English in Transition

Corpus-based Studies in Linguistic Variation and Genre Styles

Matti Rissanen, Merja Kytö
Kirsi Heikkonen
(Editors)

TiEL

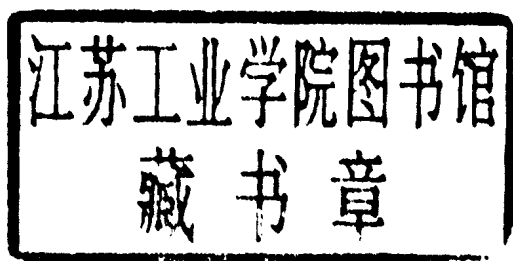W
DE
G

Mouton de Gruyter

English in Transition

Corpus-based Studies in
Linguistic Variation and Genre Styles

Edited by

Matti Rissanen
Merja Kytö
Kirsi Heikkonen

List of abbreviations

The following parameter codes included in the Helsinki Corpus appear in the present volume as such or in abbreviated versions (see also Bibliography at the end of the volume).

Prototypical text category:

EX	=	'expository'
IR	=	'instruction religious'
IS	=	'instruction secular'
IS/EX	=	'instruction secular'/'expository'
NI	=	'narration imaginative'
NN	=	'narration non-imaginative'
STA	=	'statutory'

Text type:

BIA	=	'biography, autobiography'
BIBLE	=	'Bible'
BIL	=	'biography, life of a saint'
BIO	=	'biography, other'
COME	=	'drama, comedy'
CORO	=	'correspondence, non-private'
CORP	=	'correspondence, private'
DEPO	=	'proceeding, deposition'
DIARY	=	'diary'
DOC	=	'document'
EDUC	=	'educational treatise'
FICT	=	'fiction'
GEO	=	'geography'
HANDA	=	'handbook, astronomy'
HANDM	=	'handbook, medicine'

HANDO	=	'handbook, other'
HIST	=	'history'
HOM	=	'homily'
LAW	=	'law'
LET PRIV	=	'letter, private'
LET NON-PRIV	=	'letter, non-private'
MYST	=	'drama, mystery play'
NEWT	=	'New Testament'
OLDT	=	'Old Testament'
PHILO	=	'philosophy'
PREF	=	'preface' or 'epilogue'
RELT	=	'religious treatise'
ROM	=	'romance'
RULE	=	'rule'
SCIA	=	'science, astronomy'
SCIM	=	'science, medicine'
SCIO	=	'science other'
SERM	=	'sermon'
TRAV	=	'travelogue'
TRI	=	'proceeding, trial'

Other:

PROF	=	'audience, professional'
NON-PROF	=	'audience, non-professional'
INT	=	'interaction', 'interactive'
INF	=	'informal setting'
FOR	=	'formal setting'
X, XX	=	'unspecified'

Preface

This book is one of three volumes reporting the results of the project 'English in transition: Change through variation', carried out in the English Department of the University of Helsinki. The first volume, *Early English in the computer age: Explorations through the Helsinki Corpus* (ed. by Matti Rissanen, Merja Kytö and Minna Palander-Collin, Mouton de Gruyter, 1993) is now followed by two volumes, *English in transition: Corpus-based studies in linguistic variation and genre styles* and *Grammaticalization at work: Studies of long-term developments in English*.

Both these volumes approach change in English from the angle of linguistic variation. The articles deal with processes of change in morphology, syntax and lexis, and pay special attention to the role played by textual and discourse factors across the centuries. From the methodological point of view, diachronic variation analysis and the multi-feature approach aiming at the identification of co-occurrence patterns in genres are the main frameworks adopted.

The aim of the present volume is to give new insights into the development of some central verb constructions (with *be* and *have*), expository apposition, and genre-specific features of expressions of affect and attitude in text. The *Grammaticalizaton at work* volume sheds light on the development of adverbs and indefinite pronouns and on the means of reflexivization, in relation to various grammaticalization processes.

All the studies in these volumes are based on the Helsinki Corpus of English Texts; supplementary material has been drawn from other corpora and concordances and from primary texts outside the corpora.

The 'English in transition' project was initiated in 1990 as a continuation of an earlier project which produced the Helsinki Corpus. The core team of both projects has been the same, consisting of the authors and editors of the volumes. The editors would like to express their special thanks to all research assistants of the projects and particularly to Arja Nurmi and Päivi Koivisto-Alanko for their excellent work in producing these volumes.

