word order is a dominant feature of IE narrative genres (cf. e.g. Dressler 1969), he demonstrates that simple V1 is in fact not very common in most languages. Rather, 'linkage strings' containing covered finite verbs (structures with near-initial verb) or no finite verb are more characteristic of early narrative. In addition, individual languages and particular narrative sub-genres within them differ in terms of their favorite narrative linking device and the extent to which they use it. These claims are supported with statistical data from samples drawn from three poetic Sanskrit genres, the *Iliad*, and *Beowulf*. Thus in addition to raising and illustrating a number of important genre effects in early IE syntax, Hock advances a new analysis of 'narrative V1' which identifies sub-genre as well as genre differences.

In contrast with Hock's view that the poetry vs. prose parameter is of minor methodological significance for the analysis of syntactic phenomena in early Indo-European, Herring finds that poeticality affects word order in Old Tamil, a Dravidian language (and neighbor of Sanskrit) in ancient South Asia. Almost all extant texts in Old Tamil (2nd c. B.C. - 6th c. A.D.) were composed in verse, with the exception of a small number of land grant inscriptions, most of which are fragments. However, since not all verse texts are equally poetic, Herring proposes that poeticality in Old Tamil be considered a relative dimen-



sion, rather than a binary distinction between verse and prose. She applies this notion to a corpus of the oldest available representatives of different genres including land grant inscriptions, grammatical description, epic narrative, heroic poetry, and love poetry, arranging them along a 'poeticity continuum' according to the degree to which their purpose can be considered to be primarily informative or primarily aesthetic.

Based on this corpus, Herring shows that there is a correlation between *degree of poeticity* and the rigidity of SOV constituent order in independent clauses. Verse texts in Old Tamil are less verb-final than prose texts, and verse texts whose purpose is primarily aesthetic show greater word-order variation than more informative genres such as epic narrative and treatises on grammar. A further result is that represented dialogue in verse epics, which exhibits independent properties of colloquial speech, is strongly verb-final.

These findings shed light on an ongoing debate among Dravidian linguists as to the basic word order of Old Tamil. Was Old Tamil SOV, like Modern Tamil (Zvelebil 1989), or was it relatively free in its constituent order (Andronov 1991)? Herring argues that verb-finality was the norm in ancient colloquial Tamil, but that this norm was violated in poetry for aesthetic effect. The verb-finality of Old Tamil is further evidenced in the language's preference for head-final constituent order, the predominance of SOV-compatible orders in finite clauses, and in the fact that subordinate clauses are invariably verb-final in texts of all types.

Thus in addition to shedding empirical light on a disputed question in historical Tamil syntax, Herring's study illustrates the importance of distinguishing among genres of poetry and between ordinary and poetic language more generally in analyzing the word order of older languages.

Iwasaki's chapter is a further illustration of differences in language structure and use in poetic and prose texts. His focus is the finite use of forms whose primary function is as markers of attribution, nominalization, and subordination in Late Old Japanese (9th-12th centuries A.D.). A form of this sort, the Final-Attributive, alternated with a finite Conclusive form; Iwasaki's goal is to explain this variation.

On the basis of an examination of two Late Old Japanese literary texts — one a multi-episodic prose narrative, and the other an anthology of short poems — Iwasaki identifies several distinct uses of the Final-Attributive, the distribution of which is conditioned by textual "genre". In prose narration, Final-Attributive sentences encode background information, in contrast with

Conclusive sentences, which encode the main narrative event line, or foreground information. Moreover, the two sentence types tend to occur in complementary distribution within a narrative episode, with Conclusive sentences clustering at the beginning, and Final-Attributive sentences immediately prior to the introduction of an embedded poem, where they serve a cataphoric function. In poetry, in contrast, the Final-Attributive is used to express exclamation ("non-reflective consciousness") and weak conjecture.

Iwasaki derives the functions of the Final-Attributive from a unified underlying meaning of "suppressed assertion", which he relates to the basic modifier function of the attributive form via the notion of *incompleteness*. Incompleteness creates discourse dependencies (background in relation to foreground in narrative) and makes available inferences about the speaker's affective and cognitive state (incomplete cognition leads to exclamation and weak conjecture in expressive poetry). In contrast, the Conclusive form encodes complete assertions.

Over time, the Final-Attributive form came to replace the Conclusive form in all its uses, in part, Iwasaki suggests, due to the overlap in their distribution caused by the sentence-final uses of the Attributive form. The original functions of the Final-Attributive are preserved in different attributive forms in Modern Japanese: the nominalized predicate *no (da)* has transferred the functions of the FA in prose, while the exclamatory function has been transferred to the "thetic judgment" sentence type with the case particle *ga*, which encodes "a direct response to the perceptual intake of an actual situation", e.g. *inu ga hasitteiru* 'A dog is running'. Thus different functions of the Final-Attributive that originally arose from a common source eventually came to be encoded by unrelated forms, a testimony to the extent to which they had undergone independent evolution in poetry as compared to (narrative) prose.

