Introduction

Each year brings new problems of Form and Content.
(W.H. Auden, ‘Shorts’)

The great source of pleasure is variety. Uniformity must
tire at last, though it be uniformity of excellence. We
love to expect; and, when expectation is disappointed or
gratified, we want to be again expecting.

The present book is a collection of papers illustrating the state of the
art in corpus-based research on diachronic English, by means of case-
study expositions, software presentations, and theoretical discussions
of the topic. The majority of these papers were delivered at the 25th
Conference of the International Computer Archive of Modern and
Medieval English (ICAME), held at the University of Verona on
18-23 May 2004. Corpus-based studies of diachronic English have
been thriving over the last three decades to such an extent that the
validity of corpora in the enrichment of historical linguistic research is
now undeniable. Bearing this in mind, scholars are now pondering
how far diachronic corpus linguistics may be improved in order to
further enhance our knowledge of the kaleidoscopic shifts and turns of
the English language through the centuries.

This is the issue tackled in the very first paper of the collection
(‘The Importance of Historical Corpora, Reliability, and Reading’), by
Anne Curzan and Chris C. Palmer; the authors provide a detailed
overview of ways to widen the rich contribution of historical corpus
linguistic studies to the broader field of linguistics, while recognizing
the limitations inherent to corpus-based methodologies. The argument
of the authors has two complementary strands, one focussed on the
objects of study and one on the linguistic analysis. First, historical
corpus linguistics would benefit from embracing a wider definition of a useful diachronic corpus, given the understanding that principled and unprincipled corpora can prove productive and limiting for different studies with different research goals. Second, studies in historical corpus linguistics should involve complementary methodologies and engage current linguistic theories in ways that both enrich the analysis of the corpus data and inform the development of linguistic theory.

This is true for all the main periods of the historical development of English, though Old and Middle English appear to be the most troublesome, due to scanty original documents and to the different versions of transcripts available particularly for the most ancient documents. Though scanty as the original data may be, yet these two periods yield interesting food for thought, as testified to by the four papers focussing on them in the present volume.

The very first one, by Johan van der Auwera and Martine TaeYMANS (‘More on the Ancestors of Need’) deals with the Old English origins of the Present-day English verb need; the discussion is largely based on a scrutiny of the research literature, on the entries in the standard dictionaries, and the Old English and Middle English parts of the Helsinki Corpus of Diachronic English. Bearing on these data, the authors argue that Present-day need replaces at least four earlier constructions: (i) a personal need verb meaning ‘compel’, (ii) an impersonal need verb meaning ‘it is necessary’, (iii) the verb þurfan meaning ‘need’ in negative polarity contexts, and finally (iv) a set of polarity neutral nominal constructions with the nouns nedþearf, þearf, and ned, all meaning ‘need’.

Focussing more on speech-related varieties of English, Manfred Markus convincingly comments on the affinity of alliteration to spokeness; to do so, he analyzes Middle English alliterative verse (‘Spotting Spoken Historical English: The Role of Alliteration in Middle English Fixed Expressions’), which has lately aroused new interest (cf. Brinton/Akimoto 1999; Hartle 1999; Minkova 2003). While the dominant understanding seems to be that alliteration was a specifically poetic metrical device, drawing on the Innsbruck Prose Corpus of ICAME (Innsbruck Computer Archive of Machine-Readable English Texts), Markus demonstrates that alliteration played a considerable role in common Middle English
phraseology. To illustrate his tenet, he analyzes noun-headed phrases 
(N’s N, N x N, Adj N) and verb-headed phrases of composite 
predicates (V (Prep) N). Both on statistical and on linguistic grounds, 
the paper provides the necessary background to the motivation of 
many fixed expressions partly still in use today, such as to do sports 
(< disport) and to make merry.

Moving from general English to typologically specialized 
varieties, IRMA TAAVITSAINEN, PÄIVI PAHTA and MARTTI MÄKINEN 
illustrate the corpus of Middle English Medical Texts (MEMT), 
compiled by the authors themselves (‘Towards a Corpus-Based 
History of Specialized Languages: Middle English Medical Texts’). 
MEMT forms the first part of the Corpus of Early English Medical 
Writing (1375-1750) and contains texts from c. 1375 to c. 1500 
together with a small appendix of recipes from c. 1330. The authors 
discuss in great detail the principles of corpus compilation, data 
selection criteria and database structure; moreover, they report on pilot 
studies on earlier versions of the corpus, and indicate areas in need of 
future research. Indeed, MEMT proves to be very useful for several 
kinds of analyses on linguistic developments and systematic historical 
accounts of specialized and professional languages; moreover, it 
provides a new window to the late medieval medical register and to 
genre-based language history, as medicine was the forerunner in 
vernacularization processes.