We are most grateful to the Academy of Finland for funding our project for four years. We are indebted to the University of Helsinki for giving us research premises, and to the English Department for up-to-date technical facilities, travel grants and other support. Our thanks are due to Mrs Leena Sadeniemi for giving us expert advice in computer technology and training us to use programs. Finally, we would like to thank the editors of Mouton de Gruyter for accepting the two volumes now published in their Topics in English Linguistics series.

Helsinki, June 1996

M.R.   M.K.   K.H.

Contents

Introduction

*Matti Rissanen, Matti Kilpiö, Merja Kytö, Anneli Meurman-Solin,
Saara Nevanlinna, Päivi Pahta* and *Irma Taavitsainen*

1. Methodological considerations: linguistics and philology in interaction

The last two decades have seen a rapid increase and methodological de-
velopment in the studies of variation in language. The basic assumption
in these studies is "orderly heterogeneity", i.e. variability which is not
random but affected by linguistic and extralinguistic factors or constraints
(Weinreich—Labov—Herzog 1968; Samuels 1972; Labov 1994). Lan-
guage can be seen as meaning potential which is realized in choices be-
tween forms and expressions "meaning the same thing" (Halliday 1973:
51; see also Halliday 1978). Within this framework, it has been our aim
to establish how linguistic variation is patterned not only socially, re-
gionally and temporally, but according to genres defined by extralinguis-
tic criteria. When the variationist approach is adapted to discourse studies,
comparisons of text types defined by their linguistic properties become
the key (Schiffrin 1994: 314); the last two chapters of this book extend
the application to historical stylistics with the aim of charting genre styles
and genre conventions.

   In diachronic studies, the variationist approach provides us with a good
opportunity to observe the actual process of change. We can trace the
birth and death of variant expressions, but perhaps more interestingly,
their changing frequencies and distributions within a variant field at sub-
sequent periods of time and in various genres, and we can analyse
changes in the intricate mesh of linguistic and extralinguistic factors con-
ditioning the occurrence of these variants. Within this approach our philo-

logical training has been a great asset as it emphasizes the context of expressions and the context of culture in interpreting them.

The importance of extralinguistic factors in the analysis of development and change has necessitated an intensive study and discussion of genre typologies from the point of view of the historical study of language; cf. e.g. Douglas Biber's and Edward Finegan's studies of the 'dimensions' characterizing texts and offering a new basis for their grouping. In the structure of text corpora, the labelling of genres provides a general framework for the discussion of text-related changes at different periods of time. However, the varying importance of conventions or innovative pressures in the evolution of each genre or group of genres may decrease the usefulness of such classifications, as genres are also internally heterogeneous, and the pace and direction of change may be different in individual texts representing a particular genre. A comparison of the occurrences and frequencies of variant expressions in different texts allows the reconstruction of the various levels of past expression: written and speech-based, literary and non-literary, formal and informal, etc. This method is also necessary for all attempts to describe the relationship of the standard(s) to regional or social dialects.

Of the five chapters included in this volume, three discuss morphological, syntactic or lexical questions with special attention paid to variation relating to text type, dialect etc., while two concentrate on the co-occurrence patterns of linguistic features in various types of texts and on questions of genre classification, genre markers and distinguishing features. The main focus is on long-span diachrony, mostly from Early Middle to Modern English, i.e. from the time of the radical reorganization of the structure of the language to the period of its gradual establishment.

The long time span and the wealth of primary data set specific demands for the grammatical models used in this volume. The model should make it possible to compare changing grammatical phenomena across time and, at the same time, be comprehensive enough to provide researchers with analytical tools for the very wide range of morphosyntactic issues involved in classifying linguistic features in genre studies. The model which has proved most useful in this type of research is a structurally oriented one, such as *A comprehensive grammar of the English language* by Quirk et al. (1985) for Present-day English. This type of grammar provides an adequate basis for analysis at a relatively low level of abstraction, enabling the researcher to deal even with the more problematic borderline cases. On the other hand, it is clear that, as a grammar of Present-day English, Quirk et al. cannot be directly applied to different historical phases of

English. In our case, it has provided the basis which the writers of this book have employed in different ways and to different degrees. We have made use of various approaches, from traditional grammar to semantic, pragmatic and textual theories.