This diachronic development recalls Brinton's proposal that the meaning of English *anon* evolved independently in different text types. However, while Brinton found no effect of poeticity in her data, Iwasaki's analysis shows that "poetic" notions such as expressive exclamation can become part of a construction's meaning and influence its subsequent evolution.

### 3.5 Orality

A fundamental problem for the study of languages of the past is the degree of 'orality' of a given text, that is, the extent to which it reflects the properties of

the spoken language at the time.<sup>11</sup> Oral features may be present if the text was originally produced in the spoken modality and subsequently transcribed, or they may be introduced into a text produced in writing, either unconsciously (e.g. as a reflex of the writer's spoken language practices) or deliberately (e.g. in scripted dialogue).

The parameter of orality is important in its own right, in that it can shape language use in particular, identifiable ways (cf. Benveniste 1966, 1970; Weinreich 1973). Equally or more often, however, it is linked to the broader question of textual authenticity. While contemporary linguistics privileges the spoken language as a more basic and authentic object of study than written language, historical linguists have traditionally been constrained to work exclusively with written texts.<sup>12</sup> In order to discover the most authentic — that is, the most “spoken-like” — textual data available in a text language, e.g. so as to be able to make statements concerning the colloquial, or “ordinary” language of a remote time, scholars of older languages must make use of indirect procedures for discovery. This necessity raises a particular set of issues that are addressed in several chapters in this volume.

**Kytö's** chapter raises the problem of variation and differences between spoken and written language. Given that we have only written records of older languages, how faithfully and accurately can speech-based records reflect the spoken language of the past? Kytö attacks the problem with respect to Early Modern English from two synchronic angles: (1) by analyzing the language found in manuscripts written by the merchant Robert Keayne (1588–1656) which record sermons and conversations following the sermons at meetings of the First Church of Boston, and (2) by comparing the distribution of the linguistic and textual features found in Keayne's notes to those in other contemporary texts in the American colonies.

Two of Keayne's three extant American notebooks cover a period from 1639 to 1646 and contain notes of conversations that took place during church meetings. The first impression is that the notebooks offer a fairly reliable record of spontaneous spoken language of the past, recorded in a semi-formal discourse situation. But did Keayne record what was said verbatim at the meeting-house, did he work from preliminary notes, or did he simply write down afterwards what he had heard as best he could remember?

The external evidence suggests that Keayne did not produce his notes straight from the mouths of the speakers, but rather filled in his notebooks only after the actual speech situation. Kytö draws attention to the fact that the text is

carefully and conscientiously arranged on the pages of the notebooks. Moreover, many of the corrections made by Keayne reveal a process of compilation, suggesting that Keayne followed a preliminary version of the text when entering the final version in his notebooks.

In order to assess the extent to which Keayne's notes reflect spoken language, Kytö undertakes a large-scale statistical comparison involving a computerized corpus of different genres in Early Modern English. The corpus is composed of “speech-based” texts such as trial proceedings, depositions, and sermons, and of “non-speech-based” texts such as diaries, correspondence, law texts and historiography. The investigation follows the multi-dimensional model of Biber and Finegan (1989) in using factor analysis to identify “dimensions” underlying the co-occurrence patterns of 27 linguistic features in each text. Factor scores are then used to compare Keayne's transcriptions to other texts in the corpus.

Keayne's notes of both sermons and church meetings approach other trial records in some respects but remain different in others, towards the more ‘written’ end of the textual continuum. This internal evidence confirms the results of the external evidence, in that it reveals “scribal interference”. At the same time, Keayne's notes of sermons differ from his proceedings of church meetings, suggesting that something of the genre conventions and speakers' idiosyncrasies filtered through. There is also variation in the individual speaker profiles studied. On the basis of this evidence, Kytö concludes that Keayne probably employed a note-taking method, rather than verbatim transcription. Thus his language use cannot be taken as a direct representation of spoken Early Modern English.

**Joseph** also takes on the authenticity question in his analysis of three Medieval Greek (13th to 16th c.) infinitive constructions: (1) an adverbial usage known as the ‘Circumstantial Infinitive’, (2) a periphrastic future with *thelo*: as auxiliary, and (3) a periphrastic perfect with *ekho*:. The somewhat restricted Medieval Greek infinitive, a continuation of the Ancient Greek infinitive, had been replaced by finite verb forms in virtually all its uses by the 17th century, and Modern Greek has no infinitive. Thus deciding the authenticity of these constructions can help to settle the question of the status of the infinitive in Medieval Greek, and date the loss of the infinitive from the language.

