

Advances in Historical Sociolinguistics ∞

# Patterns of Change in 18th-century English

*A sociolinguistic approach*

EDITED BY

Terttu Nevalainen,  
Minna Palander-Collin  
and Tanja Säily

John Benjamins Publishing Company

## Patterns of Change in 18th-century English

# *Advances in Historical Sociolinguistics (AHS)*

ISSN 2214-1057

Over the last three decades, historical sociolinguistics has developed into a mature and challenging field of study that focuses on language users and language use in the past. The social motivation of linguistic variation and change continues at the forefront of the historical sociolinguistic enquiry, but current research does not stop there. It extends from social and regional variation in language use to its various communicative contexts, registers and genres, and includes issues in language attitudes, policies and ideologies. One of the main stimuli for the field comes from new digitized resources and large text corpora, which enable the study of a much wider social coverage than before. Historical sociolinguists use variationist and dialectological research tools and techniques, perform pragmatic and social network analyses, and adopt innovative approaches from other disciplines. The series publishes monographs and thematic volumes, in English, on different languages and topics that contribute to our understanding of the relations between the individual, language and society in the past.

For an overview of all books published in this series, please see  
<http://benjamins.com/catalog/ahs>

## **Editors**

Marijke J. van der Wal  
Leiden University

Terttu Nevalainen  
University of Helsinki

## **Editorial Board**

Wendy Ayres-Bennett  
University of Cambridge

Martin Durrell  
University of Manchester

Agnieszka Kielkiewicz-Janowiak  
Adam Mickiewicz University, Poznań

William A. Kretzschmar Jr.  
University of Georgia, Athens GA

Mieko Ogura  
Tsurumi University, Yokohama

Suzanne Romaine  
University of Oxford

Daniel Schreier  
University of Zurich

Merja Stenroos  
University of Stavanger

Sali A. Tagliamonte  
University of Toronto

Ingrid Tieken-Boon van Ostade  
Leiden University

Donald N. Tuten  
Emory University, Atlanta GA

Wim Vandenbussche  
Vrije Universiteit Brussel

Anna Verschik  
Tallinn University

## **Volume 8**

Patterns of Change in 18th-century English. A sociolinguistic approach  
Edited by Terttu Nevalainen, Minna Palander-Collin and Tanja Säily

# Patterns of Change in 18th-century English

A sociolinguistic approach

*Edited by*

Terttu Nevalainen

Minna Palander-Collin

Tanja Säily

University of Helsinki

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/ahs.8

**Cataloging-in-Publication Data available from Library of Congress:  
LCCN 2018014190 (PRINT) / 2018040657 (E-BOOK)**

ISBN 978 90 272 0103 4 (HB)

ISBN 978 90 272 6383 4 (E-BOOK)

© 2018 – John Benjamins B.V.

This e-book is licensed under a Creative Commons CC BY-NC-ND license. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>. For any use beyond this license, please contact the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

# Table of contents

Preface and acknowledgments	IX
Contributors	XI
<b>Part I. Introduction and background</b>	
CHAPTER 1	
Approaching change in 18th-century English <i>Terttu Nevalainen</i>	3
CHAPTER 2	
Society and culture in the long 18th century <i>Terttu Nevalainen</i>	13
Infobox 1. Range of writers in the CEECE <i>Terttu Nevalainen</i>	24
Infobox 2. Polite society and rhetoric <i>Arja Nurmi and Minna Nevala</i>	25
CHAPTER 3	
Grammar writing in the eighteenth century <i>Nuria Yáñez-Bouza</i>	27
CHAPTER 4	
The <i>Corpus of Early English Correspondence Extension</i> (CEECE) <i>Samuli Kaislaniemi</i>	45
Infobox 1. Data retrieval <i>Mikko Hakala</i>	58
CHAPTER 5	
Research methods: Periodization and statistical techniques	61
5.1 Quantifying change <i>Terttu Nevalainen</i>	61
5.2 Basic methods for estimating frequencies <i>Terttu Nevalainen</i>	64
5.3 Methods for studying changes lacking a variable <i>Tanja Säily, Arja Nurmi and Anni Sairio</i>	68

**Part II. Studies**

CHAPTER 6		
“Ungenteel” and “rude”? On the use of <i>thou</i> in the eighteenth century		77
	<i>Minna Nevala</i>	
CHAPTER 7		
Going to completion: The diffusion of verbal -s		97
	<i>Terttu Nevalainen</i>	
CHAPTER 8		
Periphrastic DO in eighteenth-century correspondence: Emphasis on no social variation		117
	<i>Arja Nurmi</i>	
CHAPTER 9		
Indefinite pronouns with singular human reference: Recessive and ongoing		137
	<i>Mikko Laitinen</i>	
CHAPTER 10		
Ongoing change: The diffusion of the third-person neuter possessive <i>its</i>		159
	<i>Minna Palander-Collin</i>	
CHAPTER 11		
Incipient and intimate: The progressive aspect		179
	<i>Anni Sairio</i>	
CHAPTER 12		
Change or variation? Productivity of the suffixes -ness and -ity		197
	<i>Tanja Säily</i>	

**Part III. Changes in retrospect**

CHAPTER 13		
Zooming out: Overall frequencies and Google Books		221
13.1 Normalised frequencies of the phenomena studied	221	
	<i>Tanja Säily</i>	
13.2 Google Books: A shortcut to studying language variability?	223	
	<i>Mikko Laitinen and Tanja Säily</i>	

---

CHAPTER 14	
Conservative and progressive individuals	235
<i>Tanja Säily</i>	
CHAPTER 15	
Changes in different stages	243
15.1 Introduction	243
<i>Minna Palander-Collin</i>	
15.2 From incipient to mid-range and beyond	246
<i>Minna Palander-Collin, Mikko Laitinen, Anni Sairio and Tanja Säily</i>	
15.3 From nearing completion to completed	251
<i>Terttu Nevalainen, Mikko Laitinen, Minna Nevala and Arja Nurmi</i>	
CHAPTER 16	
A wider sociolinguistic perspective	255
<i>Terttu Nevalainen</i>	
References	271
Appendix: Editions in the <i>Corpora of Early English Correspondence</i>	295
<i>Arja Nurmi and Samuli Kaislaniemi</i>	
Index	305



## Preface and acknowledgments

The first concrete steps towards writing this book were taken in late 2008, when we drafted our preliminary research plan. Later on, we were fortunate to recruit Dr. Nuria Yáñez-Bouza to the project as a specialist in eighteenth-century grammar writing. A prolonged undertaking, the book has finally seen the light of day. The core chapters have largely turned out the way we envisaged they would, but our thinking about the broader context has been considerably enriched over the years. For example, the Google Books resource did not exist in 2008 and the “third wave” of sociolinguistics had just been mooted in a few publications. Overall, we hope that the book that we have produced is now more interesting and informative than it would have been had it appeared earlier.

During the years we have accumulated a debt of gratitude to many people. First, we would like to thank Nuria for her patience – she was the only one of us who was always on time! Our gratitude to Professor Timo Honkela and Mr. Murtaza Hazara for advising us on bootstrapping and carrying out the tests, respectively. Professor Jukka Suomela produced the beanplots and accumulation curves, and Dr. Harri Siirtola designed the density plot in Figure 4.4. Anna-Lina Wallraff formatted the Google n-gram figures in Chapter 13, Anna Suutarla and Ina Liukkonen helped us edit the final draft of the manuscript, and Lassi Saario finalized the subject index. We would also like to thank the staff of the John Benjamins Publishing Company and Anke de Looper, in particular, for their essential input into the publishing process.

We are grateful for the participant feedback we received at various conferences, including ICAME 34 at Santiago de Compostela in 2013 and HiSoN 2016 in Helsinki. We owe especial thanks to the two anonymous reviewers who made a number of valuable suggestions for improvement to the original manuscript, and to Marijke van der Wal for her careful editorial attention to both form and substance. This feedback helped us correct many errors and oversights and, we hope, make our take on historical sociolinguistics more accessible to the research community. We are only sorry we could not come up with a more interesting title encapsulated in a quote as one of the reviewers suggested. As an afterthought, it might perhaps have been: *“Slow but I believe perfectly honest”? Sociolinguistic change in 18-century English.*

This project would not have been possible without the Academy of Finland and University of Helsinki funding for the Research Unit for the Study of Variation, Contacts and Change in English (VARIENG; 2006–2011, 2012–2016, respectively), which benefited all of us and provided us with an active and inspiring research community. Other sources of funding have also in part supported our work, including project funding provided by the Academy of Finland and the Universities of Helsinki and Jyväskylä.

The Editors  
Helsinki, 20 December 2017

## Contributors

- Mikko Hakala, MA, Department of Languages, University of Helsinki
- Samuli Kaislaniemi, PhD, Department of Languages, University of Helsinki
- Mikko Laitinen, Professor of English Language, Philosophical Faculty, University of Eastern Finland
- Minna Nevala, Professor of English, Faculty of Communication Sciences, University of Tampere
- Terttu Nevalainen, Professor of English Philology, Department of Languages, University of Helsinki
- Arja Nurmi, University Lecturer, English, Faculty of Communication Sciences, University of Tampere
- Minna Palander-Collin, Professor of English Language, Department of Languages, and Director, Helsinki Collegium for Advanced Studies, University of Helsinki
- Tanja Säily, Post-doctoral Researcher, Department of Digital Humanities, University of Helsinki
- Anni Sairio, University Lecturer, Department of Languages, University of Helsinki
- Nuria Yáñez-Bouza, Lecturer in English Language, Department of English, French and German, University of Vigo



PART I

## **Introduction and background**



# Approaching change in 18th-century English

Terttu Nevalainen

## 1.1 Preamble

Unravelling the diffusion of language change lies at the heart of historical linguistics. This challenge has traditionally been pitched at a highly abstract level in social terms:

To simplify our analysis, we shall assume that the language in process of evolution is that of a strictly monoglot community, perfectly homogeneous in the sense that observable differences represent successive stages of the same usage and not concurrent usages. (Martinet 1964: 164)

By contrast, sociolinguists, who approach language from the perspective of the community, know that change does not take place in social isolation: in reality there are always “concurrent usages”. Change implies choice. Changes crucially depend on the choices that language users make in adopting or rejecting a particular new form, function or construction in the first place. The outcome of these choices can be seen in language records of different kinds, in both speech and writing, but we cannot know how homogeneous these records are in social terms without taking into account their producers and production circumstances.

Ideally, both linguistic and social factors would therefore need to be accounted for in studies of the diffusion of language change. However, as is well known, the circumstances are rarely ideal for real-time studies of change, and limited access to data lies partly behind purely linguistic studies of long-term processes of linguistic change. But this does not always mean that long-term changes cannot be observed at a fairly close distance in social terms as well. That is our aim in this volume, where we analyse a variety of linguistic changes in their social contexts in the 18th century.

In past work, overviews of 18th-century English have been provided, for example, by Görlach (2001), Beal (2004) and Tieken-Boon van Ostade (2009). *The Cambridge History of the English Language* splits the century into two, based on external criteria, discussing the language of the period in volumes 3 (1446–1776) and 4 (1776–1997), edited by Roger Lass (1999) and Suzanne Romaine (1998), respectively. There are also collective volumes that approach 18th-century usage

from various sociolinguistic perspectives, including those of normative grammar (Tieken-Boon van Ostade 2008), speakers' social roles (Pahta et al. 2010), and language ideology and change (Hickey 2010).<sup>1</sup>

As appears from this summary account, the ground covered by these volumes varies greatly. For example, Beal's undergraduate textbook (2004) provides an overview of English in its social and regional contexts between 1660 and 1945, with an emphasis on phonological developments. By contrast, the volume edited by Hickey (2010) presents individual research articles that focus on topics such as 18th-century prescriptivism, standardization, politeness, and dialect grammar.

Our book offers yet another perspective on 18th century English by presenting original research based on the same corpus of texts and a shared variationist sociolinguistic framework. One of its aims is to provide empirical and thematic continuity for the processes of change observed in the previous centuries (Nevalainen & Raumolin-Brunberg 2003). However, reflecting the changing social circumstances, the potential role played by influences such as prescriptivism is also considered in discussions of the empirical findings.

As far as long-term, real-time language changes are concerned, a number of studies analyse 18th-century developments as part of grammatical change from Old or Middle English onwards (e.g., downtoning adverbs, Nevalainen & Rissanen 2002; existential constructions, Martínez-Insua & Pérez-Guerra 2006; patterns of complementation, De Smet 2013; double object constructions, Yáñez-Bouza & Denison 2015, to mention only a few). Although studies on long-term language change may share a quantitative orientation, they are often informed by different grammatical approaches and have different aims.

As linguistic features and constructions vary in their frequency of occurrence, long-term diachronic studies need to take recourse to a variety of complementary but often discontinuous data sources, as is the case with the studies of individual grammatical changes listed above. Studies focusing on Late Modern English can benefit from large multi-genre and single-genre corpora, which enable the study of low-frequency elements in matching data sources. Among such multi-genre corpora, *The Corpus of Late Modern English Texts* (CLMET) includes some 10 million words of 18th-century English distributed over five major genres (Diller, De Smet & Tyrkkö 2011). With its subregisters, *The Old Bailey Corpus* provides some 25 million words from the 1730s to 1800 (Huber 2007; Claridge & Kytö 2014: 34).

---

1. Running since 2001, the Late Modern English Conference series has produced several thematic volumes of studies on 18th- and 19th-century English (see, e.g., Dossena 2014 and Dossena et al. 2014, and the references therein). This period has been extensively studied for other European languages as well, for example, from the sociolinguistic perspective of *ego*-documents in Elspaß et al. (2007) and Rutten & van der Wal (2014), and comparing prescriptive norms and actual usage in Rutten, Vosters & Vandebussche (2014).

However, unparalleled resources though they are, large corpora usually come with either minimal metadata or very short text samples from individuals. In both cases, analysing such data requires raising the level of abstraction in sociolinguistic terms.

In this volume we will adopt the long view but base our studies on a single-genre corpus of personal correspondence that covers the time span from 1400 to 1800 and comes with rich metadata on the letters and the letter writers. In this way we hope to minimize discontinuities in our empirical evidence and trace processes of change in as comparable materials as possible, ranging from communities and social groups down to individuals.

Our studies will build on past research on the corpus and extend the work presented in Nevalainen & Raumolin-Brunberg (2003) on Tudor and Stuart English by following up the subsequent histories of some of the linguistic changes investigated in that volume. The earlier phases of these processes have also been analysed in a series of detailed studies by Nurmi (1999), Nevala (2002) and Laitinen (2007). Moreover, we will delve into certain innovative aspects of 18th-century language variation and change, the background for which has been charted by Sairio (2009) and Säily (2014).

Aiming to compare these various processes – and hence looking for possible sociolinguistic generalizations – our analyses will largely follow the sociolinguistic framework outlined in Nevalainen & Raumolin-Brunberg (2003). Besides aggregate findings, particular attention will be paid to individual variation and new methods of analysing changes that are not easily conceivable in terms of linguistic variables.

Section 1.2 outlines the background for the correspondence corpus and the basic sociolinguistic approaches used in our earlier work and further developed in this volume. To provide a backdrop for the linguistic processes studied here, Section 1.3 discusses the time courses of the changes in Late Middle English and Early Modern English that were analysed in Nevalainen & Raumolin-Brunberg (2003) based on the corpus data. Section 1.4 specifies the aims of the present volume and the particular processes studied. The last section gives a brief overview of the materials and methods used in these studies.

## 1.2 Past work: material and method

In 2003 Helena Raumolin-Brunberg and I published a book on language change in Early Modern English entitled *Historical Sociolinguistics: Language Change in Tudor and Stuart England*.<sup>2</sup> In that monograph we used personal letters to study fourteen changes that were completed or in progress between c. 1410 and 1681. Our

---

2. A second, expanded edition of this volume came out in 2017, published by Routledge.

corpus, the *Corpus of Early English Correspondence* (CEEC), had been compiled with specifically sociolinguistic research questions in mind. Its original version (CEEC-1998) covered a stretch of 270 years from the first surviving personal letters written in English to those exchanged in the aftermath of the English Civil War, that is, roughly twelve generations.

Our aim was to study the extent to which the methods and models used in modern sociolinguistics could be applied to the past. We wanted to explore the social roots of linguistic features that were widely generalized in the Early Modern English period and later found their way into mainstream modern English dialects as well as Standard English. For that purpose we adopted a variationist sociolinguistic approach and analysed individual letter writers and letter recipients in their social contexts, paying close attention to their biographical data: year of birth, domicile, gender, social status, social and regional mobility, level of education and family circumstances. We also recorded for each letter the correspondents' mutual relations, distinguishing intimates from more distant relationships.

The linguistic changes we examined were either morphological, such as changes in pronoun use, or morphosyntactic, involving syntactic patterns such as negative concord. These changes were approached as linguistic variables, i.e. speaker choices between alternative ways of expressing the same meaning or syntactic function. By applying a variationist approach to our data we were able to ask questions about the highly variable social embedding of processes of language change and about the kinds of people who either promoted or avoided linguistic innovations.

We found that participation in ongoing changes typically correlated with the writers' regional background, and their gender and social status. The changes we studied had variable regional origins and showed different patterns and rates of diffusion. The vast majority of the incoming forms and functions were led by female writers, and those that were led by male writers, such as the disappearance of negative concord, could be related to gendered differences in educational opportunities.

At the methodological level, we also asked questions concerning the ways in which our real-time approach could be related to the apparent-time model that has consolidated its place in modern sociolinguistic studies on language change. To do that, we traced processes of change on two planes simultaneously: in real time and in apparent time. This analysis looked, for example, at the extent to which successive 20-year age cohorts participated in ongoing changes divided into successive 20-year periods. Although it was not possible to analyse all fourteen changes at this level of empirical detail, the results suggest that morphosyntactic changes can propagate both communally and generationally. Importantly, we could also show how some people changed their grammar during their lifetime while others did not.

### 1.3 Trajectories of change between 1400 and 1680

The study of fourteen different processes of change showed that an extended period would be needed to trace their diffusion across time and space, both geographic and social. Only half a dozen were completed within the 270-year time span studied using the CEEC. This was by no means an unexpected result: some of the processes began later than others, for example, the rise of the prop-word *one* in the late 16th century and the possessive determiner *its* in the 17th century. By contrast, some of those that were completed, such as the replacement of the relative pronoun *the which* by *which*, had already reached its mid-course in the 15th century.

However, an early start did not always guarantee the completion of the process in the study period as changes displayed greatly varying rates of progress. Syntactic processes typically took longer than the arguably less complex processes of morpheme replacement, for example. So the traditional pronoun *ye* was replaced by the object form *you* in the subject function (*ye go* > *you go*) in the course of the 16th century, whereas in gerund constructions, the direct object (*writing the letter*) had started to gain ground at the expense of the *of*-phrase (*writing of the letter*) in the 15th century, but the process was still under way towards the end of the 17th century.

Most of the linguistic processes analysed in Nevalainen & Raumolin-Brunberg (2003) involved two variant forms. Those with more than two tended to proceed at a slower pace. This is the case, for example, with the noun subjects of the gerund (genitive, *of*-phrase and the common case: *Mary's writing*, *the writing of Mary*, *Mary writing*) and indefinite pronouns with singular human reference (compound forms with *body*, *one* and *man*, and the simple forms *any*, *some*, *none*, *every*, and *each*).

In various ways the linguistic environment emerged as a factor affecting the rate of change even with variables that had only two variant forms. A case in point is the loss of the nasal in first- and second-person possessive determiners (*mine/thine enemy* > *my/thy enemy*), which, in linguistic terms, was phonologically determined. The change was first completed in contexts where the determiner preceded a consonant and next where it preceded an initial <h>, with prevocalic contexts taking the longest time to generalize *my* and *thy*. One particular lexeme, *own*, followed the trajectory of the prevocalic instances but at a slower pace still.

In lexically and syntactically conditioned processes, the frequency of a given context would affect the rate of change. High-frequency contexts can be expected to have what Bybee (2010: 24–25) calls a Conserving Effect and resist or delay an ongoing process of change. We traced the change of the third-person singular present-tense indicative suffix from *-(e)th* to *-(e)s* (*it maketh* > *it makes*) and observed that the *-th* form persisted longest with the most frequent verbs such as *have* (*hath*) and *do* (*doth*). Hence the process diffused lexically from verb to verb and

was not completed in these frequent verbs by the end of the 17th century (Ogura & Wang 1996).

The grammaticalization of *do*-support illustrates a syntactically conditioned process of change. The role of frequency may not be easy to account for in this case but it is interesting to observe that as the use of periphrastic *do* was diffusing in affirmative statements in the 16th century it progressed at a proportionally lower rate than in negative statements or questions and imperatives. However, since affirmative statements are the most common sentence type, the text frequency of *do* in this context was higher than in other sentence types (Nurmi 1999: 23–24). Although we may be left wondering whether the overall high frequency of affirmative statements might have had a role to play in the low relative frequency of *do*, it does not explain the demise of periphrasis in this context in the course of the following two centuries.

Modified from Nevalainen & Raumolin-Brunberg (2003: 205), Table 1.1 lists the fourteen changes studied and sums up the stage each of them had reached by the end of the study period, 1680. The possessive determiner *its* was the only one to be introduced in the 17th century; the other changes dated to the 16th century or earlier.

**Table 1.1** Stages reached by fourteen processes of change by about 1680.\*

Process of change	Stage by 1680 in the CEEC
1. Subject <i>YE</i> → <i>YOU</i>	Completed
2. Possessive determiner <i>MINE/THINE</i> → <i>MY/THY</i>	Completed
3. Possessive determiners <i>HIS, OF IT, THEREOF</i> → <i>ITS</i>	New and vigorous
4. Prop-word <i>ONE</i>	New and vigorous
5. Direct object of the gerund	Nearing completion
6. Noun subject of the gerund	Nearing completion
7. 3rd person sg. pres. ind. <i>-(E)TH</i> → <i>-(E)S</i>	Completed in most linguistic environments
8. Decline of <i>DO</i> in affirmative statements	Nearing completion
9. Rise of <i>DO</i> in negative statements	Mid-range
10. Negative concord → single negation	Completed in most linguistic environments
11. Inversion after clause-initial negators	Completed
12. Relative <i>THE WHICH</i> → <i>WHICH</i>	Completed
13. Relative adverb → prepositional phrase	Mid-range
14. Compound indefinite pronouns with singular human reference	New and vigorous

\* For a discussion of these stages, see Section 5.1.2, “Periodizing processes of change”.

It should be noted that the overview in Table 1.1 does not say anything about the time courses of the fourteen changes in regional or social terms, or indeed at the level of individuals. Various user-related parameters can be shown to differentiate their temporal trajectories a good deal further. Nevalainen and Raumolin-Brunberg (2003) specify how the diffusion of a given change can be traced both horizontally in geographic space and vertically in social space, and how these two can mix and blend in a variety of ways over time. It is only when a change comes to completion that such distinctions can no longer be observed in the data.

In the present volume the four processes shown in bold in Table 1.1 will be followed into the 18th century, including the determiner *its*. Four other processes of change will also be discussed: one that was completed for most users by the turn of the 18th century (the 2nd-person singular pronoun *thou*) and one that was only incipient at the time (the progressive aspect), as well as two productive word-formation processes used to derive abstract nouns from adjectives (*-ness* and *-ity*).

#### 1.4 Aims and scope of this volume

In view of the great variety in the time courses of linguistic changes, one of the key topics we address in this volume is the sociolinguistic embedding of long-term processes of change. We are, first of all, interested in the rate at which ongoing changes progressed in the 18th century. While most of the processes discussed in Section 1.3 followed the expected S-shaped pattern of diffusion over time (cf. Nevalainen 2015), the temporal trajectory of periphrastic *do* in affirmative statements made a notable exception to this empirical generalization.

Another question that we are interested in is the extent to which comparable social conditions obtained in the different phases of a long-term process of change. We start with the uniformitarian premise that the present can in principle be used to explain the past, provided that we remain alert to the pitfalls of anachronism (Bergs 2012). As gender differentiation is one of the most robust findings in present-day sociolinguistic studies of language change and is, as noted in 1.2, attested in real-time changes in Early Modern English as well, our hypothesis is that gender differences will also emerge in the 18th century data. Following the same line of reasoning, we will look into the roles played by other factors, such as social stratification and, where available, information on the writer's regional background and age. The extent to which these can be reconstructed naturally depends on the phase of the change: processes that are nearing completion provide less variable data than those in mid-range.

Social change also affects the degree to which comparable social conditions can be reconstructed for long-term processes of linguistic change. In concrete terms, social change has an impact on the primary data as, for example, the rate of full literacy improves over the centuries and later data provide the researcher with access to a wider range of the population than earlier data sources. We discuss various aspects of 18th-century society in Chapter 2, concluding that as a result of the fundamental social, cultural, religious and economic transformations that took place within the long 18th century, this period has closer connections with the centuries to come than those preceding it.

Importantly for linguistic research, the 18th century uniquely differs from earlier times in that it witnessed a rapid rise of normative grammar and increased language awareness. Prescriptivism – and language codification in general – constitutes a source of potential influence to be accounted for, as it could have impacted on changes in progress, especially if these were socially indexed. By way of background, the evolution of 18th-century normative grammar is discussed by Yáñez-Bouza in Chapter 3. The work she surveys suggests an impact of normative precepts on individuals, and proscription on variation in general, but little or no direct influence on ongoing processes of change. Possible exceptions are some rapid changes and those in their final stages (see also Auer & González-Díaz 2005).

To observe how these generalizations – unbroken temporal trajectories, social embedding and varying susceptibility to normative influence – materialized in the course of linguistic changes we will consider changes in their various stages, ranging from incipient to those nearing completion or practically completed. The final stages of three changes will be discussed: the repurposing of the recessive 2nd-person singular pronoun *thou* (Chapter 6), the demise of the 3rd-person singular verbal suffix *-th* in high-frequency verbs (Chapter 7), and the outgoing uses of periphrastic *do* in affirmative statements (Chapter 8). Although the processes we study in these chapters focus on outgoing forms we naturally also account for their mainstream alternatives, the 2nd-person pronoun *you*, verbal *-s* and *do*-less affirmative statements. It is shown that in all these cases the outgoing variants were associated with particular style- and register-related effects.

The outgoing forms we analyse further include the indefinite *man*-compounds and independent forms with singular human reference. Their decline will be related to the simultaneous rise of the compound indefinite pronouns with *one* and *body* (Chapter 9). We will also trace the generalization of the possessive determiner *its*, a process that diffused from a new and vigorous stage to near completion in the course of the 18th century (Chapter 10). Both these variables display several variant forms and involve a reorganization of their respective paradigms from four or more elements to one or two alternatives. The paradigm of singular indefinite pronouns is reduced to two (*-one* and *-body* compounds) and the four-member

class of possessive determiners (*his, of it, thereof, its*) comes to be represented by one prototypical form, *its*, with *of it* remaining as a minor variant. Because of their internal complexity, these two variables offer a good testing ground for the sociolinguistic hypotheses laid out above.

Incipient changes are represented by the diffusion of the progressive aspect in the course of the 18th century (Chapter 11). In its early stages, the progressive offers a window on the social roots of an innovation, which are not always easy to pin down in categorical terms. Moreover, it presents several options for identifying the variants of the relevant linguistic variable, the possible alternative ways of saying the same thing. Although well-established quantitative methods have been used in the past to show how the progressive diffuses over time, such indeterminacies have methodological implications as well. They are not unlike those encountered in the study of periphrastic *do* in affirmative statements in Chapter 8.

We will therefore also address the question of how to quantify change in processes that are not easily defined in terms of linguistic variables. For example, should a pair of productive suffixes that apply to the same types of base forms be automatically analysed as interchangeable linguistic alternatives, i.e. realizations of the same linguistic variable? This analysis is usually adopted with inflectional suffixes such as the 3rd-person singular present-tense indicative verbal endings, the justification being that they carry the same meaning and grammatical function. Similar reasoning is often applied to the derivational suffixes *-ness* and *-ity*, which derive abstract nouns from adjectives. We do not make this assumption but consider these derivational suffixes separately, showing their differential uses and applying innovative quantitative methods to examine their sociolinguistic patterning in the 18th century (Chapter 12). The same techniques are applied in the studies on periphrastic *do* and the progressive aspect.

## 1.5 Material, methods and syntheses

All our studies make systematic use of the 18th-century *Extension of the Corpus of Early English Correspondence* (CEECE), which picks up from where the earlier corpus ends and spans the “long” 18th century from 1680 to 1800. This means that we have the benefit of being able to observe how the same set of people, who represent different social strata, participate in different processes of change. Chapter 4 introduces the letter corpus and its composition in detail. The discussion includes, for example, the coverage of the social categories sampled per period. The letter collections included in the corpus are documented in the Appendix.

We use the same basic periodization in our studies, starting with six generational 20-year periods, but can also combine them into three 40-year periods,

depending on the frequency and stage of diffusion of the linguistic elements investigated. Both periodization and the different quantitative methods we have applied and adopted for the purposes of this volume are the topic of Chapter 5, where we discuss basic methods for estimating frequencies as well as quantitative methods for studying changes lacking a readily definable linguistic variable.

To further evaluate the empirical basis of our findings, we provide a comparison of the text frequencies of the changes we studied using the letter corpus in Chapter 13. Moving on to possible broader generalizations, the information provided by the *Google Books* database on these processes of change is also considered and critically evaluated in that chapter.

The last three chapters provide overviews of our findings. Focusing on variation in individual participation in linguistic change, we were interested in those letter writers who could be singled out as particularly progressive or conservative with respect to ongoing change. Chapter 14 draws together our findings on this aspect of variation in sociolinguistic terms. In Chapter 15 we move on to summarize the principal results of the linguistic changes studied. Finally, Chapter 16 relates our findings to broader sociolinguistic generalizations.

## Society and culture in the long 18th century

Terttu Nevalainen

### 2.1 Introduction

IT WAS the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way ...

(Charles Dickens, 1859, *A Tale of Two Cities*)

The opening lines of Charles Dickens's *A Tale of Two Cities* refer to the year 1775 in London and Paris. This year witnessed the beginning of the American War of Independence and the following decade led to the French Revolution in 1789. Both events transformed the history not only of their respective countries and their capital cities but also of the rest of the world.

In English historical research the “long” 18th century has various beginning and end points. The beginning has been set as early as the Restoration of the monarchy in 1660, which marked a significant cultural and political watershed (e.g. Baines 2004, Beal 2004). Another often cited beginning point is the “Glorious Revolution” of 1688, which marks the overthrow of the last Catholic monarch of the country, leading to major constitutional and political changes in the whole of Great Britain (e.g. Black 2008). The suggested end points of the century also vary considerably. For Langford (1998) and Black (2008) it is the year 1783, when the British acknowledged the independence of the thirteen American colonies that had rebelled in 1775. A literary scholar, Baines (2004) ends the century at 1790, while the end point at 1800 can be politically justified as marking the Act of Union with Ireland. Other, later years of demarcation have been considered in political, economic and cultural terms (Black 2008: 1–3).<sup>1</sup>

---

1. Continuity between the 18th and 19th centuries is emphasized by historians who, like Steinbach (2005), view the years between 1760 and 1914 as a “long 19th century”. She justifies the beginning of her period in terms of developments such as industrialization, urbanization, and dominance of the middle class, all characteristic of the 19th century, but in evidence from the 1760s onwards.

For our purposes, the period from 1680 to 1800 is motivated in practical sociolinguistic terms: apart from continuing from where we left off in Nevalainen & Raumolin-Brunberg (2003), it can be flexibly divided into six 20-year generations or into three longer 40-periods, depending on the progress of a particular linguistic change. If one overarching issue of linguistic relevance were to be sought for, it would be prescriptivism, the hallmark of the period. The immediate roots of 18th-century normative grammars can be traced back to the late 17th century and texts such as John Dryden's "Defence of the Epilogue" (1672b). Because of the unique importance of the prescriptive movement, we have devoted a separate chapter to it.

Despite the radical developments that shaped the latter half of the long 18th century, historians typically draw the line between Early and Late Modern times somewhat earlier. Forggeng (2007: ix) goes as far as to suggest that "[t]he life of an Englishman in 1700 was in many ways closer to the lives of his descendants in 2000 than his predecessors in 1500". Examining the extent to which this claim is justified, this chapter considers the period between 1680 and 1800 in terms of both continuity and change in Britain.

To outline the social and cultural background of the long 18th century, I will begin by discussing some historical events, such as the changes that took place on the throne of the Kingdom of Great Britain (prior to 1707, the Kingdoms of Scotland and England), and then continue with social and demographic issues that are harder to pin down with any one set of facts. This is the case with social structure, which can be discussed from different perspectives and at various levels of abstraction (Nevalainen & Raumolin-Brunberg 2003: 32–37). However, what makes the period between 1680 and 1800 different from earlier times is that we now, for the first time, have contemporary population statistics and assessments of issues such as the average family income.

Although the focus of this book is on "English" English, it is worth noting that the role of English as a global language gained momentum in the 18th century. Adding the estimates of English speakers in England and Wales to those in Scotland, Ireland and North America, Bailey (2010: 185) concludes that the number of English speakers worldwide increased from about nine million in 1700 to over 20 million by 1800. It is also noteworthy that immigration was intensified as the expanding foreign trade attracted foreigners to Britain, particularly to London and to the ports in the west of England and Scotland.<sup>2</sup>

---

2. Hickey (2010) presents a timeline of the various events that made history, both political and cultural, between the Glorious Revolution in 1688 and the accession of Queen Victoria in 1837.

## 2.2 Political life

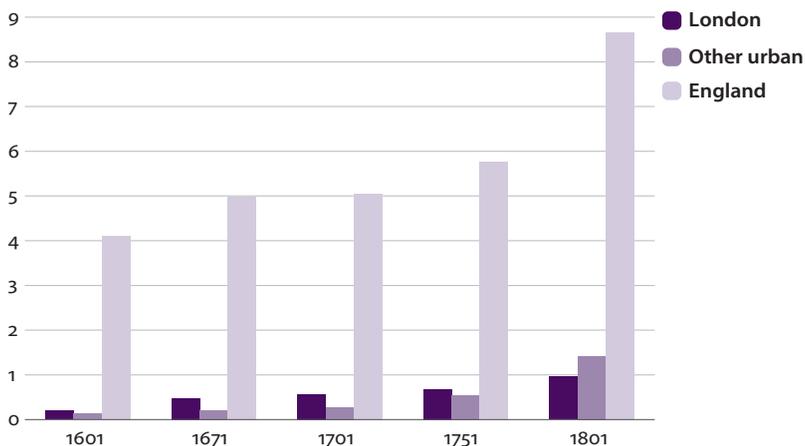
Thoroughgoing changes took place in the monarchy and in royal powers in the long 18th century. The Catholic King James II (1685–88) was deposed in the “Glorious Revolution” and replaced in 1689 by the joint protestant monarchs William, the Prince of Orange, and Mary, James’s daughter by his first marriage. Mary II died in 1694 and William III in 1702. He was succeeded by his sister-in-law Anne (1702–14), the last Stuart monarch on the English throne, in turn to be succeeded by her second cousin George I, a member of the German electorate of Hanover. During this period, the Whig and Tory parties were established in parliamentary politics. Tory supporters were associated with the church and the rural gentry, whereas Whig supporters had city connections, including businessmen, civil servants and naval officers.

In the reigns of Queen Anne’s Hanoverian successors, George I (1714–27) and George II (1727–60), Britain was governed by Whigs. During George III’s reign (1760–1820), Tories gained more influence and seats in Parliament but, despite their rivalry, the two parties were united in keeping the gentry in power. While the power of Parliament increased, royal power receded. Olsen (1999: 1) refers to 18th-century Britain as a “Nation of Prime Ministers” (Olsen 1999: 1–12; Forggeng 2007: 10–15).

In actual practice, the political life of the period was heavily influenced by the role played by the middle classes in industry, commerce, local politics and civil society at large, and the pressures they exerted on their landed representatives in Parliament. As Daunton (1995: 200) points out, representatives of the middling ranks were not only managing their individual firms but also largely running the factory towns of the Industrial Revolution. It was therefore in the interest of the powers that be to meet the needs of the country’s rapidly growing sector of industry and commerce.

## 2.3 Demography and urbanization

The long 18th century is characterized by rapid population growth in England and Wales, from five million in the last decades of the 17th century to close to nine million at the turn of the 19th century. The century was also an age of continued urbanization. The proportion of the total population living in towns with 10,000 or more inhabitants quadrupled between 1600 and 1800. It is noteworthy that by the late 18th century, London’s population was more numerous than that in all other towns with more than 5,000 inhabitants put together (Figure 2.1). In the course of the 18th century London grew to be the largest city in Europe, with 11% of the population of England and Wales living in London in 1750 (Daunton 1995: 136–138; Wrigley 2010: 61).



**Figure 2.1** Estimates of London, other urban and the national population (millions), 1600–1800 (based on Wrigley 2010: 61)

London's growth was not the result of any single factor. It was characterized by decline in child mortality over time, and internal and international migration (both economic and forced), creating a city of young people with a large proportion of women in domestic service. At the end of the 17th century specific neighbourhoods were associated with migrants from Yorkshire, Scotland and Ireland as well as from the Continent, and a hundred years later the American War of Independence gave rise to a large American loyalist community. In the last decades of the 18th century London's black population amounted to some 5,000 to 10,000 emigrants from Africa, the Caribbean and North America (Emsley, Hitchcock & Shoemaker 2013).

The rapid growth of other cities towards the end of the 18th century was largely due to industrial centres such as Manchester, Birmingham, Leeds and Sheffield, and ports such as Liverpool and Hull (Wrigley 2010: 60–63). However, despite the remarkable population growth, life expectancy at birth did not rise dramatically in England but continued to fluctuate roughly between 34 and 37 years throughout the long 18th century (Daunton 1995: 575, based on Wrigley & Schofield 1981: 230).

Connected with urbanization, industrialization was a megatrend with far-reaching consequences in the Late Modern period. Historians argue about its beginnings, but its effects and consequences are beyond dispute. A complex network of interconnected factors came to link issues like London's growth and the Industrial Revolution in England. They include three kinds of change: (1) economic changes (the creation of a national market, new methods in agriculture, changes in raw material supply, better commercial and credit facilities, an improved transport network and higher real incomes), (2) demographic changes (favourable demographic conditions leading to controlled population expansion), and (3) sociological changes

(moving from “traditional” to increasingly “rational” mindsets of individuals, new forms of social mobility and new consumption patterns) (Wrigley 1987: 149–156).

Among factors that made a major contribution to the process of industrialization, special attention should be paid to the rising agricultural productivity, which helped feed the increasing non-agricultural population, and the availability of cheap labour in the countryside supplementing their income by means of various industrial by-employments (Hay & Rogers 1997: 115–116). These labourers, cottagers and paupers made up a sizable section of the population throughout the period (see Table 2.1, discussed below). Urban growth was made possible not only by a

**Table 2.1** Contemporary estimates of distribution of average annual family income in England and Wales in 1688 and 1801–03; cumulative percentages of population, beginning with the highest income group (based on Hay & Rogers 1997: 19–21)

<b>Gregory King 1688</b>	<b>Patrick Colquhoun 1801–03</b>
Temporal lords (£6,060–), Baronets, Bishops Knights, Esquires, Greater merchants Gentlemen, Persons in greater offices (–£240)	Peers (£8,000–), Bishops, Baronets Eminent merchants, Knights, Esquires Higher civil officers, Lesser merchants (–£800)
<b>1% of population</b>	<b>1% of population</b>
Lesser merchants (£200–) Artisans and handicrafts Law, Persons in lesser offices Greater freeholders (–£91)	Capital manufacturers (£800–), Warehousemen Gentlemen, Capital ship repair, Teachers Eminent clergy, Shipowners, Lawyers Liberal arts, Lesser civil offices (–£200)
<b>5% of population</b>	<b>5% of population</b>
Naval officers (£80–), Greater clergymen Military officers, Science, Liberal arts Lesser freeholders, Lesser clergymen (–£50)	Better sort freehold (£200–) Professional skills Theatre, Shopkeepers (–£150)
<b>10% of population</b>	<b>10% of population</b>
Shopkeepers, Tradesmen (£45)	Educating youth (£150–), Tailors Naval officers, Military officers Lesser clergy, Farmers (–£120)
<b>20% of population</b>	<b>20% of population</b>
Farmers (£42.5–), Manufacturing trades Building trades, Common seamen Miners (–£15)	Dissenting clergy (£120–), Innkeepers Lesser freeholders, Clerks, Shopmen Artisans (–£55)
<b>40% of population</b>	<b>50% of population</b>
Labouring people, Outservants (£15)	Halfpay officers (£45–), Seamen, Mine labourers
<b>60% of population</b>	Hawkers, Pedlars, Labourers
Common soldiers (£14–) Cottagers and paupers Vagrants (–£2)	Soldiers, Pensioners Paupers Vagrants (–£10)

sufficient supply of food but also through available sources of energy, including large volumes of domestic fuel for cooking and heat. Wrigley (2010) even puts forward a case suggesting that industrialization in Britain was largely enabled by the country's energy supplies, notably coal.

## 2.4 Social stratification

Many developments in the social structure of the long 18th century are reflected in the figures compiled by two pioneer population statisticians, Gregory King for 1688 and Patrick Colquhoun for 1801–03, based on the first census in 1801, income tax returns and the 1802–03 survey of paupers. Table 2.1 is a comparison of the social categories that the two statisticians presented, grouped according to estimated average annual income levels into cumulative percentages of the total population in England and Wales (the highest and lowest income estimates are presented for each level).<sup>3</sup>

Both statisticians witnessed the use of new social category labels in the manufacturing section. King in fact appears to have underestimated them in technologically simple trades in the country (Hay & Rogers 1997: 115). Colquhoun's account includes capital manufacturers, warehousemen, capital ship repair and ship owners, who come high up in the income hierarchy, among the top 5% of the population. Towards the other end of the income hierarchy, we find clerks and shopmen, and lower still, mine labourers, hawkers and pedlars.

The relative positions of a number of other categories have also shifted in the income hierarchy. Upward mobility is shown by lesser merchants (to the top 1%), eminent clergy and liberal arts (to the top 5%), and farmers (to the top 20%). However, more categories are placed in lower income brackets compared to 1688: lower gentry (gentlemen) among the top 5%, better sort of freehold among the top 10%, naval and military officers and lesser clergy among the top 20%. Common seamen also show downward mobility and, most dramatically – to the extent that the categories match – artisans, who fall from among the estimated top 5% down to the 50% category.