2. New openings offered by the Helsinki Corpus and other computerized material

Scholars working on variationist studies benefit from having access to computerized collections of texts. Computerized diachronic corpora, increasingly available in international distribution, make it relatively easy to collect evidence of the occurrence of variant expressions; they also encourage the researcher to approach topics that would earlier have meant an unreasonable amount of time-consuming routine work.

The studies reported in this volume are based on the Helsinki Corpus of English Texts, the first large computerized corpus to cover the time-span of several periods in the history of English. With its 1.5 million words — c. 400 samples of texts dating from the 8th to the 18th century — it offers reliable indicators concerning the structural and lexical developments of English for over a millennium. In many cases, the results are tentative and they must — and fortunately can — be supplemented from other corpora, concordances and primary texts. In the future, the usefulness of the Helsinki Corpus will be further increased by the addition of word-class tagging and syntactic bracketing to the text samples.[1]

Each text or group of related texts of the Helsinki Corpus is equipped with a set of parameter values containing information on the text and its author, if known. In Old and Early Middle English, this information is concentrated mainly on the date and dialect of the text and on a fairly loose description of the genre. In Late Middle and Early Modern English, the genre selection is more extensive than in the earlier periods, with samples from drama texts, private letters, law court records, diaries, prose fiction, and so forth.[2] In these periods, sociolinguistic information is given on the authors of the texts (their rank, sex, and age) and, in the case of letters, on the relationship between the writer and the receiver (intimate family members are distinguished from more distant addressees; addressees superior to authors are distinguished from those inferior to them).

A large and structured corpus such as the Helsinki Corpus, equipped with textual parameter codings, has made it possible for us to combine effectively qualitative and quantitative analysis through variation. This approach has been extensively used in the analysis and interpretation of the data (cf. McEnery—Wilson 1996: 62–63). Our background knowledge of texts, also capitalized on in the compilation of the Helsinki Corpus, has been utilized to place our observations of linguistic phenomena within a larger framework. We have been able to reanalyse and re-evaluate earlier genre and text-type classifications of the older periods of English and to test the validity of the suggestions made concerning the co-occurrence of linguistic features with various types of text. The parameter coding and extensive textual basis have also allowed us to observe the role played by dialectal distribution and the differences between prose and verse texts in the survey of Old and Middle English variant forms. In the discussion of Late Middle and Early Modern evidence, emphasis has been given to the distributions shown by speech-based texts, or texts showing a relatively high degree of orality. In a few cases, it has been possible to draw conclusions from sociolinguistic variables: the genre, degree of interactiveness, level of formality, sex, age, rank or education of the author, or the relationship between the sender and receiver of letters.[3] The insights derived from recent trends in sociolinguistics have added a new angle to the discussion of these factors (cf. e.g. Romaine 1982; Milroy—Milroy 1985; Milroy 1992). We believe that, among these variables, genre is the most complex one, as the varying types of genre can be claimed to derive from a combination of other variables such as topic, audience, setting, text category, etc.

The corpus-based framework lends itself to various statistical applications by which it is possible to verify the reliability of the results obtained. In addition to traditional frequency surveys, a number of more advanced statistical analyses have been carried out for the purposes of some studies included in the volume. Thus Merja Kytö in her study on the *be/have* variation uses logistic regression analysis to assess the impact of various factors on the use of the variant forms. This analysis tests out how statistical models based on various combinations of factors account for variation in the data and assess the significance of the individual factors and their interaction. Several statistical methods are combined in Irma Taavitsainen's chapter on personal affect and genre styles. Factor analysis is applied first to identify adjoining text types and possible text-type markers. This method serves to indicate the overall patterns of genre styles and how they relate to one another. Then t-tests and f-tests are used

to evaluate the significance of the features in telling fiction apart from the adjoining genres. The combination of these methods yields results which are then assessed in a larger sociohistorical context. In the chapter by Anneli Meurman-Solin on the concept of point of view in texts, the different informative value of mean frequencies and percentages is stressed and the findings are presented by focusing on one feature or factor at a time and, after a detailed analysis of this kind, by clustering them by both syntactic and semantic criteria. This adds to the reliability of data and also allows the mapping of significant correlations in a network-like pattern of dimensions that usefully describe focal features in genre styles and text types.