The final contribution pertaining to the early periods of English 
development is by BARRY MORLEY and PATRICIA SIFT on the 
expression of speech acts in Late Middle English (‘Towards the 
Automatic Identification of Directive Speech Acts’). According to the 
authors, a challenge for Historical Pragmatics is an automated 
‘function-to-form’ study (Jucker 1995) for data identification and 
quantification. For example, speech acts are realized in a large variety 
of syntactic patterns that are difficult to trace electronically; in this 
study, the computerized identification of speech acts is shown to be 
viable, given well-defined research parameters. The paper investigates 
directive speech acts (manually identified) in a small corpus of Late 
Middle English prose sermons (1350-1500) taken from the Penn-
Helsinki Parsed Corpus of Middle English. In such religious 
instruction, directive speech acts are typically issued by preachers to 
their audiences and occur in recurring syntactic forms or formulae
qualitatively depending on audience composition (laity, educated audience, presence of a monarch, etc.). An interesting attempt has been made by the authors to quantify these formulae in terms of a set of computer filtering rules.

The following four papers exploit corpora to tackle issues pertaining to different typological and regional varieties of Modern English. HELENA RAUMOLIN-BRUNBERG studies people who led morphological changes in Early Modern England (‘Leaders of Linguistic Change in Early Modern England’); to do so she analyzes the Corpus of Early English Correspondence (CEEC, c. 1410-1680), compiled at the University of Helsinki and covering the time-span of c. 1410-1680 for a total of 2.7 million running words, representing over 6,000 letters. Three changes are discussed: the replacement of subject YE by YOU, the change from the third-person singular suffix -TH to -S, and the loss of the final nasal from the possessive determiners MINE and THINE. The leaders are singled out for two different phases of each change: the incipient (the overall proportion of the new form below 15%) and the new and vigorous (15-35%). The individual leaders can be found among the groups that are known to have led the changes on the basis of previous research. Interestingly, the incipient and the new and vigorous leaders had different social backgrounds and networks; specifically, the incipient leaders were geographically mobile middle-ranking people with a large number of weak links, whereas the new and vigorous leaders – who had a higher social status – were central people in their social networks.

Moving to 17th- and 18th-century journalistic writing, HANS MARTIN LEHMANN, CAREN AUF DEM KELLER and BENI RUEF illustrate the first public release of ZEN, the Zurich English Newspaper corpus, consisting of early English newspapers published in London between 1661 and 1791 (‘ZEN Corpus 1.0’). The authors describe the selection, transcription and format of the material and discuss the methodological decisions taken at the stages of sampling, transcribing and structuring/formatting, with specific reference to the conversion into XML format and its advantages. Finally, they provide an interesting overview of ZEN Online, a web-based search interface of the corpus, analyzing pattern-based data retrieval, the use of annotation and the integration of visual representations of the original newspaper material.
An interesting application of the ZEN corpus is provided by UDO FRIES, who investigates the genre of obituaries (‘Death Notices: The Birth of a Genre’). By the end of the 18th century – when The Times was first published – death notices introduced by an appropriate header, usually Died or Deaths, were a well-established genre. The earliest examples in the ZEN Corpus date from the early 1730s. When one extends the definition of the text class to include brief death reports without any headers, we come across many earlier examples, right from the beginning of the collection (in 1671). These show a greater variety of expression than the later ones with a header. In his well-balanced discussion, the author also warns about the limits of using computer corpora to answer text linguistic questions, such as the rise of a new text class or genre; indeed, a complete text linguistic study must always consider the possibility of texts outside the corpus, since even large corpora cannot contain all the texts of a period. Thus, the type of death notices found in newspapers should also be sought in monthly magazines, which so far have been disregarded in linguistic studies.

The final paper in this section is by FRANCK ZUMSTEIN (‘The Contribution of Computer-Searchable Diachronic Corpora to the Study of Word Stress Variation’), who presents the early results of a study carried out by the research group in Linguistics at the University of Poitiers. Such results include the making and use of computer-searchable lexico-phonetic corpora of 18th- and 19th-century English. These two centuries saw an upsurge in dictionary-making activity and were marked by the lexicographical works of famous authors such as Samuel Johnson and John Walker. In most dictionaries, the authors meticulously endeavoured to represent the pronunciation of entry words. The Poitiers research group has undertaken the task of digitizing these documents and turning them into computer-searchable corpora, in order to help the linguist to retrieve quickly and exhaustively the data she/he is looking for. Hopefully, these new tools will be of help in accounting for pronunciation variation of word entries in relation to their spelling, stress pattern and morphological structure; for word-formation tendencies and related phonological issues; and finally for the pronunciation peculiarities of learned and specialized words.
The last four contributions in the volume provide a composite, insightful overview of 19th- and 20th-century English; the former is discussed as a period of linguistic stability, variation, and change by Merja Kytö and Erik Smitterberg (‘19th-Century English: An Age of Stability or a Period of Change?’). The authors analyze lexical bundles, multal quantifiers, and the progressive and phrasal verbs as they occur in CONCE (A Corpus of Nineteenth-century English) and ARCHER (A Representative Corpus of Historical English Registers). Indeed, their survey of lexical bundles reveals a notable degree of stability across time in e.g. Letters and Science; trials display features familiar from conversation as well as from the courtroom situation; while certain bundles are genre-specific. With reference to multal quantifiers, open-class options such as a lot of are shown to increase at the expense of closed-class quantifiers like much, but only in certain linguistic contexts (especially uncountable and assertive contexts). Finally, as for progressives and phrasal verbs, they both increase greatly in frequency across the 19th century, but genre and gender are important conditioning factors. Taken together, these three case studies demonstrate that stability, variation, and change are multifaceted notions that may apply on many levels of language use. From a methodological perspective, the results illustrated by the authors are related to previous multi-feature/multi-dimensional analyses and their study testifies to the fact that such analyses can be of great help when the communicative functions of other linguistic features are interpreted.