---

3. Both accounts, but King's in particular, have been subject to later refinements and revisions; see Lindert & Williamson (1982) and Arkell (2006). The information provided by Hay & Rogers (1997) is presented here. Following the revised estimates of Lindert & Williamson, they give, for example, artisans a considerably higher income estimate than King originally did. In King's original, artisans were in the same income bracket as farmers.

One is left with the conclusion that at the end of the long 18th century neither land owning nor the army or the navy were as secure sources of income as they had been a hundred years earlier. But the categories covering the top 1%, on the one hand, and the poorest half of the population, on the other, consisted of much the same categories as a hundred years earlier. However, it is worth noting that these developments were not linear: income distribution appears to have been less unequal in the middle of the 18th century than at the turn of the 19th (Hay & Rogers 1997: 18, 225; see also Fitzmaurice 2012: 297–300).

The language of “class” came to be established in the course of the 18th century, intersecting with the older language of “rank” and “degree”. Samuel Johnson, for example, recognized the term by defining *class* as ‘rank or order of persons’ in his *Dictionary of the English Language* (1755) (Cannadine 2000: 31). Corfield (1991: 129–130) sums up this transition by saying that “[p]ower was resynthesized into active terms of acquisition, production, and display, rather than of inheritance, formal title, and ancient lineage”. Land and status no longer automatically guaranteed the gentry a position of power in society, which began to be measured in terms of personal wealth. A tripartite social division based on the sources of wealth and status – land, capital and labour – is typically found in contemporary writings (Corfield 1991: 127).

As Table 2.1 suggests, it was the trading, manufacturing and professional classes that were largely responsible for Britain’s economic growth in the 18th century (Wrightson 2002: 229). For continuity’s sake, our socio-economic classification of CEEC writers distinguishes between the gentry (including royalty and the nobility), the clergy, professionals, merchants and other non-gentry (for details, see Chapter 4). As in the previous centuries, the CEECE professionals include lawyers, government officials, medical doctors, teachers and army officers, augmented in the 18th century by numerous representatives of the liberal arts (Prest 1987).

Merchants form a separate category, consisting mostly of wealthy tradesmen. However, despite ranking high in the income hierarchy, social mobility did not automatically follow from wealth, and merchants did not have easy access to the polite society of the day. As Wrightson (2002: 290) points out, capital came in three forms, of which the economic sense of money or goods was only one. The two other forms, which could be turned into economic goods, were cultural and social capital. Cultural capital consisted of acquired skills, knowledge and demeanour, and social capital of connections to social networks of various kinds. All three needed to be nurtured by successful businessmen and professionals alike.

## 2.5 Literacy

The labouring classes are underrepresented in the CEECE largely because of their low level of full literacy, which in turn is reflected in the number of edited letter collections available. However, as pointed out in Chapter 4, more varied edited material is available from the end of the 18th century. Based on marriage register data, Cressy (1980: 176–177) estimates that in the mid-18th century some 60% of men and 40% of women in England and Wales were at least minimally literate in that they could sign their name. However, Cressy emphasizes that literacy rates remained highly stratified both socially and regionally.

The rate of what is considered full literacy, reading and writing skills, was higher in big cities than in the countryside, and much higher among the upper and middle classes than in the lower ranks. Literacy was graded in that reading skills were more widespread than writing skills across the social spectrum. Porter (2000: 346) also mentions the contemporary view that education for the poor should be strictly functional, citing the school teacher and educational writer Hannah More, who was against teaching writing to the poor (see also Steinbach 2005: 175–176; Fairman 2015: 62).

Adopting a broader view of literacy, several finer gradations emerge. Munck (2000: 47) presents the situation in more detail:

Beyond basic reading might come actual numeracy, for those who needed it for professional and occupational purposes; high-quality writing skills for those entering clerical professions; fluency in rhetoric, Latin and philosophy for those who had intellectual aspirations; additional language skills for those wanting to succeed in eighteenth-century salon society; or familiarity with appropriate local dialects for those wanting a career in public administration and the church.

He concludes that, seen in this light, literacy may appear specific to individual occupations rather than something universally desirable. What interests us in this volume in particular is *epistolary literacy*, that is, skills in letter-writing practices. These, too, were graded, and gender and class emerged as significant factors (e.g. Brant 2006: 33–34).

Whyman (2009: 32–33) notes that a classical education gave elite boys epistolary advantages over their sisters. She observes that in women's letters phonetic spelling, poor grammar and awkward presentation persisted until the mid-18th century, when the number of girls sent to school began to increase. Elite families regarded letter-writing as the kind of social and cultural capital that was required for functioning in a society heavily reliant on social networks. Whyman's case studies show that non-elite parents also appreciated letter-writing skills and that merchants, in particular, would take great care in instructing their children, sons

and even daughters, in epistolary literacy. The models they used came predominantly from authentic letters, but bibles, business records, copybooks and popular literature also provided models for young writers (Whyman 2009: 45; see also Brant 2006: 9–10; Sairio & Nevala 2013).

On the other hand, literacy and occupational status did not necessarily always correlate. Fairman (2015: 67) considers this issue for shopkeepers and artisans. Shopkeepers represent the middle social order in Colquhoun's categorization but show wide variation in schooling: those supplying the higher ranks could be grammatically schooled, whereas the majority, who supplied the lower ranks, often testified to mechanical schooling and the use of non-Standard English.

## 2.6 Cultural climate

Having surveyed the political and social divides of 18th-century Britain, it is easy to agree with Black (2008: 161) that there is no single approach to the culture of the period either. In this section only one aspect of the cultural climate of the time will be discussed, namely, the wide-ranging concept of *politeness*.

With the Restoration of the monarchy in 1660, the Royal Court regained a position as the centre of social refinement. The poet laureate John Dryden connected prestige with refined conversation and referred to the King and “the most polish'd courts of Europe” as the stylistic models to be followed by poets and critics (*The Conquest of Granada by the Spaniards* 1672a: 173). However, Dryden's own living, and that of many other writers, was already divided between court patronage and the widening reading public outside the Court. This wider audience became major consumers of cultural ideals in the long 18th century and no longer looked up to the monarchy in the same way as in the past (Adamson 1999: 597–598; Klein 2002: 879; Langford 2002: 313). Nevertheless, politeness became the dominant cultural paradigm in the course of the long 18th century.<sup>4</sup>

Klein (2002: 870) sums up the various ramifications of 18th-century politeness by observing that it was used to refer to “material and visual cultures, the organization of space, the constitution of social and political identities, the character of intellectual and artistic life, and even institutional structures”. Fitzmaurice (1998) shows in her study of the rise of “polite English”, notably manifest in *The Spectator* (1711–14) and other literary periodicals, how intimately the pursuit of politeness was related to language in the age of Enlightenment. Watts (2002) explores how

---

4. The following discussion draws on my more detailed exposition of the shifting politeness cultures in Nevalainen & Tissari (2010).

“polite English” came to be equated with “standard English” precisely because the ideology of standardization was part of the wider 18th-century ideology of politeness.<sup>5</sup>

Cultural historians distinguish two more specific modes of politeness with somewhat different agendas: the “*Spectator* mode”, propagated in periodicals, fiction and didactic literature, and the “Shaftesbury mode”, expressed in the philosophical writings of Anthony Ashley Cooper, the 3rd Earl of Shaftesbury (1670–1713). Klein (1994: 8–9) suggests that Shaftesbury advocated sociability and politeness to counter the traditional intellectual authority of the Church and the Court and to shift the guardianship of culture from the Tories to his own Whig party with its country, puritan and radical associations. Shaftesbury’s discussion was aimed at the social and intellectual elite. Langford (2002: 312) observes that, by contrast, the leading lights of the *Spectator* group – Joseph Addison, Richard Steele and their followers – used the print medium to advocate polite moralism and a code of behaviour for a wider audience, basically promoting social mobility.

However, at least to begin with, the *Spectator* mode of politeness provided a means for relatively well-placed people to empower themselves. Although people from different social backgrounds may have pursued politeness differently and for a variety of reasons, the dividing line separating polite society from the rest largely followed the boundary between the propertied and the non-propertied (Klein 2002: 896–897; Langford 2002: 312). In a personal letter included in the CEECE, excerpted in (2.1), Mary Wollstonecraft’s account of the people she had met supports the argument that the traditional gentry/professions line was not crossed even towards the end of the century: a butler can be “beyond measure civil and attentive” but, unlike a clergyman, not someone to be characterized as “polite”.

- (2.1) agreeable companion – a young Clergyman, who was going to settle in Ireland, in the same capacity as myself. He was intelligent and had that kind of **politeness**, which arises from **sensibility**. My conductor, was beyond measure **civil** and attentive to me, he is a good sort of a man, I was, at first, at a loss to guess what department he filled in the family; but I find now he is the Butler, and his wife the housekeeper. (Mary Wollstonecraft, 1786?; WOLLSTONECRAFT, 123)<sup>6</sup>

---

5. Standard language ideology permeated Western European thought from the 17th century on. For further discussion, see e.g. Deumert & Vandebussche (2003), Milroy & Milroy (2012 [1985]) and Rutten, Vosters & Vandebussche (2014). Double negation was among the constructions heavily proscribed by this normative tradition (see e.g. Nevalainen & Rutten 2012).

6. All quotations from the *Corpora of Early English Correspondence* have a text identifier consisting of the writer’s name followed by the year of writing, a short title of the collection where the letter comes from and a page reference. For detailed bibliographical information, see the Appendix.

New distinctions emerged under the general umbrella of politeness as the century wore on. The notion of “sensibility” – referred to by Wollstonecraft in (2.1) – gained ground in the mid-century, forging a strong link between inner virtue and outer manners (Carter 2001: 27–32). However, “civility”, a central cultural notion in the 16th and 17th century, was often embedded in the general discourse of politeness throughout the 18th century. *Civil* was the word used by the commentator Joseph Baretta as he predicted the improvement of the manners of the common people in England in 1770: “in about twenty years more they will become quite as civil to strangers as the French and Italians” (cited in Langford 2002: 322).

A social divide was also manifest in the extent to which the lexicon of politeness was used by literate people. Nevalainen & Tissari (2010) analysed the use of three related politeness terms in the CEECE, *civil/civility*, *polite/politeness* and *courteous/courtesy*, finding that it was stratified both by social class and gender, *polite* and its derivatives being much less frequent among merchants and the lower classes than among the gentry and professionals. In his study of the use of *civil* and *polite* in the *Old Bailey Records*, Shoemaker (2004: 291–296) found that both *polite* and *civil* were marginal in his 50-million-word corpus, *polite* being even less frequent than *civil*. He concluded that the segment of London population dealt with in these court records was rarely designated as *civil*, let alone *polite*, either by their peers or by the magistrates they came into contact with.

## 2.7 Conclusion

The 18th century marks the beginning of the Late Modern period in history books and language histories, suggesting closer connections with the centuries to come rather than those preceding it. However, the 18th century was also characterized by substantial social changes. Langford (1998: 679) closes his comprehensive survey which covers the years from 1727 to 1783 by observing that the social, cultural, religious and economic transformations that took place during that period were spectacular. He sums up his conclusion by saying that “the polite and commercial people of the 1730s was still polite and commercial in the 1780s: it did not, in any fundamental sense, inhabit the same society” (Langford 1998: 725).

In this volume we will examine the degree to which language change can be viewed against this socio-cultural background, focusing next on the rise of normative grammar as one of the major reflections of the notion of politeness as it was understood in 18th-century society.

## INFOBOX 1

**Range of writers in the CEECE**

Terttu Nevalainen

The *Corpus of Early English Correspondence Extension* (CEECE) contains royal correspondence by **King George III** and his family – he and Queen Charlotte had 15 children – as well as letters by his first love, **Lady Sarah Lennox**. Whig prime ministers are represented by the private letters of **Sir William Pitt** the Elder, some written when he was courting his wife-to-be **Hester Grenville**. Letters to and by **Georgiana Cavendish, Duchess of Devonshire**, who was influential in the Whig politics of the time, are also included.

At the opposite end of the social spectrum there are paupers seeking aid from their next-of-kin or home parish. These include the widow **Mary Pavett** writing, possibly by proxy, to her brother-in-law **Henry Pavett** asking him to contribute to his brother's burial costs by buying him a coffin. Henry declines and refers Mary to his brother's home parish. Somewhat better off is the **Clift family**, children of a Cornish miller, who all taught themselves to write. The poor writing skills of the eldest of the siblings, **Elizabeth**, were remarked upon by **William**, the youngest, who moved to London and rose up in the world, becoming the curator of the vast Hunterian collection of surgical specimens, while Elizabeth continued to work in Cornwall as a domestic servant.

The increasing numbers of merchants and manufacturers are represented by a range of writers. **William Bolton**, of Warwick, sailed to Madeira and exported Madeira wine to North America and the East and West Indies in his own ship. **Richard Champion**, a Bristol merchant and porcelain manufacturer, had trading links to North America, the West Indies and continental Europe; after losing his fortune, he moved permanently to South Carolina. Manufacturers also include small-time traders like **Nathaniel Pinney**, who was content running his lace business from his home in Dorset. The business was later taken over by his youngest sister **Hester** in London.

Foreign cultures and countries became an object of interest to correspondents across the social spectrum. **Lady Mary Wortley Montagu**, for example, travelled with her husband on a diplomatic mission in the Ottoman Empire. **Sir William Jones**, a polyglot philologist and orientalist, studied ancient India and made the connection between Sanskrit and classical Greek and Latin. **Daniel Defoe**, a trader, pamphleteer and the author of *Robinson Crusoe*, served as an intelligencer for **Robert Harley**, an eminent Tory politician, reporting to him on Ireland. The writing profession is further represented by the novelists **Frances Burney** and **Jane Austen**.

Few questioned slavery at the beginning of the 18th century but the situation changed in its course. The corpus includes letters by **Ignatius Sancho**, an African born on a slave ship in 1729, who died in London in 1780 a notable man of letters. Various CEECE correspondents in the latter half of the century opposed slavery, among them industrialists like **Josiah Wedgwood** and natural philosophers like **Erasmus Darwin**, Charles Darwin's grandfather.

---

## INFOBOX 2

### Polite society and rhetoric

Arja Nurmi and Minna Nevala

In the 18th century, one of the signs of polite society were specific rhetorical ideas and features, which were based on both the classical Roman rhetoric and the so-called New Rhetoric of the latter half of the century (McIntosh 1998: 142, 146). An essential part of politeness besides eloquence and propriety was mastery of polished language that was seen to further the process of civilization.

The New Rhetoric movement, started in 1748 in Scotland, adapted the principles of classical rhetoric into the print culture of the time, concentrating on the grammar and the 'purity' of language (MacIntosh 1998: 156–160). Precision and explicitness were encouraged, even preferred, in the writings of "the men of rank and breeding" to distinguish them from the "mean and low words" of the lower class (Smith 1762, as cited in McIntosh 1998: 166). This awareness of stylistic demands can be seen, for example, in self-corrections, where educated writers amended their letters to better meet the requirements of the elevated and correct style (Auer 2008).

While these features were common to the higher ranks of society and particularly its literary elites, letters continued to be a linguistically heterogeneous text type. Unlearned writers did not participate in the writing styles of the elite. Rather, they continued to employ oral characteristics in their writing, such as the add-on strategy, writing in a linear progression of finite units and linking them through co-ordination rather than subordination (Nurmi & Palander-Collin 2008: 23–24). There is some evidence of unschooled writers being aware of the registers of official letters (Fairman 2012: 222), which are, of course, a separate issue from the registers of polite society, and may be the influence of scribal assistance.

In England, as well as other European societies, there is 18th- and 19th-century evidence of particularly the lower strata of society taking recourse to linguistic formulae. These could be either genre-specific (e.g. relating to text organising and details of letter exchange between correspondents) or imitated from other registers of

elevated language use (e.g. the language of religion, newspapers and proverbs). (See further, e.g., Rutten & van der Wal (2014) for Dutch, Elspaß (2012) for German, Laitinen & Nordlund (2012) for Finnish, and Austin (1973) for English.) Finally, the letter-writing of the lower strata may have been guided by familiar models and what Laitinen and Nordlund (2012: 75–78) call group styles rather than by any familiarity with or aspiration towards the elevated rhetoric of the educated.

Letter writing was often taught by copying old letters (Austin 2000, Whyman 2009), which may well have contributed to both the formulae used and the general tenor of self-expression. It could even be argued that uneducated writers had their own “standard” of letter writing, separate from the one in evidence in the letters of the polite society. In addition to copying, the upper stratum of society had other means of learning the important art of letter-writing: various letter-writing manuals were increasingly printed in the 18th century.

Surprisingly many manuals, even in the latter half of the century, preferred plain style and language above the elevated New Rhetoric. *The Art of Letter-Writing* (1762: 3), for example, reminded the reader that one should avoid “these pretended Ornaments, which were formerly so studiously sought after”. More than plain or ornamental language, the key elements in 18th-century letter-writing were related to avoiding “vulgarity”, that epitome of the uneducated mass. As Sir William Pitt wrote to his nephew Thomas: “Pardon an observation on style: ‘I received yours’ is vulgar and mercantile; ‘your letter’ is the way of writing” (CEECE: William Pitt, 1755; PITT, 67).

## Grammar writing in the eighteenth century

Nuria Yáñez-Bouza

### 3.1 Introduction

The evolution of the English language during the eighteenth century has often been associated with the process of standardisation, in particular with its last two stages, *codification* and *prescription*. In the former stage, rules are laid down in grammars, dictionaries and other sorts of self-education handbooks which function as, or are seen as, authoritative manuals for its readers. Once the language is codified, the prescription stage involves the implementation of what has been codified (see in particular Milroy & Milroy 2012: 18–23; Nevalainen & Tieken-Boon van Ostade 2006: 282–286). As Tieken-Boon van Ostade (2008a: 7–8) notes, it is difficult to pin down the precise moment of the shift from codification to prescription, but she argues that as the ideology of standardisation gathered momentum in the mid-eighteenth century the two stages overlapped, from around 1740 to 1770.

McIntosh (2008: 228), amongst others, observes that the growth of language consciousness in the eighteenth century was “extraordinary”. The long eighteenth century (1660s–1800s) was a period of complex, interwoven relations between language and social, cultural and economic changes, most notably population growth, the rise of the middle classes, the rapid growth of book trade activities, the increasing interest in literacy and improvements in literacy rates, and the industrial revolution. McIntosh (2008: 228) also claims that language awareness was “accelerated by the enormous steps Britain took in this period towards becoming a fully literate print culture”. Book historians such as Raven (2009: 89) emphasise that print was “a decisive conduit for the pursuit of politeness and social advancement fostered by the expanding economy”, and that the commodification of culture was dominated by the industry of the book trade. For Feather (1985: 33), an educated readership lay at the root of the successful book trade; and education itself was becoming a vital marker of social and economic status, “a badge of respectability” not only for the social élite but also for the middling and lower middling orders (Holmes & Szechi 1993: 189).

Grammars, alongside dictionaries, newspapers, and other works on language and style, became a valuable commodity for the socially or economically ambitious.

By the end of the century politeness was not only a matter of social behaviour, manners and birth, as in earlier decades, but also of language: ‘polite learning’, as described by Klein (1994) and Watts (1999). By the early nineteenth century, correct and polite language had become “almost a national obsession” (McIntosh 2008: 228).

Given the importance of language in the eighteenth century and of grammar in particular, this chapter contextualises the practice of grammar writing, focusing on some of the elements that undergo change. Section 3.2 offers an overview of the production of English grammars. Section 3.3 explores the gradual movement towards vernacular education and changes in the target audience. Section 3.4 deals with some illustrative examples of morpho-syntactic variation and change in normative grammar books.

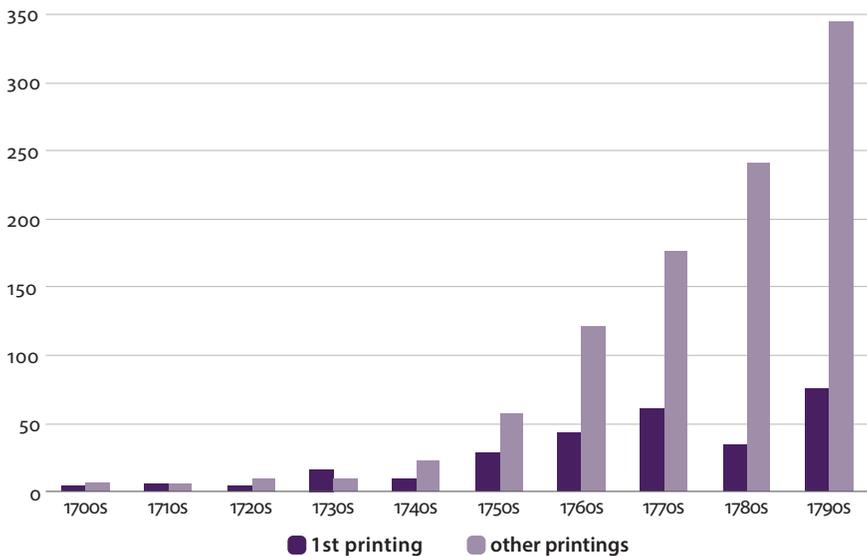
### 3.2 Grammar production

The eighteenth century, in particular the late eighteenth century, witnessed a rapid increase in the production of grammars. While by 1700 there were just twenty-one “explicitly English grammars”, by 1800 nearly three hundred grammar books had been published, counting first editions only (Michael 1970: 151). Figure 3.1 plots the diachronic distribution of grammars printed in the British Isles as recorded in the *Eighteenth-Century English Grammars* database (ECEG), with total figures adding up to 282 first printings and nearly 1,000 subsequent printings.<sup>1</sup> For a complete picture it is important to consider both types of edition, as reprinting is considered by historians to be “key to [the] burgeoning supply of books” and one of “the economic pillars of the British publishing industry” (Bonnell 2009: 709, 699; see also Michael 1997: 24). Most primary sources are grammar books, but the production of subsidiary grammars should not be underestimated, in that they comprise forty-two per cent of the total output documented in ECEG (c.120 subsidiary items versus c.165 stand-alone items). Subsidiary grammars can appear in a variety of categories, from dictionaries and self-education manuals or ‘companions’, to letter-writing manuals, spelling books and rhetorical treatises.

As we can see, there is a dramatic increase in output during the second half of the eighteenth century, in contrast to the low activity attested in the first half, which nonetheless represents a gradual increase from the incipient stages of the grammatical tradition in the sixteenth and seventeenth centuries. The figures here are telling: 40 first printings in 1700–1749 versus c.240 in 1750–1800 (14 per cent versus 86 per cent), and c.55 versus c.940 subsequent printings (5 per cent versus

---

1. The data from ECEG were retrieved as in August 2012. Unless otherwise indicated, eighteenth-century works are taken from ECEG.



**Figure 3.1** Production of grammars of the English language in the eighteenth century: first (N = 282) and subsequent printings (N = 994) (Source: ECEG)

95 per cent).<sup>2</sup> Rising figures can be equally observed in subsidiary grammars (11 per cent versus 89 per cent), which provides strong evidence for the increasing attention to and importance of language in this century. Likewise, earlier studies have also pointed to the conspicuous growth in other normative works such as rhetorical treatises, especially after the 1770s (see Michael 1987: 8–9, 381; Mugglestone 2003: 3; Suarez 2009: 44). In the 1750s and 1760s the most notable upsurge is in subsidiary grammars, which suggests that by the 1770s, not including a grammar in normative works such as dictionaries “could be seen as a shortcoming by potential customers” (Tyrkkö 2013: 189).

Amongst the various reasons for this sudden rise, Percy draws attention to early and mid-century periodical reviews as a vital source of language awareness which “intensified the existing fear and uncertainty about grammatical standards”, and which “might have catalysed the rise of prescriptivism” (2009a: 118–119; also Percy 2008). Magazines such as the *Tatler* (1709–1711), the *Spectator* (1711–1712), the *Monthly Review* (1749–1844) and the *Critical Review* (1756–1817) were “widely read and subscribed to” in town and country (McIntosh 1998: 184), and the *Spectator* in particular was “an exemplum of polite writing for a middle-class reading public” (Fitzmaurice 2000: 211), the main target audience of normative works in the century.

2. Similar findings have been reported in the literature, for instance Michael (1970: 277; 1987: 12), Tieken-Boon van Ostade (2008a; 2008b).

### 3.3 Towards vernacular education

#### 3.3.1 A practical grammar, a commodity

Grammars and schoolbooks became a commodity during the eighteenth century. In Watts' (2008) view, they were a combination of a commercial and a pedagogical project, as he explains with regard to the female grammarian Ann Fisher's occupation as a schoolmistress: "a significant link between the commercial project of making a profit out of the publication of English language teaching materials and the pedagogical project of teaching the standard (or 'polite') language to Langford's (1989) 'middling orders' of society" (Watts 2008: 49). On the one hand, there was genuine interest in educating children and adults, women and foreigners of all ranks, and thus grammars developed on more practical lines. Yet, Tieken-Boon van Ostade (2008b: 109) observes that the addition of the word *practical* to the title-pages could at times have been a marketing strategy, and thus illustrative of the commercial nature of the project. Raven (2009: 85), in particular, notes that booksellers (authors and readers to a lesser extent) were responsible for turning books into "market commodities, more as goods directed to specific audiences". Let us look into each of these roles.

To begin with readers, they are said to have created demand for self-help guides and grammars "as passports to politeness" (Percy 2008: 130). This was especially the case with the rising middle classes, as the correct use of language would not only be a key to upward social and economic mobility, but also a means of distancing oneself from social inferiors; politeness "as embodied in a notion of correctness – was a commodity that could be bought" (Fitzmaurice 2000: 197–198). It is not uncommon to find grammars particularly "fitted for trade and business", like William Loughton's *A Practical Grammar of the English Tongue* (1734: advertisement), or grammars appended to miscellaneous manuals that aimed to instruct young men to become successful, such as William Perry's (1774) *The Man of Business, and Gentleman's Assistant [...] Adapted to the Use of Gentlemen, Merchants, Traders, and Schools*.

For their part, authors were accused by book reviewers of being "particularly interested in the pecuniary success, or sale, of their books", thereby "reduc[ing] the business of education (the most important, if not the most noble, of any in civil society) to a mere mercenary job" (1765; cited in Percy 2008: 130–131). There was an "appetite" for grammars, and both authors and publishers knew it and took advantage of the fact (Tieken-Boon van Ostade 2008c: 79). Indeed, most authors were schoolmasters and many of them wrote their grammar for their own school.

It is also not unusual for changes in the approach and contents of the grammar to arise in a second or later edition, with a view to reaching a wider or more

specialised audience (see, for instance, Hodson 2008 on Joseph Priestley's *The Rudiments of English Language*, first edition in 1761 and second edition in 1768). Booksellers and publishers were clearly instrumental here too. Tieken-Boon van Ostade (2008b) noted other common marketing strategies in the book trade in general, such as adding a renowned name on the title-page, manipulating the title-pages with catchwords like *new*, *corrected*, *improved*, changing the price, format or size of the editions, or reprinting provincial grammars in the capital.

With regard to the pedagogical aim of eighteenth-century grammars, one of the main developments we can observe over the course of the century is the change towards vernacular education, that is, towards practical grammars about English and which were written in English. In a nutshell, Robins describes the history of English grammar writing as “concern[ing] mainly the degree to which, and the ways in which, the authors adjusted their work to [the] Latinate tradition, enshrined in effect since the time of Ælfric” (1986: 297). Until well into the eighteenth century, the teaching of *grammar* meant *Latin grammar*, even when the title of the work read *English* – it meant written *in* English, not *of* English, as in Edward Burles' (1652) *Grammatica Burlesa, or a New English Grammar*. The cultural dominance of Latin affected both methods and materials, and this was most notable in the treatment of the parts of speech (Michael 1987: 319).

In his study of the grammatical tradition from 1586 to 1800, Michael identified four main systems of classification and concluded that the Latin systems outnumbered the vernacular systems by far (74 per cent versus 15 per cent; see Michael 1970: 210–211, 277). The same findings emerge from the analysis of the ECEG subcorpus of British grammars: vernacular systems were used in just over thirty grammars, with a moderate peak in the 1750s. They are associated with what Michael described as a “reforming movement”, that is, a “conscious” movement of reform “which sought to give English a grammar in its own right” (1970: 210). Although implicitly initiated by John Wallis' *Grammatica Linguæ Anglicanæ* in 1653, it was Archibald Lane in 1700 who first presented the fourfold division “in deliberate opposition to the traditional Latin system” (Michael 1970: 171). It is mainly between 1730 and 1780 that authors “tried to ease their pupils' inherited burden of Latin categories by simplifying the technical terms and by modifying the Latin system of parts of speech” (Michael 1987: 343).

Despite the dominance of Latin when it comes to describing the parts of speech, English was gradually winning the battle at a higher level: authors realised the practical need for vernacular education as the readership was itself changing. Whereas in the late seventeenth century and early eighteenth century education was mainly addressed to “young males enjoying the advantages of a classical education”, the target audience soon expanded to the rising middle and lower middle classes who had had no education in classics, including women (Bax 2008: 287). A more utilitarian

approach was taken, focused on the teaching of English language.<sup>3</sup> This is well illustrated in Bax's analysis of the description of synthetic and analytic comparisons of adjectives and adverbs in twenty-seven 'practical' grammars. The descriptions whose main criteria included the Latinate origin of words are not found after the mid-1760s, an indication, Bax argues, that "the reader is no longer expected to be well-versed in Latin", and "a reflection of the type of readers that the authors had in mind", namely children, women, and (lower) middle classes (Bax 2008: 286–287).

Practical grammars can be found throughout the century, often flagging up their utilitarian nature on the title-page, but it is not until the 1770s that we find the explicit combination "practical English grammar" in title-pages, as in Isaac Hodgson's *A Practical English Grammar, for the Use of Schools and Private Gentlemen and Ladies, with Exercises of False Orthography and Syntax at Large* (1770). In Beal's view, this increasing interest in vernacular education was "[p]erhaps the most pressing and practical reason for the production of grammars in the eighteenth century" (2004: 101). Dissenting academies took the lead in promoting a school curriculum in which practical skills such as natural sciences, mathematics and accounting were taught in English, and, most importantly, a curriculum in which the English language was taught as a separate subject. The words of Joseph Priestley (1733–1804), grammarian and tutor of the Warrington Academy, serve as an example of the defence of vernacular education:

The propriety of introducing the *English Language* into *English Schools*, cannot be disputed; a competent knowledge of our own language being both useful and ornamental in every profession, and a critical knowledge of it absolutely necessary to all persons of a liberal education.

(Priestley 1761: viii; cited in Beal 2004: 103, italics in Priestley)

Michael (1970: 495ff) puts forward three main arguments in favour of vernacular education. (i) The grammatical argument defends the independence of English as a language in its own right, with its own grammar, and which should thus be taught without reference to Latin. William Loughton's title-page speaks for itself: *A Practical Grammar of the English Tongue: or, a Rational and Easy Introduction to Speaking and Writing English Correctly and Properly; Peculiarly Adapted to the Nature and Genius of the Language, and Free from the Hard and Unnecessary Terms of the Latin Rudiments*, 1734. (ii) The priority argument claims that learning must start with acquiring grammatical knowledge of the English language, as English is their mother tongue and learning Latin is "merely wasting a very precious portion of [a boy's] time to no purpose", especially in the training of trade, business and

---

3. An anonymous reviewer points to a similar development regarding Dutch grammars in the eighteenth century, for which see Rutten (2009: 57–59).

technical-mechanical knowledge (Buchanan 1770: 142; cited in Michael 1970: 506). (iii) The third argument calls for a change in the curriculum so that the vernacular should be taught before teaching Latin, as a knowledge of the structure of their own language would better prepare pupils for the acquisition of Latin or any other language.

### 3.3.2 Target audience

This section pays closer attention to the grammars' potential readership, following up McIntosh's (1998: 169) view that by the end of the eighteenth century England had become "a well-developed print society" and a consumer society.

In the ECEG subcorpus of British grammars, over half of the works which make explicit mention of their intended audience are written for children and for schools (c.150).<sup>4</sup> This is no surprise, given that most grammar writers were schoolteachers, often running their own school or academy, for which the grammar would be written; for instance, *A Short Abstract of English Grammar* written by "the Rev. Mr. R. Green of Durham, and particularly designed for the use of his academy there" (1779: title-page) (see Yáñez-Bouza & Rodríguez-Gil 2013: 152). This is also in line with Immel's (2009) observation that the market for children's fiction and non-fiction books expanded gradually during the century, especially after the 1740s, when John Newbery started publishing juvenile works, including *An Easy Introduction to the English Language*, the second volume of his successful *The Circle of Sciences* (1745). Immel emphasises that the young readership was "arguably among the most important to Great Britain's polite, commercial society", and by the end of the century school books in particular dominated the production of the Longman ledger of children's books, with new editions and numerous reprints (2009: 736, 748). The importance of writing a grammar suitable for children was such that unsuitability for that audience would be cited as a criticism in book reviews and other contemporary works, as noted by Percy (2008: 136) in a review of Anselm Bayly's (1771) *Plain and Complete Grammar*.

Gender differences in education were common, and this too is reflected in the target audience of grammar books. Women were considered intellectually inferior (so-called 'weaker sex' and 'fair sex'), and were denied access to university and grammar schools. Learning too much (or as much as men) was considered immoral and unattractive. Boarding schools for ladies were strongly objected to, even by educationalists such as Vicesimus Knox (1752–1821), headmaster of a boarding

---

4. The percentage increases to sixty per cent (c.170 items) if we factor in works targeting a mixed audience of schools and private usage.

school for boys in Kent: “girls are destined to private and domestic life, and boys to public life, their education should be respectively correspondent to their destination” (Knox 1789, *Liberal Education*; cited in Cohen 2006: 323). Despite criticism, the eighteenth century (and the early nineteenth century) witnessed a significant increase in the number of private schools for girls, boarding schools or day schools, a development described by Skedd (1997) as “commercial schooling”. With regard to grammar works, we already see complaints about the neglect of women’s education in the early part of the century (e.g. James Greenwood 1711), and by the 1730s these were “often” heard (Clarke 1731). However, by the 1760s educationalists such as John Rice (fl.1765) and grammarians such as James Buchanan (fl.1753–1773) still felt that ‘the fair sex’ was “shamefully neglected” and there were calls for a better education, including the teaching of vernacular grammar for them “to write their own Language grammatically, and to indite [i.e. compose] elegantly” (Buchanan 1762: xxxi).

Many grammars in the early and late part of the century addressed both boys and girls (c.60 in our subcorpus), as early as Lane’s *A Key to the Art of Letters*, advertised as for “the English youth of either sex” (1700: title-page). But it is not until the mid-1770s that we see the publication of a grammar book solely targeted at ladies, namely Ellin Devis’ *The Accidence*, printed in London in 1775 and “[d]esigned for the use of young ladies” (title-page). A few years earlier a grammar of English had been appended to a letter-writing manual also solely addressed to ladies, and also written by a female author: Lady Du Bois’ (1771) *The Lady’s Polite Secretary, or New Female Letter Writer*. Similarly, in 1770 a subsidiary grammar was attached to a book published anonymously in Salop (Shrewsbury), entitled *The Female Miscellany* and written “[f]or the use of a boarding school”. Overall, the subcorpus of British grammars identifies fewer than fifteen grammars written particularly for girls,<sup>5</sup> but such low numbers should not overshadow their importance: “girls’ education was important enough to be a commodity” (Percy 2009b: 79). Female grammarians in particular were well aware of the importance of language correctness, of grammar, as a social marker, and offered examples which would help ladies acquire social, moral and intellectual knowledge: “learning grammar might cultivate both virtue and learning” (Percy 2009b: 77). The school curriculum expanded beyond domestic skills but was still limited to “a little learning” of each subject (Cohen 2006: 330–331).

As suggested elsewhere in this chapter, social status was a determining factor behind the writing of eighteenth-century grammars. Cressy (1980: 177, 186) states that (il)literacy “remained highly stratified” and, more specifically, that educational

---

5. In the ECEG subcorpus there are more grammars written solely for a male audience, including William Perry’s *The Man of Business, and Gentleman’s Assistant*, written for and adapted to “the use of gentlemen, merchants, traders” (1774: title-page).

achievements were closely related to rank and occupation. There were different types of schools for different social groups, as the poet George Crabbe (1754–1832) noted (Stone 1969: 71). Immel clearly illustrates this point by comparing two books for children, where the distribution of literary and educational materials was determined by the child's social class:

A speller for charity schools like Henry Dixon's *English instructor* (1725) would devote far less space to polite literature than to the articulation of religious principles because lower-class pupils needed to be reconciled to their lot, not tempted to aspire above their stations. Robert Wharton's elementary reader *Historiae pueriles* (1738), on the other hand, consisted of retellings from Plutarch's *Lives*, biographies of great English kings and Addison's fiction from the *Spectator*. This text was supposed to give a seven- or eight-year-old middle-class boy 'a notion of the strength and purity of his own language' to lay the foundation for correct English usage, a sign of full literacy. (Immel 2009: 739)

Stone (1969: 69–118) describes the social/occupational-educational context from c.1635 to 1822 as far as men are concerned, and we see that, while squires and gentlemen were one hundred per cent literate – they received education in classics, and attended university or the Inns of Court –, labourers and servants, the lower orders, were the least literate social group, with stable literacy rates around forty-five/forty-eight per cent. They had access to elementary schooling in charity schools, Sunday schools, or evening schools, where they would be trained in just one of the three Rs, namely reading: they would learn a little reading and to sign their own name – “the most that the poor could aspire to” (1969: 70) – but they would often leave school before writing skills were taught. In other words, their education consisted of “[h]eavy indoctrination and mechanical rote learning” (1969: 113). As a whole, the provision of educational opportunities and facilities during the period 1660 to 1780/90 shows major improvements in quantity and quality, but social differences continued.

### 3.4 Morphology and syntax in eighteenth-century grammars

#### 3.4.1 Divisions of grammar

Eighteenth-century grammar books usually define grammar as “the art of writing or speaking the English language properly” (Burn 1766: 19), where ‘art’ implies “a rule-governed technique, rather than an innate competence” (Vorlat 1996: 165). The use of expressions such as “correctly” and “with propriety” carried a prescriptive connotation even in early grammars, since the ultimate aim of all grammarians was ‘standard’ English.

Grammar, the system, is traditionally divided into four parts: orthography, etymology, syntax, and prosody. Sometimes there is a fifth part named orthoëpy. Etymology in the eighteenth century had two meanings: ‘word-provenance’, as in the modern sense, and ‘word-variation’, including ‘word-formation’; the latter was in fact “by far the more common” meaning, and thus the term *etymology* was the common label to refer to what is nowadays understood as *morphology* (Michael 1970: 185). This is the part of grammar that received most attention in stand-alone grammars as well as in subsidiary grammars, probably because grammar writing in the Early Modern English period (1500–1700), and still during the eighteenth century, was largely influenced by the Latin tradition, as pointed out in Section 3.3.1 of this chapter, and Latin is a morphologically rich language. Grammarians often tried to force Latin morphology onto the morphology and syntax of English. Examples of extreme cases include Wells Egelsham (1780) declining the article for case and number, and William Ward (1765) drawing up the paradigm of nouns in six Latin cases (see Finegan 1998: 547–550).

While the discussion on etymology/morphology would go into detail on the parts of speech, syntax received less attention until the late eighteenth century, to the extent that at times it was simply integrated into the discussion about the parts of speech. Some early eighteenth-century grammarians still echoed the common view of Early Modern English writers who quietly discarded any discussion of syntax, as in Jeremiah Wharton’s *The English Grammar*: “Here should follow the Rules of Concord, and Construction: but because there is little, or no variation in the parts of the English tongue: they are altogether needless” (1654: 60). Syntax was given “steadily increasing attention” over the course of the eighteenth century (Michael 1970: 468), possibly due to the general move towards vernacular education – the focus is on English, and English does have a syntactic structure worth discussing – and by the new utilitarian/practical approach to the study of grammar – the knowledge of grammar is important in society, and thus English must be explained fully. Michael (1970: 468) estimates that syntax featured in sixty per cent of the works published before 1740, up to seventy per cent between 1740 and 1770, and up to eighty-five per cent between 1770 and 1800.

Besides, during the second half of the century, the treatment of syntax was often accompanied by exercises of ‘false syntax’ or ‘bad English’, a very popular pedagogical method introduced by Ann Fisher in *A New Grammar* (?1745). These exercises would typically present the incorrect form, with pupils expected to (a) identify the mistake, and (b) provide the correct form.

Syntax also features in the discussion of ‘grammatical figures’, which combined grammar and rhetoric. It was usually presented as ‘figurative syntax’ or ‘artificial syntax’, in contrast to ‘natural syntax’; the latter refers to grammatical and logical rules of word order, whereas the former was understood as a ‘departure’ or a

‘deviation’ from natural word order. This approach can be seen as “descriptive and appreciative” (Sundby et al. 1991: 12), but there was often a certain nuanced implication that figurative syntax was disfavoured in that it might easily alter the sense of the sentence, resulting in imprecise or inaccurate expressions.

For instance, the syntactic construction of preposition stranding was often discussed as an example of ‘transposition’, understood as “the putting the Words in a Sentence, or Sentences, out of their natural Order” (Greenwood 1711: 218). Thus, the placing of a preposition at the end of a sentence, separated from the word it governs, was criticised as a deviation of the natural word order of a *pre*-position, that is, preceding its object. Authors such as James Greenwood (1711: 218) discussed transposition partly in a neutral way, yet also acknowledged the risk of violating the principles of perspicuity and unity, whereas others such as Hugh Blair (1783: I.xii.285–287) strongly condemned the transposition of elements in the late eighteenth century: “Such conclusions are always enfeebling and degrading. [...] it is disagreeable for the mind to be left pausing on a word, which does not, by itself, produce any idea, nor form any picture in the fancy”.

### 3.4.2 Awareness of variation and change

Language change was indeed a concern in the eighteenth-century context. As noted in the introduction, this is a period of the codification and prescription of the English language, two key stages in the process of standardisation. Standardisation as “an ideology” aims ultimately at uniformity and stability in language, so that a standard language is “a set of norms to which actual usage may conform to a greater or lesser extent” (Milroy & Milroy 2012: 19). It naturally follows that standardisation is “inimical to change and variation” (2012: 22), and these attitudes would go against change and variation at all levels.