3. Variation on the level of morphology, syntax and lexis: the verbs *be* and *have*

The studies by Merja Kytö and Matti Kilpiö in this volume focus on the use and development of the verbs *be* and *have*, which have played a vital role in the shaping of the English verb phrase. The paths of *be* and *have* are parallel, both having occurred as lexical and auxiliary verbs providing variant expressions in the auxiliary function (e.g. compound tenses with mutative verbs, and expression of obligation).

Owing to complexities in the development of these verbs, there is no consensus about the role played by such crucial processes as grammaticalization. With the verb *be*, for instance, the status of the verb in progressive forms has clearly changed from the copula in Old English into an auxiliary from Middle English onwards. The uses of *be* and *have* grammaticalize at different stages in the history of English, and the two studies only concentrate on certain key periods in these developments.

The study of the *be/have* variation with mutative intransitives examines the process by which *have* finally supersedes *be* in present and past perfect constructions; the study on *be* focuses on the forms and functions of the verb, with an eye on the developments in its functional load. The former study covers the period from Late Middle to Modern English, the latter from Old to Modern English. Among the extralinguistic factors included in the two studies are chronology, region and dialect, and foreign influence; moreover, with the *be/have* variation, such factors as text type, relationship of the text to spoken language, level of formality, orality and

authorial properties are taken into account. Among the linguistic factors observed in both studies are tense and certain verb constructions (finite/ non-finite forms, *-ing* constructions). Furthermore, with the study of the *be/have* variation, attention is paid to the status of the verb (stative and mutative; action and process; frequent and rarer verbs), durative, iterative and conditional contexts, negation, and object-like and other comple- ments. With *be*, additional factors include developments in the morphol- ogy of *be*, the function of the verb (its use as an auxiliary or a lexical verb, copular or non-copular), and semantic-syntactic functions of *b*-forms as against non-*b*-forms (for Old English).

These studies have brought up new evidence pinning down trends of development in greater detail than found in previous research, thanks to the combination of quantitative and qualitative corpus linguistics and philological assessments. With the *be/have* variation, the use of *have* is shown to increase gradually from the Late Middle to the Early Modern English period, gain in momentum in the late 1700s and supersede *be* in the early 1800s. The more powerful extralinguistic factors influencing the process of change include chronology and text type, and a number of lin- guistic factors (relationship to tense, aspect, complementation etc.). In the study dealing with *be*, the most striking morphological developments are the rapid disappearance in the Early Middle English period of the co- existent Old English present tense paradigms (*beon/wesan*), the diversifi- cation of forms in Middle English and the subsequent process of simplifi- cation and regularization as Early Modern English is reached. In the sur- vey of the main functions of *be*, the remarkable stability seen in the rela- tive share of the three main functions of *be* over the centuries is the most important finding. Within the auxiliary category, the predominance of the use of *be* as a passive auxiliary is the most noticeable feature.

These two studies have shown that the Helsinki Corpus, supplemented by other diachronic corpora and other standard reference works (*LALME, MED, OED*), offers a solid basis for the empirical approach aimed at ac- counting for variation and change.

4. Variation in re-phrasing: expository apposition across the centuries

The chapter by Päivi Pahta and Saara Nevanlinna forms a bridge between the studies concentrating on the syntactic and morphological developments described above and those discussing the characteristics of genres and text types. It examines characteristics of expository apposition — the grammatical category connected with re-phrasing. Re-phrasing as a communicative phenomenon occurs in both written and spoken media, and in both planned and unplanned discourse. It can generally be analysed as the writer's or speaker's attempt to reformulate an utterance in order to achieve successful communication. On closer inspection, the decision to re-phrase may be based on various factors, including stylistic and didactic considerations, the author's assessment of the addressee's ability to process given information, or the author's wish to add to the flow of discourse by providing more information about the topic of discussion.

This chapter focuses on the development and use of expository apposition with an explicit marker in Late Middle and Early Modern English. Apposition is seen as a broad notional category containing both nominal and non-nominal phrases, clauses and sentences. There are no previous detailed studies of apposition in this period, and the theoretical framework adopted as the starting-point in this study is the recent discussion of apposition in Present-day English by Meyer (1987 and 1992), where apposition is seen as a syntactic, semantic and pragmatic relation. The main emphasis in Pahta and Nevanlinna's study is on the semantic characteristics of appositional constructions and their distribution across different text types. Attention is also paid to the devices used in linking appositional units, i.e. explicit markers of expository apposition.