Clemens Fritz (‘The Conventions’ Spelling Conventions: Regional Variation in 19th-Century Australian Spelling’) draws attention to Australian English spelling by analyzing the minutes of the Federation Debates of the 1890s, comprising ca. four million words transcribed in Adelaide (1897), Melbourne (1890, 1898) and Sydney (1891, 1897). Despite its comparative uniformity, Australian English appears to exhibit a number of regionalisms, which are not restricted to lexis but cover the area of spelling as well. Currently, spelling variables are typically associated with differences between British and American English, while in the 19th century they were not yet codified to the same extent. Therefore, English in Australia could still choose e.g. between -our and -or, and choices differed by writer, period and region. The investigation of the Federation Debates shows
that the state parliaments of Adelaide, Melbourne and Sydney had fixed certain spelling variables differently. On account of this, regional standards were established which can sometimes still be found today.

The last two papers start from the 20th century but go well beyond the mere illustration of a "short-term change in diachrony" (Kytö/Rudanko/Smitterberg 2000: 85), which is supposed to focus only on the language of the previous century; indeed, both authors embrace also the older periods of the history of English. Tine Breban does so to test the hypothesis that adjectives of comparison – such as other, identical and similar – display a polysemy explainable as the result of an ongoing process of grammaticalization (‘The Grammaticalization of the English Adjectives of Comparison: A Diachronic Case Study’). Specifically, the author lays the results of her study on the diachronic development of six adjectives of comparison from 750 to 1920, representing three semantic subgroups – other and different for ‘difference’, same and identical for ‘identity’, and similar and comparable for ‘similarity’ – and compares their development to the situation in Present-day English. The reference corpora consist of eight random samples, covering the periods 750-1050, 1050-1250, 1250-1500, 1500-1710, 1710-1780, 1780-1850, 1850-1920 and a Present-day English sample containing material from 1990 onwards, for each of the six adjectives. The first four samples (750-1710) are from the Helsinki Corpus of English Texts, the next three (1710-1920) from the Corpus of Late Modern English Texts, while the Present-day English texts are extracted from the COBUILD Corpus (Bank of English). The resulting analysis encompasses both a qualitative and a quantitative perspective, since it thoroughly investigates both the development of new meanings and the changing distribution of the different meanings for each of the adjectives.

The final article is by Göran Kjellmer (‘Panchrony in Linguistic Change: The Case of Courtesy’) and focuses on the development of the word courtesy. The author remarks that, in a normal pattern of linguistic change, one phase is succeeded by another one, which in turn is succeeded by a third, etc.; the older stages vanish successively as the later ones establish themselves. However, with the help of data from the Oxford English Dictionary and from the CobuildDirect corpus, it can be observed that this is not the only
pattern in language change. The case of Present-day English *courtesy* illustrates a pattern where it is possible to trace a number of consecutive stages – logically developed from an original stage – which have all remained in the language, none of them showing any signs of being supplanted by the others. Moreover, the final phase is well on its way to becoming grammaticalized. The author concludes that the simultaneous existence in the current language of all the stages of the development of *courtesy* is seen, perhaps paradoxically, as an instance of panchrony in linguistic change.

This collection terminates with the slightly provocative claim of panchrony in linguistic change and the insightful, perceptive exploitation of Present-day English corpora for the study of diachronic issues. With such premise, the ground is open for a second book of corpus-based studies to be published out of the conference ‘Corpus linguistics: The state of the art twenty-five years on’; in the book, currently in preparation, theoretical issues and individual case studies will be dealt with reference to contemporary English. The two books are intended as complementary since, as we hope to have testified to in the present collection, past and present are so strongly interlocked and so inextricably entwined that it proves hard – if not preposterous – to fully understand present-day English structure and peculiarities without turning back to the previous centuries for an in-depth knowledge of the ‘whys’ and ‘hows’ of the current state of the art.

References