As early as the seventeenth century, John Dryden (1631–1700) had expressed his concern about change in language, describing English as “in a manner barbarous” (1693: 110), while praising Latin as “a more stable” language (1679: A3<sup>v</sup>). In the early eighteenth century, Jonathan Swift (1667–1745) called for an Academy that would “ascertain” the language and “fix” its improprieties, in the belief that “it is better a Language should not be wholly perfect, that it should be perpetually changing; and we must give over at one Time, or at length infallibly change for the worse” (1712: 31). In the mid-eighteenth century, Thomas Sheridan (1719–1788) shared similar concerns about the contamination of English with improprieties from ‘the vulgar’: “many pronunciations, which thirty or forty years ago were confined to the vulgar, are gradually gaining ground and if something be not done to stop this growing evil and fix a general standard at present, the English is likely to

become a mere jargon” (1780; cited in Görlach 2001: 90). Lexicographers such as Samuel Johnson (1709–1784) echoed the same feeling in the definition of certain words: “a word that has crept into conversation and low writing but ought not to be admitted into the language” (1755; cited in Card et al. 1984: 57).

For their part, grammar writers would tend to pick on old usage and contrast that with contemporary usage. Typically, a reference to ‘old’ language would carry a negative connotation, and there was no shortage of critical terms for that: Sundby et al. (1991: 50) list forty-nine terms under the cover term *obsolete*, implying that (a) “X was used in the past but not nowadays”; (b) “X, not now in use, was used earlier, but erroneously”; or (c) “X should not be used now because it is of the past” (1991: 42). It is noteworthy that some expressions refer to change associated with social class and literary style, for instance “long abandoned by the polite world”, “now confined to the conversation of the vulgar” or “now never used by good authors” (1991: 42).

Language variation caused much anxiety in the eighteenth century, too. Strictly speaking, standardisation “involves *the suppression of optional variability in language*” (Milroy & Milroy 2012: 6; original italics); hence, there is a strong feeling that in codifying and prescribing language norms, if two or more forms are available to choose, one will be preferred over the other(s). The former will be considered standard, correct or more proper; the latter non-standard, incorrect or less proper, less precise, less idiomatic, etc.<sup>6</sup> The reason behind evaluative judgements is at times arbitrary: the *ipse-dixit* attitude that ‘he himself said it, just because’. Sometimes the selection of a particular variant is justified, or at least accounted for, by analogy with Latin, because of logic, etymology, or by resorting to the canons of rhetoric, notably purity and perspicuity, but also unity, clarity, harmony, strength or simplicity.

A good example of a prescriptive grammarian who combined grammar and rhetoric in his argumentation is Lindley Murray (1745–1826), in *English Grammar* (1795) (see Vorlat 1996). For instance, on the grounds of logic Murray condemned double negation because “[t]wo negatives, in English, destroy one another, or are equivalent to an affirmative” (1795: 121), and he condemned the use of the plural with the ‘conjunction disjunctive’ *or* in sentences such as “Ignorance or negligence has caused the mistake”, because the noun “is referred to the preceding terms taken separately, [hence] it must be in the singular number” (1795: 93–94). Murray rejected certain expressions that he considered unidiomatic as “not so proper”, such as “A beautiful field and woods” for “Beautiful fields, and fine woods” (1795: 134–135). Murray also believed that “there is harshness” in using plural verb agreement with ‘nouns of number’ because, he argues, “the ideas they represent seem not to be

---

6. See Sundby et al. (1991: 44–53) for an exhaustive list of negative terms.

sufficiently divided, as it were, in the mind”, for instance “The House of Commons were of small weight” should say “The House of Commons was of small weight” for the sake of elegance and harmony (1795: 95).

Although intolerance to variation was the majority feeling in this century, there were also grammarians who adopted a more descriptive approach to usage and who presented variants in a neutral way, without making evaluative judgments on good/bad English. An illustrative example is John Shaw’s (1778) passage on preposition stranding and pied piping in *A Methodical English Grammar*: “it [the preposition] **may either** stand immediately before them, or after the verb in the same clause; as, [...] John, *with whom* I conversed; **or** *whom* I conversed *with*” (Shaw 1778: 94–95; bold added).

### 3.4.3 Case studies

In the late 1980s complaints were raised that the field of grammaticology and historiography needed more and more cross-disciplinary research. Thirty years later, the study of early grammars is thriving in many ways, and their potential for the study of the history of the English language has come to be recognised (Tieken-Boon van Ostade 2000; Beal et al. 2008; Tieken-Boon van Ostade 2008). The common approach in the discipline of normative linguistics is to study the correlation between precept norms and actual usage from various perspectives, the shared line of investigation being the assessment of the influence of the former on the latter (Yáñez-Bouza 2016). The analysis of the original sources has shed more light on the external factors that can affect the history of English, putting prescriptivism on the map, not only because of its effect on stigmatising certain linguistic features, but also because on some occasions empirical findings have challenged traditional myths about the negative influence of eighteenth-century prescriptivism. The paragraphs that follow examine several case studies concerning language variation and change affecting morpho-syntactic features.

A number of scholars have observed that individual authors who were exposed to ‘correct’ and ‘proper’ English changed their language use in order to adhere to what was considered ‘standard’. To give an example, William Clift (1775–1849) was born to a modest working class family in Cornwall; he seems to have received some education at a writing school, but there is no evidence that he read grammar books. Clift came in contact with the use of good English after he moved to London in 1792 to work as an apprentice to John Hunter, a famous surgeon and anatomist of London’s ‘polite society’, and the letters written by Clift from 1792 and 1801 show that his usage “moved increasingly towards the accepted norm” (Austin 1994: 287). This was already apparent in the early letters and at all levels: grammar, spelling,

idiom and dialect use. For instance, Clift's early letters showed variation between the dialectal *do/don't* and the proper *does/does not* in the third person singular form. The form *does not* was used as early as 1792, and the 1795 letters show *does* and *does not* only. Similarly, Clift gradually favoured the use of *wh*-relative forms over relative *that*. The latter had been subject to strong criticism in a number of grammars: *that* was "inelegantly" used for *who*, *whom*, *which* (Buchanan 1767: 74), and *wh*-pronouns were "better" forms of expression (Hornsey 1793: 24). Clift's early letters show an even distribution of *that* and *which*, and *who* is barely used; however, in 1796 and thereafter, the *wh*-forms increased "dramatically" and "almost eclipse[d] *that* altogether", including instances of hypercorrection (Austin 1994: 294).<sup>7</sup>

Large-scale studies of precept and usage corpora have confirmed that prescriptive judgements, whether positive or negative, have triggered change in the development of certain linguistic features during the Late Modern English period (1700–1900), whether permanent or temporary. It is often the case that change results in the suppression of variation, as pointed out above.

A good illustrative example is the loss of variation in the preterit and past participle forms of irregular verbs such as *write*. In present-day English the standard forms are *wrote* for the preterit and *written* for the past participle. In the seventeenth century and early eighteenth century there was a wide array of variants in both functions, with "possible regional, sociolectal and formal/informal distinction" (Görlach 1988: 127; cited in Oldireva Gustafsson 2002: Section 1). Oldireva Gustafsson reports that in private correspondence the letters of literate, educated writers showed a preference in usage for *wrote* as past participle in both the 1680–1710 and the 1760–1790 periods; the forms *writ* and *writt* were frequently attested in the early period, but receded gradually and fell from twenty-one to ten per cent and from twenty-four to zero per cent, respectively (Oldireva Gustafsson 2002: Section 4). Clearly, "none of the letter-writers was a consistent user of *written*", the standard form nowadays (2002: Section 5).

On the other hand, grammarians were well aware of morphological variation already in the mid-seventeenth century, and so recorded it in their grammars, usually in lists of irregular verbs (Oldireva Gustafsson 2002: Section 2, Table 2). The precept data point to a strong preference for the variant *written* in the past participle, and this is the form that gradually became *the* preferred form. The frequent use of *wrote* as past participle drew much criticism in grammar books, and it was stigmatised as "corrupt", "improper", "inelegant", "solecism", "barbarous", "unidiomatic", "inaccurate", "ungrammatical" and even "vulgar" (Sundby et al. 1991: 233–234).

---

7. CEECE contains a number of letters written by members of the Clift family and which are discussed in subsequent chapters of this book; for instance, Nevalainen on the diffusion of verbal *-s*, Säily on the productivity of the suffixes *-ness* and *-ity*, and Sairio on the progressive aspect.

The variant *writ* was also proscribed, though to a lesser extent, as “inelegant” and “improper” (1991: 233). Here, prescriptivism seems to have inhibited the natural process of morphological levelling that was taking place between the preterit and the past participle forms in the late seventeenth century and early eighteenth century, and it imposed rather a three form-function distinction for the sake of greater clarity and perspicuity (e.g. Fenning 1771; cited in Oldireva Gustafsson 2002: Section 6).<sup>8</sup>

Rather than triggering change, on occasions grammarians became aware of variation or change and their recurrent comments and norms resulted in a reinforcement of an ongoing trend in usage. Usually, this involves the selection of a particular variant as the correct, standard form, in detriment of others as incorrect, stigmatised forms.

For instance, double marking of degree is stigmatised in present-day standard English with regard to double periphrastic forms (*more lovelier*). This has often been associated with eighteenth-century prescriptivism, but González-Díaz (2008) has shown that there is a temporal mismatch between its decline in usage and its criticism in normative works. Double periphrastic forms were not infrequent in the Middle English period, but usage decreased drastically during Early Modern English; the decline was very pronounced from Early to Late Modern English, and continued dropping in written standard English until the present day. Double comparatives and superlatives were indeed subject to criticism in eighteenth-century works, but González-Díaz’s data show that change in usage began in the pre-normative era. Besides, there is some early criticism by grammarians such as Paul Greaves in *Grammatica Anglicana* (1594) and by literary writers such as John Dryden (1672b); these authors branded double periphrastic forms as a “barbarous”, “ignorant”, “impure”, and “gross” manner of expression (cited in González-Díaz 2008: 292). It seems, then, that the “social downgrading” of this feature started in the late sixteenth century, and by the end of the seventeenth century it had already spread. The effect of criticism in eighteenth-century grammars was to reinforce the process with more and more nuanced labels, such as “bad English”, “obvious redundancies”, “mistakes”, “a greater impropriety”, “nonsense”, or “peculiar to the Old Vulgar translations” (2008: 293–294). The correct form is the simple inflected or periphrastic variant, as recommended by Ann Fisher in *A New Grammar*: “A Comparative Adverb must not be set before a Quality compared by *er* or *est*; as, *wiser*, *wisest*, and not *more wiser*, or *most wisest*, &c.” (1753 [?1745]: 117).

Another approach to the study of variation and change in early grammars is to explore the correlation between real-time change and evaluative judgements

---

8. See Sundby et al. (1991: 225–236) for criticism related to other irregular verbs whose tendency towards morphological levelling was likewise disrupted by normative rules.

on language usage. Anderwald's (2012, 2016) investigation follows these lines by comparing precept and usage in relation to the rate and phase of the S-curve, and to salience. With regard to salience, she argues that features associated with informal or colloquial language are likely to raise criticism, as do features that tend to occur in a particular genre. For example, one of the strongest criticisms of the progressive passive, as in *The house is being built*, was that it was perceived as a construction imposed on the language "by the newspaper press" (Bingham 1867; cited in Anderwald 2016: 201–202).

With regard to the S-curve pattern of language change, it has been observed previously that the typical pattern is "slow-slow-quick-quick-slow" (Denison 2003: 56). The rate of change may play a role in that slow changes are likely to go unnoticed or elicit positive attitudes, such as the rise of the active progressive, whereas rapid, drastic changes are likely to face strong criticism, as happened with the progressive passive, condemned in the eighteenth century as "bad", "improper", "solecism", and in the nineteenth century as "uncouthness and deformity", "an absurdity", or "not very elegant" (Anderwald 2016: 205–214; Sundby et al. 1991: 236–237).

Similarly, the phase of the S-curve may play a role: very new changes at incipient stages may not be observed, whereas nearly completed changes are bound to draw grammarians' attention and, typically, the old form would be considered less proper (Anderwald 2012: 43–44). A good example is the *be/have* variation in the perfect, as in *I am arrived/I have arrived*. Variation between auxiliary *be* and auxiliary *have* is a well known characteristic of early English in the context of mutative intransitive verbs. In present-day English *have*-perfect is used almost categorically, but at the start of the eighteenth century the proportion was around twenty per cent. By 1800 it had increased to forty per cent and continued rising rapidly during the early nineteenth century, when it enjoyed "paradigmatic majority"; the change towards *have*-perfect came to completion towards the end of that century (Rydén & Brorström 1987; in Straaijer 2010: 64). Anderwald (2012: 39–41) argues that during the eighteenth century, while change in favour of *have*-perfect was slowly progressing, attitudes towards the old variant *be*-perfect were not strongly critical, with authors acknowledging variation between the two auxiliary verbs. For instance, James Buchanan (fl.1753–1773) noted that both auxiliaries were used and pointed out that *be*-perfect was the variant "frequently" used (1762: 142; in Straaijer 2010: 65). However, in the nineteenth century, when *have*-perfect was rapidly gaining ground and the change was in its completion stages, grammarians condemned *be*-perfect as obsolete and incorrect, while the incoming *have*-perfect was the preferred, correct variant. Thus, Goold Brown (1791–1857) also acknowledged variation, but made the remark that *have*-perfect was the preferred variant at the time of writing his grammar: "with a few exceptions, present usage is clearly in favour of the auxiliary *have* in preference to *be*, whenever the verb formed with

the perfect participle is not passive: as, They *have* arrived – not, They *are* arrived” (1857 [1823]: 72). In the second half of the nineteenth century, voices were raised against the use of *be*-perfect as “a departure from grammatical correctness” (Parker & Fox 1835 [1834]: 19), and as a “now rarely used” form that “should be avoided” (Quackenbos 1868 [1862]: 123).<sup>9</sup>

### 3.5 Postscript

The case studies in the present volume aim to contribute to the above-mentioned approaches of investigation by analysing linguistic features in relation to language change triggered by internal processes from below and/or by external processes from above, such as the influence of eighteenth-century normative precepts. By following up the developments already observed for the Early Modern English period (Nevalainen & Raumolin-Brunberg 2003), their findings will shed more light on the interwoven relationship between real-time change and evaluative (pre/proscriptive) judgements on language usage. At the same time, the long-span overview will help us gain a broader picture not only of the potential influence of late eighteenth-century works, but also of the latent prescriptivism of the late seventeenth century and early eighteenth century.

### Acknowledgments

I am grateful for generous financial support to the Spanish Ministry of ‘Economía y Competitividad’, to the Ramón y Cajal Scheme (RYC-2011-07863); the State Secretariat for Research, Development and Innovation (FFI2011-25683); the European Regional Development Fund (FFI2013-44065-P); the Autonomous Government of Galicia (Secretary General for Universities, GPC2014/060).

---

9. Nineteenth-century sources cited in this paragraph are taken from Anderwald (2012: 40–41).



## *The Corpus of Early English Correspondence Extension (CEECE)*

Samuli Kaislaniemi

### 4.1 The CEEC project and the CEEC family of corpora

The *Corpus of Early English Correspondence Extension*, or CEECE (pronounced ‘cease’), expands the CEEC family of corpora to cover the eighteenth century from 1681 to 1800. The CEECE was compiled with two aims in mind: First, the idea was to provide diachronic continuity to the first part of the CEEC (finished in 1998, called CEEC-1998), in order to follow through the linguistic changes that were still ongoing in 1681, where the CEEC-1998 terminates (Nevalainen & Raumolin-Brunberg 2003); and second, to track linguistic change in the long eighteenth century. The CEEC-1998 and CEECE allow tracking linguistic features over nearly four centuries from Late Middle English into Late Modern English (the entire CEEC family is consequently referred to as CEEC-400; see Kaislaniemi 2006; Nevala & Nurmi 2013).

The compilation of the CEECE followed the completion of the CEEC-1998, which has its roots in the compilation of the Early Modern sections of the *Helsinki Corpus of English Texts* (or *Helsinki Corpus* for short; HC). After the completion and release of the HC, the compilers of the Early Modern section, Terttu Nevalainen and Helena Raumolin-Brunberg, decided to pursue investigating language change in Early Modern English in more detail. To this end, they would compile a separate corpus of Early Modern English; and due to their interest in testing the applicability of modern sociolinguistic methods on historical materials, they decided to compile the corpus from personal correspondence. The project was initiated in 1993, and with the help of student assistants, the corpus was finished by 1998. The CEEC-1998 covers the period from about 1410 to 1681 and contains some 2.6 million words in nearly 6,000 letters from almost 800 informants. Due to copyright reasons (the CEEC corpora are compiled from published editions) the CEEC-1998 could not be released without acquiring permissions from the publishers. Unfortunately, not all publishers granted permission to publish their texts as part of the corpus, and the version released publicly in 2006, the *Parsed Corpus of Early English Correspondence*

(PCEEC), contains some 2.2 million words (5,000 letters from 650 writers; see the CEEC entry in CoRD for various CEEC corpora).<sup>1</sup>

From around the time of completion of CEEC-1998, the CEEC team had begun to think about extending the letter corpus to cover the eighteenth century. At the time, the only corpus projects containing eighteenth-century correspondence were the *Corpus of late 18c Prose* (CLEP) being compiled at the University of Manchester under David Denison (see the CLEP entry in CoRD), and the ARCHER corpus (version 1b: see the ARCHER entry in CoRD) – but the former was to cover only the years 1761–1790, and the latter had only just under 29,000 words of LModE letters. In the light of the results coming from studies using the CEEC-1998, it was clear that there was a call for a corpus of eighteenth-century letters (Kytö & Rissanen 1997: 14).

We had several expectations regarding letters from the eighteenth century. Literacy increased dramatically during the period, and we expected this to be reflected in an increase in the amount of material from the lowest social ranks – poorly represented in the CEEC-1998 – and hoped to get ample evidence of language change from below. For much the same reasons, we expected to find editions of 18th-century correspondence to contain more letters from women. And finally, we expected to see the effects of the proliferation of prescriptive grammars in the personal letters of the period.

## 4.2 Corpus compilation

The CEEC *Extension* is based on the same principles as CEEC-1998. Both corpora were designed for historical sociolinguistics: for the application of sociolinguistic methodology on historical texts. For this purpose, correspondence is a useful data source because, firstly, language change from below first appears in written form in private writings (ego-documents), such as personal letters, and secondly, personal letters are arguably the text type closest to speech (Biber 1995; see also Byrne 1964; Culpeper & Kytö 2010). Further, personal letters are eminently suited for sociolinguistics as by definition they (tend to) contain information about the language users and the usage context (names, dates, relationships, places, etc.). And more than anything, personal letters allowed us to access the language of individual writers in their daily lives.

---

1. The part-of-speech tagging of the PCEEC was carried out by Arja Nurmi (University of Helsinki) and the syntactic annotation by Ann Taylor (University of York). The corpus is distributed through the Oxford Text Archive (OTA).

CoRD is the acronym for the Corpus Resource Database: [www.helsinki.fi/varieng/CoRD/](http://www.helsinki.fi/varieng/CoRD/).

In terms of coverage – whose letters to include, and how many? – like CEEC-1998, the CEECE consists of “judgement samples selected on the basis of extralinguistic criteria” (Nevalainen & Raumolin-Brunberg 1996: 41). The aim of all the CEEC corpora is at social representativeness – while acknowledging the fact of “bad data” (Labov 1994: 11) and the underrepresentation of the lower classes. But within possibilities, the CEEC corpora have been compiled aiming at a balanced corpus, in which there is equivalent coverage of all social parameters (Leech 1993: 13). The social parameters are explained below.

The CEECE contains 77 collections, compiled from 81 different sources (see the Appendix for a list of the corpus collections with their word counts and source details). The compilation process was begun in 2000, and effectively finished in 2006. The external databases have been further augmented since that time.

### 4.3 Coverage (representativeness and balance)<sup>2</sup>

The CEECE contains writers from all social parameters employed in CEEC corpora. In terms of balance it is thicker at the end of the period covered, and its regional coverage is not quite satisfactory. However, as it is compiled according to a model designed for quantitative sampling of Early Modern English society, it is worth remembering that 18th-century England was in some ways remote from what in essence was a post-medieval world. In other words, some of the inherited categories are not necessarily relevant for making sense of 18th-century England and Late Modern English. For instance, East Anglia was highly important in the Tudor period, but by the Georgian era its cultural and political relevance had waned. Similarly, the clergy were not nearly as important in 18th-century England as they had been in the 16th century, thanks to advances in literacy and the availability of the Bible in English. So, the fact that a high proportion of CEECE data comes from the professional classes and from Londoners is rather a reflection of changes in English society than of skewed data. Perhaps the greatest strength of the CEECE is its gender division. More than a quarter of the words in CEECE are by female informants, up from c. 17% in CEEC-1998.

---

2. This section is by Samuli Kaislaniemi and Mikko Hakala.

### 4.3.1 Diachronic and quantitative coverage

The CEEC Extension covers the long 18th century from 1681, where CEEC-1998 ends, to 1800 (inclusive). It contains 4,923 letters by 308 writers, amounting to just over 2.2 million words.<sup>3</sup> The distribution of data is not equal: as can be seen in Figure 4.2 below, there is a dip in the word count in the period 1720–1739; and a full third of the material falls in the last period (1780–1800). Although initially the aim was to choose 10 letters per writer (of ‘average’ length, i.e. excluding very short or long letters), this figure was soon expanded – in part due to abundant sources, in part in order to include good coverage of writers whose letters span a long period of time. Increasing sample size was also one way to attempt at variety in addressees of a single writer, to allow for individual variation according to recipient; and also attempt was made to include letters between two writers, in both directions.

In some cases, it was felt reasonable to include writers from whom there were less than the ideal 10 letters available. These concern primarily writers from the lower social ranks, and also women. Particularly good examples of collections with letters from the lower social ranks include *CLIFT* and *PAUPER* (see the Appendix for details), both of which contain letters from poorly literate writers.

### 4.3.2 Gender balance

One of the aims of CEECE, as noted above, was to compile a corpus in which the proportion of female informants was higher than in the CEEC-1998. Figure 4.1 tracks the number of male and female informants for the years 1680–1800 in 20-year periods. Figure 4.2 shows the number of words in our data set divided into 20-year periods.

---

3. The studies in this volume focus on the period 1680–1800 and the CEECE provides the vast majority of the material studied. There is, however, some overlap between the original CEEC-1998 and the Extension. CEEC-1998 contains altogether 29,199 words written between the years 1680 and 1681, and CEECE 31,600 words from 1653–1679. For the sake of accuracy, only material dating after 1679 was included in our data set, which in all comprises 4,945 letters by 315 writers, coming to 2,216,119 words. To give an accurate characterization of the material used for the studies in this volume, the tables and graphs in this chapter are based on this latter data set (all letters in CEEC-400 spanning 1680–1800) rather than on the CEECE proper. However, as more than 98% of the material is from the CEECE, the figures may also be taken as a fairly accurate description of the structure of the CEECE. See the Appendix for more detailed information on the collections included in CEEC-1998 and CEECE.

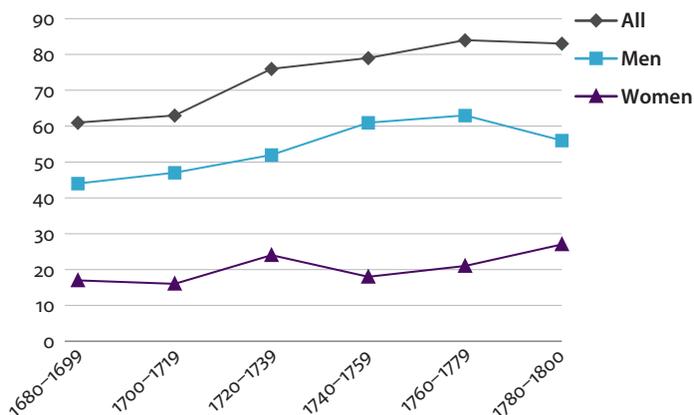


Figure 4.1 Number of informants per period: gender division (absolute frequencies)

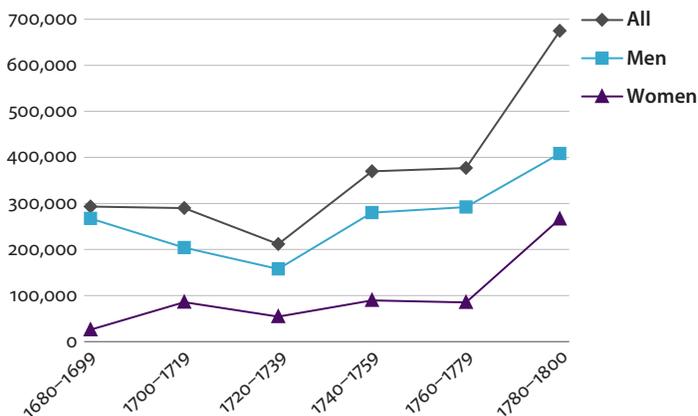


Figure 4.2 Number of words per period: gender division (absolute frequencies)

Apart from a minor slump in 1720–1739, there is relatively little fluctuation between 1680 and 1779. The last period reflects a major increase in the overall availability of data. One major component of the last period are the family letters of the kings George III and IV. The first Hanoverian monarchs were not native English speakers, but from George III their letters are included in the corpus, and happily the extensive family correspondence of the House of Hanover has been published and could be included in the corpus (GEORGE 3A, GEORGE 4; see the Appendix).

As seen in Figure 4.3, there are no dramatic changes in the proportion of female informants, which remains at around 30% over the entire 120-year period. Figure 4.3 reveals that the proportion of data from women is mostly between 20–30%, but the first and last period, 1680–1699 and 1780–1800, are exceptions. For the first period, for some reason there are simply less letters per writer in the source editions. For the last period, the rise in material is due in part to royal family letters, which include letters from four princesses as part of the family correspondence of the House of Hanover. The word count is also bolstered by the letters of writers such as Fanny Burney, Hester Piozzi (Mrs Thrale) and Sarah Duchess Lennox, and also by those of Jane Austen and Mary Wollstonecraft.

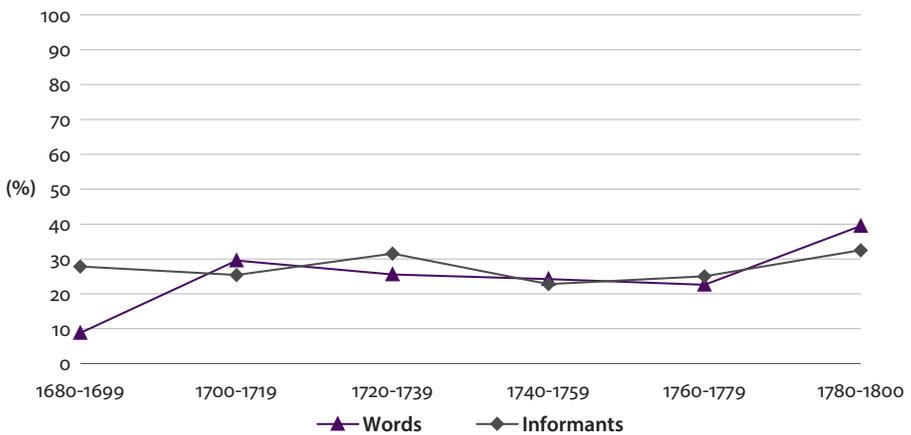


Figure 4.3 Proportion of running words and female informants per period

### 4.3.3 Social ranks

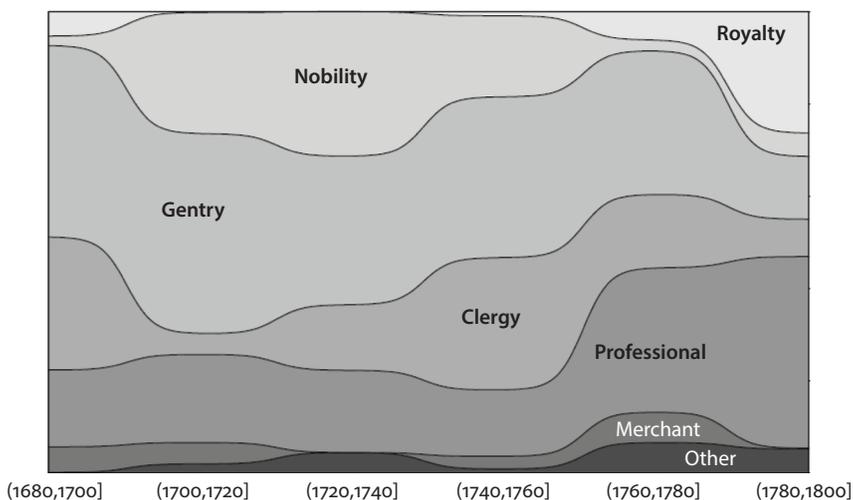
Table 4.1 provides a breakdown, based on running words, of the data according to the social status of the writer.

English society in the 18th century was greatly different from previous centuries; therefore the categories are not directly equivalent with those in CEEC-1998. In particular, the development of the wealth and (political) power of the middling ranks – merchants and professionals – makes them a highly influential part of society in the eighteenth century (see further 2.4). Their rise to prominence is particularly reflected in their share of the data – a full third of the words are by professional or merchant writers (even though the ratio of merchants is much smaller than in CEEC-1998, cf. Nevalainen & Raumolin-Brunberg 2003: 46).

**Table 4.1** Social status of informants, percentage of running words per gender

	Men	Women	Total
Royalty	7%	6%	6%
Nobility	6%	34%	14%
Gentry (Upper)	9%	3%	7%
Gentry (Lower)	21%	23%	21%
Clergy (Upper)	6%	0%	5%
Clergy (Lower)	14%	10%	13%
Professional	30%	21%	27%
Merchant	3%	1%	3%
Other	4%	4%	4%
Total	100%	100%	100%

In general, there is more data from the upper than the lower ranks, but the social coverage of men's letters is somewhat more even than women's. This is an artefact of the sources, but also of English 18th-century society: on the one hand, most letter editions publish the correspondence of the elite; on the other, literacy strongly correlated with social rank, and the spread of literacy down the social ladder was slow. Figure 4.4 shows the distribution of the different social status groups over time in CEECE.



**Figure 4.4** Density plot of the distribution of letters by social rank (courtesy of Harri Siirtola)

Figure 4.5 shows the proportion of data from the highest and lowest ranks in 20-year periods – royalty and the nobility on the one hand, and non-gentry on the other. There is a small increase in the proportion of data from the lowest ranks in the last two subperiods, but otherwise their proportion remains quite low throughout. There is some fluctuation in the amount of data from the highest ranks, the most notable feature in Figure 4.5 being the spike in 1720–1729. This however is an artefact caused in part by the slump in the overall amount of data (Figure 4.2), coinciding with a period where there are more letters from the nobility than from other social ranks.<sup>4</sup> The rise in the last period, on the other hand, is explained by the inclusion of royal family letters, as explained above.

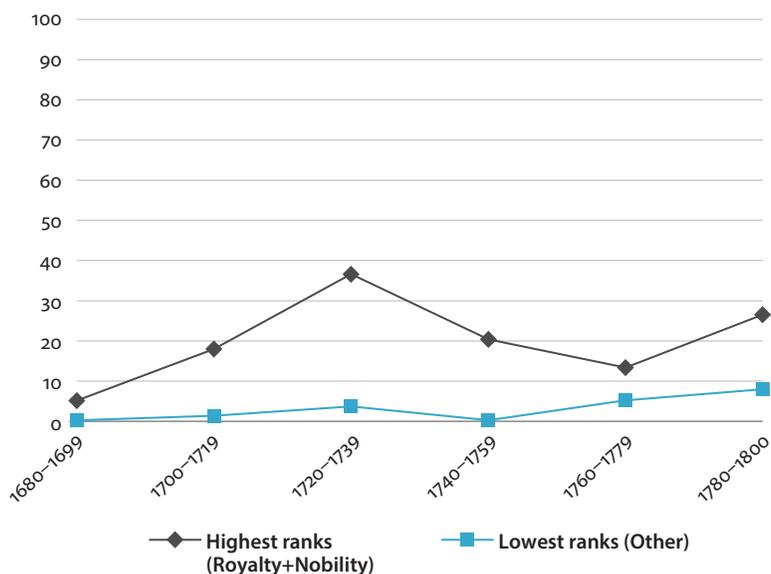


Figure 4.5 Proportion of running words, highest and lowest ranks

#### 4.3.4 Regional coverage

The CEECE follows the same division into regions as is used in CEEC-1998. The encoded regions are: the North, East Anglia, London, the Court, the Home Counties, Other Areas in England, and Abroad (see Figure 4.6). Key characteristics of the regional coverage of CEECE are the rise of the proportions of informants living in London or so-called Other Areas. Between 1650 and 1800, London more than

4. To illustrate the skew another way, the word counts of letters from nobility for the periods 1720–1739 and 1740–1759 are the same (just over 81,000).

doubled in size, and the proportion of the population living in London grew from around 8% to 12% (Boulton 2000: 316; Schwartz 2000: 648; Wrigley & Schofield 1981: 207; see 2.3, above). This increased importance of London is reflected in the contents of the CEECE. At the same time, increased regional mobility makes it hard to pinpoint where some informants come from. This uncertainty is on the other hand counterbalanced in CEECE by registering domicile at a county level.



**Figure 4.6** Map of regions covered by CEECE (1 North, 2 East Anglia, 3 London, 4 Court, 5 Home Counties, 6 Other areas)

Table 4.2 gives the regional distribution of the data according to the informants' domiciles. The most striking features here are the dominance of London and the dearth of data from East Anglia. The former is closely linked with increased domestic migration and the rise of the professional classes. The lack of data from East Anglia, on the other hand, is a reflection of what has been edited – it turned out that there are very few editions of letters from eighteenth-century East Anglian writers.

**Table 4.2** Regional distribution, percentage of running words per gender

	Men	Women	All
The Court	5%	3%	5%
London	38%	58%	43%
East Anglia	1%	0%	1%
Home Counties	17%	12%	16%
North	10%	1%	8%
Other areas	25%	20%	23%
Abroad	4%	5%	4%
Total	100%	100%	100%

#### 4.4 Coding

The coding of CEECE has followed that of CEEC-1998 (Nurmi 1998), which was derived from the coding used in the *Helsinki Corpus* (Kytö 1996).

There are two kinds of coding: text-level coding, and parameter coding. The former includes information contained in the source edition, such as special characters and textual commentary, but also information encoded by the CEECE compilers, such as omitted passages and foreign language strings. Parameter coding on the other hand contains extratextual information, including the file name and bibliographical information, page numbers, and importantly, information on the letters and their writers.

Each letter is preceded by a letter header consisting of four encoded lines. The first L-line contains the unique **letter ID**. The second is the Q-line: it contains a code letter indicating the **authenticity** of the letter, the **year** of writing, a code letter indicating the **relationship** between the writer of the letter and the addressee, and then the **correspondent code** – a unique identifier of the writer of the letter. The third line is the X-line, which contains only the **name of the writer** in human-readable form, as opposed to the opaque correspondent code. The fourth and final line is the P-line, containing the number of the **page** on which the text begins in the source edition (further P-tags are also inserted in the text at every page break).

For more detailed information on the coding used in CEEC-400, see Nurmi (1998) and Raumolin-Brunberg & Nevalainen (2007).

#### 4.4.1 Letter quality

Since the CEECE was designed to gain access to the language of individuals, letter quality was a primary concern. Authorship is the most critically relevant feature here: only letters that were demonstrably written by an identifiable individual are included in the corpus. Scribal letters and copy-letters are marked as such. Linked to this, the extent to which a writer's social background was recoverable was important. And thirdly, only editions that retained original spelling were used – although this in practice meant including works which have normalized u/v-variation (“vp” > “up”, “ouer” > “over”), expanded abbreviations (“w<sup>ch</sup>” > “which”, “y<sup>e</sup>” > “the”), and modernized capitalisation and punctuation. Because different editions have followed differing practices, and furthermore as nearly all editions have to some extent normalized spelling, as a rule, the CEEC corpora cannot be used to study orthographical features.

These three aspects have been combined to give each letter a rating of **authenticity**. The authenticity codes are:

- A = autograph letter in a good original-spelling edition; writer's social background recoverable
- B = autograph letter in a good original-spelling edition; part of the writer's background information missing
- C = non-autograph letter (secretarial work or copy) in a good original-spelling edition; writer's social background recoverable
- D = doubtful or uncertain authorship; problems with the edition, the writer's background information, or both.

In practice, the proportion of non-A letters in CEECE is so small that authenticity does not play an important role in analysing results of corpus queries. More than 96% of the letters in CEECE are quality A.

#### 4.4.2 Relationship between writer and recipient (register)

Social relations between correspondents are encoded into the corpus texts. Addressee relations allow tracking register variation, as distance and power relations manifest in the language of the letters (see Nevalainen & Raumolin-Brunberg 2003: 189–191 and 191ff.; also Nevala 2004a (reprinted in 2004b)). There are five basic relationships:

- FN = Nuclear family members (close)  
 FO = Other family members (more distant)  
 FS = Family servants (distant)  
 TC = Close friends (close)  
 T = All other relationships (not close)

Establishing these categories is fairly straightforward, although on occasion it can be difficult to identify close friendships, and sometimes determining kinship can require some genealogical archaeology. The distribution of the letters in CEECE according to relationship of writer and recipient is shown in Table 4.3.

**Table 4.3** Relationship of writer to recipient, percentage of letters per gender

	Men	Women	All
Nuclear family (FN)	27.6%	45.1%	32.0%
Other family member (FO)	8.5%	11.1%	9.2%
Family servant (FS)	0.2%		0.1%
Other (T)	36.2%	21.1%	32.4%
Close friend (TC)	27.4%	22.7%	26.2%
Total	100%	100%	100%

As can be seen, there are proportionally more family letters from women than from men in CEECE. This is because relatively more family letters survive from women than from men, which in turn is a reflection of the patriarchal structure of society as well as of literacy rates (for women letter-writers, see Daybell 2001, 2006).

#### 4.5 Corpus formats and external databases

What makes the CEEC corpora stand apart from other historical corpora are their supporting databases: the **letter database** contains information regarding each individual letter, and the **correspondent database** contains information on each writer and recipient.

The letter-specific information includes information on the textual source (bibliographical details of source edition, page numbers, letter number in edition if any), details on the contents of the letter (date and place of writing, content type, word count), and brief information on the writer and recipient (unique correspondent IDs, names, social status and rank at the time of the letter, and the relationship of the writer and addressee). We have also recorded information of particular linguistic relevance (letter authenticity; presence of opening and closing formulas), and

metadata relating to the corpus itself (unique letter ID, collection name, copyright permissions status, state of completion of database entry, last update). Most of this information is stored in free text fields, but some (such as social rank of correspondents, their relationship, and letter authenticity) is recorded as code.

The correspondent database contains more in-depth information on each correspondent. Although it was begun as a sender database (e.g. Raumolin-Brunberg & Nevalainen 2007: 160–164), it has been updated to include recipients as well, so that it is now possible to find all letters in CEEC-400 written to recipients matching given parameters. Much of the information in the correspondent database is stored in both codes and free text fields to facilitate quick searches on the one hand, but also to retain more detailed information that is difficult to codify, such as specifics relating to the education, migration or career of any individual.

For reasons primarily to do with the state of computing at the time the CEEC project was planned (the early-to-mid 1990s), the sociohistorical information required in this kind of a sociolinguistic corpus was entered into external databases rather than encoded into the corpus texts. But in 2007, the CEEC-400 was uploaded into a relational database on an online server, accessed by a bespoke search tool called CEECer (pronounced ‘seeker’), used with a web browser. This integration now allows us to search the corpus using as query parameters any information in the databases, greatly enhancing the process of performing corpus searches. In essence, with CEECer we can create subcorpora from CEEC-400, which can then be extracted and queried in a concordancer or other corpus tool.

The current base format of the corpus is no longer by collection or even by writer, but by individual letter. The base texts remain ASCII plain text files (a format inherited from HC), but these have been supplemented with the POS-tagged and syntactically parsed texts of PCEECE, and also with the standardised-spelling version of the corpus (see Palander-Collin & Hakala 2011). All these versions have been stored in CEECer.

#### 4.6 Copyright and publication

The publication of a corpus built from previously printed sources requires acquiring copyright permissions from the editors and publishers of the source texts. Of course, this is an old concern, voiced for instance by Kytö & Rissanen (1997: 18–19). The process for acquiring permissions to publish the CEEC-1998 corpus was begun in the early 2000s. This turned out to be more difficult than envisaged, and for some texts the permission was not received, as noted above.

When the copyright-clearing process was started for CEECE in 2010, we expected similar difficulties. Yet over a few years, publishers' attitudes to corpus compilation and similar work had changed greatly. The launch of mass digitization projects like Google Books (in 2004) had raised publishers' awareness of the power of digitization, and resulted in what can only be called confusion and panicked protectionism. In sum, even publishers that had formerly kindly allowed using texts published in their books in linguistic corpora, now were reluctant to grant permission, or denied it outright. As of late 2017, we have not received permission for some 2/3 of the texts in CEECE, and publication of the corpus remains uncertain.

---

## INFOBOX 1

### Data retrieval

Mikko Hakala

This section provides a short description of the data retrieval process for each study in this volume. Although the search parameters were specific to individual studies, the retrieval process usually followed the same basic outline: a research assistant first performed the corpus queries and sorted the results and each researcher then checked and further pruned their own data. The main tools used in retrieving the data were WordCruncher 4.30 and WordSmith 5.0, with the former mostly preferred with single-word search terms and the latter with search terms containing multiple words or wildcards.

Associating the results with relevant sociolinguistic metadata is an important consideration in our research. Both programs provide significant help with this task; WordCruncher automatically retrieves letter-specific ID-tags from the corpus files and adds them to the extracted data, while WordSmith can be instructed to add the same ID-tags with a separate tag file. Biographical information about the authors and the recipients can then be extracted from a separate database and associated with examples on the basis of the ID-tags. As spelling variation is still common in 18th-century letters, dictionaries (mainly the OED and the MED) and wordlists generated from the corpus (using WordSmith and WordCruncher) were consulted to cover the different spellings of each search term as thoroughly as possible. All results were extracted from the concordance programs with a context of five lines of text before and after each search term. The results were always checked manually for false hits.