The study shows that the use of expository apposition links up with some of the most central lexical phenomena of the Middle English and Early Modern English periods, such as dialectal variation and the adoption and accommodation of loan-words. It also indicates a clear tendency for some text types to favour the use of appositional constructions in general, and certain semantic and syntactic types in particular. This is so throughout the period, although there is internal variation within most text

types. Most of the markers of expository apposition available in Present-day English were found to exist even in the late medieval period, with many others which have since gone out of use. In this respect, the results obtained point to a difference in the use of coordinative apposition, particularly binomial constructions with the markers *and* and *or*.

In the course of the analysis, the use of traditional philological tools (knowledge of textual background, cultural status of texts, etc.) proved helpful. With certain limitations, the corpus-based approach offered a fruitful way of collecting data for the study of appositional constructions. Considering the open-class nature of apposition as a linguistic phenomenon, the results obtained in this study show the way for further work on the topic.

5. Expressions of personal affect and stance marking: identifying genre-specific choices

The last two chapters of this volume, by Irma Taavitsainen and Anneli Meurman-Solin, have a somewhat different problem-setting, but the approach combining the quantitative and qualitative methods applies here as well. Their main topic is identifying genre-specific features in the linguistic choices related to participant roles by analysing expressions of affect and attitude in texts. Taavitsainen discusses the use of personal pronouns, exclamations, direct questions and other expressions of personal affect, while Meurman-Solin's study focuses on the frequencies and distributions of adjectives and open-class adverbs as stance markers.

In recent years genres have been looked at from many different perspectives. Besides thorough comparative studies of features of individual genres (such as fiction in Fludernik 1993 and 1996) or a wide range of genres in a particular time period (such as Renaissance genres in Lewalski 1986), we find the interdisciplinary approach (for example in Sell and Verdonk 1994) and the important advances in discourse analysis (Coulthard 1994) particularly relevant. Taavitsainen and Meurman-Solin approach the problems of genre studies from a variationist's and sociolinguist's point of view. Rather than restrict the focus to the dimension of written genres as against genres reflecting usages more typical of spoken language, they aim at pointing out clusters of features which position texts on a number of other dimensions, particularly those which reflect

idiolectal or genre-specific characteristics of participant roles, or focus on genre markers.

Traditional genre labels such as 'history', 'private letter', 'autobiography' or 'sermon' are used in the majority of recently produced text corpora. A typology of this kind has also been adopted as a working tool in the Helsinki Corpus and its supplement, the Helsinki Corpus of Older Scots (see note 1). The labelling in these corpora is based on extralinguistic factors such as the social and communicative function of texts or their subject matter. In a number of studies based on the Helsinki Corpus and/ or its supplement of Scots (see the bibliographies in the chapters by Taavitsainen and Meurman-Solin in this volume), medieval and Renaissance prose genres have been shown to be linguistically relatively heterogeneous. This heterogeneity is partly due to the compilers' decision to polarize the samples: they have intentionally selected the representatives of a genre from different stylistic traditions when such stylistic variation has been established in earlier research (Nevalainen—Raumolin-Brunberg 1989: 99). This should always be taken into consideration when generalizing from the results.

Patterns of co-occurring features, illustrated in Taavitsainen's and Meurman-Solin's studies, provide evidence for a classification of texts into text types. Each text type may thus comprise texts which represent different genres; in addition, intertextuality phenomena between genres complicate the issues. Biber's pioneering work (1988) in corpus-based stylistics with its multifeature and multidimensional statistical assessments is strictly linguistic; our innovation is the firm philological anchoring, limiting the comparisons to texts that share common features and that belong to related genres (Taavitsainen) and recategorizing multifunctional linguistic features by means of a thorough analysis of their varying syntactic and semantic properties in different time periods (Meurman-Solin). We have also profited from other studies which tackle related questions. The two studies aim at making it applicable to the analysis of early prose texts by selecting features other than those in Biber's factors, and by introducing a more detailed semantic subcategorization of a more comprehensive set of realizations of some features included in his factor analysis, and by developing the research tasks for different aims. Because of the emphasis on semantic features, the relevant examples are carefully selected by qualitative reading and analysed in the wider context of running text. Both studies thus highlight the importance of combining the quantitative approach of corpus linguistics with a detailed analysis of discourse function and meaning, central in the philological tradition.