Joseph's definition of ‘authenticity’ differs somewhat from that of other authors in the volume: an authentic form is one whose use corresponds to

some part of the *grammar internalized by native speakers of the language*, as opposed to being 'inauthentic' — having no basis in actual usage, but rather being an artificial aspect of the language of a given text or register (e.g. an error, an invention, a literary convention). While in most cases authentic language is found in actual spoken language usage, this need not be the case; Joseph cites the example of 'quotative inversion' in Modern English as an authentic usage limited to a written, literary register.

With regard to the "suspect" Medieval Greek infinitives, Joseph considers a range of evidence bearing on their authenticity that includes their frequency and systematicity of use, their distribution across textual "genres" (with greater weight being assigned to the evidence of prose texts over poetic ones, and to quoted direct speech over narration), statements by contemporary grammarians, and the continuation or lack of continuation of the constructions in later stages of the language. All three infinitives are frequent, systematic in their form and syntax, and are amply represented in a wide variety of Medieval Greek literature, as well as being mentioned in contemporary grammars. However, the first construction, the Circumstantial Infinitive, has no exact counterpart in Ancient Greek and no continuation in Modern Greek. Joseph posits that it was restricted to written occurrence only, i.e. as a literary construction similar to quotative inversion in English. As for the second and third constructions, their authenticity is corroborated by modern Greek evidence and by the naturalness of the language changes involved.

Joseph interprets this collective evidence as showing that the infinitive was a vital grammatical category in Medieval Greek. More generally, his chapter demonstrates the range of methodologies one can — and in some cases, must — employ to decide matters of authenticity in text languages.

Gvozdanović's chapter presents a different perspective on the orality question. She investigates changes in punctuation habits from medieval Russian (13th century) up to the modern Russian norm established in the 18th century, hypothesizing that punctuation units in older Russian reflect units of spoken discourse production similar to what Chafe (1994) calls 'information units'. In medieval Russian, the structure of such units depends on text pragmatics, whereas modern punctuation involves a more standardized practice based on division of a text into structural (syntactic) units.

An investigation of two original secular prose manuscripts attests different stages of punctuation in the evolution towards modern Russian punctuation. Punctuation in the 13th c. *First Chronicle of Novgorod* uses periods to

separate phrases and clauses which convey the idea of unrelated events/states or referents, afterthoughts, and pieces of information which are salient, even if related to the preceding information. Such units average just over three words in length — as Gvozdanović points out, well within the limits of short-term memory, and within the range of intonation units in speech — and generally obey the pragmatic principle of "one chunk of homogeneous new information" per unit. The punctuation of the 16th c. *Domostroj*, in contrast, consists of both commas and periods. Commas separate information units which may consist of single or coordinated complement phrases, appositional phrases, or clauses or sequences of clauses, provided they are united by "sameness of implicature". Periods are used for indicating topic discontinuity. Overall, punctuation units in the 16th c. text are considerably longer than those in the earlier text — nearly 10 words in length — and are more likely to be syntactically-complete units, comprising subordinations and conjunctions of clauses.

Modern punctuation rules for Russian are similar to those for other European languages in that they mark off sentential boundaries and within them clauses and phrases of unequal syntactic and/or pragmatic status. The modern system is thus argued to be the result of a continuous development from predominantly speech-based to longer, more complex units, and of the increased need for syntax as a basis for processing the latter. In this respect, Gvozdanović's analysis spotlights orality in Old Russian not as an indicator of 'textual authenticity', but as a source of linguistic structure in its own right.

### 3.6 *Dialect*

Older texts were often produced in languages for which there was no single written standard norm. The term 'dialect' is used broadly here to refer to both regional and social varieties of language. Regional dialects are defined by geographical location, including whether speakers of the variety in question live in a rural or urban area (Kurath 1972), and social dialects are defined in terms of the age, sex, socio-economic class, etc. of groups of speakers (Labov 1966). Each of these factors has the potential to condition language use, and where sufficient evidence is available, should be considered as separate parameters in historical linguistic analysis.<sup>13</sup> However, since only two chapters in this volume address dialect issues, we group them together here for purposes of presentation.



Related to dialect is the question of contact across languages or language varieties. Language contact has potential consequences for the textual parameters approach in several respects. 1) Portions of texts or entire texts may have been translated from another language, compromising the text's authenticity. 2) A linguistic structure under investigation may have been introduced through contact with another language, suggesting an external, rather than an internal, explanation for its origin and patterns of diffusion. 3) Different dialects may have been differentially affected by contact with other varieties (Weinreich 1964), necessitating that texts produced in those dialects be analyzed separately from those in other varieties, even though they are nominally from the "same" language.

An example of the third type of situation is investigated in the chapter by **Kroch, Taylor and Ringe**. The northern and southern dialects of Middle English differed in the way that they implemented the verb-second (V2) constraint common to the Germanic languages. Kroch, Taylor and Ringe propose that this difference was a syntactic consequence of contact-induced simplification in the verbal agreement paradigm of the northern dialect.