Data retrieval for the studies dealing with individual lexemes was relatively straightforward. In her chapter, Nevala focuses on the personal pronouns *my*, *mine*, *thou*, *thee*, *thy*, *thine*, *you*, *your*, and *yours*. All spelling variants of these pronouns were searched for using WordCruncher. For Nevalainen's study of the third-person indicative suffixes *-(e)th* and *-(e)s*, the relevant forms of the verbs *do*, *have* and *say* were searched for. And for Laitinen's study, the search terms were the indefinite pronouns *-one*, *-body* and *-man*. For Nurmi's study on *do*, the various forms (*do*, *did*, *don't*, *didn't*) and their spelling variants were retrieved from the data. The instances of periphrastic *do* in affirmative statements were then identified and manually sorted.

Palander-Collin's chapter deals with *its*, *of it*, *of the same*, and *thereof*, all instances of which were retrieved using WordSmith. For the search terms *of it*, *of the same* and *thereof*, the search parameters were further specified to include a definite article to the left of the target word/phrase and up to six words intervening (e.g. *the sulliage and infection of it*), which had proved sufficient during trials.

For Sairio's study of the progressive aspect, all instances of BE + *-ing* were extracted from the corpus with, on the basis of trials, up to four words intervening. Forms of the present participle incorporating the clitic *a* (e.g. *a going*) were taken into account. Different spelling variants of both the verb and the past participle were searched for using wildcards in WordSmith.

Säily's study deals with the derivative suffixes *-ness* and *-ity*. The data for her study were retrieved with WordSmith by using a search list of potential spelling variants, including plural forms, for both suffixes. The search list was created on the basis of previous research, the OED and close readings of the corpus. The results were connected to the relevant sociolinguistic metadata with Python scripts.

## Acknowledgments

For a more detailed overview of the history of the CEEC project, see Nevalainen & Raumolin-Brunberg (1994a; 1996 (eds.); 2003), Raumolin-Brunberg & Nevalainen (2007), and the CEEC entry in CoRD. The premises of the CEEC *Extension* were first outlined in Laitinen (2002).



## Research methods

### Periodization and statistical techniques

#### 5.1 Quantifying change

Terttu Nevalainen

##### 5.1.1 Need for multiple methods

Past work has shown that the letter genre differs from other written genres in several respects linguistically. Analysing persistent register features in the ARCHER corpus, Biber (2001: 100) concludes that 18th-century personal letters, for example, are “nearly as involved as drama”. However, in a socially stratified corpus these broad register findings do not preclude internal variation. Apart from the usual speaker variables that include age, gender and socio-economic status, issues such as the relationship between the sender and recipient of a letter have a role to play. Chapter 4 looked at the CEECE in extralinguistic terms and estimated the proportions of the various social categories sampled over time. This information provides the basis needed for interpreting in social terms the linguistic variation found in corpus data.

This chapter introduces the quantitative methods that we have adopted for analysing and comparing the processes of linguistic change investigated in this volume. Our aim is to accumulate evidence for real-time processes of change by using data sources and analytic techniques that provide a close match with past work while taking into account both group and individual patterns (cf. Tagliamonte 2012: 108–110, Brezina & Meyerhoff 2014: 23–24). However, new approaches are also called for to arrive at more adequate quantitative descriptions of real-time changes. As noted in Section 1.4, in the long term, most of the processes we analyse can be described using the S-curve model, although periods of stable variation are also found (cf. Labov 2001: 85–86). For both these alternatives, we propose a method of periodization that allows the researcher to make comparisons across the time spans of changes in the long 18th century. Section 5.1.2 discusses this approach in relation to the general issue of periodization in historical corpus studies.

The changes we study can be roughly divided into those that are based on a linguistic variable, often defined as alternative ways of “saying the same thing”, and those that less easily lend themselves to a variationist approach. We adopt different

techniques for the quantitative analysis of these two kinds of process. Studies analysing variables are carried out using basic methods of estimating frequencies that utilize as much of the primary material as possible and allow direct comparison with past work in the field. Standard methods such as pooling frequency scores and averaging individual averages are implemented in the chapters discussing indefinite pronouns, the second-person singular *thou*, third-person neuter possessive *its*, and verbal *-s*. These methods are discussed and evaluated using some more sophisticated techniques in Section 5.2.

Quantifying processes of change that do not have a readily definable linguistic variable call for more advanced quantitative methods. We introduce and illustrate two such techniques in Section 5.3. They enable flexible sociolinguistic comparisons across different groups of people and provide an exploratory approach to the study of variation. These methods are applied in chapters that focus on periphrastic *do*, the progressive aspect and the nominal suffixes *-ness* and *-ity*. They could naturally also be applied to processes that are discussed in terms of linguistic variables and will therefore be useful in future work on such topics as well.

### 5.1.2 Periodizing processes of change

As shown in Chapter 4, it has been our aim as corpus compilers to facilitate the study of linguistic processes across time and the social space by presenting the material in twenty-year periods and, to enable diachronic comparisons, by using the social categories developed for earlier research on the CEEC. These sampling principles allow the researcher to compare social groups' and individual letter writers' simultaneous participation in ongoing processes of change and to identify the loci and leaders of change in social and regional terms (Nevalainen, Raumolin-Brunberg & Mannila 2011).

However, despite our efforts, the material sampled for the CEEC is unevenly distributed over time in that the corpus does not contain equal amounts of data for each social category in each 20-year period. This may become an issue with fine-grained social distinctions or shorter subperiods. Using longer periods and less fine-grained social status schemas is of course possible, depending on the research question at hand. Whatever the issue, a balance needs to be struck between data granularity and the generalizations to be made on the basis of it. Our studies therefore aggregate the data over 40-year periods as well. Similarly, a five-class model of the social order is adopted, for example, to account for the final stages of the generalization of verbal *-s* in social terms.

Some automated methods for identifying stages in the temporal dimension of linguistic changes are also available. Gries & Hilpert (2010) used variability-based neighbour clustering (VNC) to identify the temporal stages to which the diffusion

of verbal *-s* could be divided in the PCEEC data between 1417 and 1681. They arrived at five stages: 1417–1478, 1479–1482, 1483–1609 (excluding the years 1509 and 1544), 1610–1647, and 1648–1681 (excluding 1649) (Gries & Hilpert 2010: 302). This division shows that a sample bias in the corpus is replicated by a bottom-up approach of this kind. The period 1479–1482 is largely due to one family and one particular individual who contributed a large sample to the corpus but deviated from the mainstream southern English usage of the day. This bias is naturally also reflected in basic-level periodizations of the data, as in Nevalainen & Raumolin-Brunberg (2003: 68), but duly accounted for in discussions of the regional diffusion of the innovation (2003: 178). In sum, the level of abstraction often varies in periodizations. While no historical sociolinguist would set great store by a three-year period in late Middle English data as an indicator of real-time change, this brief stretch of time can nevertheless indicate significant variation in the corpus at that point.

Since linguistic changes progress at different paces and along their unique paths over time, a common yardstick is useful in comparing processes in their own terms. In this volume we adopt a more process- and corpus-aware approach to periodization than a bottom-up analysis of the kind used by Gries & Hilpert (2010) would offer us for that purpose. To achieve that, we link the five stages of linguistic change sketched by Labov (1994: 79–83), i.e. incipient, new and vigorous, mid-range, nearing completion and completed, to the gradual diffusion of the incoming form. Labov's model reflects speakers' age levels in synchronic apparent-time analyses of phonological variables but it is also extendible to real-time analyses of processes of change. The five stages proposed in Nevalainen & Raumolin-Brunberg (2003:55) divide the slope of the real-time S-shaped curve as follows:

Incipient	below 15%
New and vigorous	between 15% and 35%
Mid-range	between 36% and 65%
Nearing completion	between 66% and 85%
Completed	over 85%

This approach allows us to compare changes and their social embedding both within and across stages. It makes it possible, for example, to compare the rate of real-time change with patterns observed in apparent time. We can ask, for example, whether it is true for real-time processes as well, as Labov (1994: 82) found to be the case for apparent time, that the rate of change is fastest at the new and vigorous stage, and slowest when the change is almost completed. Other points of comparison include the sociolinguistic patterns that characterize changes at midpoint, where we may expect contact between speakers and their individual differences to be greatest (Labov 1994: 65–66, Kurki 2005: 239–240).

Labov (1994: 82–83) further argues that in phonological changes the level of social awareness is maximal for changes nearing completion and minimal for changes at the new and vigorous stage. We can find out whether this is also the case with real-time morphological changes by comparing the stages of the changes we have analysed with comments found in normative grammars, on the one hand, and with actual usage, on the other. Another question concerning the completion of changes relates to outgoing variants and the extent to which they recede to certain limited functions or may be reappropriated in new uses.

## 5.2 Basic methods for estimating frequencies

Terttu Nevalainen

One of the issues with small-to-medium sized corpora is what Rissanen (1989) called “the mystery of vanishing reliability”. By this telling label he referred to the problem arising with small corpora that have been divided into and annotated for various use- and user-related categories.

The number of parameter values is, of course, inversely proportional to the amount of evidence in each information area sampled. For this reason, particularly in a corpus divided both according to chronology and text type, it may be difficult to maintain the reliability of the quantitative analysis of less frequent syntactic and lexical variants. The problem becomes even more obvious if attention is paid to sociolinguistic parameters. (Rissanen 1989: 18)

Issues also arise from the quantitative methods selected. In corpus studies, pooling subgroup data and calculating the average frequency from the total number of occurrences is commonly used in assessing both chronological developments and social variation. Subgroup data are aggregated by pooling because it is easy to carry out and provides a methodological basis for comparison with earlier studies. However, subgroups consist of individuals and, unless quota sampling is used by fixing the variable total for each individual, this method suffers from unequal contributions made by individual samples of different sizes. As Nevalainen & Raumolin-Brunberg (2003: 214–217) illustrate with CEEC data, the problem becomes acute with individuals with large samples contributing considerably more data than others and hence skewing the overall result of the analysis.

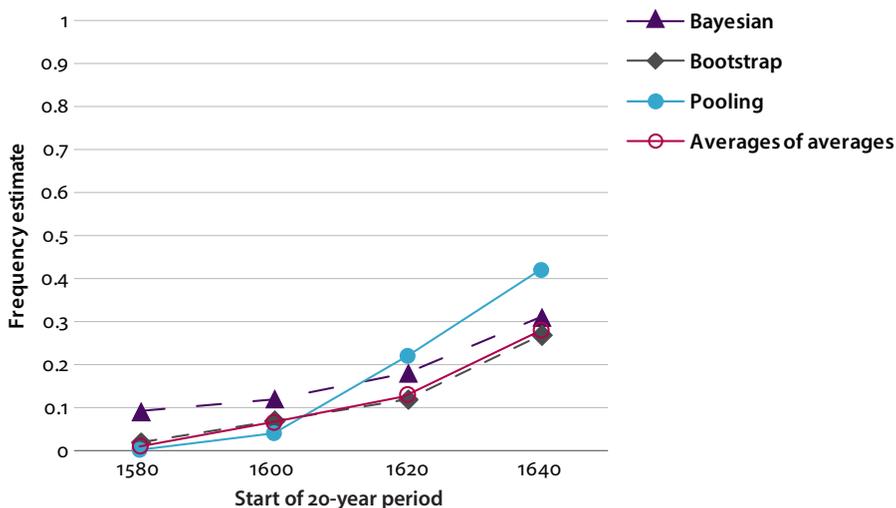
Another basic method of calculation, which avoids this problem, is computing the mean frequency of averages, that is, aggregating subgroups by calculating each individual’s average frequency separately and then calculating the average of these

averages. This way each informant is given equal weight in the outcome. But sample size and the number of attestations may again affect the outcome: if one writer has one instance of the incoming form and none of the outgoing one, the contribution of this individual would be considered equally important as that of another writer who produced no instances of the new form but as many as 40 of the old one. Using this method, we will need to set minimum frequencies for the number of observations for each informant to prevent such extreme disparities.

An issue related to both these methods, pooling and averaging of averages, is poor *dispersion*, a situation in which there is only a small number of informants, who show large differences in the amounts data they have produced. Hence some minimum frequencies for the number of contributing individuals will also need to be set. Ideally, adequate frequency levels can be set by testing how well subgroups match the larger datasets that they form part of.

Pooling and averaging averages are both maximum likelihood methods in that they try to obtain one estimate for the value of the unknown frequency of the incoming form. There are other, more sophisticated methods that can be used for estimating frequencies. In earlier work we compared pooling and averaging of averages with both *bootstrapping* and the *Bayesian approach* (Hinneburg et al. 2007, Mannila, Nevalainen & Raumolin-Brunberg 2013). In bootstrapping, the uncertainty in the frequency estimate is examined by resampling the original data a large number of times (cf. also permutation testing discussed in Section 5.3.2). For each bootstrap sample a method such as pooling can be used for estimating the frequency of the incoming form, and the median of these estimates is then taken to represent the bootstrap estimate for the frequency of the incoming form. The Bayesian method is rather more complicated taking into account both the population frequency and each individual's frequency in order to produce probable values for the parameter studied.

One way to assess the degree to which the simple and more sophisticated methods converge is to apply them to the same data set. Figure 5.1 compares the frequency estimates obtained using all four methods to study the gradual replacement of the third-person singular suffix *-th* by *-s* in *have* (*hath* v. *has*) between 1580 and 1660. The calculations are based on a dataset consisting of over 400 people and the total of 2,464 instances of *hath* and 472 of *has*. The four curves all indicate the relatively slow process made by *has* in the 80-year period, basically suggesting that it progressed from the incipient to the new and vigorous stage.



**Figure 5.1** Results given by four estimation methods for the frequency of *has* in four 20-year periods from 1580 to 1660 in the CEEC

All four methods produce quite similar results with the exception, from the 1620s onwards, of pooling, which indicates higher frequencies than the other three. Because of the prior information required by the Bayesian method, the first period starts on a higher level with that technique than with the others. Averaging of averages and bootstrapping produce closely matching results. This was the general conclusion that we came to applying these four methods to different datasets in Mannila, Nevalainen & Raumolin-Brunberg (2013).

Unlike pooling and averaging of averages, the bootstrap and Bayesian methods produce intervals representing the degree of uncertainty in the estimates. With bootstrap methods, the standard deviation of sample frequencies can be used to yield a confidence interval for the frequency in the original data. These confidence intervals directly reflect the amount of data analysed. For the data in Figure 5.1 the confidence intervals were very narrow and for the bootstrap method, for example, overlapped only partly in the first two periods and not at all in the last two. In Mannila, Nevalainen & Raumolin-Brunberg (2013) we found that having at least 15 persons with at least 10 instances per variable in a bootstrap estimate gave a good fit between a subgroup and the full dataset from which it was drawn. Even fewer instances, such as five per person, could give a reasonable fit, provided that the number of informants was large enough.

However, it is a common occurrence that the researcher is left with smaller subgroups especially with low-frequency linguistic features or when studying short time periods or a range of socio-economic categories. This was the case with the

analysis we carried out to determine whether individuals were progressive or conservative with respect to an ongoing change in Nevalainen, Raumolin-Brunberg & Mannila (2011), where we applied the bootstrap method to the CEEC data in a number of linguistic changes. To minimize skewing, we only examined periods for which there was data on the use of the study variable for at least five individuals. Moreover, the procedure for determining whether an individual was progressive or conservative was applied only if the person had at least six occurrences of the variable, that is, the sum of the occurrences of the outgoing and the incoming variant for the person was at least six.<sup>1</sup>

In Part II, the variation studies of linguistic changes make use of the basic methods discussed above, bearing in mind their limitations especially with low-frequency variables. Some chapters also consider the differences produced by pooling and averaging of averages, and those that analyse linguistic variables use bootstrapping to assess the degree of uncertainty in their estimates. Most of the studies also raise the level of abstraction by aggregating data over longer periods and over broader social categories than those specified in our sampling frame. These measures help diminish the degree of “vanishing reliability” and make for more reliable results. The measures used naturally impact on the generalizations that can be made on the basis of the quantitative findings. Using multiple methods, we hope to offer empirical baseline information of different kinds on the sociolinguistic contexts of language change in 18th-century English.<sup>2</sup>

Finally, recognizing the internal heterogeneity of subgroups, we also discuss outliers, individuals who deviate from others, either leading the process of change or lagging considerably behind their contemporaries. They are of particular interest both as unique individuals and as representatives of intersecting sociolinguistic categories and communities. Since we are always analysing the same set of people, the ways in which these individuals pattern with respect to linguistic changes in their different stages make interesting comparisons. This issue is addressed with all the processes studied, regardless of whether they can be construed in basic variationist terms or not.

---

1. For further discussion of these and other quantitative methods applied to the CEEC, see Hinneburg et al. (2007), Nevalainen, Raumolin-Brunberg & Mannila (2011), Mannila, Nevalainen & Raumolin-Brunberg (2013) and Nevalainen (2014a).

2. Nevalainen & Raumolin-Brunberg (2003: 193–200) used a Variable Rule Analysis application (GoldVarb) to study the relative impact of the various social variables on processes of change. In this volume it would have been possible to update the traditional Varbrul toolkit and resort to more recent techniques such as Rbrul (e.g. Tagliamonte 2012: 138–157). We decided not to pursue that line of inquiry but to explore a variety of measures and make more transparent the basic principles of counting frequencies instead.

### 5.3 Methods for studying changes lacking a variable

Tanja Säily, Arja Nurmi and Anni Sairio

#### 5.3.1 Introduction

Some linguistic structures and forms do not present an obvious variable to analyse. In this volume, we look at three main types of such cases. Firstly, there are cases where it is reasonably simple to identify the possible variable but unfeasible to retrieve it from an untagged corpus. So, in the case of periphrastic *DO*, the variable would include all instances with a bare main verb that could be used with *DO*, called the “simple form” by Ellegård (1953). There are some borderline cases regarding verbs to include, but the main principle is clear, and it is merely the problem of data retrieval that prohibits the use of the variable. (But see Chapter 8 for a more detailed discussion.)

The second type of structure without a variable is one where identifying the form is more problematic. This is illustrated by the progressive, where the varying complexity of the verb phrase associated with the form does not allow for a clear definition: all VPs do not accept the progressive aspect, and a simple normalisation of the frequencies does not indicate how much possibility there is for variation to occur. An added difficulty is the spread of the form into new grammatical contexts over the period studied: should the variable change accordingly? There would also be the choice of treating all finite verb phrases as the variable, but this would again require a tagged corpus. For a discussion of the problems associated with defining the variable for the progressive, see Smitterberg (2005), Kranich (2010), and Chapter 11 below.

The third type of change without a variable concerns a different type of problem, where defining a reasonable variable would not be solved with the help of a tagged corpus. The study of derivational morphemes *-ness* and *-ity* leaves even more problems open for question. Should the variable include all abstract nouns? All nouns with any derivational morpheme? What about the bases: Should we include all adjectives (and other word classes) which could be used as bases for these morphemes? Only those bases where each of the morphemes in question is genuinely possible? For example, *-ity* can generally only be used with bases of a French or Latinate origin, whereas *-ness* takes both foreign and native bases. For further discussion of defining the linguistic variable in derivational morphology, see Cowie (1999: 183–189) and Säily (2014: 33–34).

Owing to the issues outlined above, we abandon the idea of constructing a linguistic variable for three of our changes in this volume: periphrastic *DO*, the progressive, and the nominal suffixes *-ness* and *-ity* (Chapters 8, 11 and 12 below). Instead of expressing the frequency of these forms as a proportion of the overall frequency of a variable, we can normalise it as a proportion of the number of running words

in the corpus (Nurmi 1996). But how do we know whether the observed change in the frequency is statistically significant? Recent research (e.g. Kilgariff 2001; Lijffijt et al. 2012; Bestgen 2014) has shown that many widely used tests of statistical significance, such as the chi-square and log-likelihood ratio tests, are in fact inappropriate for comparing word frequencies in texts, because they assume that all words occur independently from each other, which is never true. Moreover, change is often visualised using simple line graphs, which hide the variability within the corpus.

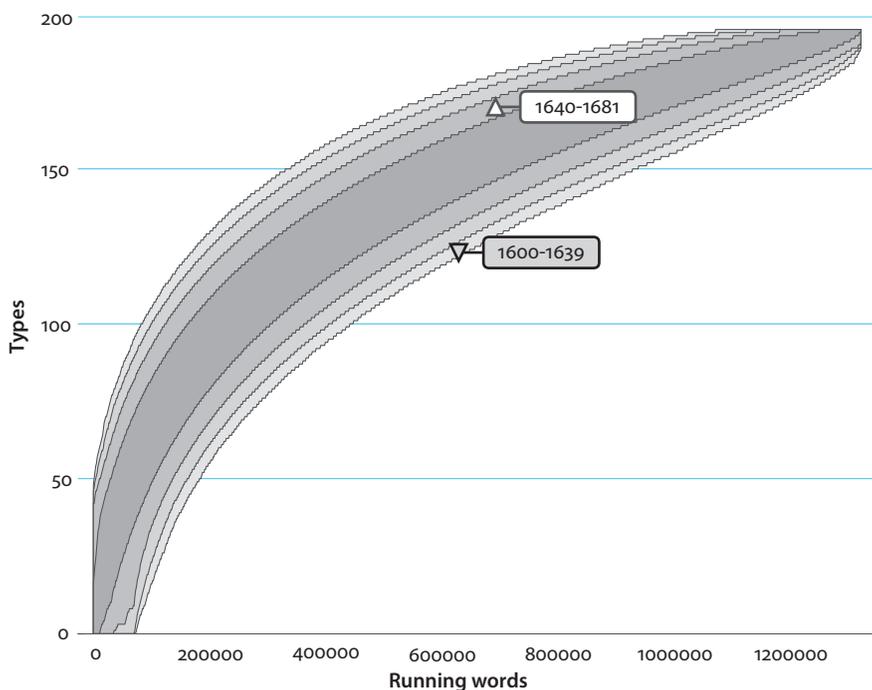
### 5.3.2 Method 1: accumulation curves and permutation testing

We solve the problems of visualising change and determining its statistical significance by using two robust methods. The first of these was proposed by Säily & Suomela (2009). Involving accumulation curves and the statistical technique of permutation testing, this method assesses significance without resorting to simplifying assumptions. It was originally developed for comparing type frequencies in the study of morphological productivity, as in our third type of change above. *Type frequency* refers to the number of different words formed using the element under study in the corpus, whereas *token frequency* refers to the number of all occurrences of the words under study in the corpus.

Unlike token frequencies, type frequencies present the additional complication that they cannot be normalised, because normalisation presupposes that the growth rate of the frequency is linear, which is not the case with type frequency. As corpus size increases, the number of types increases in a nonlinear manner, with more new types being encountered when the corpus is small, and the growth rate decreasing as the corpus gets larger. Therefore, type frequencies obtained from corpora of different sizes, such as subcorpora representing different time periods, cannot be compared through normalisation. Our method eliminates the need to normalise, and it can be applied to both type and token frequencies.

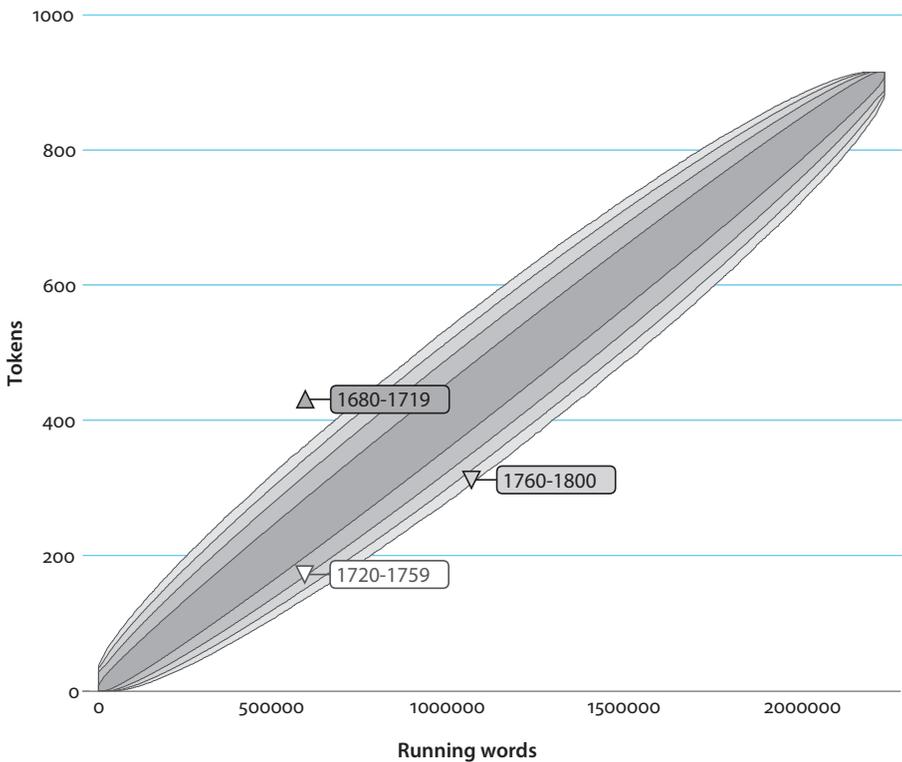
The idea behind the method is as follows. Instead of trying to compare subcorpora of different sizes, we compare actual subcorpora with randomly composed subcorpora of the same size. The randomly composed subcorpora are obtained by dividing the entire corpus into samples and by randomly sampling these using a statistical technique called permutation testing. The samples need to be large enough to preserve discourse structure; they can consist of individual texts or e.g. all texts written by a person during a time period. For each corpus size, a million random subcorpora are sampled by a computer program (Suomela 2014). These random subcorpora give upper and lower bounds for the type or token frequency of the form in question. If the frequency observed in the actual subcorpus is higher than, say, 99.9% of the observations in random subcorpora of the same size, we can say that the frequency is significantly high at a probability of  $p < 0.001$ . For a more detailed description of the method, see Säily (2014).

Figure 5.2 illustrates the results of applying the method to type frequencies of the suffix *-ity* in the 17th-century section of the CEEC, 1600–1681. The *x*-axis shows corpus size in running words, while the *y*-axis shows type frequency. The shaded areas display the range of type frequencies of *-ity* in the random subcorpora: the darkest area in the middle covers most of the subcorpora, and when the next darkest area is added, the coverage increases to 90%, then 99%, 99.9% and finally 99.99% (pure white). We can see that type frequency increases with corpus size in a nonlinear manner: the shaded plot is curved, not straight. We call these figures “type accumulation curves”, although “cumiform plot” has also been suggested to describe the shape. If we look at the actual subcorpus of letters written during the first 40-year period in the corpus, 1600–1639, we can see that its type frequency falls in the lightest grey area. This means that fewer than 0.1% of the randomly composed subcorpora of the same size have such a low type frequency, making the productivity of *-ity* significantly low in this period at  $p < 0.001$ . The second period, 1640–81, is not significantly different from randomly composed subcorpora of the same size, as up to 10% of them have a similarly high type frequency ( $p < 0.1$ ). From these results we may deduce that the productivity of *-ity* increases over time in the corpus.



**Figure 5.2** Bounds for the type frequency of the suffix *-ity* in the 17th-century section of the CEEC, 1600–1681

Figure 5.3 demonstrates the use of the method with token frequencies, namely those of periphrastic *DO* in affirmative statements in the 18th-century section of the CEEC, 1680–1800. Note that token frequency does grow linearly with corpus size, resulting in a straight rather than curved plot. Nevertheless, we call these figures “token accumulation curves”; the term “cucumiform plot” is also applicable here as cucumbers can be either curved or straight. We can see that the first 40-year period, 1680–1719, has a significantly high token frequency, as practically all of the randomly composed subcorpora of the same size have a lower token frequency than it ( $p < 0.0001$ ). The middle period of 1720–1759 is not significantly different from random subcorpora, but the last period, 1760–1800, uses *DO* significantly less frequently than random subcorpora of the same size ( $p < 0.001$ ). We may thus state that the use of affirmative *DO* declines significantly over time in this corpus (see further Chapter 8 below).



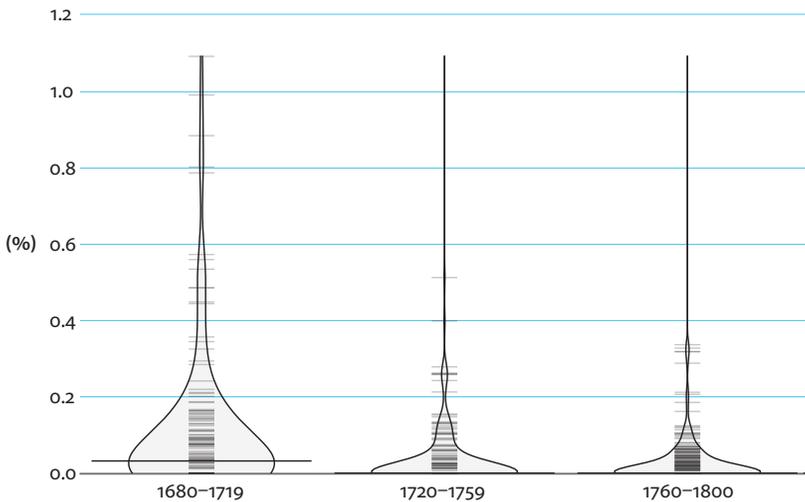
**Figure 5.3** Bounds for the token frequency of periphrastic *DO* in affirmative statements in the 18th-century section of the CEEC, 1680–1800

### 5.3.3 Method 2: beanplots and the Wilcoxon rank-sum test

While the accumulation curves technique is a marked improvement over many earlier methods, it does have some drawbacks. Firstly, as a measure of significance it can sometimes be overly conservative (Lijffijt 2013: 35, 38), so that some genuine differences may be classified as non-significant. Secondly, the visualisation is perhaps unintuitive for studying change over time, as the *x*-axis displays corpus size rather than time. Therefore, it is here complemented with another method applicable to token frequencies (in our case, affirmative DO and the progressive). First applied to historical sociolinguistics by Vartiainen et al. (2013), this method visualises change using beanplots (Kampstra 2008) and assesses statistical significance using the Wilcoxon rank-sum test, also known as the Mann–Whitney U test (Wilcoxon 1945; Mann & Whitney 1947).

Some scholars in diachronic corpus linguistics have been moving from line graphs to boxplots because the latter provide more information on the variability within subcorpora. As a further improvement, Säily et al. (2011) introduce the beanplot to the field (for a comparison of boxplots and beanplots, see Säily 2014: 57–59). Figure 5.4 presents a beanplot view of change in affirmative DO, which was illustrated using accumulation curves in Figure 5.3. The *x*-axis shows time period, while the *y*-axis shows normalised token frequency. In the middle of each of the three “beans” is a vertical scatterplot, where each thin tickmark represents the normalised frequency of affirmative DO in one sample; the tickmarks give an indication of the variability and amount of data within the time period. The samples consist of a person’s letters to a specific kind of recipient (nuclear family, other family, family servants, close friends or other acquaintances) during a 20-year period. This allows us to study the change in sufficient detail while ensuring that no one person contributes more than a few samples and thus cannot easily skew the results. The thicker line going horizontally through each bean represents the median frequency of the samples – while the original beanplot (Kampstra 2008) uses mean frequency, Vartiainen et al. (2013) change this to median, noting that the latter is more robust to outliers. The shape of the bean reveals the distribution of the samples, which is crucial to identifying outliers. Here the shape is mirrored on either side of the scatterplot, but beanplots can also be formed so that they consist of two different subcorpora, e.g. women on the left and men on the right, allowing for easy comparison between the two.

The beanplot shows that the normalised frequency of affirmative DO decreases over time in the corpus, the median practically dropping to zero after the first period. To assess whether the decrease between the first two periods is statistically significant, we use the Wilcoxon rank-sum test (Wilcoxon 1945; Mann & Whitney 1947). Like permutation testing, this test is assumption-free; furthermore, Lijffijt



**Figure 5.4** Beanplot: normalised frequencies of periphrastic DO in affirmative statements in the 18th-century section of the CEEC, 1680–1800

et al. (2016) find it to be among the best-performing methods for comparing word frequencies. We formulate the hypothesis that the normalised frequency of affirmative DO is smaller in samples produced in the period 1720–1759 than in 1680–1719. We then make a list of the samples from both of these periods and order the list by normalised frequency. The null hypothesis is that the samples from both periods are distributed evenly within the ordered list. The test measures how surprising the actual distribution is compared to the null hypothesis, and the  $p$ -value produced by the test tells us the probability that we are wrong in rejecting the null hypothesis. In this case,  $p \approx 0.000000$ , so the probability is extremely low. We may thus say that the use of affirmative DO declines significantly between these two periods in the corpus.

### 5.3.4 Addendum: multiple hypothesis testing

There is an additional component to both of the methods presented above. The  $p$ -value yielded by significance testing indicates the probability that we are wrong in rejecting a single null hypothesis. However, we need to test multiple hypotheses, as we wish to compare many social categories and time periods. The more hypotheses we test, the greater the probability that we are wrong in rejecting the null hypothesis in at least some of the cases. Therefore, we need to adjust the significance level (which is conventionally set to  $p < 0.05$ ) to reflect the number of hypotheses tested. There are many ways to do this; we have chosen to control the **false discovery rate**, or the proportion of false positives out of all positives, using

Benjamini & Hochberg's (1995) procedure. For a simple description of the procedure, see Säily (2014: 50–51). We have chosen an acceptable false discovery rate of 10%, which leads to a different significance level for each of our changes and methods, depending on the number of hypotheses tested and the  $p$ -values gained. For instance, in the case of affirmative DO and the Wilcoxon rank-sum test (194 hypotheses tested), the significance level is  $p < 0.002$ . In the chapters analysing our changes, we will use these adjusted significance levels.

## Acknowledgments

The authors would like to thank Jukka Suomela for methodological assistance and CSC – IT Centre for Science, Finland, for computational resources.

PART II

## **Studies**



## “Ungenteel” and “rude”?<sup>1</sup>

### On the use of *thou* in the eighteenth century

Minna Nevala

#### 6.1 Introduction

Think not, my Maria, my beloved wife, yt. I have deceived ym. or mean to deceive ym. in the least by what I have said. No, I have look'd upon myself as married for above this year & half, ever since I made to **thee & thou** madst to me in the face of Heaven, a vow mutually to regard one another as man & wife, & never to belong to anyone else but to each other, & wh. vow **thou** hast so lately confirm'd by ye. consent **thou** hast given to *become mine*. Dost **thou** think, **thou** only beloved of my heart, yt. if it had been possible for **thee** to disregard such bonds, to have broken yr. vows & to have thrown yourself into the arms of another, first of all I neither cd.<sup>2</sup>

The above passage is taken from a love letter. If we had to guess when the letter was written and by whom, we would probably say that it was sent from a loving seventeenth-century husband to his newly-wed wife. The use of *thou* and the recipient's first name would reveal to us that the relationship was extremely close and passionate.

In reality, the relationship between the writer and the recipient was indeed close and passionate at the time of writing. George, Prince of Wales, had become infatuated with the twice widowed Maria Fitzherbert a year earlier, in 1784, and pursued her into marrying him. Secretly, the couple went through a form of marriage on 15 December 1785, in the drawing room of her house in London. The marriage was of course invalid because it had not received the prior approval of King George III and the Privy Council, and because if it had been declared valid, the Prince would have been removed from the succession to the British throne.

A young man, very much in love, poured out all his passionate love in his letter to his “wife”. In order to convey his innermost feelings, he sought for the language and style of writing modeled in past love letters. This was not unheard of in the

---

1. These descriptions appear e.g. in Loughton 1744 and Farro 1754.  
2. CEECE: George IV, 1785; GEORGE 4, I, 196.

eighteenth century, although the use of *thou*, in particular, was considered not only archaic, but also ungentle and rude by the contemporary grammarians.

This chapter will concentrate on the use of *thou* in eighteenth-century letters. The purpose is to take a closer look at the textual context in which *thou* is used in personal correspondence, since, as the above example shows, the Late Modern English use of *thou* was not restricted to religious discourse or literary quotations. In fact, there appear to have been some “real” users of *thou* even in the later decades of the century. My purpose is also to examine the pronoun from the point of view of its history: why did its use decline so rapidly, and who were the ones who kept *thou* alive in a time when its use could so easily be associated with such negative characteristics as poor breeding and impoliteness.

I will first give a brief summary of the history of *thou*, both from a diachronic and synchronic point of view. I will then go more deeply into the use of *thou* in the *Corpus of Early English Correspondence Extension* (CEECE), more particularly the eighteenth century, looking more closely at the textual contexts in which *thou* occurs in the data. Finally, I will discuss the reasons for the instances of *thou* in the material, drawing conclusions based on, for one, the most prolific user, Ignatius Sancho.

## 6.2 A short history of the rise and fall of *thou*

### 6.2.1 The pre-eighteenth-century use of *thou*

The history of the second-person pronoun *thou* began in the Old English period. The address pronoun system distinguished between singular (*þū, þīn, þē, þē*), dual (*git, incer, inc, inc*) and plural forms (*gē, eower, eow, eow*). By Late Middle English, after the disappearance of the dual form, *thou* and *you* became the only second-person pronouns, which were distinguished from each other by number. *Thou* was used in second-person singular and *you* in second-person plural (*3ē/yē, 3our, 3ou, 3ou*).

The role of *thou* changed in the course of the Early Modern English period. It was first used as a status marker, in that it was mostly used to address either a social inferior (e.g. a servant or a member of a lower rank), or an inferior in power within a family (e.g., wife and children). Members of the lower ranks also used reciprocal *thou*. This status-bound use of *thou* is said to have been influenced by French (Barber 1976: 208; Görlach 1991: 85; Blake 1992: 536). In early medieval France, the ruling orders had developed a convention of addressing members of the upper ranks in the plural, whereas members of the lower ranks were addressed in the singular. It was, however, not until the Norman Conquest that the use of these forms began to spread also to the Anglo-Norman court in England (Mustanoja

1960: 126–127; also Finkenstaedt 1963: 59–62; Lutz 1998: 191–192). By Late Middle English, *you* was associated with the learned and polite register of courtly style, whereas *thou* was connected with informal speech. On the other hand, *thou* also became marked by distance, so that by using it, the speaker could signal either emotional distance (e.g. contempt, anger) or proximity (e.g. love, liking) to the addressee.<sup>3</sup>

In practice, both *you* and *thou* were used alongside of each other, and reasons for the fluctuation between the two pronouns have recently been explained on both a social and situational basis. Contextual factors or style shifts may explain the use of *thou* and *you* in literary texts, but in material like letters, they cannot fully explain pronominal variation. Both pronouns mostly appear in similar grammatical and stylistic contexts. The use of *thou* could also be considered a matter of religious ideology. In 1650 George Fox, the founder of the Quakers or the Religious Society of Friends, claimed he had got a revelation from God saying that he should use *thee* and *thou* to all men and women, regardless of their status and wealth. *Thou* became a common feature in the speech of the Quakers, who declined to use the plural *you* to a single person, because it was against the plain and pure speech of the Bible (Brown & Gilman 1960: 265–266; Finkenstaedt 1963: 174ff.; Jucker 2000a: 94; Burnley 2003: 27). In a conduct book intended for the Quakers, *The Pure Language of the Spirit of Truth*, Richard Farnworth (1655: 2) warned the reader that a person who does not use *thou* to everyone is “ignorant of the Life of the Scripture and also ignorant of the Language of the Spirit of Truth, and knows not the Life of the Holy men of God”. He called for the “simple and pure language of God and the holy men”. In addition to *thou*, the Quakers were to call another friend of the order by first name and the term *friend*.<sup>4</sup>

It has been suggested that the reasons for the decline of *thou* in the seventeenth century were influenced by both formal and social factors. Linguists have suggested that the overall trend in Early Modern English towards simplified verbal inflection had its effect also in the use of the pronouns (Brown & Gilman 1960: 266). On the other hand, there have been said to be more socially governed reasons for the increasing dominance of *you*: both Barber (1976: 210; see also Quirk 1971: 70) and Wales (1983: 117–119) suggest that because *you* was the unmarked pronoun, it was

---

3. See Bergs (2004, 2005: 130–131) for a discussion on the representation of *thou/thee* in spoken vs. written language in Late Middle English.

4. This was commonly mocked in several contemporary pamphlets and books, like *The Quakers Art of Courtship* (1689). It includes many ‘instructions’ on the use of address forms to friends, both when the speaker is in a good mood (*O King, Ruler, Magistrate*, etc.) and when in a bad mood (*Tyrants, Wolves, Evil Shepherds, Foxes, Vipers, Baal’s Priests, Blind Guides*). The Quakers’ use of address was considered humorous also in other contemporary genres, as e.g. in plays.

also more freely usable in situations where the speaker either wanted to imitate the speech of the upper ranks, or where expressions of social superiority or emotion were not preferred. It may also be that decrease in the use of *thou* happened as a counteract to its use among the Quakers, and by the eighteenth century, *thou* was only in marginal use (Lass 1999: 11). Today, the use of *thou* as an address pronoun is considered archaic and poetic. In addition to the Quakers' language, it is found in non-standard, dialectal speech, where it is used by parents to children, and between intimates. The OED lists its current use as characteristic of addressing God or Christ, as well as typical of homiletic language, poetry, apostrophe and elevated prose.

The second-person plural form *you* was originally used for the dative and the accusative case. In the course of the fourteenth century, it began to be used also for the nominative *ye*, and by the end of the sixteenth century, *you* had almost entirely replaced *ye* in the subject position (for studies on the correlation of social variables like gender with the *ye/you* variation, see e.g. Nevalainen & Raumolin-Brunberg 2003 and Nevalainen & Raumolin-Brunberg 1996). Reasons for this change have been said to include a confusion caused by similar pronunciation (*ye* [jə] and *you* [jə]), as well as the general tendency of oblique forms to assume the subject position (see e.g. Nevalainen 2000).