Two different implementations of the verb-second constraint are found in the Germanic language family. The first is that of German, Dutch and Mainland Scandinavian, which move the tensed verb to the COMP position (abbreviated CP-V2). The second is that of Yiddish and Icelandic, which move the tensed verb to a lower position, named INFL in Chomsky 1986 (abbreviated IP-V2). All V2 languages exhibit verb-second word order in main clauses. The difference between the two implementations is found in subordinate clauses, as the former subtype of Germanic languages only allows verb-second order in embedded clauses that in some way have the structure of matrix clauses. In contrast, the latter subtype allows verb-second word order in a broad range of subordinate clauses. The southern dialect of Middle English appears to have developed into the IP-V2 type, in contrast to the northern dialect of Middle English which, like Mainland Scandinavian languages, has developed into a CP-V2 type. Kroch et al. attribute the pattern in the northern dialect to the influence of Old Norse, and specifically, to the Viking invasions and subsequent settlement of Northumbria, Lincolnshire and East Anglia in the 8th through 11th centuries.

Interestingly, however, Old Norse itself was most probably not a CP-V2 type, but an IP-V2 language like Modern Icelandic. The authors propose that the transition from IP-V2 to CP-V2 in the north was an effect of substratum

influence and of imperfect learning of English by the substantial number of Scandinavian immigrants in this area. The borrowing of not only lexical, but also grammatical (e.g. function word) elements from Scandinavian into Northern Middle English is further evidence of the intimacy of the contact that must have existed between the native English and the invading Scandinavians. Quantitative evidence of word order patterns in text excerpts from the Penn-Helsinki Parsed Corpus of Middle English (1994) is presented as evidence of CP-V2 order in the north.

Finally, Kroch et al. propose that the eventual loss of V2 syntax in Modern English is the result of competition between the grammars of the northern and southern dialects in the speech of people who had been exposed to both systems. Thus dialect differences in V2 type would have contributed to undermining the V2 system more generally, a finding which illustrates the importance of taking dialect into account when analyzing diachronic developments in syntax.

**Gregersen and Pedersen** also investigate a problem of word order variation and change. In Modern Danish main clauses, a sentence adverbial follows the finite verb which directly follows the pronominal subject, whereas in subordinate clauses, a sentence adverbial follows the subject and thus precedes the finite verb. The subordinate word order is not original in Danish but rather has evolved to become a criterion for subordinate clauses. Furthermore, the subordinate word order distinguishes modern written Danish from modern spoken Danish in that some clauses have main clause word order in the spoken language, whereas this is not normally the case in written language. Where did the subordinate clause order come from, and how did it develop?

In an attempt to answer these questions, Gregersen and Pedersen undertake a diachronic study of a corpus of representative Danish texts from different periods — texts that represent the "dominant" text types of the culture at the time. Their approach constitutes a methodological departure from that advocated in other papers in the volume, in that they argue in favor of holding the "socio-psychological" status of the text, rather than the type of text, constant. Their corpus nicely demonstrates that between the 16th and early 20th centuries, the old (main clause) word order in subordinate clauses steadily lost ground to the new word order, the former increasingly coming to be seen as a spoken language feature. Conversely, the authors argue that the subordinate clause word order (SCWO) was from the outset a written lan-



guage phenomenon, innovated as a formal means of distinguishing main from subordinate clauses on the model of other literary languages such as Latin and German. As such, it would have been reinforced through literacy instruction in schools, as the written dialect of Danish was elevated to the status of the national standard language.

Synchronic support for this analysis is adduced from two dialect studies of modern spoken Danish carried out by the authors. The first, a study based on sociolinguistic interviews with natives of Copenhagen, reveals that use of the two word orders is variable in spoken Danish, although the variation does not appear to correlate with speaker age, sex, or social class. However, the syntactic type of the subordinate clause conditions the choice of word order, in a distribution which preserves the pattern of the older written texts. The second study, a regional dialect survey, shows that the MCWO is most frequent and is attested in the greatest range of clause types in western Denmark, the geographical region farthest from urban Copenhagen, and hence the most conservative. Thus modern spoken Danish dialects can be seen to preserve evidence of the past status of the MCWO.

By making use of both historical and synchronic sociolinguistic methods, Gregersen and Pedersen add considerable detail to what they term the "still unfinished picture" of the use of subordination signals in Danish. Their eclectic approach embraces a variety of textual parameters, including orality, dialect, cultural status and diachrony, and demonstrates the interactions among them in a well-documented case study. As such, their chapter stands as a fitting conclusion to the volume as a whole.

#### 4. Directions for Future Research

Taken together, the papers in this volume show that older languages attested in written records are variable, and that aspects of that variation are systematically conditioned by textual parameters. These facts have implications for how research into "text languages" should be carried out in the future.