From the fourteenth century onwards, *you* started to appear as a substitute for *thou* and *thee* and, as stated above, it finally became the dominant second-person form in the seventeenth century. As already discussed, the function of *you* was originally different from that of *thou*, and whereas *thou* was used to express the addressee's inferiority, *you* was generally used as a token of respect in addressing a superior, or between upper rank equals – hence its later definition as a 'polite' pronoun (as opposed to the 'familiar' pronoun *thou*). Brown & Gilman (1960: 256) suggest that the singular *you* was first a form used to address a superior in power in the state, but was later generalised as a form used to a superior of power in a nuclear family. *You* was not originally in reciprocal use between people of great social distance, but as it began to be semantically more mixed with *thou*, the importance of social status as a determiner of pronoun usage gradually became bleached. So, later in the eighteenth century, the use of *thou* was increasingly restricted into different textual contexts: although it still appeared a distinctive feature in Northern English dialects, *thou* was even more so used as a marked feature of such specific registers as biblical and literary language.

### 6.2.2 Eighteenth-century grammars on the use of *thou*

The rules of using *thou* and *you* appear to have mostly related to case and number in Early Modern English grammars.<sup>5</sup> According to Walker (2007: 99), it is not until the latter half of the seventeenth century that we can find comments on the discursive use of the pronouns. The most prominent users of *thou*, the Quakers, published a book called *A Battle-Door for Teachers & Professors to Learn Singular & Plural; You to Many, and Thou to One* in 1660, in which they even attacked the Pope for his “degenerated” use of *thou* and *you* as a status-marker in Italian:

And now you that say [*tù*] *thou*, to your servants of low degree, and [*voi*] *you*, to your servants of better account (as *you* call it) and [*vo signoria*] that is *your* Lord-ship to one another; are not *you* degenerated from your own Mother-tongue, and through *your* Pride and Ambition, do speak contrary to your own Grammers, and Bible? And so are false into the respect of persons, saying, [*tù*] that is *thou*, to your servants of mean account, (as *you* call it) and not, [*voi*] that is, *you* to them, but [*voi*] *you*, to your servants of better esteem, and to Artificers, and [*V.S.*] *your* Lord-ship to one another. Is not this the Anti-christ, who is exalted above all that is called God?

(Fox 1660: 24–25)

Since *thou* was already marked by the end of the seventeenth century, most eighteenth-century grammars condemned its use as “vulgar”, “ungenteel”, and “rude” (see e.g. Loughton 1744 and Farro 1754). In addition to the grammatical factors of singular and plural usage, Greenwood (1722 [1711]: 114) also refers to the use of the pronouns in expressing mood, emotion or distance towards the hearer:

It is customary among us, (as likewise among the *French*, and others) tho’ we speak but to one particular Person, to use the *Plural Number*: But then we say *You*, and not *Ye*. So likewise out of complaisance, as we use *You* for *thou* and *thee*, so we frequently say *Your* for *thy*, and *Yours* for *thine*. When we speak in an Emphatical manner, or make a distinct and a particular Application to a Person, we often use *Thou*; as, *Remember O King, thou art a Man*. Otherwise if any one speaks to another, in the *Singular Number*, as, *Thou Thomas*, it is reckoned a Sign of Contempt or Familiarity.

Although eighteenth-century grammarians like Duncan (1731: 55; also Brittain 1788) thought in general that *thou* was “quite out of Date in the polite World”, they might still refer to its use in status-related contexts, such as that between members of the lower social classes, close friends or when talking to social inferiors. So by the middle of the century, the distinction between *thou* and *you* started to be increasingly related to the style of writing and literary genre: the use of *thou* was more

---

5. I am grateful to Dr Nuria Yáñez-Bouza for her help in this section.

and more connected with the “solemn style”, the language of poetry and religious prose and prayer. Johnson, in his 1755 grammar, declares that “It is used only in very familiar or very solemn language. When we speak to equals or superiors we say *you*; but in solemn language, and in addresses of worship, we say *thou*”. Similarly, Lowth (1762: 34) refers to the solemn style which “admits not of *you* for a single person”. Lowth, like Baker (1770) and Murray (1795), insists on the *-st* forms of verbs agreeing with *thou*, and blames most of the contemporary poets “of the first rank” for mixing *thou* and *you*. Here again, a clear distinction between the proper, solemn, style and the improper, vulgar, style is made.

Some writers have used *ye* as the objective case plural of the pronoun of the second person; very improperly and ungrammatically. ... It may perhaps, be allowed in the comic and burlesque style, which often imitates a vulgar and incorrect pronunciation, as “by the Lord, I knew *ye*, as well as he that made *ye*.” Shakespear, I Hen. IV. But in the serious and solemn style no authority is sufficient to justify so manifest a solecism. The Singular and Plural form seem to be confounded in the following sentence: “Pass *ye* away, *thou* inhabitants of Saphir.” Micah, i. II.

(Lowth 1762: 22)

Oddly enough, there were also those late-eighteenth-century grammarians who still preferred *thou* over *you* (see e.g. Fogg 1792–96; Hornsey 1793). The distinction in the use of the pronouns was mostly made according to grammar (singular/plural), but the comments concerned the “unnatural” use of *you* used in the same context as *thou*. Baker (1770: 96) puts the blame on famous authors:

The using *you* and *thou* in the same Period (and more especially so very near together) is an unnatural Way of writing. And yet we have many Authors guilty of it. Pope is not a little faulty in this Particular.

Letter writers themselves made some remarks about the use of *thou*, warning of the different contexts they should appear in. One of these comments can be found in John Warton’s letter from 1785: “I don’t agree with you about *have* in ... – tis really false English – & *you* & *thou* must not be promiscuously used” (CEECE: John Warton, 1785; WARTON, 508). It seems then that by the end of the eighteenth century the difference between the “faulty” and the “right”, the “rude” and the “polite”, as well as the “vulgar” and the “solemn”, was already clear in the minds of the language users.

### 6.3 *Thou* in eighteenth-century letters

In this study, I will account for and further analyse the instances of *thou* in private correspondence between family members (meaning nuclear family and other kin) and between close friends. The reason for doing so is based on the material alone; as the majority of instances of *thou* occurs in letters written to either family members or friends, the rare cases of *thou* in letters between more distant correspondents do not allow proper analysis, and thus justifiable results.

The CEECE was searched for all instances of *thou*, *thee*, *thy*, *thysself* and *thine*. Table 6.1 shows the number of all users of *thou* in the entire CEEC-400, incorporating information from my previous study (Nevala 2002). The table shows that, not surprisingly, the number of *thou* users appears to be at its highest in the seventeenth century, declining in the course of the eighteenth century. Those who use *thou* in the eighteenth century use it more in letters to their family and kin than to their friends.

**Table 6.1** The users of *thou* in CEEC-400 (FN = nuclear family; FO = other family members; TC = close friends)

Time period	Number of users in FN/FO letters	% of all writers per century	Number of users in TC letters	% of all writers per century
15th century	2	1	0	0
16th century	6	2	0	0
17th century	25	6	7	2
18th century	17	6	8	3
<b>Total</b>	<b>50</b>	<b>4</b>	<b>15</b>	<b>1</b>

The three most prolific users of *thou* in the eighteenth-century data are Ignatius Sancho (72 instances), William Cowper (39 instances) and John Banks 2 (23 instances). Of these three, Sancho and Cowper wrote their letters in the latter half of the eighteenth century, whereas Banks' correspondence dates to the first half of the century. As already stated, most writers used *thou* when writing to their nuclear family members or other relatives, as well as their close friends. In addition, there are twelve users of *thou* in the eighteenth-century letters to recipients outside family and friends, but their use is mostly so marginal (i.e. only one or two instances per writer) that they have not been counted in the figures in Table 6.2.

The data from the eighteenth-century part of the CEECE includes all the instances of *thou* found (in brackets). For the sake of comparability to earlier centuries, the table also shows the number for *thou* used in actual address to the recipient, excluding all instances of *thou* in, for example, biblical and literary quotes, as well as in passages of reported speech from other than the letter writer him/herself. As expected, the number of instances radically increases in the seventeenth century,

**Table 6.2** The number of *thou* in letters to family and friends in CEEC-400  
(see also Nevala 2004b: 181–182)

Time period	Instances of <i>thou</i> in FN/FO letters	Normalized per 10,000 words	Instances of <i>thou</i> in TC letters	Normalized per 10,000 words
15th century	55	1.38	0	0
16th century	50	0.51	0	0
17th century	1,484	7.68	82	0.42
18th century	122 (152)	0.63 (0.79)	107 (110)	0.56 (0.57)
<b>Total</b>	<b>1,711 (1745)</b>	<b>3.27 (3.34)</b>	<b>189 (192)</b>	<b>0.36 (0.37)</b>

after which it decreases during the eighteenth century. The numbers are somewhat skewed because of the three most prolific writers, without whose instances the use of *thou* would, particularly in the latter half of the eighteenth century, be too marginal to mention.

## 6.4 *Thou* on closer view

### 6.4.1 The contextual use of *thou*

Today, the use of *thou* is most often related to poetry and religious discourse. As already discussed, eighteenth-century grammarians like Brittain (1788) already considered *thou* a language feature used only by the Quakers. In the data, there are a few writers who repeatedly resort to biblical references or style in their letters, particularly when they are giving advice to the recipient. In this context, the use of biblical language often works as a powerful tool in stressing the gravity of the issue at hand.

In the first example, Mary Prior quotes a phrase, “thou art a stranger”, from *Genesis*<sup>6</sup> and continues using *thou* in the rest of the sentence. By doing so, Prior emphasizes how little her friend knows about the Lord in question, i.e., how much of a stranger he is to the people in question, and how much he needs Prior’s help in adjusting. In the second example, William Pitt takes a quote from *Ecclesiastes*.<sup>7</sup> He is giving advice to his young correspondent, nephew Thomas Pitt, about life in

6. *Genesis* 17:8: “And I will give unto thee, and to thy seed after thee, the land wherein thou art a stranger, all the land of Canaan, for an everlasting possession; and I will be their God.”

7. *Ecclesiastes* 12:1: “Remember now thy Creator in the days of thy youth, while the evil days come not, nor the years draw nigh, when thou shalt say, I have no pleasure in them;”

general and university life in particular, and wants to remind him of trusting in God for guidance. Pitt uses this same phrase in another letter to his nephew a year later.

- (6.1) **Thou** art a Stranger in Israel, my good Friend, and seemest to know no more of this Lord than **Thou** didst of the Condé de P- when first I construed Him to **Thee** at the Coffee house (Mary Prior, 1720; PRIOR, II, 332)
- (6.2) Remember **thy** Creator in the days of **thy** youth, is big with the deepest wisdom: The fear of the Lord is the beginning of wisdom; and, an upright heart, that is understanding. (William Pitt, 1754; PITT, 26)

Another context in which *thou* is used, although to a much lesser extent than others, is that of reference to literature. This is most typical towards the end of the century, and most often the main purpose of quoting passages from well-known works of drama and poetry is to display the writer's vast knowledge of literature, among other topics. Examples (6.3) and (6.4) show instances of this kind of usage. In the excerpt from Hester Piozzi's letter, *thou* is embedded in a quote from a piece of literature or poetry. The style is similar to that of Shelley, although the origins of the passage has not been identified. Example (6.4), on the other hand, contains two lines from *Othello*.<sup>8</sup> Quoting Shakespeare was popular at the time, and there are other instances from, for example, *Julius Caesar* that appear in the data. It is not surprising, therefore, that especially contemporary authors like Samuel Crisp should embellish their language with literary references.

- (6.3) you were A Sweet Creature to send her such a play thing: it has had its full effect, and her Husband who will be stone-blind in about a Quarter of a Year, has seen this Sight before he took Leave of *Life and Light* and *thou blest Sun serene*. (Hester Piozzi, 1796; PIOZZI, II, 412)<sup>9</sup>
- (6.4) therefore as Iago says to Roderigo – *put money in thy purse! fill thy purse with money!* Need I say any more to induce you, immediately, and positively to insist on your money from [LETTER TORN] ? Lose no time. (Samuel Crisp, 1780; CRISP, 50)

*Thou* still appears to be used in formulaic contexts by certain writers, such as John Banks 2, William Cowper and David Garrick. Examples (6.5) and (6.6) show instances of *thou* in subscription formulae. Cowper ends his letter to his cousin Harriet Heskett with a short blessing and the subscription “thine ever”; similarly,

---

8. *Othello*, Act I, sc. 3, l. 341, 347.

9. Hester Piozzi quoted the same phrase in her journal entry from 1777, in the form “Farewell light and life, and thou blest sun serene!” (Black et al. 2006: 741). According to Google Books, this is an excerpt from Dr. Marriot.

Garrick uses the phrase “I am thine ... most affectionately” when writing to his brother George. Both writers’ use of *thou* is not, however, restricted to formulaic use. As discussed above, Cowper, being the second most prolific user in the eighteenth century, uses *thou* rather fluently in all kinds of contexts, and, respectively, Garrick’s dramatic and flamboyant style is complemented by his use of *thou* as a familiarity booster.

- (6.5) May God bless **thee** and preserve us both to another happy meeting! **Thine** ever Wm Cowper. (William Cowper, 1787; COWPER W, III, 8)
- (6.6) I write in a very Noisy Coffee house, & having Nothing more to Say I am **thine** D<sup>r</sup> George, most Affectionately D. Garrick – (David Garrick, 1747; GARRICK, I, 91)

Although it seems first that in the eighteenth century data the use of *thou* can already be considered archaic, the majority of those writers that use the pronoun throughout their letters do so in other than formulaic or literary contexts. Similarly to its use in earlier centuries, *thou* in letters to family and friends relates to differences in power between the correspondents. In most cases, the relationship between the writer and the recipient is that of a parent and a child, a husband and a wife, or an older relative and a younger one. In (6.7) we can see an example from a letter written from a parent to a child. Mary Banks, like her husband in his letters, uses *thou* to address her daughter Mary. Banks is also one of the few eighteenth-century writers who still use the phrase *pray thee*.

- (6.7) Sunday morne. I hope **thou** would recave **thy** father’s and myne by Silcock last night. If not, pray **thee** send forth to Sheffield for it, and **thy** father desires **thou** will observe the orders in it. (Mary Banks, 1710?; BANKS, 14)
- (6.8) How much pleasanter an employm<sup>t</sup> would it be, if I had time for it, my ever dear & darling Susey, to scribe nonsense to **Thee**, than either to write Hist<sup>y</sup> or read lectures on its alphabet, from door to door, as I now begin to do! (Charles Burney, 1783?; BURNEY, 402)

The last Example (6.8) is an excerpt from a letter written by Charles Burney to his daughter Susanna Phillips. In general, Burney’s relationship with his daughters was very close and loving (see e.g. Palander-Collin & Nevala 2010). His use of nicknames and terms of endearment to the daughters is very typical of his letters, as can be shown here by “my ever dear & darling Susey”. *Thou* in this context adds to the deep level of intimacy and shows an atypically low difference in power between a father and a child.

### 6.4.2 The most prolific “thouer”: Ignatius Sancho in focus

(Charles) Ignatius Sancho was born on a slave ship on its way to the West Indies in c. 1729.<sup>10</sup> His mother soon died in the Spanish colony of New Granada, and soon after that, his father committed suicide. As a two-year old, Sancho was taken by his owner to England where he was then given to three maiden sisters in Greenwich, who surnamed him Sancho after the fictional Don Quixote’s squire. Later on, John, second duke of Montagu, who had met Sancho by accident, got so impressed by Sancho’s intelligence that he began to encourage him to read by giving him books. Shortly after the duke’s death in 1749, Sancho fled from Greenwich and sought protection from the duke’s widow. The duchess hired him as her butler, and, at her death in 1751, left him an annuity of £30. Combined with his own savings, he had a sum of £70 beyond the annuity. He soon spent all his money on gambling, women, and the theatre, where he spent his last shilling to see David Garrick play the role of Richard III.

After failing to become an actor because of his speech impediment, Sancho sought service in the household of the late duke’s son-in-law George, who made him his valet in 1766. Gout and obesity eventually forced Sancho to quit Montagu’s service. Sancho and his wife, Ann, established a grocery shop in 1774 in Westminster. They got seven children, the ‘Sanchonettas’ as Sancho himself called them, of whom only four survived. To supplement his income, he also published compositions of music, a *Theory of Music*, and two plays. His position as a male householder in Westminster qualified him to become the only eighteenth-century Afro-Briton known to have voted in parliamentary elections for the representatives of Westminster in 1774, as well as in 1780 when he died of gout.

Sancho gained celebrity when one of his letters appeared in the posthumously published *Letters of Laurence Sterne* (1775), with whom he had initiated a correspondence in 1766, encouraging Sterne to continue writing about the oppression of Sancho’s fellow Africans. But even though his own status rose, he never lost contact with those who had been his fellow servants, some of whom were also fellow Afro-Britons. In 1782, Frances Crewe, one of Sancho’s younger correspondents, edited and published the two-volume *Letters of the Late Ignatius Sancho, an African*, sold by subscription. The subscription list for the first edition of the *Letters* shows the vast range of Sancho’s social circle, which included men and women, aristocrats, servants, artists, businessmen, country squires, and prominent politicians.

Sancho’s surviving letters portray him as a religious and patriotic man, as well as an avid reader of eighteenth-century literature. Sancho gives the impression of

---

10. All biographical information here is from the *Dictionary of National Biography* (online) (Carretta 2004) and Edwards & Rewt 1994.

being almost wholly assimilated into the lifestyle and values of the polite society of eighteenth-century England. Many of Sancho's letters discuss public affairs, as do his newspaper essays, under both his own name and the pseudonym *Africanus*. In his writings, he displays constant concern for his friends and his country, often using humour when touching grave topics like discrimination, illness, political failure, and death.

Sancho's correspondence with Laurence Sterne, as well as his knowledge of Sterne's works, appears to have had a big impact on his own language use and style of writing. According to Sandhu (1998), Sancho wrote his letters in the register and grammar of *Tristram Shandy*. They show his "education to imitate Sterne's prose style, punctuation, Shandean multilingual jokes, and word play" (Carretta 2004). Sancho was particularly impressed by Sterne's typography, grammar and form. He also, similarly to Sterne, coined new words like 'bumfiddled' in place of 'befuddled', as well as exclaimed "alas! an unlucky parciplepliviaplemontis seizes my imagination", and described his friend, John Ireland, as an "eccentric phizpoop".

Whether imitations of Sterne's language or not, Sancho's letters show certain linguistic features that could be considered archaic already in the eighteenth century. One of the most prominent ones is his use of *thou*. Knowing his deeply religious background, it is not surprising that his writings contain long passages discussing the Bible or giving advice in the biblical style. In Example (6.9), Sancho addresses his friend John Meheux with *thou*, together with the future form *shalt*. He also uses allusions to historical or biblical things, such as 'writing a philippic' or 'the day of retribution'. Similarly, in (6.10) Sancho first refers to William Stevenson's father as a true believer in Christ and soon after addresses Stevenson himself with *thou (wilt)*.

(6.9) A heart and mind impressed with a firm belief of the Christian tenets, must of course exercise itself in a constant uniform general philanthropy – Such a being carries his heaven in his breast – and such be **thou!** Therefore write me a bitter Philippick against the misusers of Jack-asses – it shall honor a column in the Morning Post – and I will bray – bray my thanks to you – **thou** shalt figure away the champion of poor friendless asses here – and hereafter shalt not be ashamed in the great day of retribution.

(Ignatius Sancho, 1777?; SANCHO, 102)

(6.10) Your good father feels a satisfaction in doing – I think a wrong thing – His motive is right – and, like a true servant of Christ, he follows the spirit – not the letter – He will be justified in a better world – I am satisfied in this – and **thou** wilt in **thy** feelings be gratified. – Huzza! – we are all right – but your father pays the piper.

(Ignatius Sancho, 1779?; SANCHO, 200)

Sancho's deep knowledge of drama and literature, both eighteenth-century and earlier, also shows in his letters. In addition to other writers, he has a few references

to Sterne's works, as can be seen in Example (6.11). The Yorick Sancho is referring to a fictional clergyman, supposedly a descendant of Hamlet's jester, invented by Sterne for *Tristram Shandy*. *Thou* is excessively used also by Sterne, and the style in this passage appears to imitate his style in other ways as well. This can be seen by the use of exclamations and innovative words such as *namby pamby* (meaning, 'silly sentimentality'). In (6.12), Sancho alludes to Shakespeare: "as the Jew says" is a reference to Shylock in *The Merchant of Venice*.<sup>11</sup> The passage, also containing *thou*, is written as banter to his friend Meheux, 'complaining' about his "abominable" characteristics – namely, conscience, honesty and good nature.

(6.11) By St. Radagunda! quoth I – (ramming my nostrils with Hardham) he has caught the mantle. – Alas, poor Yorick! – Oh! that **thou** hadst, by divine permission, been suffered a little – little longer, amongst the moon-struck children of this namby pamby world! – Father of light and life! **thy** will be done: – but surely – half the wit – half the good sense – of this present age – were interred in Sterne's grave. (Ignatius Sancho, 1777?; SANCHO, 111)

(6.12) Why then **thou** art a silly fellow – incumbered with three abominable inmates; – to wit – Conscience – Honesty – and Good-nature – I hate **thee** (as the Jew says) because **thou** art a Christian. (Ignatius Sancho, 1779?; SANCHO, 194)

Banter, or comic insults, like the one in Example (6.12), appear to be very typical of Sancho's style of writing. Very often this happens in letters written to his younger correspondents. In (6.13), Sancho has been asked for advice by young Julius Soubise, which he gladly gives, but in a comic tone. He alters between the two pronouns, quickly changing into *thou* when he begins his complaint about Julius Soubise's "demand" for advice. It seems that by switching into *thou*, Sancho diminishes the distance between himself and the recipient further, as if to make sure that the recipient understands accusations like "thou art a flatterer" or "thou canst not discern wood from trees" as banter.

(6.13) Your letter gave me more pleasure than in truth I ever expected from your hands – but **thou** are a flatterer; – why dost **thou** demand advice of me? Young man, **thou** canst not discern wood from trees; – with awe and reverence look up to **thy** more than parents – look up to **thy** almost divine benefactors – search into the motive of every glorious action – retrace **thine** own history – and when you are convinced that they (like the All-gracious Power they serve) go about in mercy doing good – retire abashed at the number of their virtues – and humbly beg the Almighty to inspire and give you strength to imitate them. (Ignatius Sancho, 1772?; SANCHO, 56)

---

11. Act I, sc. 3, l. 43: "I hate him for he is a Christian."

- (6.14) Poor blundering M[eheux], I pity **thee** – **thou** art a bungler in every thing – ask the girls else. – You know nothing of figures – you write a wretched hand – **thou** hast a nonsensical style – almost as disagreeable as **thy** heart – **thy** heart, though better than **thy** head – and which I wish from my soul (as it now is) was the worst heart in the three kingdoms – **thy** heart is a silly one – a poor cowardly heart – that would shrink at mere trifles – though there were no danger of fine or imprisonment: (Ignatius Sancho, 1779?; SANCHO, 194)

In a similar vein, Sancho first uses *thou* in Example (6.14) to set the ‘insulting’ mode (“*thou* art a bungler in every thing”), and then alters between *thou* and *you* when describing how bad he thinks his friend Meheux’s characteristics are. From the outside, the passage seems a critical one – very harshly so – but knowing Sancho’s penchant for these kinds of comic insults, it only shows his close relationship to his friend.

## 6.5 The use of *thou* in CEECE

### 6.5.1 The influence of social and linguistic norms

Reasons for the appearance of *thou* in the material should be examined in the light of the social background and environment of its users. Eighteenth-century English society gave rise to a new way of describing social hierarchy: the middling sort spread among the upper and the lower sort. Even so, society continued to be a matter of status and hierarchy. Proper behaviour, and politeness in the social sense of the word, was to be learnt, and this enlightened society was further divided into the common and the élite – a cultural differentiation of “the informed” and “the ignorant” already introduced in the seventeenth century. In general, social relationships between strangers and distant people could be characterised in terms of formality. Informal social contacts across barriers of wealth and status were not encouraged, and people spent time within their own social groups.

Burke (2000: 39–40) also notes that in eighteenth-century England the middle class used polite speech as a way of showing their closeness to the upper class, alongside other means, such as accent or vocabulary. To be a polite individual in eighteenth-century English society meant also being a member of a social network and fulfilling one’s social and intellectual aspirations by belonging to societies and clubs. This showed in the use of address terms and pronouns, and, for example, the use of the term *friend* was largely extended and, by the end of the eighteenth century, it had acquired two meanings, a literal and an ideological one.<sup>12</sup>

12. For more information on the use of *friend* in early letters, see Nevala (2009).

The Enlightenment was a new era, and the ideals of polite society dictated not only what kind of people would be accepted into polite society, but also prescribed the linguistic norm of the period. The contemporary view supported a divergence from Latin, and educated speech was seen as a cause of decay in the language, as was the use of non-standard varieties like slang, cant and dialect. “Correct English” was supported by moral reform, and grammarians were affected by the dominant ideologies of the time, such as the elitist theory. As Tiekens-Boon van Ostade (2006: 541) notes, the unpopularity of eighteenth-century grammars among later structural linguists allowed prejudice about the grammars and their authors to grow. Another factor possibly working against them concerned the alleged fact that the grammarians focused on the language of gentlemen in order to decide what was grammatically correct and what was not.

Against this social background it is understandable that *thou* as a rapidly declining address pronoun was generally considered a remnant of a previous era. Eighteenth-century grammarians like Duncan (1731) claimed that *thou* was outdated, but they still referred to its marked use in status-related contexts, such as that between members of the lower social classes, close friends or when talking to social inferiors. By the middle of the century, the distinction between *thou* and *you* started to be increasingly related to the elevated style of writing and literary genre: the use of *thou* was more and more connected with the language of poetry and religious prose and prayer. My material seems to corroborate this to a certain degree, particularly in the case of Ignatius Sancho.

For people like Sancho, who aspired for a membership in the literary or educated circles, knowing how to switch between the two address pronouns must have been a matter of specific learning. He must have been aware of the fact that his use of *thou* was already considered archaic and, for a common letter-writer, even odd. Still, his fascination for imitating the elevated literary style of Laurence Sterne probably gave Sancho as a self-made man a certain sense of gentility – precisely of what eighteenth-century grammarians were so eagerly trying to deny men like him.

### 6.5.2 A marker of status and intimacy – and of interpersonal identity?

As stated in previous research, the use of *thou* has generally been considered to be marked for status and power, in that it is said to be used to address the inferior. This has been the case at least in studies of pre-eighteenth-century data. For example, my earlier study of *thou* in the CEEC (Nevala 2002) confirmed the hypothesis to a certain extent, since *thou* mainly occurred in letters which were written to recipients inferior in power. Also in the eighteenth century *thou* appears to be used as a status-marker by some writers. The meaning of a ‘status-marker’ seems to have changed, however. *Thou* is not that much, if at all, used as a marker of social status

at large; it is rather used to mark the recipient's status within a certain group, like a family unit or close kinship. Among those who use the pronoun, there are certain writers who use it to address their wives, daughters, sons, other younger family members and relatives. One of these is Mary Banks who, like her husband, mostly uses *thou* when writing to her daughter. *Thou* can, of course, be also used as a marker of intimacy and affect from one correspondent to another. The fact that *thou* is mainly used in letters between family members, especially in those within the nuclear family, and ties between close friends alone indicate the presence of an increased level of intimacy, and it is not surprising that *thou* is taken as an additional tool for expressing it.

One might of course argue that, since it also appears in formulaic subscriptions and allusions to the Bible and major works of literature and drama, the instances of the "authentic" use of *thou* are too few to make any definite conclusions about its main functions. The fact that writers use these passages from literature can, in general, have another function, i.e. that of constructing a social identity or an interpersonal role.<sup>13</sup> This kind of usage is most typical towards the end of the eighteenth century, and very often the main purpose of quoting passages from well-known works of drama and poetry is to display the writer's vast knowledge of literature, among other topics. In this case as well, the use of passages containing *thou* can be seen to work on an interpersonal level. The recipients of the letters where *thou* is mostly used, i.e. close friends and family members, are well-known to the writer. This means that the writer is often well aware of, for example, the recipient's personal interests and his/her level of schooling and knowledge in such specific subjects as literature or drama. By inserting literary and biblical quotes, the writer can boost, or even create, his/her status or identity in the relationship as a learned and scholarly person, and thus either diverge from or converge to the recipient, who may or may not share the writer's position.

In my data, constructing one's identity by switching between *thou* and *you* is particularly shown in Sancho's pronominal use. As discussed earlier, his letters contain long passages discussing the Bible or giving advice in the biblical style. Similarly, he displays his deep knowledge of literature by inserting passages from Sterne or Shakespeare in his letters. In these cases, the function of *thou* could be called 'instrumentally interpersonal', since it works as a tool for building up his interpersonal identity by way of increasing the distance between the two interactants. Instrumentality could here be described as 'coded' for status, and it is more driven by power, either social or relative. Instrumentally interpersonal *thou* can

---

13. See e.g. Nurmi & Pahta (2010) on the use of code-switching as a tool for constructing social roles in eighteenth-century correspondence.

also, of course, be used in textual contexts to convey a certain level of familiarity, as is the case of subscription formulae. In these instances, the writer ends the letter using *thou* in the specific purpose of diminishing the distance between the correspondents.

Sancho's comic insults to his close friends, however, seem to have a function other than instrumentality. It seems that by switching into *thou*, Sancho brings the recipient even closer to himself, as if to make sure that the recipient understands accusations like "thou canst not discern wood from trees" as banter. In this context, *thou* is used in the sense which might be called 'intimately interpersonal', i.e. approaching the recipient on a deeper level of friendship and familiarity. Similarly, intimacy allows the exercise of power by means of, for example, criticism. As already discussed, this is particularly the case when *thou* is used by parents to their children. The momentary shift from *you* to *thou* can increase the distance between a father and a son, for instance.

A simplified chart of the contextual use of *thou* in conveying instrumental and intimate interpersonalities is shown below in Figure 6.1.

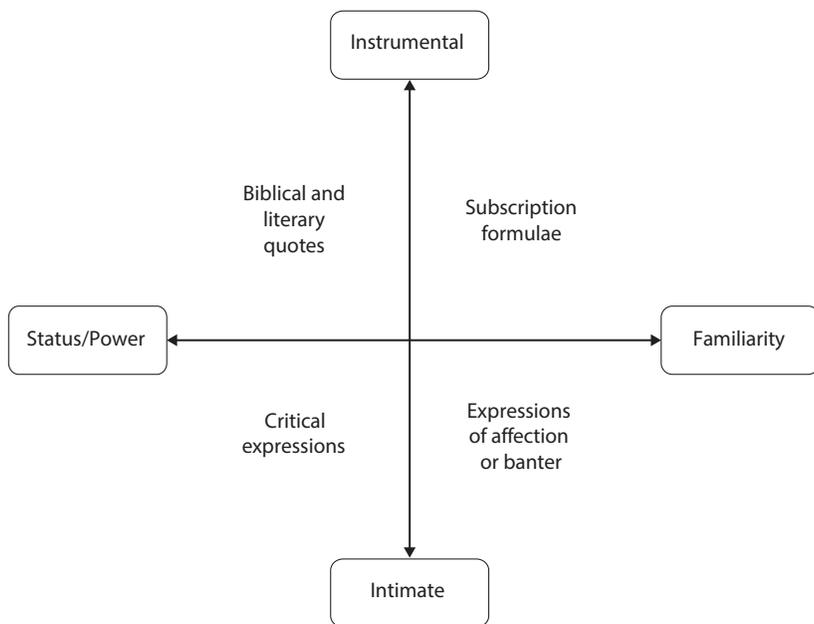


Figure 6.1 Instrumental and intimate interpersonalities in the eighteenth-century use of *thou*

Many contemporary scholars (e.g. Jucker 2000b, Honegger 2003, to mention a few) have noted the importance of taking the situational context, with its temporary balances of power, into account. In Jucker's (2000b: 159) opinion, the situational context may overrule the general effects, brought on by the interactant's social status and social role, which govern the use of address pronouns.<sup>14</sup> This is the type of analysis that Lass (1999) also aims for in his examples of the use of *thou* and *you* in letters. What seems to be central here in the deeper interpretational level is the tone of the exchange. Minor shifts in the subject matter may trigger the change from one pronoun to another. These kinds of changes are what Lass (1999: 152) calls "interpersonal" to "impersonal" type of narrative. Judging from my material, this may be partially true, since it seems that, in general, the use of *thou* did become more varied in the eighteenth century. It cannot, however, be the reason behind every pronoun shift. In my data, as in my earlier one (2002), examples remain where there seems to be no apparent or embedded reason for the writer's choice of *thou* over *you*.

## 6.6 Conclusions

This study has aimed at shedding new light to the textual context in which *thou* was used in eighteenth-century letters. Already by the time *thou* was described as 'ungenteel' and 'rude' in various eighteenth-century English grammars, had its markedness started to shift from status-related to register-related. *Thou* was increasingly used in personal letters in more specific textual contexts of the literary, elevated, style in order to imitate great authors of the time. Apart from setting the overall stylistic register, *thou* could also be used to adjust the emotive register. Whether it was a mother scolding her son, or a father praising his daughter, *thou* appeared, for instance, a perfect tool for expressing changes in the tone or mood of the letter.

As the case study of Ignatius Sancho's use of language shows, *thou* was one of those features which could be used for different discourse functions. Firstly, *thou* was used to describe 'instrumental interpersonality' as a tool for increasing the distance between the two interactants. Secondly, it seems to have appeared in the function of 'intimate interpersonality', for example, when approaching the recipient on a deeper level of friendship and familiarity. In this case, the intimacy also allowed a certain level of criticism or banter between the correspondents.

---

14. Hope (1993: 95) points out that switching between the two pronouns should not be approached from the norm-deviant point of view, but on the basis of changes in a micro-macro, or social-nonsocial model.

Since *thou* was so marginally used from the 1680s onwards, studies on later language have mostly concentrated on its disappearance in status-related use. Taking a closer look on the register-related contexts in which *thou* appears seems, however, to be equally important, if not more important, in its eighteenth-century usage. Uncivil or not, *thou* was still used at least in personal letters of the time. How much this is influenced by, for example, literary ideals or social conventions is a topic for further studies on *thou* in other text types and genres.



## Going to completion

### The diffusion of verbal -s

Terttu Nevalainen

#### 7.1 Introduction

One occasionally comes across sentences like (7.1) and (7.2) in English-language media, both traditional and digital. They represent a creative use of the once productive third-person singular present tense indicative suffix *-(e)th*. In (7.2), the use of the suffix has been extended to the plural. Minugh (1999) suggests that this pseudo-archaic element is kept alive by English speakers' familiarity with Shakespeare and the 1611 *Authorized Version* of the Bible, but it also now appears to lead a life of its own.

(7.1) **Cometh** the hour, **cometh** the man.

(*The Guardian Weekly*, 18 July 2014, p. 21)

(7.2) But what the tabloids **giveth** they may also **taketh** away, and Charles must watch his step.

(*Time*, 19 June 2000, p. 33)

Although variation is found in many registers, in the 17th century the *-(e)th* suffix became the hallmark of formal and religious usage, for example, in the *Authorized Version* and the 1662 revised edition of *The Book of Common Prayer*. High-frequency verbs such as *do*, *have* and *say* retained their *-th* forms much longer than other, less frequent verbs, which had largely gone over to *-(e)s* in mainstream use in the course of the late 16th and early 17th centuries (see e.g. Kytö 1993, Nevalainen & Raumolin-Brunberg 2003: 67–68, 177–180). This chapter discusses the social contexts of the demise and occasional survival of the *-th* suffix in *do*, *have* and *say* in personal correspondence in the long 18th century.

The generalization of *-(e)s* was a process of *lexical diffusion*, the incoming suffix spreading from word to word, rather than being applied in a rule-like fashion to all relevant contexts at once (Ogura & Wang 1996). The examples in (7.3) to (7.8) illustrate the use of *-th* and *-s* forms in *do*, *have* and *say* towards the end of the 17th century. Even one and the same person could use the incoming and outgoing forms in the same context, as Thomas Browne does in (7.5), and Daniel Fleming in (7.6).

- (7.3) I wonder y<sup>e</sup> Secretary should say he **doth** not know him.  
(Humphrey Prideaux (b. 1648), 1681; PRIDEAUX, 110)
- (7.4) I bless god our Deare Tomey **dooes** still mend, but is so leane as there will be need of a good stomack to recrutt him  
(Dorothy Browne (b. 1621), 1681; BROWNE, 199)
- (7.5) Our Tommy **has** had a grievous Cof and feavor ... These 3 dayes hee **hath** been in better temper and prettie chearly, God bee prayed.  
(Thomas Browne (b. 1605), 1681; BROWNE, 193)
- (7.6) This Letter I hope y<sup>e</sup> Doctor **hath** received ... / I am glad that your Book **has** rectifyed the late Omission; (Daniel Fleming (b. 1633), 1692; FLEMING 2, III, 67 / 1695; FLEMING 2, III, 206)
- (7.7) She **saith** now her brother is gon, she speakes courageously, she will have her husband at Denton; (Elizabeth Oxinden (b. 1638), 1681; OXINDEN, II, 352)
- (7.8) but Mr Hewers **hath** don like a wise man to get his own which he **ses** is 700 lb.  
(Esther St. Michel (birth year n.a.), 1681?; PEPYS, 184)

*Have* and *do* come in two functions: main verb and auxiliary. The auxiliary, illustrated in (7.5), (7.6) and (7.8) with *have*, is by far the more frequent of the two. The main verb *have* is exemplified in (7.9).

- (7.9) For it **hath** no clause to save any Rights of any kind to that Church:  
(Thomas Secker (b. 1693), 1748; SECKER, 168)

The next section provides more background for the replacement of the southern suffix *-(e)th* by the originally northern *-(e)s* in mainstream use. Section 7.3 zooms in on *do*, *have* and *say* and presents the results of my empirical study by relating the overall development of these verbs to the two major social variables analyzed in this volume, gender (7.3.2) and social status (7.3.3). Section 7.4 goes deeper into individual variation, analyzing outliers (7.4.1) and looking at two of them in more detail (7.4.2). The role that normative grammar might have played in the retention of the outgoing forms is examined in Section 7.5, and Section 7.6 sums up the complementary sociolinguistic scenarios that emerge from the study.

## 7.2 Verbal *-s* before the long 18th century

### 7.2.1 Interconnected processes

The extent to which the replacement of *-(e)th* by *-(e)s* can be seen as a unitary morphosyntactic process depends on the scope of one's analytic framework. There is evidence to suggest that the success of *-s* generalization was also contingent on

phonological factors. Ferguson (1996: 247) argues that the change from *-th* to *-s* ought to be viewed as a special long-term sound change that went back to Old English and affected verbal morphology. However, this analysis would also need to take account of early dialect variation, as *-(e)s* is only attested in Old Northumbrian and diffuses to southern dialects later, in the course of the Middle and Early Modern English periods. Hence it is necessary to look at dialect contact and competition between suffixes rather than simply accounting for the process as a sound change.

Apart from the change in the final consonant, *-s* generalization is intimately connected with vowel deletion in the suffix. This syncope is part of a morphophonemic drift as a result of which English lost the unstressed preconsonantal vowel in inflectional endings, first in nouns and later in verbs and participles. The vowel loss happened earlier in the north of England than in the south, and it was faster in colloquial speech than in other registers. Based on orthoepic evidence, Nevalainen & Raumolin-Brunberg (2000) suggest that the early phase of the change could be analyzed in terms of dialect variation between the two consonants, while the later stage, gaining momentum in the mid-16th century, was chiefly a rivalry between the syllabic *-eth* suffix and the syncopated *-s*. A related phonological factor is the delaying effect of stem-final sibilants on the introduction of *-s*, the one context in which the unstressed vowel continues to be pronounced today (see 7.2.2).

Orthoepic sources show that *-(e)s* had lost its vowel in non-sibilant contexts by the mid-16th century. However, many commentators viewed *-s* as a colloquial contraction of the syllabic *-eth* even in the 18th century (see Section 7.5). Syncope did not apply to *hath/has* variation, and here regional associations persisted longer. *Has* continued to be identified as a northern form and *hath* as its southern equivalent, for example, in Alexander Gil's *Logonomia Anglica* (1619: 17, 59).

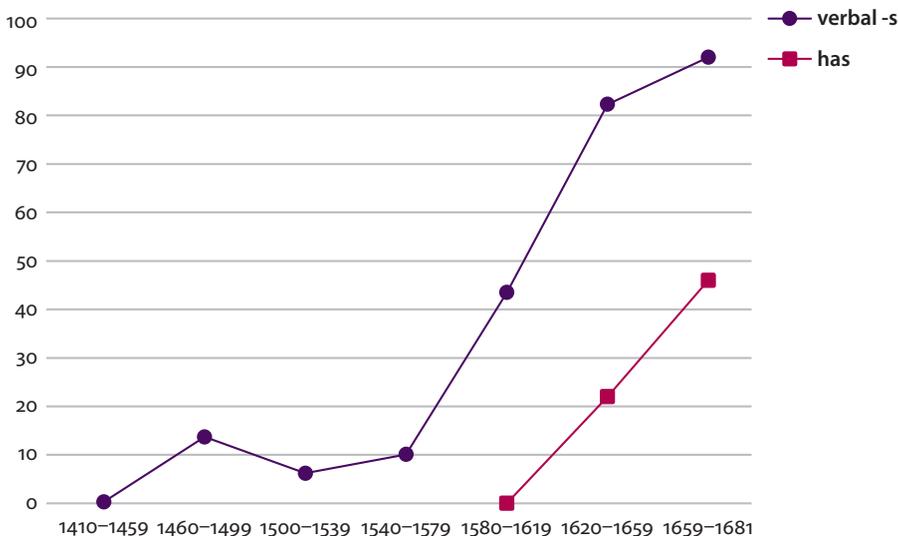
In view of the slow process of lexical diffusion, could the use of *-th* in *hath* and *doth* in the 18th century represent uniform *-s* pronunciations, as suggested by Strang (1970: 146)? She goes on to note, however, that “when 19c writers use *-th* as a deliberate archaism, after a period of complete break with the old tradition of representation, they understand by the symbols the same as we do today”. One way to approach the issue is to chart the dimensions of variation in the 18th century by carrying out a sociolinguistic study of the variant forms – something that Strang did not have the means to do in the late 1960s.