First and foremost, more studies are needed which identify the effects of textual parameters on linguistic variation. Such studies could conceivably reanalyze previously analyzed historical linguistic data for which no textual or contextual distinctions were drawn, asking new questions of old data. The dimensions of variation investigated could include the ones mentioned in this

introduction — including those not addressed in the present volume, such as writer demographics and translation effects — as well as additional parameters yet to be discovered. For example, diglossia and register phenomena such as level of formality are likely candidates to be added to the list of parameters relevant to the description of older languages. We hope that the investigations of parameters in this volume will invite other linguists to detect supplementary parameters.

This in turn raises a second issue: what, if anything, constrains the identification of textual parameters? What do the various parameters have in common, and how do they differ, e.g. in their effects on language structure and use? In this early stage of research, there is as yet no theory of textual parameters *per se*; the study of textual parameters is still a largely exploratory, inductive enterprise. So far, it has produced a list of contextual and co-textual factors that potentially exercise a significant influence on linguistic phenomena, and that historical linguists would thus be well advised to take into consideration in their research. In this sense, it is not unlike the list of situational factors identified for research into the ethnography of speaking by Hymes (1974), commonly referred to by the acronym SPEAKING (setting, participants, event structure, acts, key, instrumentality, norms, and genre). Like Hymes' list, the current list of textual parameters has practical consequences for how research should be conducted, but both are unordered, diverse in what they include, and largely pre-theoretical constructs. What is needed, in the textual parameters approach as well as in the ethnography of speaking, is a *theory of context*, a principled account that allows for predictions about what kinds of situational factors make a linguistic difference; when, how, and to what degree; and how the factors interrelate. The papers in the present volume contribute towards answering these questions with data from specific cases. Future research into textual parameters should address the need for an integrated theory, with particular reference to the situational contexts of written, older texts.

The findings reported in this volume also have implications for *methodological practice* in historical linguistics. Two areas of practice that are especially impacted if one adopts a variation-sensitive approach are data sampling and interpretation of research results. Regarding sampling, in keeping with the tenets of modern sociolinguistic studies of variation in speech, the researcher might seek to control for those variables (parameters) hypothesized to influence linguistic choice by selecting an equal number of texts of different

'types' (broadly defined) and subjecting data from those types to systematic comparison. This approach is feasible for older language varieties for which many texts, and many varieties of text, are available to choose from. The larger the sample, the more reliable the patterns found, and the more likely it is that tendencies or less frequent patterns will be uncovered.

The management of large amounts of textual data is facilitated by new tools of research in computer science. Several of the chapters in this volume made use of computer-assisted corpus analysis, and those by Kroch et al., Kytö, and van Reenen & Schöslér benefited from the availability of large computerized corpora; future research can be expected to make more extensive use of computing resources. Future generations of historical linguists also stand to benefit from the increased availability of sources of data enabled by new recording and communications technologies. Already linguists have begun analyzing old recordings of speech, some of which date back one hundred years (Bailey et al. 1991; James et al. 1999); as the time depth of such records continues to deepen, their historical value will increase accordingly. Most recently, typed computer-mediated communication such as takes place on the Internet has started to be analyzed for evidence of linguistic change in progress (Herring 1998, 1999). Internet data have the advantage of being archived automatically as they are produced, making available potentially vast databases from which linguists can extract samples of language in use.

The problem remains, however, that for most older language varieties, the available textual data are limited. This makes the ideal of balanced sampling difficult or unattainable in many historical linguistics studies. The challenge to researchers then becomes one of responsible interpretation of the research findings in light of the parametric properties of the data sample. That is, researchers must account for the effects of textual parameters, even if this limits them to making qualified claims about subsets of data, rather than generalizations about the language as a whole.<sup>14</sup> Ultimately, the researcher may be led to a rather modest position vis-à-vis the possibility of unveiling the grammar of an older language in all its details. What is possible, however, by means of thorough and systematic empirical investigation, is to strive for the identification of classificatory patterns that provide homogeneous subsets of data. When such homogeneous subsets are found, the researcher is entitled to suppose them to reflect genuine distinctions of the old language.

It is our hope that the ways in which we have presented our ideas and the results of the case studies described in this volume will stimulate further study

along similar lines, contributing to the study of textual parameters in older languages in both theoretical and methodological domains.