### 7.2.2 Past corpus evidence

A good deal of empirical research has been done on the history of verbal *-s*, beginning with Holmqvist's detailed study (1922) on the regional distribution of the alternative inflections. More recently, Kytö (1993) and Ogura & Wang (1996) studied the progress of verbal *-s* using the Early Modern English section of the *Helsinki*

*Corpus* (1500–1710). Apart from lexical diffusion, they show the relevance of the phonological structure of the verb to the diffusion of the incoming form. At the lexical level, a final sibilant and sibilant-final affricate had a retarding effect on the process, while a final dental stop promoted it (for similar findings, see Gries & Hilpert 2010). Kytö (1993) also showed the extent of register variation in the diffusion of verbal *-s*, and that speech-related genres, such as personal letters and comedies, promoted the incoming form.

Nevalainen & Raumolin-Brunberg (2003) traced the overall diffusion of verbal *-s* using the original version of the CEEC and its *Supplement* up until 1681. We found that the incoming form spread in an S-curve fashion across the language community in less than a hundred years in the course of the 16th and early 17th century. The effects of lexical diffusion were evident in the data, and the verbs *have* and *do* were therefore excluded from this analysis. Figure 7.1 shows the progress of the change in verbs other than *have* and *do* (upper line) and in *have* (lower line). The overall figures are based on quota sampling but those on *have* on the pooling of raw frequencies, over 4,000 instances of the variable between 1600 and 1681. (For other methods of calculating the distribution of *has* vs. *hath*, see Figure 5.1 in Chapter 5.)

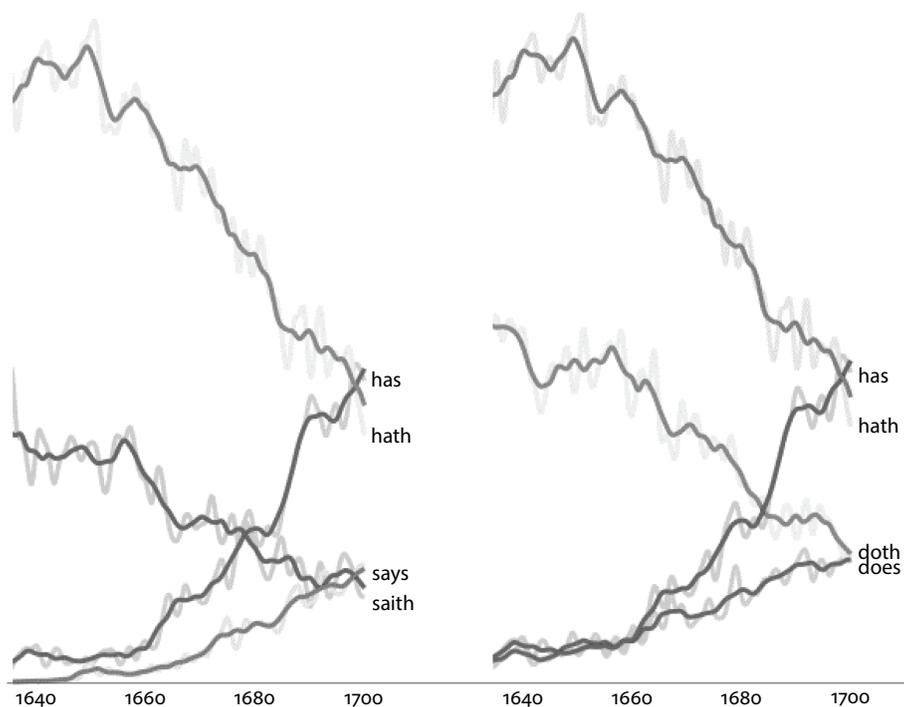


**Figure 7.1** The spread of verbal *-s* (%) in verbs other than *have* and *do* (upper line; redrawn from Nevalainen & Raumolin-Brunberg 2003: 68) and in *have* (lower line). CEEC and CEEC *Supplement*

Nevalainen & Raumolin-Brunberg (2003) found significant regional and gender differences in the diffusion of verbal *-s*. A systematic gender difference receives support from Kytö's (1993) study of the correspondence included in the *Helsinki Corpus*. In the CEEC data, the process was generally led by women and diffused

from north to south with East Anglia as a relic area. These regional differences are in keeping with sources such as LALME that mark a clear north–south divide in the distribution of the suffix in Late Middle English (McIntosh, Samuels & Benskin 1986), and they agree with the East Anglian lag Joby (2014) found in his empirical work. The CEEC data also reveal the inroads that the southern *-(e)th* made to the north in the 16th century. As shown by Moore’s case study (2002) of the northern Plumpton family correspondence, this was a change from above which was reflected more in the usage of upper-ranking men than women, and more in writing than in speech.

Although the vast majority of verbs completed the change in the 17th century, as shown in Figure 7.1, *have* clearly did not, nor did the two other verbs, *do* and *say*, studied in this chapter. To get an idea whether the bulk of the change took place before 1700 in all three, we can examine the text frequencies of the variant forms in the *Early English Books Online* database (EEBO).<sup>1</sup> The results are shown in Figure 7.2.



**Figure 7.2** EEBO-TCP n-gram browser graphs showing the text frequencies of *hath* and *has* (both panels) 1640–1700, and *saith* and *says* (left panel) and *doth* and *does* (right panel)

1. For this purpose, the online n-gram browser of the EEBO-TCP version was used (<http://earlyprint.wustl.edu/>).

By the latter half of the 17th century the spelling of the different variants was largely standardized in print and the results therefore comparable. In Figure 7.2 the two forms of *have* are shown in both panels to serve as a yardstick for the other two, less frequent items. The figures suggest that the crossing-over from *hath* to *has* took place in the 1690s. *Says* also seems to have become more frequent than *saith* around the same time but, despite its steadily rising frequency, *does* had not quite overtaken *doth* by 1700.

If we take a look at the post-1700s in a large database such as Google Books, the final stages of *s*-generalization are found to be prolonged. As shown in Section 13.2, such massive but unmonitored data sources can be suggestive of overall trends but should be tested against carefully constructed corpora, which is what will be done in the remainder of this chapter. Both the EEBO-TCP and Google Books data agree that *has* overtook *hath* in print in the 1690s. Whether this was also the case in personal correspondence will be looked at next.

### 7.3 New results

As discussed in Chapter 4, the CEEC data are not evenly distributed over time. To strike a balance between data granularity and the generalizations to be made on the basis of it, I aggregated my data over 20- and 40-year periods, and adopted a five-class model of the social order to account for the final stages of the generalization of verbal *-s* in social terms.

To provide a basis for comparison with earlier studies, both pooling subgroup data and averaging individual averages are used in this chapter in assessing social variation. As noted in Section 5.2, aggregating subgroup data by pooling is widely used in corpus studies because it is easy to carry out but, unless quota sampling is used, the method suffers from unequal contributions made by different-sized subgroups. Another method, which avoids this problem, is aggregating subgroup data by averaging of individual contributors' mean frequencies; using this method, some minimum frequencies for the number of observations and individual writers need to be set for the data to be representative of the subgroup studied. Pooling is used in Section 7.3.1, and both pooling and averaging are referred to in Sections 7.3.2 and 7.3.3. Bootstrapping provides confidence intervals for assessing the statistical significance of subgroup differences in 7.3.2. When the incoming forms become (nearly) categorical, comparing methods is no longer meaningful.

### 7.3.1 Time course of change

The corpus returns the following variable totals for the period 1680–1800: *have*: 8,538 instances (7,673 of *has*, 865 of *hath*), *do*: 1,147 instances (1,049 of *does*, 98 of *doth*) and *say*: 1,038 instances (1,000 of *says*, 38 of *saith*). Overall, *have*, the most frequent of the three, is also shown to be the most variable, as might be predicted in a process of lexical diffusion.

Figures 7.3 to 7.5 present the generalization of the three forms *has*, *does* and *says* in decreasing order of verb frequency with the number of writers (N) contributing

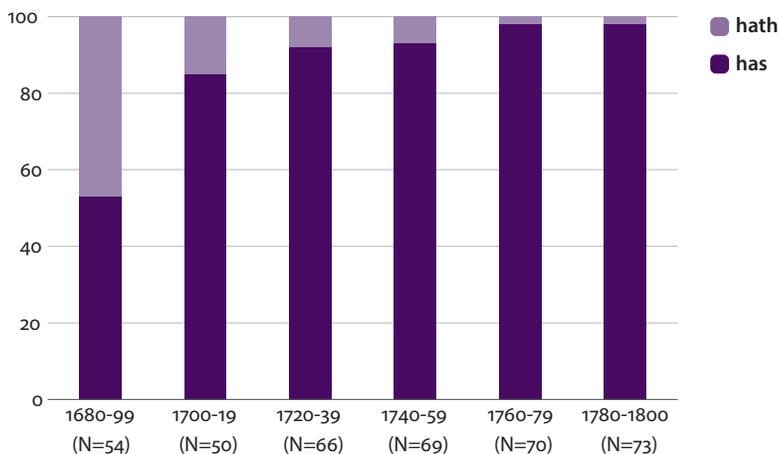


Figure 7.3 Relative frequencies (%) of *has/hath* (N = number of writers)

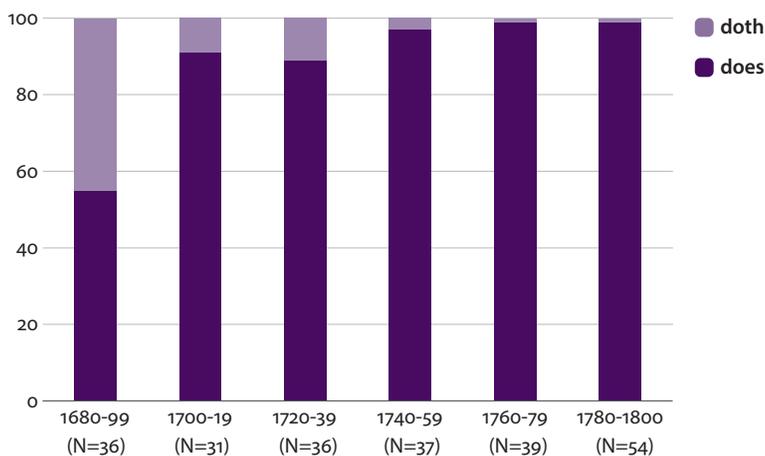


Figure 7.4 Relative frequencies (%) of *does/doth* (N = number of writers)

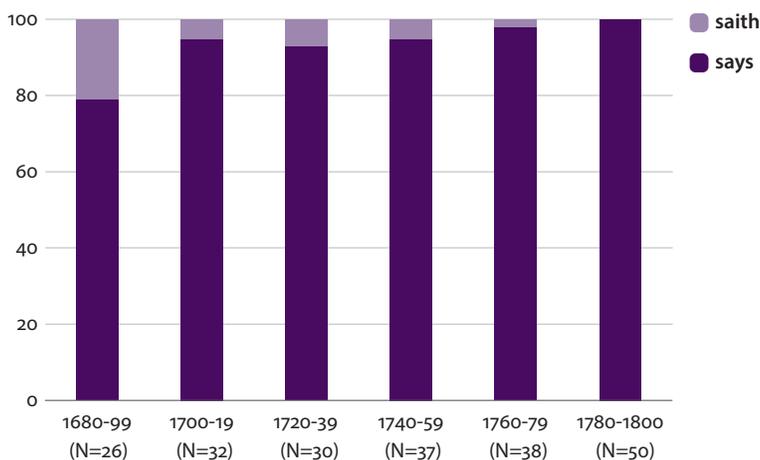


Figure 7.5 Relative frequencies (%) of *says/saith* (N = number of writers)

to the mean scores shown in brackets. These aggregate figures indicate that the last two decades of the 17th century were markedly different from the 18th century, especially with *have* and *do*, as they display well over 40 per cent frequencies of the outgoing form in that period. *Say* is more advanced and the *-th* form only occurs in some 20 per cent of the cases between 1680 and 1699. In all three verbs *-th* forms virtually disappear by the mid-18th century. The CEEC findings on *have* agree with both the EEBO and the Google Books data presented in Figures 7.2 and 13.5, respectively, in that the crossing over from *hath* to *has* appears to have taken place around the turn of the 18th century.

Referring to the stages in the S-curve discussed above in Section 5.1.2, we can see that *has* and *does* were both in mid-range (between 36 and 65 per cent) in the last two decades of the 17th century, whereas *says* was already nearing completion (between 66 and 85 per cent). Passing the 85 per cent mark, the change was completed with *does* and *says* in the first two decades of the 18th century, while *has* was only nearing completion at the time.

A closer look at the spread of verbal *-s* to the main verb and auxiliary functions of *have* provides further evidence for lexical diffusion based on frequency. The auxiliary, over five times as frequent as the main verb, has 51 per cent of *-s* in the first period, 1680–1699, where the main verb has reached the 63 per cent level. According to the chi-square test, this is a significant difference ( $p < 0.01$ ). As the process advances, the gap narrows. In the next period, 1700–1719, the main verb has completed the change at 88 per cent, and the auxiliary shows the average of 85 per cent. In the following period, the difference can no longer be detected in the data with 93 and 92 per cent frequencies of the incoming form in the main verb and auxiliary, respectively (Figure 7.6).

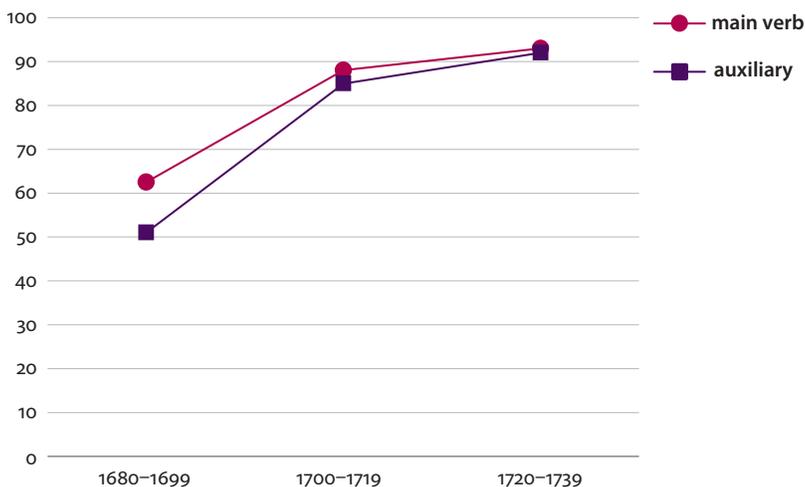


Figure 7.6 Relative frequencies (%) of *has* as main verb and auxiliary

### 7.3.2 Gender variation

The writer's gender proved one of the key social variables in the diffusion of verbal -s in the 16th and 17th century data studied in Nevalainen & Raumolin-Brunberg (2003: 122–124), and of *has* and *does* in the period 1640–1680 (Nevalainen 1996). In both cases, women led the process.

Table 7.1 presents the relative frequencies of the -s forms for *have*, *do* and *say* by gender in 20-year periods in the 18th-century data (see the Appendix for absolute frequencies). The table shows the results based on pooling the subgroup data but also records, in brackets, the results obtained by averaging individual writers' average frequencies if they differ from those obtained by pooling by more than two percentage points. For both methods the minimum number of individuals contributing to the subgroup total per period was set at five and for averaging individual averages, following Nevalainen, Raumolin-Brunberg & Mannila (2011), the minimum frequency of the linguistic variable per person was six.<sup>2</sup>

2. For *says*, the female-authored material was insufficient in the first subperiod, consisting of a total of only 12 observations with 4 of the 5 instances of *saith* by a single individual.

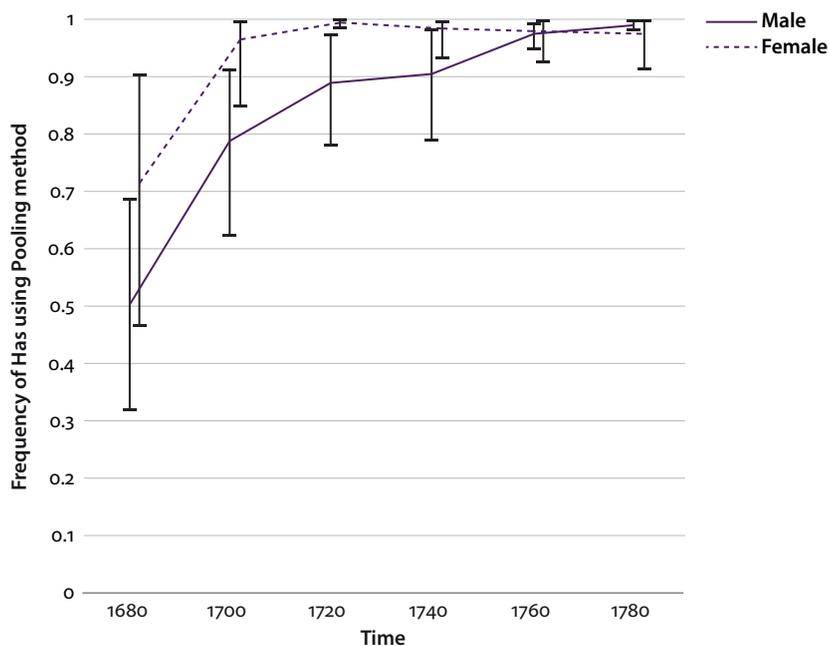
**Table 7.1** Completion of the change in male and female writers (> 85%, cells highlighted in grey); N = number of individuals contributing to the totals

Period	Men						Women					
	<i>has</i> (%)	N	<i>does</i> (%)	N	<i>says</i> (%)	N	<i>has</i> (%)	N	<i>does</i> (%)	N	<i>says</i> (%)	N
1680–1699	50.2	39	51.8	27	85.0	20	71.2 (66.7)	15	68.2	9	58.3*	6
1700–1719	78.6	40	86.7	23	90.5	24	96.4 (89.4)	10	98.0	8	100.0	8
1720–1739	88.5	48	83.3	24	89.5 (77.2)	22	99.6	18	100.0	12	100.0	8
1740–1759	90.3	55	96.6	31	93.3 (87.5)	29	98.6	14	100.0	6	100.0	8
1760–1779	97.7	52	98.0	30	96.9	29	97.9	18	100.0	9	100.0	9
1780–1800	99.3	49	97.5	34	99.6	30	97.4	24	100.0	20	100.0	20

\* (of the total of 12 instances)

The overall picture remains largely the same with both methods. Table 7.1 suggests that gender difference remains relevant in the final stages of the diffusion of verbal *-s*, women leading in all three cases. Considering the change completed when the incoming form exceeds the relative frequency of 85%, we find that with *has* this is the case for women in the first 20-year period of the 18th century, but only a generation later, in the next 20-year period, for men. With *says*, women's usage becomes invariable at 100% between 1700 and 1719, and with *does* in the next 20-year period. The most frequent form, *has*, does not reach the categorical 100% level during the 18th century for either gender, and men show some degree of variation with both *does* and *says* throughout the period.

As always, mean frequencies abstract away from subgroup internal variation. This aspect of the data was examined using bootstrapping, a resampling method which gives information on the uncertainty of mean frequency estimates. Figure 7.7 presents the gender diffusion of *has* with confidence intervals using this method (see Section 5.2). It suggests that the range of variation, and hence uncertainty, is the widest in the first subperiod for both genders and that there is a fair amount of overlap between them. The pooled figures for the period 1680–99 in Table 7.1 should therefore be interpreted as more uncertain than those for the later periods. By contrast, in the period 1720–39 the confidence intervals for two genders do not overlap, showing that women writers had completed the change but that this cannot be predicted confidently about male writers at the 85% level until 1760–79.



**Figure 7.7** Gender variation in the distribution of *has*, bootstrap estimates (start of each 20-year period)

### 7.3.3 Social status variation

The diffusion of verbal *-s* to the south was originally a process from below in social terms. In its final stages, social status distinctions were largely levelled out in the verbal category as a whole (Nevalainen & Raumolin-Brunberg 2003: 140–141, 144–145). The relevance of social status for *-s* generalization in the last three verbs to complete it proved harder to assess than that of gender. This was due to the fast completion of the process in *do* and *say* and to fine social categorization reducing subgroup sizes. *Have* was the only verb to provide more data, but even with *have* the analysis was complicated by the increasingly polarized use of the alternative forms (see 7.4, below).

The analysis of *have* was carried out in stages, the first based on the 20-year and 11-rank template used in the tabulation of the raw frequencies of the verbs. This level of detail resulted in a number of empty cells and very low numbers in some categories. Aggregating the six 20-year periods into three 40-year periods and reducing the social status categories to five enabled a more systematic quantitative comparison: now there was only one social category in one period with fewer than five individuals with the required minimum of six instances. The broad social

categories examined were: nobility (including royalty), gentry (upper and lower), professionals, clergy (upper and lower), and other ranks (merchants, tradesmen and those below them in the social hierarchy).

However, aggregating individual scores by pooling proved problematic in that the lower social orders included a few persons with large contributions, thus skewing the analysis in their favour (cf. 5.2). Averaging individual writers' average scores was therefore resorted to instead, although it draws on fewer informants. The resulting implicational scale in Table 7.2 suggests that none of the five status categories reached the completion stage in the first period: the nobility and the gentry were nearing completion while the rest were at mid-range. The table also suggests that those representing the other ranks, notably merchants, did not complete the change in the course of the 18th century. But it should be noted that in that category the numbers are small and hence subject to individual fluctuation. The conclusion drawn from Table 7.2 is that the diffusion of *has* was first completed by the upper ranks, the nobility showing invariable use of the incoming form from the middle period on.

**Table 7.2** The diffusion of *has* by social status, averages of average scores (> 85%, cells highlighted in grey); N = number of individuals with more than 6 instances of the variable

Social rank	1680–1719		1720–1759		1760–1800	
	<i>has</i> (%)	N	<i>has</i> (%)	N	<i>has</i> (%)	N
Nobility	82.1	11	100.0	22	100.0	24
Gentry	69.6	28	99.7	28	98.1	21
Professionals	51.5	11	89.6	9	99.1	31
Clergy	62.7	11	78.6	18	93.3	16
Other ranks	60.0	5	75.0	4	79.8	8

#### 7.4 Polarization of individuals

Throughout the generalization of verbal *-s*, letter writers showed a fair amount of variable use of the two suffixes. In the category as a whole (excluding *have* and *do*) this was the case from the mid-range of the process to its final stages, as most writers had variable grammar, using neither form less than 10% or more than 90% of the time (Nevalainen & Raumolin-Brunberg 2003: 92–98). However, when the diffusion of *-s* in *have*, *do* and *say* was nearing completion, remarkable polarization took place, as not only were those who preferred *-th* forms reduced to a tiny minority but variable use also radically diminished. Analyzing all those who provided six or more instances of the linguistic variable, we find that, with *have*, variable use

fell from 16% in the first subperiod (1680–1699) to 7% in the next (1700–1719), falling from 5% to 2% during the remaining four 20-year periods. The process of polarization was equally rapid with *do* and *say*.<sup>3</sup>

#### 7.4.1 Conservative minority

In 1680–1699 there were more writers with less than 10% of the incoming forms than in the later periods and, overall, many more with *have* than with *do* and *say*. An intriguing sociolinguistic question is whether these conservative individuals had anything in common. Table 7.3 lists all those who provided six or more instances of the linguistic variable and hardly used the incoming forms at all.<sup>4</sup>

A number of issues emerge from Table 7.3. First, most people (shown in bold) are consistent in their linguistic choices. Those whose writing spans over two periods are conservative in both. With *have*, this is the case with three clergymen: Humphrey Prideaux, Francis Blomefield and Theophilus Hughes. Several writers are conservative in their use of both *have* and *do* (Daniel Fleming, Nathaniel Pinney, Humphrey Prideaux,) or *have* and *say* (William Burbidge and Thomas Secker). Thomas Browne was conservative in his use of *do* and stylistically divided with *have* (19% of *has* overall, see 7.4.2). Only one person, John Gay, who writes in two periods, shows an increase in his use of *has*, from 8% in 1700–1719 to 37% in 1720–1739.

Various factors could account for a writer's conservatism, age being an obvious one. Looking at the first period, it could be a relevant factor in the case of Thomas Browne, who was born in 1605, but probably not with Humphrey Prideaux (b. 1648), or even less with somebody like Nathaniel Pinney (b. 1659). Education does not seem to correlate directly with conservative use either: Sir Thomas Browne, a medical doctor, had a university education, and so did Humphrey Prideaux, a clergyman. By contrast, Nathaniel Pinney was a merchant with no classical education. This pattern of social variation is repeated with women: Anne Montague was a noblewoman, Elizabeth Hatton a member of the lower gentry and Esther St. Michel the uneducated wife of a professional. The three men used the outgoing form when writing to their immediate family or a close friend; Elizabeth Hatton and Esther St. Michel to their in-laws, and Anne Montague to her niece.

---

3. Due to the small number of writers with the minimum of six instances of the variable, with *do* the proportion of those who had variable grammar fluctuated between 0 (four periods) and 20% (one period), and with *say* between 0 (two periods) and 17% (one period).

4. Following Nevalainen & Raumolin-Brunberg (2003: 97), if a person's score (%) of a variant form fell between 0 and 10 or 90 and 100, his or her grammar was considered invariable with respect to that form.

**Table 7.3** Conservative individuals (the relative frequency of *-s* forms and social status shown in brackets).\*

Period	<i>has</i>	<i>does</i>	<i>says</i>
1680–1699	N = 37 Anne Montague (sr) (0%; N) +++ (Thomas Browne (19%; G) +++) William Petty (0%; G) ++ <b>Daniel Fleming</b> (2%; G) ++ Roger Fleming (0%; G) Charles Hatton (0%; G) -- Elizabeth Hatton (0%; G) William Longueville (0%; G) -- Philip Henry (0%; C) -- John Pinney (0%; C) <b>Humphrey Prideaux</b> (0%; C) ++ Balthasar St. Michel (8%; P) Esther St. Michel (0%; P) William Bolton (0%; M) Nathaniel Pinney (0%; M)	N = 6 <b>Thomas Browne</b> (0%; G) +++ <b>Daniel Fleming</b> (0%; G) ++ <b>Humphrey Prideaux</b> (0%; C) ++ <b>Nathaniel Pinney</b> (0%; M)	n.a.
1700–1719	N = 29 William Berkeley (3%; N) <b>Humphrey Prideaux</b> (0%; C) ++ John Gay (8%; P)	N = 5	N = 6
1720–1739	N = 41 <b>Francis Blomefield</b> (5%; C) ++ <b>William Burbidge</b> (0%; O)	N = 4	N = 8 <b>William Burbidge</b> (0%; O)
1740–1759	N = 40 <b>Thomas Secker</b> (2%; C) ++ <b>Francis Blomefield</b> (0%; C) ++ <b>Theophilus Hughes</b> (8%; C)	N = 10	N = 10 <b>Thomas Secker</b> (0%; C) ++
1760–1779	N = 48 <b>Theophilus Hughes</b> (0%; C)	N = 14	N = 9
1780–1800	N = 52 John Clift (0%; O)	N = 21	N = 14

\* N = number of individuals with the minimum of 6 instances of the variable in each period; social classes: N = nobility, G = gentry, C = clergy, P = professional, M = merchant, O = other; upward (++) and downward (--) social mobility.

Some of the conservatism may have its roots in regional dialects. Previous research shows how slow East Anglians were in adopting verbal -s (Nevalainen, Raumolin-Brunberg & Trudgill 2001; Nevalainen & Raumolin-Brunberg 2003: 177–180; Joby 2014). Born in London, Thomas Browne lived most of his adult life in Norwich, and Francis Blomefield, a native of Norfolk, never left the region. Perhaps a stronger case could be made for the south-west. The sources cited in Ihalainen (1994: 224–225) indicate that forms in *-th* were preserved in rural dialects in the south-west as late as the 19th century and sporadically even later. Nathaniel Pinney was born and lived in Dorset, and John Clift was from Cornwall and lived there all his life. Balthasar St. Michel was born in Devon, as was John Gay, the poet and dramatist, and the family of William Berkeley hailed from Somerset.

Yet there are also many writers whose usage cannot be explained by rural dialect influence: a group of *hath*-users consists of individuals whose origins were in the Home Counties and the capital region, including William Petty, a natural philosopher and administrator in Ireland (Hampshire), William Longueville, a lawyer (Buckinghamshire), and Philip Henry, a clergyman (Westminster). Apart from the Flemings (Yorkshire), conservative writers with northern roots – and therefore expected to use *has* rather than *hath* – include William Burbidge (Lincolnshire) and Thomas Secker (Nottinghamshire). However, family ties were not always a clue to conservative usage in the 17th century: among the three Fleming brothers, George Fleming used *has* in 26% of the cases, and Henry in 98%, whereas the youngest, Roger, did not use it at all. A 17-year-old Oxford undergraduate writing to his conservative father Daniel Fleming, Roger was probably more mindful than his older brothers of the register associations of *hath*.

However, one thing many upper-ranking conservatives shared was either upward or downward social mobility with respect to their father's social status. Individuals who moved several steps up on the social ladder (marked by plus signs in Table 7.3) are found among the nobility, gentry and clergy. They include the *-th* users Sir Thomas Browne, Daniel Fleming, Humphrey Prideaux, Francis Blomefield and Thomas Secker. Those who moved downwards (minus signs) were fewer but came from nobility and upper gentry backgrounds or were like Philip Henry, the son of a courtier. A similar phenomenon was found in our previous research as the rise of verbal -s was shown to be one of the changes which social aspirers, in particular, were sensitive to (Nevalainen & Raumolin-Brunberg 2003: 144–145).

The growing frequencies of *has* suggest that its social connotations were neutralized over time. However, as Table 7.2 suggests, it was the clergy and professionals, on the one hand, and the merchant and other ranks, on the other, who retained the outgoing form the longest. Educated men and members of the clergy may have favoured *hath* because it continued to be associated with official and

religious registers. As far as people with little formal schooling are concerned, the use of the outgoing form could have been a reflection of their vernacular, which converged with the biblical registers, written and spoken, they were familiar with. To find out whether this is a plausible generalization, two writers will be analyzed in more detail.

#### 7.4.2 Two case studies: Thomas Browne and John Clift

Two men from different social and regional circumstances were selected for closer analysis, Sir Thomas Browne and John Clift. Browne wrote in the late 17th century and Clift towards the end of the 18th. Sir Thomas Browne (1605–1682), a medical doctor, had been knighted for his merits but came from the merchant ranks and was raised by his uncle, a grocer. He was born in London and studied in Oxford and on the continent, settling in Norwich in 1637. In his correspondence Browne invariably used *doth* and almost invariably *hath*. *Has* only appears in his letters to his daughter Betty (b. 1648), while *hath* is the sole form found in letters addressed to his other correspondents, including his son Edward (1644–1708). The passages in (7.10) and (7.11) are dated to the same day and partly recount the same news but with different verb forms (third-person singular forms shown in bold).

##### (7.10) Dearest Betty,

Thy letters are still our best diversion, to hear you and all that belong to you gat so well to Portsmouth was very wellcom to us, yr thoughts for us are Equall with ours for you. I am sure their **Passes** not a day that we are not severall times talking of you. Poor Tomeys Cough **have** brought him in to a great heat, but I hope it will not be so bad as that feavor was wch you were so helpfull to him; his stomack very bad, we are this after noon goeing to bath [him] by his own desire.

Our Tommy **has** had a grievous Cof and feavor, yr sister Frank **has** dun more for him then I could have thought; he was bled and bathd and I bless God he had got down amongst us again and is much delighted with yr Letters ...

(Sir Thomas Browne to his daughter Mrs. Lyttelton, 6 June 1681;  
BROWNE, 194)

##### (7.11) D. S.,

Mr Deane Astley went to London on this day was sevenight, and sayd hee would call upon you. [...] This day, God bee thancked, wee had a fine showre of raine; the spouts of our howse have not Runne for 8 or 9 weekes before. I had a cough for 6 weekes very feirce in the night and it held mee till within these 12 or 14 dayes, most persons in my howse had it or have it except my wife. Frank **hath** it and it **hath** been with hooping and vomiting, butt is persua[dable] to

take litle and will not abstaine from going to morning and evening prayers which wee daylie have at our owne parish church. Tommy **hath** had it with some hooping and vomiting, butt now **vomits** butt seldome, butt **sleepes** prettie well in the night and at any time when hee **lyeth** downe in the day; hee **hath** been very hot and so that hee beggd to bleed a litle and to goe into a *balneum dulce* which [he] had used in a sicknesse before.

(Sir Thomas Browne to his son Dr. Edward Browne, 6 June 1681;  
BROWNE, 192–193)

Browne's use of *hath* and *has* corresponds to his other rhetorical choices. He responds to his daughter's affectionate thoughts but in his son's letter dispenses with such intimacies, beginning with news about visiting relatives. The letter to Betty opens *Dearest Betty* and *thy*, the one to Edward with the abbreviation *D.S.* (*Dear(est) Son*). In the letter to Betty the other family members are qualified by possessive determiners (*our Tommy*, *yr sister Frank*). With frequent references to *you*, the style is more interactive than in the narrative in (7.11), which relies on plain third-person references: *our Tommy* in (7.10) is referred to as *Tommy* or *Tom*, and *yr sister Frank* as *Frank*. In (7.10) Browne uses the first-person pronoun *I* to express stance (*I am sure*, *I hope*), whereas in (7.11) it only marks him as the experiencer of an ailment (*I had a cough*). In (7.11), the first person possessive *my* appears in *my house* and, strikingly, *my wife*, referring to Edward's mother, Dorothy Browne.

These two excerpts suggest that for somebody like Sir Thomas Browne, who grew up in the first decades of the 17th century, the choice between *has* and *hath* was a stylistic one: *hath* was more associated with detachment and *has* with involvement, potentially varying with the East Anglian zero variant (*Cough have brought*; Nevalainen et al. 2001). It is noteworthy that the other family members did not follow this pattern but used verbal *has* throughout. Example (7.4), above, comes from Dorothy Browne's letter to her son Edward.

John Clift (1759–1819), the son of a Cornish miller, was born at Glynn, near Cardinham, and lived all his life in Cornwall. He was apprenticed to carpentry and building but there is no record of him having received any formal education apart from some elementary schooling. The excerpt in (7.12) reproduces parts of a letter John wrote in his late 30s to his younger brother William (1775–1849). Compared to Browne's, Clift's spelling is more standardized and does not appear to contain any distinct dialect forms. He uses verbal -s with all verbs except *have* and *do*. But his use of *hath* does not suggest detachment; on the contrary, the letter strikes the modern reader as involved and even chatty. Clift refers extensively to himself, and the recipient *you* is included in the discussion, as is news and gossip of sisters and cousins. The first-person pronoun *I* is found in a variety of expressions, including stance markers (*I am certain*, *I am verry happy*).

(7.12) Dear Brother,

I am certain by this time you think that I never intend to write you any more I shoud have writ long before now but I have been at St Kew for some time Painting for our Steward [...] I am verry happy to hear that you enjoy your health which I cannot say I do at this time so well as I coud wish I caught a Great cold about aMonth ago in Geting wet in going to wood after Timber which I have not Quite rub'd out yet I have a smart pricking pain in one side but if it **doth** not leave me verry soon I intend to get some Blood taken off which us'd to relieve me, I have rec<sup>d</sup> a Letter from Both Sisters since I rec<sup>d</sup> yours Sisters are both in good health they both complain of your not writing to them I find Joanna **intends** to Get Married as soon as her Sailer (as she is pleasd to call him) **returns** from Sea she **hath** desidr to know if I am Just going to be married if not till spring she **says** she will come home and make one of the party [...] Cousin peggy Bate is married to a Soldier and is sent abroad she **hath** had a Child and **hath** been home with her parents for some time but is now gone to Glynn to be nurse Mrs Glynn Died in Child Bed some time ago about 3 Months ...

(John Clift to his brother William Clift, 8 January 1798; CLIFT, 168)

The results of this comparison suggest both continuity and change over time. Towards the end of the 17th century, an elderly educated gentleman could exploit the choice between *hath* and *has* for stylistic effect even in his family correspondence. A hundred years later, outgoing forms still appear in family letters written by a young working-class man in the West Country. It is worth noting that these forms were not limited to men but that John Clift's sister Elizabeth also used *hath* 40% of the time in her letters to William.

## 7.5 Normative grammar

Perusing *A Dictionary of English Normative Grammar, 1700–1800* (Sundby, Børge & Haugland 1991: 303–304, 392), suggests that the three verbs analyzed did not receive much attention from grammar writers. On occasion, they were subsumed under general discussions of *-s/-th* variation, which took on different forms. In general, grammarians appear to have been unaware of the process of lexical diffusion in the outgoing suffix, which was practically completed in mainstream usage by the mid-18th century. Yáñez-Bouza's observation (Chapter 3) that early grammars were not particularly normative but referred to stylistic variation is also borne out here. Introducing archaisms, Lowe (1737: 7), for example, notes that "(till of late, and even now, chiefly in law and divinity) *-eth*, or *-th* is us'd for *s* of the 3d person singular: as *he fill-eth*, *he love-th*". No mention is made about any individual verbs.

In the mid-century the two suffixes begin to be discussed in less neutral terms. Hammond (1750: 62) appeals to the notions of homophony and euphony in his general preference for the *-th* form. He writes: “*Eth* is often changed to *s* or *es*, both in *Discourse* and *Writing*; as in *wants* for *wanteth*, *teaches* for *teacheth*: But this Contraction is apt to confound the Names of *Acts* with Names of *Things* ... and multiplieth the hissing Sound of *s*, which is disagreeable to nice Ears, and makes our *Language* the Jest of Strangers.”

Bayly (1758: 100), in turn, treats the sibilant ending as an instance of what he calls “Abbreviations, allowable enough in the Hurry of Conversation”, but which “ought undoubtedly to be avoided in correct Writing”. But a closer inspection shows that, unlike Hammond, Bayly is not targeting the final fricative as such but the contraction of the syllabic suffix *-eth*: “Thus, *es*, *s*, *th*, for *eth*, as *loves*, *leads*, *doth* for *loveth*, *leadeth*, *doeth*.” That he is, however, not fully consistent in his presentation is shown by his next example: “So, *has* for *hath*.” A similar distinction between the syllabic forms “doeth” and “sayeth” and the contracted forms *doth* and *saith* was made by Kirkby (1746: 136, as cited in Sundby, Børge & Haugland 1991: 303).

The icon of 18th-century normative grammar, Robert Lowth, follows his contemporaries in listing both *-th* and *-s* in his paradigm of third-person singular verb suffixes. Discussing auxiliaries, Lowth notes that “*Hath* properly belongs to the serious and solemn style; *has* to the familiar. The same may be observed of *doth* and *does*” (1763: 37). Lowth testifies to both stylistic considerations and lexical diffusion in that in his grammar he alternates between *-th* and *-s* in auxiliaries but avoids the outgoing forms in his personal correspondence (Ihalainen 1994: 225, Tiekens-Boon van Ostade 2005: 145).

The two auxiliary variants continued to live side by side in grammars until the end of the century. Nicholson (1793: 15) declines *do* in the singular: *I do*, *thou dost*, *he doth*, or *does*; and *have*: *I have*, *thou hast*, *he has*, or *hath*. What matters to him is that agreement and the sequence of forms are not violated. So the students of his grammar should correct sentences like *He that forgetteth his duty*, *does wrong*, and *The years passeth away* (Nicholson 1793: 35). How common such instances were in real life, remains an open question. In general, 18th-century grammars upheld a distinct register and medium difference (conversation vs. “correct” writing) between the two suffixes without paying much attention to individual verbs. In doing so they no doubt reinforced the use of the outgoing suffix in writing. At the same time, the regional use of the frequent forms *hath* and *doth* appears to have gone unnoticed by normative grammarians, most of them writing in the capital region.

## 7.6 Conclusions

The results of this study point in two directions: the outgoing *-th* forms could be fostered, in terms of Trudgill (2011: Ch. 6), both in “societies of strangers” and in “societies of intimates”. This duality had to do with formal or archaic registers, on the one hand, and with regional use, on the other. *Hath* and *doth* had become associated with religious and legal language. Although diminishing numbers of writers made use of them in their personal letters, *-th* forms continued to be recommended by normative grammars for written language throughout the 18th century. *Hath* survived longest in the letters of clergymen and educated male professionals, indicating that they were more conservative in this respect than male writers coming from other social backgrounds or female writers in general.

However, *-th* could also be fostered in “societies of intimates”. Several *hath* users lived or had their roots in regions where the local people continued to employ forms that had long since disappeared from mainstream usage. This was the case with the nonmobile members of the Cornish Cliff family writing in the 1790s. In comparison, in the last two decades of the 17th century, writers from more varied backgrounds continued to use the outgoing forms, especially *hath*, with both intimate and more distant recipients. In both cases, it may have represented a mere polite spelling convention for some and a genuine speech form for others.

One characteristic of the choice between the *-s* and *-th* forms may lend support to the spelling convention hypothesis: few letter writers had variable grammars even with the most frequent verb *have*, but increasingly preferred one form only. As shown by Kytö (1993), Nevalainen & Raumolin-Brunberg (2003: 92–98) and Gries & Hilpert (2010), this was not the case earlier when verbal *-s* was spreading and most people varied their choice of suffix according to both linguistic and social factors. These constraints no longer obtained in the 18th century as the outgoing *-th* forms were rapidly becoming a nonproductive alternative.

## Appendix

Frequencies of verbal *-s*/total of *-s* plus *-th* (N) for male and female writers.

Period	Men			Women		
	<i>has</i> /N	<i>does</i> /N	<i>says</i> /N	<i>has</i> /N	<i>does</i> /N	<i>says</i> /N
1680–1699	429/854	57/110	34/40	79/111	15/22	7/12
1700–1719	544/692	78/90	57/63	378/392	50/51	63/63
1720–1739	545/616	65/78	68/76	253/254	37/37	35/35
1740–1759	878/972	115/119	112/120	348/353	33/33	33/33
1760–1779	1001/1025	147/150	123/127	334/341	65/65	42/42
1780–1800	1644/1655	196/201	232/233	1240/1273	191/191	194/194

# Periphrastic DO in eighteenth-century correspondence

## Emphasis on no social variation

Arja Nurmi

for he has quite convinced me that having no children it is rather a laudable than a wild scheme, for him to occupy his mind in some pursuit from which he secures the absence of ennui, a chance for fortune, & a certainty that if he *does* obtain any it will be in a pleasant way; (Sarah Lennox, 1783; LENNOX, II ,39)

### 8.1 Introduction

In her letter, Sarah Lennox presents a use of DO-support typical for Present-day English: DO is stressed – at least if we are to trust the italics – and it presents a supposition in contrast to what has gone before. By the end of the eighteenth century, this was the case more often than not. While the use of italics (or, more properly, underlining in the handwritten letters) does not appear in every instance, the role of auxiliary DO in affirmative statements does seem, in general, to follow the principles of emphatic DO attested in our contemporary usage.