## Notes

1. Cf. the technical definition of a parameter as "a variable or arbitrary constraint appearing in a mathematical expression each value of which restricts or determines the specific form of the expression" (*The American Heritage Dictionary*, New Second College Edition, 1983).
2. See, however, Koch (1995).
3. This difficulty is relative. Epistolary correspondence tends to preserve indications of the identity of the writer, the recipient, and the social relations between them, making limited socio-historical analysis of such texts possible (Nevalainen & Raumolin-Brunberg 1996). In non-epistolary genres, social relations can sometimes be inferred from forms of address in direct discourse.
4. A similar observation was made 100 years ago as a criticism of the Neogrammarians by specialists in dialectology such as Schuchart (1885) and Gilliéron (1917), but the point was swept aside in the wake of Saussure's ideas on the strict separation of synchronic and diachronic linguistics (Reynolds 1994). See also Fleischman (1996).
5. But cf. Ball (1995), Kroch (1989), and Pintzuk (1995).
6. But cf. work in what is sometimes termed 'socio-historical linguistics', e.g. Biber & Finegan (1989, 1992); Nevalainen & Raumolin-Brunberg (1996); Romaine (1982).
7. A good example is Fleischman's (1986) analysis of 'overlay structures' in Medieval French. Notable exceptions are Biber and Finegan (1989, 1992) and Taavitsainen (1993, 1997), who focus squarely on the role played by textual genre in the evolution of Modern English.
8. An additional paper was originally solicited from Charles Li which analyzed Late Archaic Chinese, but unfortunately it could not be included in the volume because of length considerations. The paper has since appeared as Li (1996).
9. Indologists commonly date the Vedic hymns to ca. 1500 B.C., the time when the Aryans are thought to have arrived in India (Masica 1991). However, the manuscripts that preserve the hymns were not written down until the 10th C. A.D. or later (Hock, personal communication). The oldest written manuscripts analyzed in the present volume date from the 2nd C. B.C., and are in Old Tamil.
10. But cf. Banniard (1989) and Wright (1982), who consider the Romance languages to be modern continuations of Latin.
11. See e.g. Koch (1995), Bakker & Kahane (1997), Jucker et al. (1999).
12. More recently, linguists have begun to have access to voice recordings of earlier stages of

modern languages. For an interesting discussion of the implications of a set of early wire recordings of U.S. former slaves for questions relating to the historical development of African American Vernacular English, see Bailey, Maynor and Cukor-Avila (1991). However, such recordings are limited in their availability, and raise some of the same problems of representativeness and lack of social context as older written documents (Rickford 1991).

13. See, for example, a recent body of work that considers the parameter of social class in analyzing texts from 19th century Europe (Mihm 1998; Vandebussche 1999).
14. Alternatively, the researcher must present an argument that in a given analysis, a parameter or parameters that might otherwise be expected to exercise an influence (e.g. the fact that the data are poetic, cf. Hock's chapter in this volume) are not relevant to the analysis at hand.