Periphrastic DO as a construction has received a great deal of attention in earlier research (see e.g. Ellegård 1953; Nurmi 1999), but the final stages of the shift from periphrasis to present-day DO-support have been less studied. For the eighteenth century, Tieken-Boon van Ostade (1987) is one notable exception. There are also case studies of individual authors, most recently Tieken-Boon van Ostade (2014), which touch upon the use of DO.

This study investigates, firstly, whether there still is social and linguistic embedding of auxiliary DO in affirmative statements in the eighteenth century, and, secondly, presents the construction's final stages of change towards present-day usage through an analysis of some of the linguistic conditions of its use. The data comes from the *Corpus of Early English Correspondence Extension* (CEECE), which is described in more detail in Part I, Chapter 4 of this volume.

## 8.2 From periphrastic DO to DO-support

In Present-day English DO-support appears in connection with Negation, Inversion, Code and Emphasis, the so-called NICE-properties (see e.g. Quirk et al. 1985). By 1700, the use of DO was already quite standardised to the present-day format, although some variation still occurred; DO in negated statements, for example, continued to avoid the company of certain high-frequency verbs (Ellegård 1953). The focus of this chapter is on the continuing variation in the use of auxiliary DO in affirmative statements without inversion. Below, the status of DO is first described as it appeared in the sixteenth and seventeenth centuries, then the existing evidence (including contemporary grammars) of eighteenth-century usage is collected, and finally the relevant elements of present-day usage are outlined.

### 8.2.1 Periphrastic DO before the eighteenth century

In the *Corpus of Early English Correspondence*, the frequency of DO in affirmative statements had declined from its peak of 37.2/10,000 words in 1590–1599 to 16.4/10,000 words in the 1660–1681 subperiod (Nurmi 1999: 139, 169). Women seem to have used the construction slightly more often than men at that stage. Ellegård's (1953: 162) data show that after reaching the height of approximately 10% of all possible contexts of use in the late sixteenth century the frequency of DO approached zero in the early eighteenth century. There was still variation depending on genre. Bækken (1998, 1999, 2003) found in her multigenre corpus of Early Modern English that the frequency of DO in non-inverted structures with more than one initial element was 1.0%. The last subperiod of her study (1680–1730) coincides with the early part of the CEECE data used in this chapter.

For earlier stages of English, several reasons for the use of DO in affirmative statements have been hypothesised. Dahl (1956: 11) lists seven cases where DO might have been useful. Of these, tense marking for avoiding ambiguity with non-varying verbs (*put*) and “euphony” in second person singular, where the inflectional ending might produce sound combinations difficult to pronounce (*thoughtst* vs. *didst think*) have found little support in corpus-based studies. As Rissanen (1985: 165) points out, this may be due to the fact that these seem more likely to appear in spoken language, which in historical contexts is not readily available for study. Some other cases, like the presence of an adverbial near the predicate – what Ellegård (1953) divides into SdAV and SAdV – and co-ordinating simple tense with compound tenses (or other auxiliaries) seem to have played a role at most stages of the development of DO.<sup>1</sup> Both are still attested in the CEECE data, and the latter is mentioned in descriptions of Present-day English as a means of contrasting tense or aspect.

---

1. Ellegård's abbreviations refer to S(ubject), d(o), A(dverbial) and V(erb).

The last three of Dahl's reasons for use of DO relate to the present emphatic DO: (1) signalling a statement's truth value, (2) intensity of feelings and (3) expressing solemn statements. The last can also be connected to the legal, performative uses still available. Stein (1990: 32–34) argues that from the fifteenth century onwards at least to the latter half of the seventeenth century one of the uses of DO is at “the informational peak of the discourse”, appearing in “the high point of the episode” and “marking semantic prominence” (see also Rissanen 1985: 168–169). That is, DO was a “marker of saliency or foreground” (Stein 1990: 41). Stein (1990: 50) goes on to identify another use of DO as a “hallmark in Euphuistic style and Ciceronian rhetoric”, associated among other things with antithesis and emphatic language. Related to potentially emphatic uses of DO, Bækken (2002: 328) finds that intensifying and emotional adverbials are quite notably present in her seventeenth-century data. According to Stein, Rissanen and Bækken then, it seems emphatic use has been a characteristic of the auxiliary for a long time.

Denison (1993: 266–267) notes that identifying emphatic periphrastic DO in early texts is very difficult, as all estimations include a subjective element. He nevertheless cites definite examples from the fifteenth century, and points out that there are no earlier definitively acceptable examples. Visser (1969: 1512), on the other hand, claims that it is only “well on in the 19th century, when stressless, periphrastic *do* is obsolete, that instances occur of which one can be absolutely sure that they contain an emphatic *do*”. He goes on to mention that early grammarians mention emphatic DO, but they, in his opinion, do not provide evidence (for further discussion, see Section 8.2.2 below). In the eighteenth century, Tieken-Boon van Ostade (2014: 180–181) finds DO in all its present-day uses in the letters of Jane Austen. She points out that rather than solely for emphasis, DO is also used for contrast (1987: 48–49). This would seem to confirm the presence of emphatic DO from the earliest stages of DO use onwards.

There are attested tendencies related to the verbs used with DO in Early Modern English. Rissanen (1985: 174–175) shows that for some genres multisyllabic verbs tend to appear more commonly with DO than monosyllabic ones (for further discussion of this, see Section 8.4.2 below). This trend is confirmed by Bækken (1999: 118). While there is a difference between the use of DO in favour of transitive rather than intransitive verbs in the earlier centuries (particularly with inverted word order), by the late seventeenth and early eighteenth centuries this difference is no longer significant (Bækken 1999: 111–112). Non-inverted structures were fairly evenly divided between NP and pronoun subjects with transitive verbs, which “suggests that the use of periphrastic *do* is sensitive to the realization of the subject” (Bækken 1999: 119).

According to Sweet (1903: 89), non-emphatic DO continued to appear with “verbs of requesting [...] and asserting” after other uses had disappeared. It is interesting to note that the examples given by Sweet (*I do entreat you* and *I do assure you*)

are both in first person singular. Ard (1982: 461) suggests that these verbs themselves, rather than *DO* alone, convey emphasis. In a similar vein, Ziegeler (2004) discusses the perfective uses of *DO*, first identified by Denison (1985, 1993), and proposes that *DO* was becoming a grammatical marker of perfectivity through association with lexically perfective verbs. Since this use was redundant, *DO* itself later became redundant (2004: 544–545). Ziegeler also provides a suggestion linking the performative uses of *DO* (such as those still attested in legal contexts even today) to this perfective usage. She concludes her analysis by suggesting that periphrastic *DO* was in fact not a semantically empty form, but rather that, through a process she calls hyperanalysis, the verbs with which *DO* co-occurred in its causative and perfective senses have gained these senses and thus rendered *DO* redundant.

### 8.2.2 *DO* in the eighteenth century

There is very little research on periphrastic *DO* in affirmative statements in the eighteenth century. The only substantial study I know of is by Ingrid Tieken-Boon van Ostade (1987), comparing authors writing in a range of genres in a small corpus. She has divided the uses of *DO* into ten different types; here relevant are type 5: plain unemphatic *DO*, and type 6: *DO* with an adverbial and infinitive. The other types deal with negative statements, questions, imperatives and subject-auxiliary inversion, and those constructions follow their own history.

Tieken-Boon van Ostade (1987: 203–225) also carried out a detailed survey of early English grammars and their views on the use of auxiliary *DO*, including altogether 37 works published in the eighteenth century. Since Tieken-Boon van Ostade's review is so comprehensive, I will mostly report on her findings rather than proceed to duplicate her work. For this study, the most relevant sections are the one discussing "plain unemphatic *do*" (Type 5; 1987: 209–211), and the one outlining mentions of emphatic *DO* (1987: 212–213). I refer readers interested in contemporary grammarians' opinions on *DO* in e.g. negated statements to Tieken Boon van Ostade's excellent survey of those constructions in other sections of her (1987) volume.

According to Tieken-Boon van Ostade's survey, all through the eighteenth century some grammarians mentioning *DO* (e.g. Brown 1700; Lane 1700; Brightland and Gildon 1711; Jones 1724; Lowe 1737; Bayly 1772; Fell 1784) recognise the appearance of *DO* in affirmative statements, while others also add a comparison between plain and emphatic *DO* and yet others include a discussion of *DO* in other constructions (e.g. negations). How much of this acceptance of unemphatic *DO* is due to the practice of copying from earlier grammars is up to conjecture. There are, however, also authors who clearly note the archaic nature of *DO* in affirmative

statements. So, for example, Coote (1788) mentions that only emphatic DO is contemporary usage, and Webster (1789) agrees with him. The first grammarian to denounce DO is Samuel Johnson (1755), but there is evidence of resistance even before him. Tieken-Boon van Ostade cites Alexander Pope, who felt unemphatic DO had no place even as a metrical filler in verse. Other grammarians opposed to the construction include White (1761), Burn (1766) and Horne Tooke (1798).

Some authors give reasons for the use of DO. So, Dilworth (1751), Fell (1784) and Loughton (1734) suggest it is an imperfective marker, but Tieken-Boon van Ostade (1987: 211) found little support in her data for this assumption. The use of DO for emphasis was first mentioned by Wallis (1653: 106), but, as Ard (1982: 460–461) points out, this does not mean the emphatic use had not been available even earlier. Emphatic DO keeps being mentioned by various authors in the eighteenth century (Tieken-Boon van Ostade 1987: 212), and the construction is often described in terms of forcefulness, fullness, passion and energy besides emphasis (Tieken-Boon van Ostade 1987: 212–213). Elphinson (1765: 74) styles some uses of DO as “solemn”, which – given the examples he provides – seem to relate to notions of emphasis as well as the legal style.

According to quotations found in Sundby et al. (1991), some early grammarians suggest that using DO for polarity contrast is the preferable form. So, Fogg (1796: I.155) gives the example “Many people suspect that I do not believe the scriptures, but I believe them” and suggests that the latter clause is better expressed as “but I do believe”. A similar example is “I think you honest now, but I thought differently”, with the suggested form “I did think differently” (Murray 1795: 88). On the other hand, plain DO without polarity contrast often receives suggestions for amendment. Thus, Martin (1754: 121) suggests “I do write” is better expressed as “I write”, and Brittain (1788: 75) amends Bacon’s “No man doth accompany with others but he learneth ere he is aware, some gesture, voice, or fashion” into “No man accompanies”. Similarly, Fogg (1796: I.155) finds the periphrastic construction in Burnet’s “If he did this he should find the effects of the love that the blessed virgin did bear him” inelegant and obscure, and suggests the simple form (*bore*) in the concluding clause.

### 8.2.3 Present-day English DO-support

Of the NICE-properties associated with DO (Negation, Inversion, Code, Emphasis), the relevant case here, emphasis, is often termed the emphatic positive (Quirk et al. 1985: 124). It is a use of DO which appears when a speaker or writer denies a negative. Huddleston & Pullum (2002: 97–98) prefer to regard this in the larger framework of “emphatic polarity”, where “stress is placed on the primary verb to emphasize the positive or negative polarity of the clause”. While their view

obviously covers the emphatic DO in negated clauses as well, the focus in my study is solely on the affirmatives. The negative polarity in question can be either stated or implicit (Quirk et al. 1985: 124; Biber et al. 1999: 433; Huddleston & Pullum 2002: 98). In addition to polarity contrast, DO can also be introduced to bear nuclear stress to draw contrastive attention to tense or aspect (Quirk et al. 1985: 171–172; Tieken-Boon van Ostade 1987: 50). Gerner (1996: 247) finds confirmation for polarity and tense contrast in his data, but notes he found no cases where there was a contrast in aspect. Biber et al. (1999: 433) frame the question of contrast in less detailed form, simply suggesting that the use of DO usually marks a contrast, whether explicit or implicit. The contrast can occur across clauses without an overt connector. Huddleston & Pullum (2002: 97–98), in addition to emphatic polarity, mention emphasis on the lexical content, where stress is on the main verb. Gerner (1996: 242–243) creates a more general definition of contrast, calling it contrastive contextualisation, where the speaker has reason to assume that, at the time of the utterance, there is, in the mind of the hearer, a proposition which the speaker is contrasting. In cases like these DO is not necessarily the only part of the utterance indicating contrast.

Nevalainen & Rissanen (1986), in their study of the *London-Lund Corpus of Spoken Present-day English* (LLC), identified three different uses of periphrastic DO in affirmative statements: explicit opposition (DO1, 18% of instances), implicit contrast (DO2, 28%) and no obvious opposition or contrast (DO3, 54%). The first two are covered by the descriptions discussed above, but the third one is not discussed in the standard reference grammars. With the help of the prosodically annotated data of the LLC, Nevalainen & Rissanen were able to take stress into account. The most frequent uses of DO in their data involved stress and pitch prominence, and only approximately 4% showed neither (Nevalainen & Rissanen 1986: 37). This clearly demonstrates that there is no necessary link between prosody and functions of DO. The DO3 of their classification seems to appear typically when the course of discourse is negotiated between speakers: introducing a new discourse topic or concluding a previous one, as well as elaborating or illustrating the topic under discussion. These uses can be summed up as textual emphasis and avoidance of disruptions in communication (Nevalainen & Rissanen 1986: 42–43); some of these may well be related to the marking of semantic prominence identified by Stein (1990) for Early Modern English. Furthermore, DO3 has a clear connection with intensifying adverbials, most frequently *really*, *well*, *actually*, *of course*, with *I think* and *you know* as frequent pragmatic markers (1986: 43–44). The verbs associated with DO3 typically express mental processes (49%), frequent cases are with *want*, *think*, *seem*, *say* and *know*. Finally, the frequency of DO in spoken language is clearly higher than in written language: 0.8 vs. 0.2/1,000 words respectively (Nevalainen & Rissanen 1986: 47).

Kleiner & Preston (1997) suggest that in Present-day English DO has a special function in discourse disputes, particularly where sensitive topics are discussed. The cases they have analysed do not carry nuclear stress and the main verb is not weakly stressed, so they are not prototypical emphatic positives. There is some underlying contrast, which is not necessarily related to the previous conversation but rather to an opinion believed to be generally held. Furthermore, they link this usage to instances where the main verb is stative, typically mental, and to first person singular subjects (*I do honestly feel that a person ...*). Even when the main verb *feel* is included, the function does not refer to the speaker's emotional state, as it seems to be used in the sense 'believe'. They distinguish this use of DO as separate from emphatic and emotive DO, because of the lacking "first-pair part", i.e. an explicit preceding argument to be contradicted (1997: 120–121). In many ways, this seems to be a subcategory of DO3 as described by Nevalainen & Rissanen (1986). Also Gerner's (1996: 243–244) non-contrastive contextualisation seems to be related to this usage, although he argues that even in these cases the speaker presents his or her utterance in a way to imply a contrastive assumption although none is present.

Finally, there are two uses which seem less frequent, the purely emphatic DO, said to indicate the strength of one's beliefs or feelings (Quirk et al. 1985: 124; Huddleston & Pullum 2002: 98), and the archaic style of legal documents, where DO is still found in the empty periphrastic sense as an alternative to the simple present or past, receiving only brief mentions (Quirk et al. 1985: 134; Huddleston & Pullum 2002: 98; Schütze 2004: 496–497).<sup>2</sup>

According to Biber et al. (1999: 433–434), present tense uses of DO are more frequent in conversation (250/1 million words in present tense and 150 in past tense) than in fiction (150/1 million words for both present and past tense), with news and academic writing coming last (100/1 million words in present tense and 50 in past tense). The frequency they give for conversation is clearly lower than that found by Nevalainen & Rissanen (1986). The verb most commonly appearing with DO in all registers is *have*, but other common lexical verbs also appear. Biber et al. list *get, know, go, look, say, want, come, feel, see, think, make, like, need, take, tell, work, believe, happen, hope, seem* and *occur*, some of them with particular associations with specific registers.

---

2. DO also has dialectal uses in Present-day English, for example a habitual meaning in the dialect of East Somerset (Ihalainen 1976).

### 8.2.4 The construction studied

The focus of this study is solely on **DO** in affirmative statements with no subject–verb inversion. All environments where the use of **DO** is triggered by linguistic requirements in present-day English have been excluded. That is, whenever sentence-initial adverbs have caused subject-auxiliary inversion (see Examples 8.7–8.9) and **DO** functions as a support for the construction in the absence of some other auxiliary, this is regarded as mandatory, and therefore is left outside this study.

Examples (8.1)–(8.3) show some typical uses of **DO** in the CEECE. Italics in examples correspond to italics in the letter editions (and most likely underlinings in the manuscripts, although it was not possible to check the manuscripts for this), while boldface has been added to the examples to make the instances of **DO** stand out.

- (8.1) Time & Treasure are more precious, than many young men **do** esteem them  
(Daniel Fleming, 1684; FLEMING X, II, 119)
- (8.2) I am determin'd to write to you, though those dirty fellows of the Post Office **do** read my Letters  
(John Gay, 1730; SWIFT, III, 318)
- (8.3) He knows I *do* love him, & being certain of that he laughs at every objection that is started  
(Sarah Lennox, 1781; LENNOX, II, 5)

As mentioned above, one linguistic feature identified by Ellegård (1953) is the place of adverbial in relation to **DO** and the main verb. Examples (8.4)–(8.6) show some instances of these. In (8.4) **DO** appears without an adverbial (SdV, to use Ellegård's abbreviation). In (8.5) the adverbial appears between subject and **DO** (SAdV) and in (8.6) between **DO** and the main verb (SdAV). In (8.6) the adverbial can also be regarded as intensifying the expression.

- (8.4) I **do** assure you the Bankruptcy is not fraudulent  
(Henry St. John, 1731; SWIFT, III, 421)
- (8.5) I always **did** love it better than London  
(Hester Lynch Piozzi, 1798; PIOZZI, II, 541)
- (8.6) I **do** seriously protest to your Grace, that I am under deeper apprehensions of mischief to myself and family from the partisans of these Men  
(William Nicolson, 1716; ORIGINAL 4, 375)

Excluded cases are ones involving negation and inversion. Negation with *not*, as well as with negative adverbs *never*, *neither* and *nor* (which also cause inversion when appearing at the beginning of a clause) has been excluded (e.g. Example 8.7), as have instances where some other clause-initial element has caused inversion

(8.8). Similarly exclamations, since they also involve inversion,<sup>3</sup> and in fact structurally follow questions, are excluded (8.9). Finally, imperatives are also left outside the scope of this study, mostly because of their rarity in the corpus data.

- (8.7) & *never* **does** one syllable of religion arise to cause the least dispute.  
(Sarah Lennox, 1796; LENNOX, II, 125)
- (8.8) while I was present with you Time bore me on his rapid Wing, *so swiftly* **did** the delightful hours pass on (Jeremiah Bentham, 1745; BENTHAM J, I, 2)
- (8.9) but how often **does** a hard and unbelieving Heart deprive me of its proper Effect!  
(William Cowper, 1771; COWPER W, I, 241)

Periphrastic DO in affirmative statements is one of the linguistic structures and forms which do not present an obvious variable to analyse. It is even difficult to decide whether the variable should include all finite verb forms (*I have been think-ing about this*), as suggested by Frank (1985), all finite verb forms that do not have another auxiliary (*I am uncertain*) or just cases with a bare main verb that could be used with DO (*I think so* vs. *I do think so*), called the “simple form” by Ellegård (1953: 157–158). If we choose the last option, how can we be sure which verbs to include? How do we count verbs for any of these in an untagged corpus? As this would have been far too labour-intensive, the question of a variable was left outside the scope of this study, and the results are presented in normalised frequencies/1,000 words. Furthermore, Nurmi (1999: 65–69) compared the results obtained from Frank’s and Ellegård’s models to normalised frequencies and found that for frequent items like periphrastic DO in affirmative statements all three gave sufficiently similar results for comparisons between informants.

### 8.3 General development of DO in CEECE

The frequency of periphrastic DO in affirmative statements over the course of the period studied shows a steadily declining trend. While the last two decades of the seventeenth century display a frequency somewhat reminiscent of the heyday of DO, the decline in the eighteenth century is rapid and clear. Table 8.1 presents the absolute and normalised frequencies for twenty-year subperiods.

---

3. Inversion here refers to the inversion of auxiliary and subject so that the word order is auxiliary-subject-main verb.

Table 8.1 Periphrastic DO in affirmative statements in CEECE

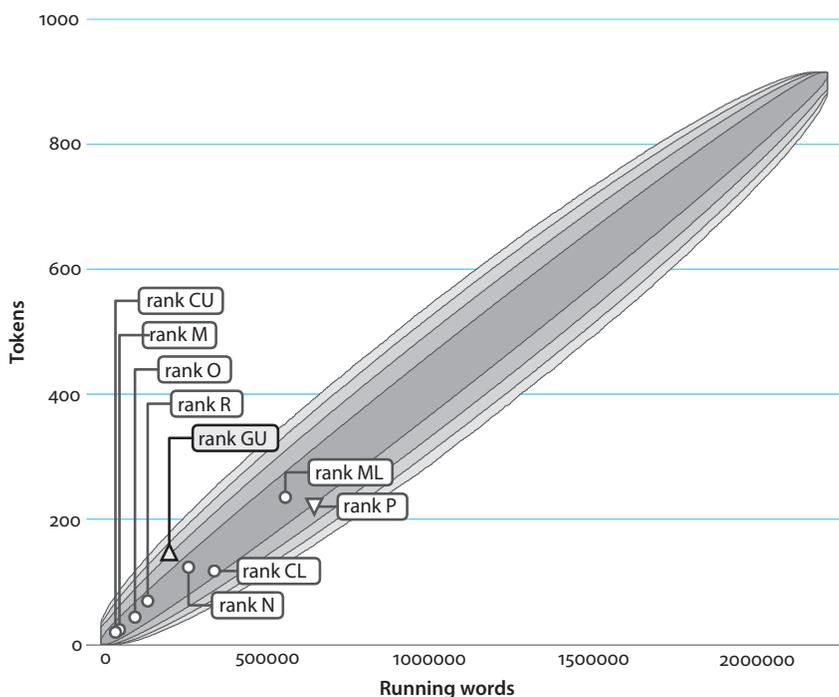
Subperiod	N	/1,000 words
1680–1699	280	0.95
1700–1719	151	0.52
1720–1739	83	0.39
1740–1759	89	0.24
1760–1779	104	0.28
1780–1800	208	0.31

In the latter half of the eighteenth century the frequency of DO resembles that attested in Present-day English spoken language, which is 0.4/1,000 words according to Biber et al. (1999: 433–434) and 0.8/1,000 words according to Nevalainen & Rissanen (1986: 47). Biber et al also record a frequency of 0.3/1,000 words for DO in fiction. Since correspondence has some of the interactive and informal characteristics of spoken language as well as personal narratives similar to fiction these points of comparison seem well suited. Gerner's (1996: 85) results show similar frequencies: 0.195/1,000 words in the London-Lund Corpus of spoken language and 0.228 for written language in the Lancaster-Oslo-Bergen Corpus of written language. Gerner's figures include imperatives, but as their frequency with DO is so low, the effect is probably not notable.

### 8.3.1 DO and social variation

The social embedding of DO in affirmative statements in the CEECE turned out to be nearly non-existent. The only clearly significant factor analysed was change over time. Social variation was first tested using the methods of Säily & Suomela (2009), comparing actual subcorpora with random subcorpora.<sup>4</sup> The results of this test of significance showed that the social variables (gender, social status, relationship with recipient) seem to be non-significant with regard to the use of DO in affirmative statements. As Figure 8.1 shows, there is a slight overuse of DO by members of the upper gentry, but in the absence of any other patterns of significance, this result is possibly due to corpus imbalance, since there is more text from the upper gentry in the first 20-year subperiod of the corpus, when DO-use is also at its highest. Given the small number of informants in this category, it can be assumed that idiolectal variation may well play a role as well.

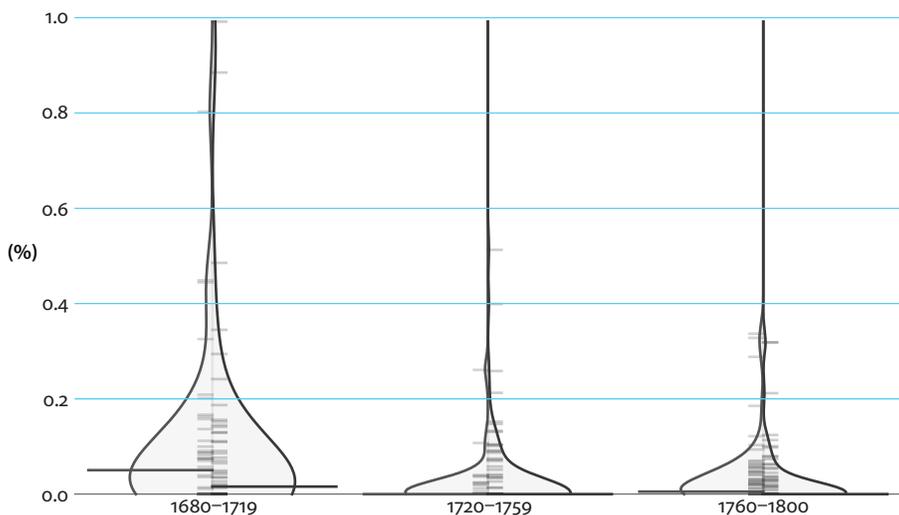
4. I am grateful to Dr Tanja Säily for not only carrying out both tests of statistical analysis for this data, but also for creating the images in Figures 8.1 and 8.2 and generously advising me in their interpretation.



**Figure 8.1** The statistical significance of social status in the use of DO

An additional test was carried out using the Wilcoxon rank-sum method and a modified version of the beanplot, introduced in Vartiainen, Säily & Hakala (2013). This test provided indications that there is a non-significant tendency to use more DO towards the intimate end of our recipient relationship cline and less DO towards the more formal end. That is, our writers tended to use more DO when writing to their nuclear families and close friends, less when writing to their extended families and least DO when writing to strangers. The underuse with distant (T) recipients is the clearest finding of these. Figure 8.2 illustrates the tendency in recipient types.

Finally, four social outliers could be tentatively, although not significantly, identified using the Säily & Suomela (2009) method. Mary Wortley Montagu and David Garrick were using less DO than other writers in the corpus, while Hester Piozzi and Humphrey Wanley were using more DO. In Wanley's case the frequent use of DO can perhaps be partly explained by his age: born in 1672, he may have consistently used DO in the manner and frequency found in his youth. Furthermore, in the letters he was writing to his patron, or to others in superior social positions, he seems to have employed a politeness strategy of emphasising his requests with DO (as in Example 8.10). In other words, this seems to have been a personal mannerism perhaps brought on by both upbringing and the type of correspondence found in the corpus.



**Figure 8.2** Beanplots illustrating the difference in letters written to nuclear family (FN, left) as opposed to more distant recipients (T, right)

(8.10) Therefore I **do** earnestly intreat you, Sir, to enquire after any Books or Inscriptions in their Antient Language. (Humfrey Wanley, 1701; WANLEY, 157)

In many ways, Hester Piozzi is the only genuine outlier, as she writes at the end of the eighteenth century yet uses a great deal of **DO**. Considering her usage patterns, Piozzi seems to be rather an early adopter of the present-day pattern of emphatic **DO** than an old-fashioned writer. As seen in (8.11), Piozzi also used underlining (seen as underlining in the edited volume and italics in the example) to signify that many of the **DO** auxiliaries she used were indeed emphatic. Again, her letters contain a great deal of text written in moments of heightened emotion (regarding, for example, the marriage of her daughter, which Piozzi strongly opposed), so this may have been a combination of temperament and situation rather than any reflection of her status as a linguistic innovator or conservative.

(8.11) Things do not go quite right you find, and if my young Turkeys *do* come out while this Weather lasts – they must die. (Hester Piozzi, 1791; PROZZI, I, 351)

As to the two underusers of **DO**, there seems to be very little to explain their patterns of usage. David Garrick, as an avid Shakespearean (besides acting in numerous roles in the Bard's plays he also built a Temple to Shakespeare at Hampton and staged the Shakespeare Jubilee in Stratford-upon-Avon), would certainly have been familiar with the use of **DO** in affirmative statements. Furthermore, he had some tendency to use linguistic features which could be regarded as archaisms at least in formulaic contexts (for *thou*, see Nevala of this volume). It appears, however,

that he found very little use for DO in his correspondence (only three instances in over 43,000 words). Mary Wortley Montagu was one of the informants studied by Tieken-Boon van Ostade (1985, 1987). Her findings concern negative statements and questions and indicate that Montagu's use of DO in those environments was close to the present-day norm. This may suggest that Montagu's infrequent use of DO in affirmative statements was also a sign of a speaker whose language was ahead of the norm. It may also be that the kind of emotive and emphatic use of DO seen in the letters of e.g. Piozzi was not characteristic of Montagu's writing in general.

It is interesting to note that not even the most robust social variables, region and gender (according to Nevalainen & Raumolin-Brunberg 2003: 198–201) are significant in connection with such a late stage change.<sup>5</sup> During the course of the sixteenth and seventeenth centuries, DO appeared to be a change led by women, and the increase of DO originated in the south, particularly in London and the Court. In the late seventeenth century and over the eighteenth century there is no longer more than a trace of this social embedding.

## 8.4 Frequent linguistic contexts

Since the use of DO in affirmative statements has been linked to various specific linguistic environments in Present-day English, these seemed a suitable focus for further study. I will discuss subject type, and particularly the role of first-person subjects, the types of verb appearing with DO in CEECE and the role of adverbials. In addition, I will present some cross-tabulations in Section 8.4.4.

### 8.4.1 Subject type

The type of subject used with DO, and particularly the construction's connection with first person singular was briefly discussed in Section 8.2. In addition, e.g. Ando (1976) finds a connection with first-person subjects. It should be noted, however, that correspondence tends to show a high frequency of first person subjects in general, since most of it is related to the interpersonal interchange of writer and recipient; this subject type appears, for example, in request markers (*I beseech you*) and attitude/stance phrases (*I think*) as well as phrases relating to typical topics in letters (*I send you*) (Palander-Collin, Nevala & Sairio 2013: 303).

---

5. Region was not tested using the Säily & Suomela (2009) or the Vartiainen, Säily & Hakala (2013) methods, but the differences between regions in DO-use measured by normalised frequencies were negligible.

As Table 8.2 shows, the most common subject type with *DO* in CEECE is first person singular (52%). Given the genre of letters, this is not necessarily noteworthy, but it could also be connected to the “discourse dispute” *DO* in Present-day English described by Kleiner & Preston (1997: 113), as well as the tendency of eighteenth-century grammarians (and Sweet 1903) to use first-person examples when describing the use of *DO*. The other two subject types worth notice are noun phrases (25.9%) and 3rd person singular (9.8%). The usually suggested link between 2nd person singular and use of *DO* is not attested in this data, but it should be noted that instances with *thou* are rare. The other subject types were not frequent enough to be worth mentioning here. As to Bækken’s (1999) suggestion that at least in the seventeenth century the use of *DO* with transitive verbs was sensitive to subject type, this will be explored further in Section 8.4.4.

**Table 8.2** Subject types of clauses with *DO* in CEECE\*

Subject type	N	%
1st person singular	476	52.0
2nd person singular	48	5.2
3rd person singular	90	9.8
Noun phrase	237	25.9
Other	64	7.0
<b>Total</b>	<b>915</b>	

\* For 3rd person singular only pronoun subjects were included, NP subjects were included in the Noun phrase category.

### 8.4.2 Type of main verb

As to the verbs appearing frequently in connection with *DO*, there are some trends in earlier research. Tieken-Boon van Ostade (1987: 118–119) lists *assure*, *believe*, *beseech*, *entreat* and *promise* as appearing in her eighteenth-century correspondence data; these seem to correspond well with the “verbs of strong emotional content” mentioned by Ellegård (1953: 172) in connection with *DO* in affirmative statements. Tieken-Boon van Ostade (1987: 119) further mentions that half the verbs found in her data with *DO* seem to accord with Nevalainen and Rissanen’s (1986: 44) category of “mental processes” found in Present-day English.

Similarly, Ando (1976: 310) notes that in Marlowe’s language there seems to be a pattern of *DO*, first person singular and verbs of strong emotional content. He lists *believe*, *beseech*, *assure* and *confess* in this context. Kekäläinen (1971: 101–102), in her study of popular ballads, found more than 400 different verbs used with *DO* in affirmative statements. These were most often verbs of motion, utterance

and perception, while also rest, emotion, mental activity and sound were found in reasonable quantities. The most common verbs appearing with DO in Kekäläinen's data were *say* and *see*, with *go*, *fall* and *call* following. Stein (1990: 93–95) suggests that the use of DO with speech act verbs is linked to the ideas of foregrounding and intensifying with adverbials, which he identifies as some of the main elements in the use of DO. He further identifies the main groups of verbs appearing with DO to be “speech act verbs, performative verbs and verbs of perception, such as *consider*, *discern*, *hear*” (1990: 94–95), some of which would fall in the category of mental process. All these earlier studies look at main verbs specifically appearing with DO, and do not take into account the overall frequency of the verbs they list in their data.

Table 8.3 shows the most frequent verbs found with DO in CEECE. The two first, *assure* and *believe* were also on Tieken's (1987) list, but the rest do not overlap. *Beseech*, in fact, does not appear at all with DO in CEECE, and *entreat* (4 instances) and *promise* (8 instances) are relatively rare. From Ando's list *confess* appears only 3 times.

Table 8.3 Most frequent verbs found with DO in CEECE

Main verb	N	% of all instances of DO
<i>assure</i>	58	6.3
<i>believe</i>	43	4.7
<i>think</i>	30	3.3
<i>love</i>	23	2.5
<i>intend</i>	18	2.0
<i>come</i>	15	1.6
<i>desire</i>	15	1.6
<i>know</i>	15	1.6
<i>say</i>	15	1.6

Table 8.4 Types of main verbs associated with DO in CEECE

Type of main verb	N	%
Speech act	201	21.7
Cognitive process	165	17.8
Emotion	124	13.4
Other	435	47.0

In order to compare the findings from CEECE to the suggestion by Ellegård of the link between DO and verbs of strong emotion, the main verbs used with DO were classified into three groups. Speech act verbs such as *assure* and *say* form the largest group (21.7%), verbs expressing cognitive processes (*believe*, *think*) come next with 17.8% and verbs of emotion (*love*) come third (13.4%). The group

Other is quite large with 47%, and contains many high-frequency verbs (*come*). Of Kekäläinen's (1971: 101) categories, "utterance" seems to coincide with speech act verbs, and mental activity is similar to my category of cognitive process (and similar to the mental processes discussed by Nevalainen & Rissanen 1986 and Kleiner & Preston 1997).

A further point related to the type of main verb is the length of it. This is brought up both by Rissanen (1985: 174–175) and Baekken (1999). Both distinguish between monosyllabic and multisyllabic verbs used with *DO*. In Rissanen's data *DO* is most frequently found with long verbs (57% to 73% depending on text), with the exception of sermons, where it is equally common with long and short verbs. In the CEECE data 55% of the instances of *DO* are with multisyllabic verbs and the remaining 45% with monosyllabic verbs. It would seem that the difference is not notable.

#### 8.4.3 Adverbials with *DO*

As one of the possibly influential features of *DO* use, the place of adverbials according to Ellegård's (1953) scheme was tracked. Of the 915 instances of *DO* in the data, 67% were without an adverbial between the subject and the main verb. The construction SdAV was the more frequent one (29%), with only 3% appearing in the SAdV format. Compared to the results in Rissanen (1985: 173) for seventeenth-century American English, where the SdAV construction ranged from 23% to 38% depending on genre and author, it is clear that there is no great change in this respect. There is no comparison material for Present-day English, but the unmarked position for adverbials is between auxiliary and main verb, which seems to suggest that the placing of the adverbial may have more to do with word order than any special relationship with *DO* at this point.

The types of adverbials appearing with *DO* have been studied either in relation to the two structures described above or to the clause in general. Tieken-Boon van Ostade (1987: 120) lists the adverbials she found in her data in the SdAV structure, but most of them appear only once, and there is considerable variation between the genres she studied. The only adverbial to appear with any great frequency is *but*, which is most common in direct speech. In Bækken's (2002) data, intensifying and emotional adverbials appearing with *DO* are not limited to these two positions. For Present-day English, Nevalainen & Rissanen (1986) identify a connection between their unstressed *DO*3 and intensifying adverbials, but they do not suggest a specific place for them. In CEECE, 333 instances of *DO* (36%) appear accompanied by an intensifying adverbial. The most commonly appearing ones are *really* (21 instances), *indeed* (19) and *but* (17); otherwise, there is a great variety of expressions.

#### 8.4.4 Cross-tabulating subject type and main verb

One connection between subject type and type of main verb became apparent during the analysis. It would seem that the three categories of frequent verb types (speech act, cognitive process, emotion) typically occur with first person subjects while the verbs in the category Other co-occur more commonly with noun phrase subjects.

Based on Table 8.5 it can be seen that 73% of clauses with 1st person singular subjects have a verb from the first three categories, while 77% of NP subjects have a verb from the category Other. In other words, it seems that a noun phrase subject could be one of the triggers of using DO in eighteenth-century English, being perhaps weightier and therefore requiring some fortification for the verb phrase as well.

**Table 8.5** Cross-tabulated subject type and main verb with DO in CEECE

	Speech act	Cognitive	Emotion	Other	Total
1st sg	139	129	82	126	476
2nd sg	3	14	5	26	48
3rd sg	11	14	10	55	90
NP	31	8	9	188	236
Other	15	11	0	39	61
<b>Total</b>	<b>199</b>	<b>176</b>	<b>106</b>	<b>434</b>	<b>915</b>

### 8.5 Towards DO-support

As DO in affirmative statements seems to have no social stratification and judging by the slow rate of decline towards the end of the eighteenth century, it seems plausible to assume that this is a change close to its very end. This supposition is supported by Tieken-Boon van Ostade (1987), who mentions that in her eighteenth-century data half the instances fit into Nevalainen and Rissanen's (1986) four categories of DO-use in Present-day spoken English (for a description of these, see Section 8.2.3 above). This last subsection looks at the differences between the first and last sub-periods of the CEECE, 1680–1699 and 1780–1800 respectively.

Emphasis is indicated by underlining in 36 instances of the 915 in the CEECE data. All instances are found in the three last subperiods (after 1740), and only seven appear before 1780. While the role of editors should not be forgotten, it seems that the need to indicate emphasis has risen. Another 36 instances show DO co-ordinated with another auxiliary, as in Example (8.12). This usage appears through the whole time period covered by CEECE.

(8.12) I have felt, & **do** feel, so much unwillingness.

(Fanny Burney, 1775; BURNEY F, II, 64)

In the first subperiod, 108 of the 280 instances (39%) are in the first person singular and 111 occurrences (40%) have NP subjects. Of the main verbs 49 (18%) are speech act verbs, 42 (15%) express cognitive processes and 29 (10%) emotions. In the last subperiod, the most diagnostic subject types show 114 of 208 (55%) in the first person singular, 40 (19%) with an NP subject. The main verbs have 27 (13%) speech act verbs, 38 (18%) expressing cognitive processes and 39 (19%) emotions. In other words, the role of first-person singular subjects in the use of **DO** in affirmative statements has increased considerably. This is consistent with the present-day usage, particularly the trends identified by Kleiner & Preston (1997). The verb types associated with **DO** show two main changes. Firstly, in the last subperiod a larger proportion of the verbs (from 43% to 50%) belong to one of the three categories of frequent use. Secondly, there seems to be a shift to using more verbs of emotion with **DO**, with the proportion of speech act verbs declining. This could be regarded as opposing the descriptions of Present-day English **DO**-support with an inclination towards verbs of mental processes. On the other hand, it could arguably also be an indication of the status of **DO** as a marker of emphasis.

## 8.6 Conclusion

This study has shown that in the eighteenth century periphrastic **DO** in affirmative statements demonstrates no notable sociolinguistic variation, with the exception of the upper gentry employing the form more than other social strata. The probable explanation for this state of affairs is the stage of change **DO** is in: it is already in the last two decades of the seventeenth century clearly a declining variant, and in the eighteenth century its low frequency may have made it less salient as a social variable.

It can be argued that **DO** is no longer a semantically empty tense carrier, as it was characterised in the sixteenth and seventeenth centuries, but rather has found its role in the specific functions that are still in evidence in Present-day English. In the words of Ihalainen (1982: 3), the “perfectly good change that did not quite make it” as part of the tendency of English moving towards a more analytic language is clearly finally done (Denison 1993).

If we consider the development of **DO** in affirmative statements as an s-curve, we have to separate the two parts of the change. The original spread of **DO** went on until the early seventeenth century, but never seemed to reach the rapid spread stage (Ellegård 1953; Nurmi 1999). The decline of **DO** can be regarded as an inverted

s-curve, with the long tail of the curve never quite ending. DO has been repurposed and assigned fairly specific duties, and this change seems to have taken place during the course of the eighteenth century. All the present-day uses of DO (emphasis, contrast, emotion) have, of course, been evident from the earliest stages onward, but more as accidental components of usage, not as the main meal. Now they are the central ground of DO-usage in English.