## References

- Andronov, M. S. 1991. "Word Order: Causality and Relations". *Studies in Dravidian and General Linguistics: A Festschrift for Bh. Krishnamurti*, ed. by B. Lakshmi Bai & B. Ramakrishna Reddy, 91–97. Hyderabad, India: Osmania University.
- Bailey, Guy, Natalie Maynor & Patricia Cukor-Avila, eds. 1991. *The Emergence of Black English*. Amsterdam & Philadelphia: John Benjamins.
- Bakker, Egbert & Ahuvia Kahane, eds. 1997. *Written Voices, Spoken Signs: Tradition, Performance and the Epic Text*. Cambridge, MA: Harvard University Press.
- Ball, Catherine N. 1996. "A Diachronic Study of the Relative Markers in Spoken and Written English". *Language Variation and Change* 8(2), 227–258.
- Banniard, Michel. 1989. *Viva Voce. Communication écrite et communication orale du IV<sup>e</sup> au IX<sup>e</sup> siècle en Occident Latin*. Paris: Etudes Augustiniennes.
- Benskin, Michael & Margaret Laing. 1981. "Translations and Mischsprachen in Middle English manuscripts". *So Many People, Longages and Tonges. Philological Essays in Scots and Mediaeval English Presented to Angus McIntosh*, ed. by M. Benskin & M.L. Samuels, 55–106. Edinburgh: The Editors.
- Benveniste, Emile. 1966, 1970. *Problèmes de linguistique générale* 1–2. Paris: Gallimard.
- Bergner, Heinz. 1995. "The Openness of Medieval Texts". In Jucker, ed., 1995, 37–54.
- Biber, Douglas. 1988. *Variation Across Speaking and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas & Edward Finegan. 1989. "Drift and the Evolution of English Style: A History of Three Genres". *Language* 65, 487–517.
- Biber, Douglas & Edward Finegan. 1992. "The Linguistic Evolution of Five Written and Speech-based English Genres from the 17th to the 20th Centuries". *History of Englishes: New Methods and Interpretations in Historical Linguistics*, ed. by Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen, 688–704. Berlin: Mouton de Gruyter.
- Blank, Andreas & Peter Koch, eds. 1999. *Historical Semantics and Cognition*. [Cognitive Research 13]. Berlin & New York: Mouton de Gruyter.
- Brinton, Laurel. 1996. *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin & New York: Mouton de Gruyter.
- Cerquiglini, Bernard, Jacqueline Cerquiglini, Christiane Marchello-Nizia & Michèle Perret-Minard. 1976. "L'objet 'ancien français' et les conditions propres à sa description linguistique". *Méthodes en grammaire française*, ed. by Jean-Claude Chevalier & Maurice Gross, 185–200. Paris: Klincksieck.
- Claife, Wallace L. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.
- Dressler, Wolfgang U. 1969. "Eine textsyntaktische Regel der indogermanischen Wortstellung". *Zeitschrift für vergleichende Sprachforschung* 83, 1–25.
- Fleischman, Suzanne. 1986. "'Overlay Structures' in the Song of Roland: A Discourse Pragmatic Strategy of Oral Narrative". *Proceedings of the 12th Annual Meeting of the Berkeley Linguistics Society*, 108–123. Berkeley: Berkeley Linguistics Society.
- Fleischman, Suzanne. 1996. "Medieval Vernaculars and the Myth of Monoglossia: A Conspiracy of Linguistics and Philology". *Literary History and the Challenge of Philology*, ed. by Seth Lerer, 92–104, 267–274. Stanford: Stanford University Press.
- Gilliéron, Jules. 1917. *Généalogie des mots qui désignent l'abeille*. Paris: Champion.
- Givón, Talmy, ed. 1983. *Topic Continuity in Discourse*. Amsterdam & Philadelphia: John Benjamins.
- Herring, Susan C. 1998. "Le style du courrier électronique: variabilité et changement." *Terminogramme* 84–85, 9–16.
- Herring, Susan C. 1999. "Actualization of a Counter-change: Contractions on the Internet." Paper presented at the 14th International Conference on Historical Linguistics, Vancouver, Canada.
- Hock, Hans H. 1997. "Chronology or Genre? Problems in Vedic Syntax". *Inside the Texts — Beyond the Texts: New Approaches to the Study of the Vedas*, ed. by Michael Witzel, 103–126. Harvard Oriental Series, Opera Minora, 2. Cambridge, MA: Harvard University.
- Hymes, Dell. 1974. *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.
- Jacobs, Andreas & Andreas Jucker. 1995. "The Historical Perspective in Pragmatics". In Jucker, ed., 1995, 3–33.
- Jakobsen, Roman. 1960. "Linguistics and Poetics". *Style in Language*, ed. by Thomas A. Sebeok, 350–377. Cambridge, MA: MIT Press.
- James, Deborah, Sandra Clarke & Marguerite MacKenzie. 1999. "Dreams and Surprises in Cree/Montagnais/Naskapi: The Semantic History of a Modal Inflection". Paper presented at the 14th International Conference on Historical Linguistics, Vancouver, Canada.
- Jeffers, Robert & Ilse Lehist. 1979. *Principles and Methods for Historical Linguistics*. Cambridge, MA: MIT Press.
- Jucker, Andreas, ed. 1995. *Historical Pragmatics: Pragmatic Developments in the History of English*. Amsterdam & Philadelphia: John Benjamins.
- Jucker, Andreas, Gerd Fritz & Franz Lebsanft, eds. 1999. *Historical Dialogue Analysis*. Amsterdam & Philadelphia: John Benjamins.

- Kayne, Richard S. 1996. "Microparametric Syntax: Some Introductory Remarks". *Microparametric Syntax and Dialect Variation*, ed. by James R. Black & Virginia Motapanyane, ix-xvii. Amsterdam & Philadelphia: John Benjamins.
- Kemenade, Ans van & Nigel Vincent. 1997. "Parameters and Morphosyntactic Change". *Parameters of Morphosyntactic Change*, ed. by Ans van Kemenade & Nigel Vincent, 1-25. Cambridge: Cambridge University Press.
- Koch, Peter. 1995. "Une langue comme toutes les autres: latin vulgaire et traits universels de l'oral". *Latin Vulgaire-Latin tardif IV: Actes du 4e Colloque International sur le Latin vulgaire et tardif*, ed. by Luis Callebat, 125-144. Hildesheim: Georg Olms AG.
- Kroch, Anthony. 1989. "Function and Grammar in the History of English: Periphrastic *do*". *Language Change and Variation*, ed. by Ralph Fasold & Deborah Schiffrin, 133-172. Amsterdam & Philadelphia: John Benjamins.
- Kurath, Hans. 1972. *Studies in Area Linguistics*. Bloomington: Indiana University Press.
- Labov, William. 1963. "Social Motivation of a Sound Change". *Word* 19, 273-309.
- Labov, William. 1966. *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- Lehmann, Winfred. 1973. *Historical Linguistics: An introduction*. 2nd ed. New York: Holt, Rinehart & Winston.
- Li, Charles N. 1996. "A Cryptic Language with a Minimal Grammar: The Confucian Analects of Late Archaic Chinese". *Lexical Structure and Language Use*, ed. by E. Weigand & F. Handsnurscher, 53-118. Tübingen: Max Niemeyer.
- Lightfoot, David. 1991. *How to Set Parameters: Evidence from Language Change*. Cambridge, MA: MIT Press.
- Lightfoot, David. 1999. *The Development of Language: Acquisition, Change, and Evolution*. Malden, MA & Oxford: Blackwell.
- Longacre, Robert E. 1979. "Why We Need a Vertical Revolution in Linguistics". *The Fifth LACUS Forum 1978*, 247-270. Columbia, SC: Hornbeam Press.
- Longacre, Robert E. 1996. *The Grammar of Discourse*. 2nd ed. New York: Plenum Press. (1st ed. 1983)
- Masica, Colin P. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.
- Mihm, A. 1998. "Arbeitersprache und gesprochene Sprache im 19. Jahrhundert". *Sprache und Bürgerliche Nation*, ed. by D. Cherubim, S. Grosse & K.J. Mattheier, 282-316. Berlin & New York: Mouton de Gruyter.
- Nevalainen, Terttu & Helena Raumolin-Brunberg, eds. 1996. *Sociolinguistics and Language History: Studies Based on the Corpus of Early English Correspondence*. Amsterdam: Rodopi.
- Pintzuk, Susan. 1995. "Variation and Change in Old English Clause Structure". *Language Variation and Change* 7(2), 229-260.
- Reynolds, William. 1994. *Variation and Phonological Theory*. Unpublished doctoral dissertation, University of Pennsylvania.
- Rickford, John R. 1991. "Representativeness and Reliability of the Ex-slave Materials, with Special Reference to Wallace Quarterman's Recording and Transcript". In Bailey et al., eds., 1991, 191-212.
- Romaine, Suzanne. 1982. *Socio-Historical Linguistics. Its Status and Methodology*. [Cambridge Studies in Linguistics 34.] Cambridge: Cambridge University Press.
- Schösler, Lene. 1995. "New Methods in Textual Criticism: The Case of the 'Charroi de Nîmes'". *Medieval Dialectology*, ed. by Jacek Fisiak, 225-276. Berlin & New York: Mouton de Gruyter.
- Schuchart, Hugo. 1885. "Über die Lautgesetze: Gegen die Junggrammatiker". Translated and reprinted in Schuchart, the Neogrammarians, and the Transformational Theory of Phonological Change, ed. by T. Vennemann & T. H. Wilbur, 39-72. Frankfurt: Athenäum Verlag.
- Swales, John. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Taavitsainen, Irma. 1993. "Genre/subgenre Styles in Late Middle English". *Early English in the Computer Age: Explorations through the Helsinki Corpus*, ed. by Matti Rissanen, Merja Kytö & Minna Palander-Collin, 171-200. Berlin: Mouton de Gruyter.
- Taavitsainen, Irma. 1997. "Genres and Text Types in Medieval and Renaissance English". *Poetica* 47, 49-62.
- Traugott, Elizabeth Closs. 1982. "From Propositional to Textual and Expressive Meanings: Some Semantic-pragmatic Aspects of Grammaticalization". *Perspectives on Historical Linguistics*, ed. by Winfred P. Lehmann & Yakov Malkiel, 245-271. Amsterdam & Philadelphia: John Benjamins.
- Traugott, Elizabeth Closs & Bernd Heine, eds. 1991. *Grammaticalization*, vols. 1 & 2. Amsterdam & Philadelphia: John Benjamins.
- Vandenbussche, Wim. 1999. "'Arbeitersprache' in Bruges During the 19th Century". *Beiträge zur Historischen Stadtsprachenforschung*, ed. by Helga Bister-Broosen, 49-65. Vienna: Praesens.
- Virtanen, Tuija. 1992. "Issues of Text Typology: Narrative — A 'Basic' Type of Text?". *Text* 12(2), 293-310.
- Virtanen, Tuija. 1995. "'Then I Saw To Antique Heddes': Discourse Strategies in Early Modern English Travelogues". In Jucker, ed., 1995, 499-513.
- Weinreich, Harald. 1973. *Le temps*. Paris: Seuil.
- Weinreich, Uriel. 1964. *Languages in Contact*. London/The Hague: Mouton.
- Weinreich, Uriel, William Labov & Marvin Herzog. 1968. "Empirical Foundations for a Theory of Language Change". *Directions for Historical Linguistics: A Symposium*, ed. by Winfred P. Lehmann & Yakov Malkiel, 95-188. Austin: University of Texas Press.
- Wright, Roger. 1982. *Late Latin and Early Romance*. Liverpool: Francis Cairns.
- Zvelebil, Kamil V. 1989. *Dravidian Linguistics: An Introduction*. Pondicherry, India: Pondicherry Institute of Linguistics and Culture.