# Indefinite pronouns with singular human reference

## Recessive and ongoing

Mikko Laitinen

### 9.1 Introduction

This chapter discusses indefinite pronoun variability, and the results illustrate how the diachronic developments, studied up to the end of the seventeenth century in previous studies, continue in the eighteenth. The focus is on four paradigmatic variants, the compounds with *-one*, *-body*, *-man* and the independent forms, such as *some (other)*, *any (other)*, *every* and *each*. The diachronic developments of these variants have been documented in Raumolin-Brunberg (1994) and Raumolin-Brunberg & Kahlas-Tarkka (1997) who trace their variation in the diachronic part of the *Helsinki Corpus of English Texts*. Rissanen (1997) examines the indefinite pronouns in Middle English and investigates the emergence of *some* and *any* from the Old English *wh*-pronouns, *man*, *aught* and *sum/ænig* (also Los 2002 on the loss of *man*). In addition, the data presented in Nevalainen & Raumolin-Brunberg (2003) offer a refined view of the sociolinguistic variability of these pronouns in Early Modern English correspondence. Their results indicate that, on the one hand, gender was the only background variable that showed correlation with the variables, and on the other hand, the change consisted of three processes: (1) the rise of the forms in *-body* and *-one*, and (2) the gradual loss of *-man*, and (3) the rise and fall of the independent forms in Early Modern English. However, much less is known of the developments of these compound indefinites during the late modern period. Denison (1998: 101–106) only briefly discusses the variation between *-one* and *-body* forms in his extensive presentation of Late Modern English syntax. D’Arcy et al. (2013) investigate stylistic nuances between two of the variants, *-body* and *-one* in Modern English, excluding the *-man* and the independent variants. Their results from the *Penn Corpora of Historical English* suggest two important diachronic developments. Firstly, by 1700 the forms in *-body* were starting to develop “indications of a vernacular variant associated with men and oral genres” (2013: 293). Secondly, after an increase, the share of *-body* started to decrease in edited prose

in the mid-nineteenth century, and the “more ubiquitous” *-one* became the more prestigious variant more closely associated with formal and literate genres. They suggest lexical conditioning as the underlying factor since the determiners *any* and *every*, which are more likely to appear in communication between close communicators, show greater tendencies for *-body*. The present study, using correspondence material accompanied with detailed socio-demographic information, puts both of these conclusions to the test.

As for Present-Day English, the variation between *-one/-body* is not completed, but the *-man* forms are lost except in proverbs and in archaic style. Cheshire (2013) however suggests that pronominal *man* has emerged in young urban dialects in London. As for textual and regional variation, the forms in *-one* are typically associated with more literary texts, and they are said to be associated more with British English, whereas *-body*, with its more casual overtones, tends to be used more in spoken uses and in American English (Quirk et al. 1985: 378; Biber et al. 1999: 352–353; D’Arcy et al. 2013: 296–298). However, Svartvik & Lindquist (1997) find no convincing evidence of distinctions between British and American English in the late twentieth century. In addition, Bolinger (1976) suggests that *-one* is more frequent in American English, and the forms in *-body* tend to be used more in the Scottish dialect(s) than in the dialects in England. He also presents evidence of functional differences between the variant forms (supported by Svartvik & Lindquist 1997; Laitinen 2007a).

This study presents evidence of sociolinguistic variability of four morphological variant forms. The variable in the long eighteenth century consists of the forms in *-body* (9.1), *-one* indefinites (9.2)–(9.3), those in *-man* (9.3), and the independent forms (9.4).

- (9.1) According to promise I send you Mr. Frankland’s further account of what I proposed. You see how it is, and so can judge of the matter as well as **any body**. I’ll say nothing anyway. (William Steer, 1710; BANKS, 5)
- (9.2) and as to the young gentleman I do not know any objection that can be made against him but that we had up at Ferrybridge, and you seemed to be satisfied in it, and though he may have used some freedom that way formerly, yet I hear of late that his conduct has been such that **no one** need be uneasy on that score. (William Steer, 1711; BANKS, 15)
- (9.3) However if you know your own intrest take no notice of this letter to **anyone**, [this ... anyone, INSERTED] & if he offers to pay you for the 2 cutts done tell him you have sent me your specimens & fear I should take it ill if you should take [it. You may depend on it, no more shall come into **any mans** hands but your own, DELETED] (Francis Blomefield, 1736; BLOMEFIELD, 126)

- (9.4) I have taken No money upon the botem off Mr. Sheldon, nor **any other**, save 1000 Rupees off Sir Edward Littleton, which Ingageth or Enlargeth the Ships Cargo.  
(Joseph Tolson, 1707; BOWREY, 291)

To ensure comparability with the previous studies by Raumolin-Brunberg (1994) and Raumolin-Brunberg & Kahlas-Tarkka (1997), the results exclude partitive structures, which are possible with *-one* and the independent indefinites (*anyone of them*, *none of them*), but not with *-body* (*\*anybody of them*) and *-man* (*\*no man of them*). In addition, only singular generic human references are included. Deciding what counts as a human generic reference occasionally involves high degree of subjective interpretation, in particular with the disappearing *-man*, which, at the final stages of change, can be used to refer to a male human being or human beings irrespective of the referent's gender. A similar difficulty arises in deciding generic references, and, therefore, great effort has been made to exclude specific and known references. The criteria for genericness follow the definitions of the concept developed in Raumolin-Brunberg & Kahlas-Tarkka (1997) and Laitinen (2007b: 128–132).

On the basis of indirect evidence from eighteenth century grammars, the changes in the indefinites seem to have taken place below the level of linguistic awareness. Sundby et al.'s (1991) examination of 187 grammars, spelling books, manuals and dictionaries shows that the variation between the nominal elements is only sporadically commented upon in grammars. They are mentioned only in the concord error category, by and large concerned with the relationships between an antecedent and co-referential pronouns, heads and modifiers in noun phrases, and subject-verb agreement (1991: 19–25, 103–104). If one uses the illustrations as a diagnostic tool for exploring variation, most forms appear in *-one*, as in (9.5).

- (9.5) “Let **every one** mind their own concern.”  
(from Richard Postletwhite's late 18th century grammar,  
quoted in Sundby et al. 1991: 107)

The fact that the indefinites in *-one* dominate is hardly surprising since most of the authentic illustrations in the eighteenth century grammars cited in Sundby et al. (1991) are from literate texts. There are also occasional independent forms such as in (9.6) among the illustrations, but the forms in *-body* are infrequent (9.7):

- (9.6) “Let **each** esteem other better than themselves”  
(James Buchanan's grammar from 1760s, quoted in Sundby et al. 1991: 106)
- (9.7) “What is **every-bodies** business, is **nobodies**”  
(a late eighteenth century grammar by Peter Fogg,  
quoted in Sundby et al. 1991: 300)

The structure of this chapter is such that Section 9.2 presents a diachronic overview of the previous results, drawn from the *Corpus of Early English Correspondence* (CEEC-1998), and shows how the trajectories of change continue in the long eighteenth century. The evidence shows that the changes consist of two overlapping processes, the first involving the increase of *-one* and *-body*, and the second consisting of the decrease and subsequent loss of two forms (*-man/the independent*), thus presenting an interesting setting for sociolinguistic analyses. Section 9.3 illustrates the sociolinguistic embedding of the variant forms and contributes to understanding the stylistic and regional variability of indefinite pronouns in Present-day English. Section 9.4 discusses the results and examines to what extent the observed results could be used to understand linguistic conservatism. Conservatism is, according to Chambers (2009: 215–217), typically associated with age, as older people tend to participate in ongoing change less often than the younger members in the society. In addition, previous studies in historical sociolinguistics have shown that upwardly mobile people, typically male, tend to avoid using new linguistic forms until their social values become established (Nevalainen & Raumolin-Brunberg 2003: 152–153). Section 9.5 concludes the chapter.

## 9.2 Diachronic overview

This section combines the results presented in Nevalainen & Raumolin-Brunberg (2003: Figure 4.16) with those in CEECE. Their results show that by the late seventeenth century the independent forms were the dominant variant but had decreased sharply since the turn of the century. Of the three compound forms, two (*-body* and *-one*) were on the increase, and the chronologically earlier form *-body* had a greater share. The forms in *-man* had decreased steadily since Late Middle English and were in decline. Table 9.1 shows the variant proportions in correspondence material and the directions of change at the end of the 17th century.

**Table 9.1** The indefinite pronouns at the end of the 17th century (Nevalainen & Raumolin-Brunberg 2003: Figure 4.16)

Variant	Proportion	Direction	Description
Independent	37%		Steep decrease
-BODY	26%		Vigorous growth
-ONE	21%		Vigorous growth
-MAN	16%		Decrease

Figure 9.1 below visualizes the changes and combines the results from CEEC-1998 with the CEECE (for the total CEECE figures see the Appendix). The recessive *-man* forms are marginalized gradually throughout the four centuries depicted here, and the change is completed by the early 18th century (i.e. the variant proportions < 15%); the forms in *-body* are the competing variants for *-man*. The independent forms first increase towards the 17th century but start losing ground as the forms in *-one* increase their share.

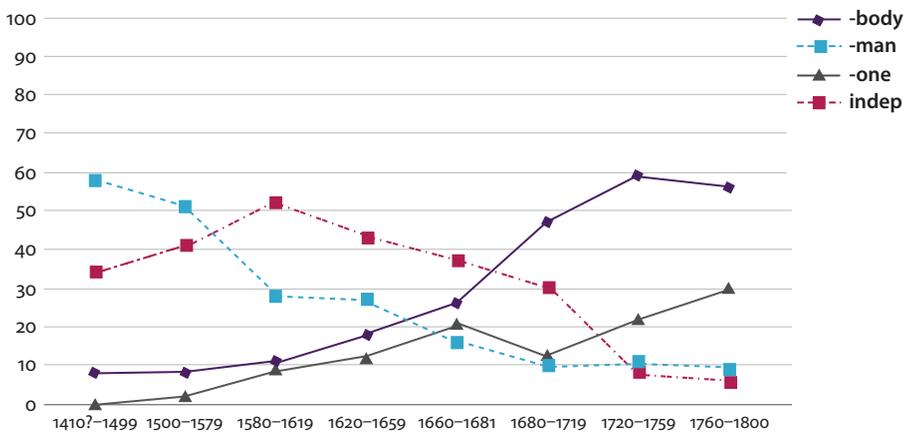


Figure 9.1 The percentages of the variants in the indefinite paradigm in CEEC-400

In addition, there are three noteworthy developments, all of which could be complemented by sociolinguistic information (see Section 9.3 below). Firstly, the decline of the independent forms is rapid and is completed roughly within a century. Secondly, the 17th century increase of *-one* is stalled for a few decades in the early 18th century because of the rapid increase of *-body*. Thirdly, once the *-one* forms start increasing in the early 18th century, there is competition with *-body* whose spread is halted in the mid-18th century. This incipient decline of *-body* in correspondence data takes place a century earlier than in the more formal literary texts studied in D’Arcy et al. (2013).

As for language internal developments, these results are illustrative of semantic bleaching and grammaticalization (cf. Raumolin-Brunberg & Kahlas-Tarkka 1997: 69–77). The two increasing variants (i.e. *-one* and *-body*) can be traced back to meaning a singular human being, but since the semantic weight of singularity is less transparent in *-body* than in *-one* forms, the former enters the grammaticalization path earlier. This development is seen in Figure 9.1 as the vigorous increase of *-body* slows the increase in *-one*. Even though the *-one* indefinites have a long history, dating back to the 13th century, it is only in the mid-eighteenth century when they lose enough semantic weight and become grammaticalized.

Before moving on to social embedding, I will make use of the possibilities in our material and examine the changes by looking at our informants. Figure 9.2 illustrates the share of individuals who have completed the change in their lect (Nevalainen & Raumolin-Brunberg 2003: 55), i.e. those whose total proportion of the incoming variant forms (*-one/-body*) exceeds 85%. It shows that at the turn of the century only 35% (27/77) used the *-one* and *-body* variants predominantly, but the proportion rises to nearly two-thirds by mid-18th century and remains at that level towards the end of the century.

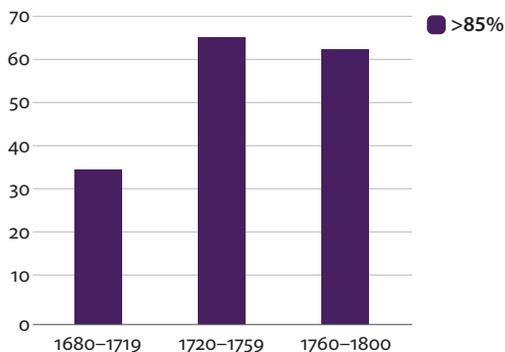


Figure 9.2 Individuals whose total share of *-one* and *-body* variants is > 85%

In the first 40-year period, nearly two-thirds of the informants still use the recessive variants as part of their repertoire. However, a great majority of them also resort to the other variants, and only ten writers (13%) use them exclusively. One of them is Henry Fleming, a clergyman, whose letter to his father Sir Daniel Fleming is shown in (9.8).

(9.8) ...a man reputed to be very sober and civil, exceeding industrious, and careful of his Schollars, and I question not but he'l be as careful of him as of any.

(Henry Fleming, 1688; FLEMING 2, 219)

In the next two periods (1720–1759 and 1760–1800), the number of informants resorting to the recessive independent forms drops considerably.

As shown in Figure 9.1 above, *-man* had lost ground earlier than the independent forms, which is also seen in the fact that there are only five individuals who resort to the *-man* forms solely (see Section 9.4 below); one of the late ones is Richard Champion, a Bristol merchant (born 1743) who keenly reports on the events in the North American colonies in the 1770s.

- (9.9) And however wanting any of the Americans may have been in support of their own Cause, **Every Man** who appeared in their defence here, was supporting in their persons the Liberties of the whole Empire

(Richard Champion, 1775; CHAMPION, 46)

As for the increasing variants, *-body* is used by the great majority of our informants throughout the century (nearly three out of four informants use it in their lect). The share of letter writers for whom it is the sole variant remains constant (circa 25%) for the first two time periods (1680–1719 and 1720–1759). One of them is Anne Stuart (Queen Anne), whose letter to Frances Apsley is shown in (9.10):

- (9.10) ye reason I will tell you when I writt next as for what you desire to know about Sir Benjamin **no body** has ever said anything against him to me

(Anne Stuart, 1683; ROYAL 4, 162)

This proportion drops to 19% in the third period (1760–1800) brought about by the rise of *-one*. The share of the individual writers for whom *-one* is an option in the CEECE material rises from 34% (26/77) in 1680–1719 to 62% (63/103) in 1760–1800.

The number of informants for whom *-one* is the only indefinite pronoun is very low in the early parts of the long eighteenth century (2/77). One of them is Sarah Pinney, a teenaged daughter of a clergyman who primarily uses the recessive independent variants. The illustration in (9.11) is an extract of a letter to her father John Pinney.

- (9.11) I hope thofe I have done evely in one thine you will not excemt me fro~ bing achild as wille as my brother ho never ofended you before to my noleg nor wronged you of one farden that I no of thofe I have bine moch wrangd & abused you will no I woos never woos thin my woord to **any one** in my Life moch more woold i not to you hoping I may find others so to me thof

(Sarah Pinney, 1686; PINNEY, 37)

This share of individuals using *-one* solely rises to 17% in the second period, and eventually levels to circa 10% in 1760–1800 when the two variant forms even out. The next section turns the attention to the social embedding of the change.

### 9.3 Social embedding

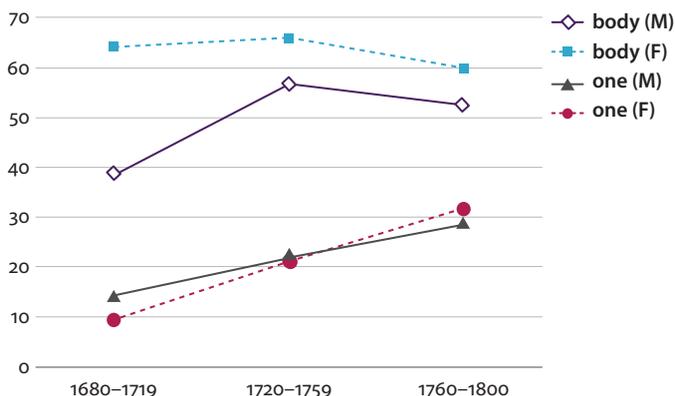
The preceding section shows considerable changes in the long eighteenth century. There is a cross-over from the independent forms to *-body* in the late seventeenth, but *-body* starts declining by the mid-century. As pointed out in the introduction

(Section 9.1), the incipient stages of this decline take place a century earlier in correspondence than in the literate texts investigated in D’Arcy et al. (2013), and it constitutes a process whereby *-body* indefinites become stylistically marked as more informal and casual, associated with spoken genres. The forms in *-one* remain minor variants until the early eighteenth century once the decline of *-man* to a minor variant is completed.

In Nevalainen & Raumolin-Brunberg’s (2003) description of the variable in Early Modern English, the change towards *-one* and *-body* is led by women, whereas men continue to use the decreasing *-man* forms later than women. Their results, however, show no significant correlations with writers’ social status, but they however suggest some regional variation (see Section 9.3.3 below).

### 9.3.1 Gender

Figure 9.3 illustrates the proportions of the two incoming variants correlated with writer’s gender. Showing the proportions in 40-year periods, it illustrates that women are ahead of men in the use of *-body*, as was the case in the CEEC-1998 results. The differences are statistically significant in the first (chi square value 82.6,  $p < .001$ ) and the third period (15.3,  $p < .001$ ), but not in the second. In the increase of *-one*, there are no considerable differences in the first two periods, and it is only in the third period (1760–1800) when statistical testing suggests that women use this form more than men (7.4,  $p < .001$ ), but the bootstrap estimates presented below shed additional light on the issue.



**Figure 9.3** The percentages of *-body* and *-one* over time for male and female writers ( $n = 1,334$ ; for each gender and variant, the value shown is the percentage of the variant out of all indefinites used by writers of that gender)

These results and the bootstrapping evidence in Figure 9.4 offer a socio-cultural context for understanding the stylistic developments associated with the *-body* indefinites, which are suggested by D'Arcy et al. (2013). The evidence here does not however corroborate their observations that the forms in *-body* would have, at this early stage, been associated with men. On the contrary, the results indicate that the change towards *-body* was promoted by women, which is also confirmed below. In the bootstrapping estimate in Figure 9.4, the frequency of *-body* in the male sample is well below 0.5 in 1680–1719, and it increases to over 0.5 in the later periods. The estimate for women is higher. In addition, the confidence intervals show that the eighteenth-century male evidence differs considerably from the previous century, and the intervals overlap only after the early decades of the century as men catch up on women in the use of *-one*.

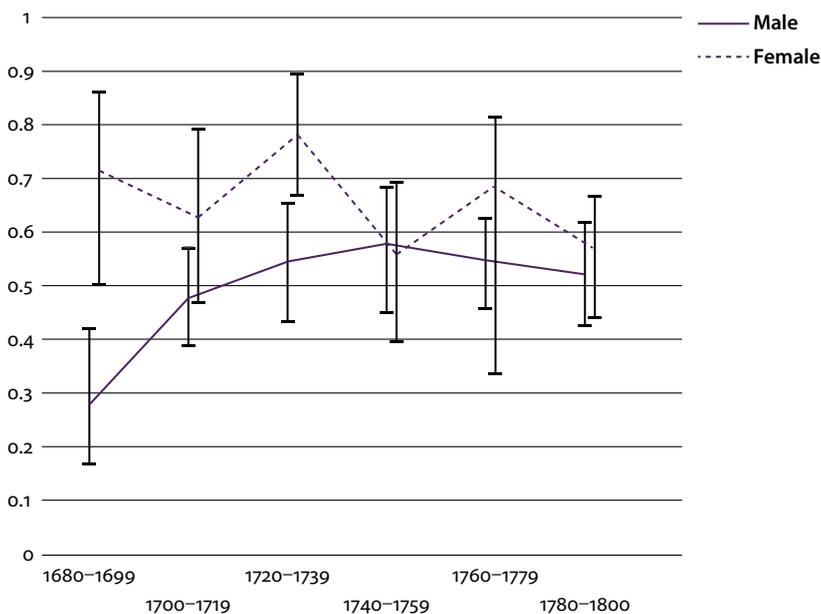


Figure 9.4 Bootstrapping estimates of the frequency of *-body* in 20-year periods

As for *-one*, the gender stratification is less considerable. The correlations in Figure 9.3 suggest that the early development towards *-one* indefinites in the first half of the century is promoted by men and by women once the change progresses. Yet, the bootstrapping estimates in Figure 9.5 indicate that there are no significant differences between men and women, as the confidence intervals overlap for all the 20-year periods. The estimation method used here illustrates that there is clearly less uncertainty in the male sample as the confidence intervals in second half of the century differ from the first part for men, but not for women.

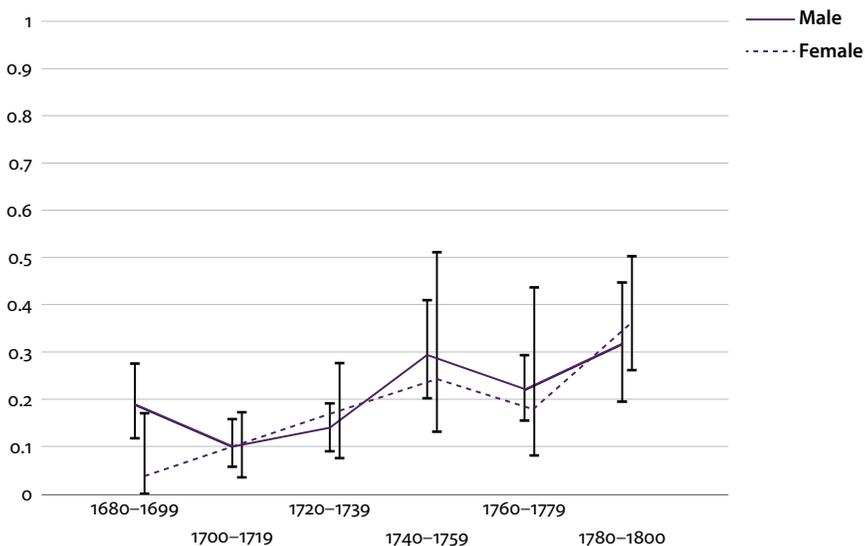


Figure 9.5 Bootstrapping estimates of the frequency of *-one* in 20-year periods

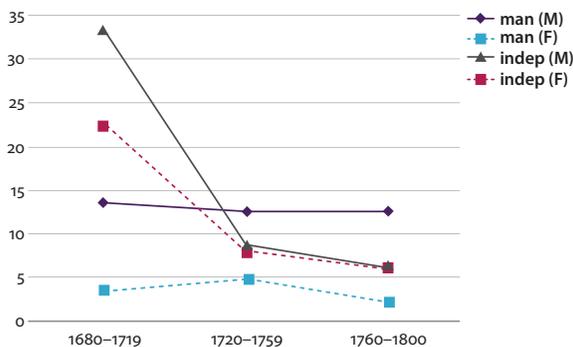


Figure 9.6 The percentages of *-man* and the independent forms over time for male and female writers ( $n = 393$ ; for each gender and variant, the value shown is the percentage of the variant out of all indefinites used by writers of that gender)

As for the recessive variants (i.e. the forms in *-man* and the independent indefinites), Figure 9.6 illustrates two developments. On the one hand, the proportions of *-man* had already decreased to the point that the change was as good as completed by late seventeenth century. By then it was primarily used by men, and women resorted to extremely few *-man* indefinites. These results further confirm Nevalainen & Raumolin-Brunberg's (2003) observation that the direction of the change away from *-man* was from below. However, as a sociolinguistic variable, the compound indefinite pronouns are complex, since one of the variant forms (*-man*) is semantically ambiguous, and the variable may have been above the level of linguistic

awareness for at least some writers. It is fair to assume that this semantic ambiguity contributed to the loss of the indefinites in *-man*, which is illustrated by the fact that the few *-man* indefinites from women are borderline cases between generic references and gendered ones at the early eighteenth century, whereas similar ambiguities only arise in men's data towards the end of the time frame here. This is the case in (9.12) which illustrates an extract from Anne Clavering's (aged 22) letter to her uncle, Sir James Clavering. The topic here deals with local events and even though the writer mentions several men in the preceding context and the masculine anaphoric reflexive pronoun, the sentence with *any man* is a generic reference since it draws from a reference to an indefinite (unknown) set of objects among which one (or more) objects could be potential referents. The illustration has therefore been interpreted as relevant.

(9.12) I have charity enough yet to look upon that as malice, but at the same time the ground of this quarrell must be laid at his door. How can **any man** have so mean a soul as to make himself a meer tool to the high-Church?

(Anne Clavering, 1708; CLAVERING, 8)

On the other hand, women lead the decrease of the independent forms at the turn of the century, but not at statistically significant levels. Once the decrease of the variant forms is completed, the proportions between men and women even out.

### 9.3.2 Age and social status

This section presents evidence of how other social variables correlate with the changes. The results are firstly presented using the informants' birth years, which are known for nearly 90% of the informants (235/268 individuals throughout the period). The informants are divided into groups are: (1) those born before 1639, i.e. those who had reached the age of 40 by the 1680s, (2) those born between 1640 and 1679, (3) 1680–1719, (4) 1720–1759 and (5) those born after 1760.<sup>1</sup>

In Figure 9.7, the dominant variants for group (1) are the independent forms and *-body*. The proportion of the independent forms declines heavily, and this decrease is made up by the increase of *-body* only. The proportion of *-body* peaks in age group (3), and it takes place before the forms in *-one* enter a period of vigorous growth (age groups (3)–(5)). This increase in *-one* is even more considerable than that of *-body* a few decades earlier, as the share doubles from circa 20% to 42%. These results confirm the language internal development whereby the decrease in the independent forms left a gap that was then filled with the increase in *-one*.

---

1. According to Wrigley (2004: 436) the expected life span at birth in the late seventeenth century was roughly 35 years, gradually rising to nearly 40 by the end of the next century.

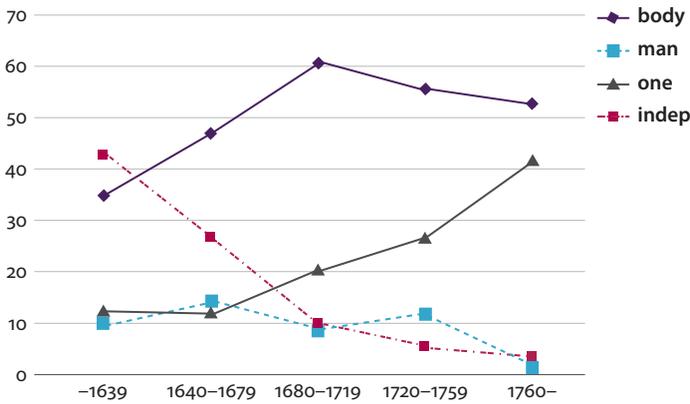


Figure 9.7 The percentages of the variants per age group

Figure 9.8 below, showing the bootstrapping estimates for the frequency of *-one*, confirms that its significant increase is brought about by age groups born after the early eighteenth century.

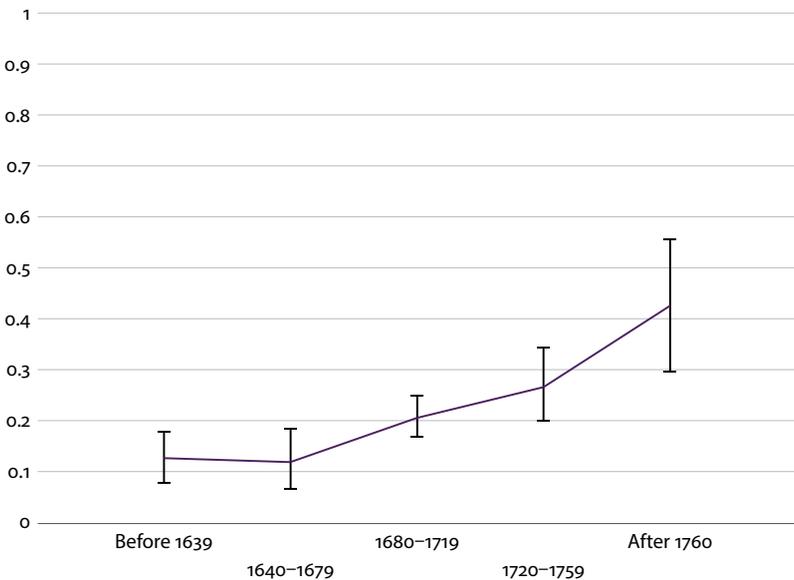


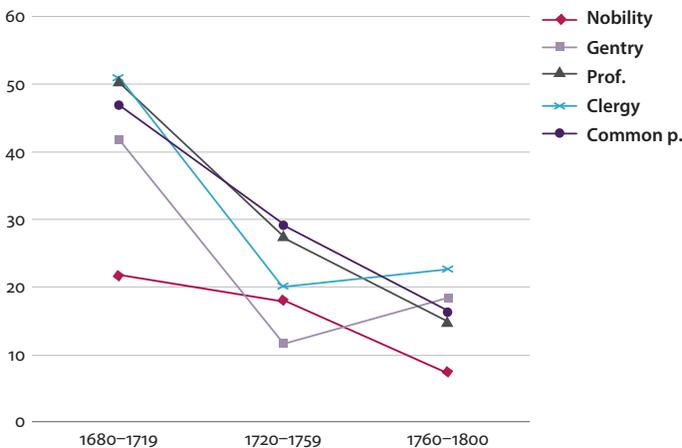
Figure 9.8 Bootstrapping estimates of the frequency of *-one* for the age groups

As for the youngest age group (5), the compound indefinite pronoun paradigms are strikingly different from age group (1), and they use the forms in *-body* and *-one*.

I will next move on to investigate the writers' social status, and the aim is to complement the CEEC-1998 results in which no correlation was found between social status and the variant forms. The larger size of the CEECE should offer insights

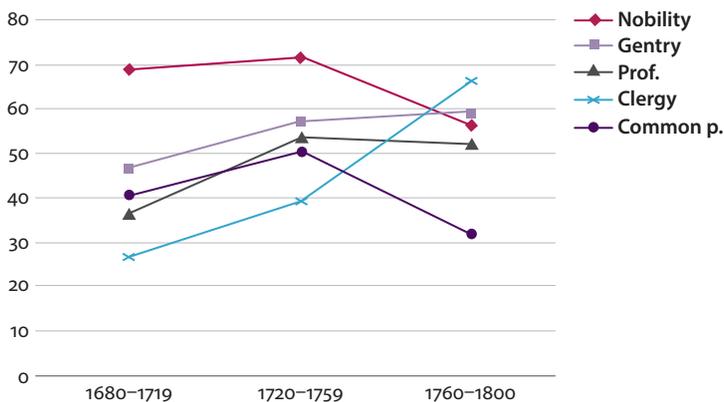
of social stratification, which, together with the results presented in two previous sections, could enable understanding the direction of the change better.

The results are divided into five social categories (cf., Chapters 2 and 4 of this volume): (a) nobility, consisting of the top layers, *viz.* the royals and the nobility, (b) land-owning gentry, both upper and lower, (c) professionals, (d) upper and lower clergy and finally (e) the common people of merchants and other groups (manual laborers, paupers, etc.). Figure 9.9 illustrates the aggregate proportions of the recessive variants for the social groups. It shows that the process is led by the highest social layer. For instance, in the first time period, the only group nearing completion was nobility (with proportions nearing < 20%), and the other four groups lag behind. After the rapid decline in 1720–1759, the change is completed by two social groups, nobility and gentry, and professionals and common people trail behind, but at statistically significant levels only between gentry and professionals (5.737,  $p < .05$ ). Finally, in the last period, the change is completed by all groups except clergy who stand out slightly from the other groups, which will be discussed below. This result complements the developments described in Nevalainen & Raumolin-Brunberg (2003) and suggests that the changes are led by women (as seen above) and by the highest social layers.



**Figure 9.9** The loss of the independent and *-man* indefinites by social status (for each social group, the value shown is the percentage of the independent and *-man* indefinites out of all indefinites used by that group)

There is considerable variation linked with the informants' social status, and I will therefore deal with three of the variants separately. In the incoming *-body* and *-one*, the spread of *-body* is socially conditioned in the first two periods, only leveling towards the end of the century, whereas there is less social variation in *-one*. Figure 9.10 shows that the forms in *-body* are first adopted by the nobility,



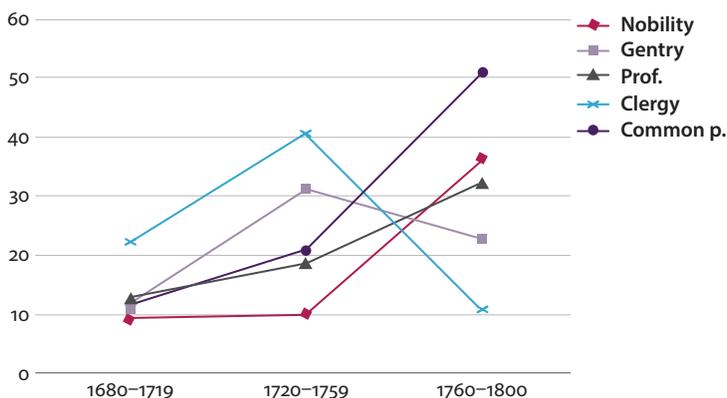
**Figure 9.10** Social variation in *-body* over time (for each social group, the value shown is the percentage of *-body* out of all indefinites used by that group)

and these results together with those presented in the sections above point to the fact that the change towards *-body* was driven by women and by the highest social layers.

The next largest shares for *-body* are found in the gentry, professionals, and common people. The informants in the clergy are the last to adopt the forms in *-body* in the second half of the century, when its total share had already started to decrease (cf. Figure 9.1 above). There are three groups whose share of *-body* increases consistently throughout the long eighteenth century; they are the gentry, professionals and clergy who seem not to be the driving forces behind the rise of *-one* which is an indication of emerging stylistic differences between *-one/-body*. As said, the shares are highest for the nobility, most likely because they adopt the *-one* forms later than the others (see Figure 9.11 below). For the nobility, the share of *-body* decreases for the third period. This decrease could be interpreted as the first stage in the process whereby *-body* develops into a more “vernacular variant” (cf. D’Arcy et al. 2013: 293). This process is initiated by the highest social layer.

The largest increase of *-body* takes place among the clergy. As for the common people, their absolute frequencies are considerably lower, which makes generalizations difficult. These two changes are clearly not driven by them, and the higher share for *-one* in Figure 9.11 is brought about the high variant frequencies in one letter collection (*viz.* the Clift collection in which Elizabeth and Joan Clift account for nearly half of the instances for this social group in 1760–1800).

Figure 9.11 illustrates the social variation in the increase in *-one*. It is noteworthy that the proportions for the nobility remain low until the second half of the century, and the change is clearly led by other groups in the beginning. Figure 9.11 also shows that it is the clergy who are ahead of the other groups at statistically

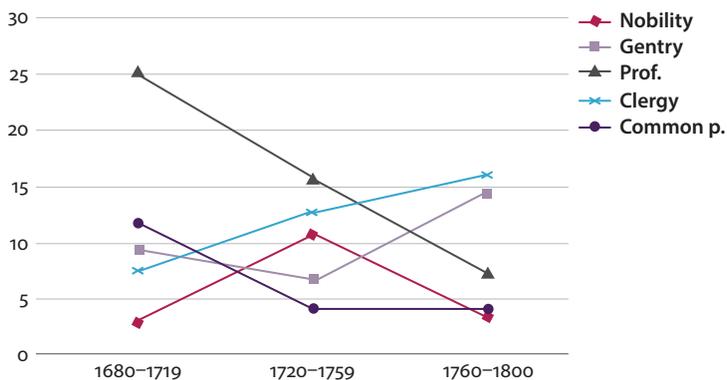


**Figure 9.11** Social variation in *-one* over time (for each social group, the value shown is the percentage of *-one* out of all indefinites used by that group)

significant levels in adopting *-one* in the first part of the century. One possible explanation for the prominent role for the clergy is the fact that the main variants in the *Authorized Version of the Bible* of 1611 are *-man* and *-one*, the latter being the main variant with the universal *every* indefinites. The concept of *body* tends to be associated with the body of Christ in Biblical texts, which might partly explain why clergy are slow in adopting it and resort to *-one*. Their share for the forms in *-one* drops in the third period, and they finally adopt the forms in *-body* when the other social groups show incipient decreases (as shown in Figure 9.10 above).

Altogether, four social groups show consistent increase in *-one* throughout the period. The only exception is the nobility who are late comers in adopting them. It is the strong presence of *-body* among the nobility that prevents them from adopting the variant form *-one* up to the mid-century. These results suggest that the spread of *-one*, once it enters the vigorous growth in the mid-century, is a change from above in which the driving force is the highest social layer (and men at the incipient stages of change as seen in Figure 9.3 above). This social stratification may be reflected in present-day distribution of *-one* and *-body*, as the forms in *-one* are typically associated with formal genres and written communication.

The absolute frequencies for *-man* are low, but they still allow sociolinguistic generalizations since there are three groups whose results show consistent developments. Figure 9.12 illustrates that the nobility is the first group to lose this variant form. The professionals clearly lag behind the others in the first two periods, but their results show consistent decline in the proportions. The same holds for the common people, but what is surprising is that the clergy show a slight increase in the uses for *-man*. This increase is consistent and goes against the long-term diachronic trend.



**Figure 9.12** Social variation in *-man* over time (for each social group, the value shown is the percentage of *-man* out of all indefinites used by that group)

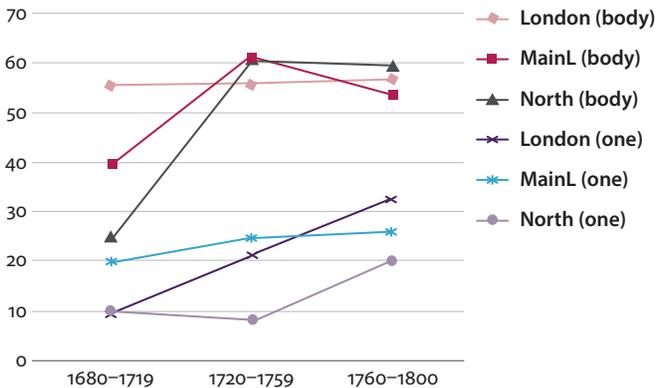
As pointed out above, the size of CEECE and its extensive social coverage also enable investigating how linguistic variables might be used to index social characteristics. One social characteristic related to the recessive variant forms is linguistic conservatism, which seems to be particularly valid for the forms in *-man*, because of its semantic load. If one uses the results in *-man* the two linguistically conservative groups in correspondence data are the professionals and the clergy. The professionals stand out from the other social groups and retain this recessive variant longer than the others. As for the clergy, the situation is more complex. On the one hand, they are conservative as they are the last group to adopt *-body* (Figure 9.10) and indeed seem to go against the trend in increasing their share of the recessive *-man*. On the other hand, they are the first to adopt *-one* forms. I will return this question in Section 9.4 below.

### 9.3.3 Region

This section looks into regional differences and is motivated for two reasons. Firstly, the regional differences between American and British English in Present-day English (Biber et al. 1999: 352–353; Bolinger 1976) could be attributed to diachronic variability. Secondly, Nevalainen & Raumolin-Brunberg (2003: 181) show that despite low variant frequencies, the forms in *-body* were clearly favored in London, the Royal Court and in East Anglia in the early seventeenth century, but were only occasionally (< 5%) used in the North, where *-one* dominates. In addition, the spread of *-one* to all regions took place much earlier than the spread of *-body*. The results are here presented in three regional categories that follow the division

in Nevalainen & Raumolin-Brunberg (2003), and I will use the names London (including the City, the Royal Court and East Anglia), Main (which consists of the Home Counties and other counties except the six Northern counties), and North (consisting of the counties of Cumberland, Durham, Lancashire, Northumberland, Westmorland and Yorkshire).

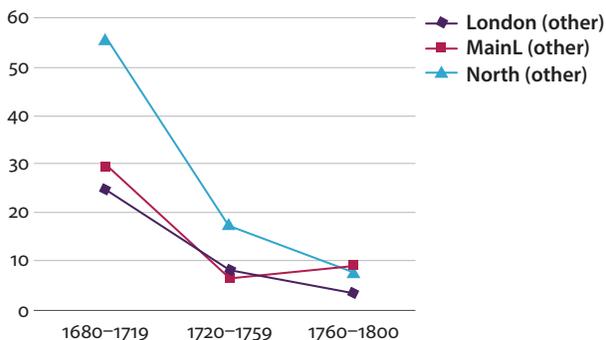
Figure 9.13 below shows the regional distribution of the two incoming forms. It confirms the tentative conclusion from Nevalainen & Raumolin-Brunberg's (2003: 181) Early Modern English results, as the *-body* forms in the first period are clearly associated with London.<sup>2</sup> The other areas lag behind; there is however a steep increase of *-body* in these two areas by the mid-century. The forms in *-one* are not associated with North in the long eighteenth century, and indeed the form is adopted there last. The same holds for London where these indefinites are picked up in use later than in Main, but once they start increasing in frequency, the increase is led by London.



**Figure 9.13** Regional variability of *-body* and *-one* over time (for each region and variant, the value shown is the percentage of the variant out of all indefinites used by the informants of that region)

Figure 9.14 depicts the decrease of the independent forms in the three regions, and it shows that the decrease is led in London followed by those in Main, whereas the informants in the North clearly retain this recessive variant latest.

2. The proportions remain similar even if East Anglia is excluded from the London/Court results.



**Figure 9.14** Regional variability of the independent indefinites over time (for each region, the value shown is the percentage of independent indefinites out of all indefinites used by the informants of that region)

The regional pattern is consistent in all the three areas. In London and in Mainland, the independent indefinites are the second-most frequent form in the first time period, and they decrease to a marginal variant form in a few decades. The same holds true for the North, but since the change takes place later, the pattern is considerably steeper.

#### 9.4 Discussion on the new evidence from correspondence

This section draws the results together, and by making use of roughly 400 years of correspondence data with extensive background information on the informants, it first focuses on broad patterns of variability after which it zooms into linguistic conservatism and individuals who prolong change.

The forms in *-one* start spreading in the early eighteenth century irrespective of the steep increase in *-body*, but rather as a result of the loss of the independent forms. As shown in Figure 9.1, *-one* becomes an alternative only in the eighteenth century, and its proportion in correspondence reaches roughly 30% by the end of the century, and this rise slows down the spread of *-body*, whose share starts to decline after mid-century.

As for the developments whereby the forms in *-body* become more informal variants and the related emergence of *-one* as more standard or prestige forms, D’Arcy et al. (2013: 293) suggest that there are signs of that development by 1700. Their argument is that “the *-body* forms were developing indications of a vernacular variant associated with men and oral genres” (2013: 293). The CEECE results only partially support this conclusion. Rather, the results show, on the one hand, that

the forms in *-body* are clearly more frequent in spoken-like written genre of correspondence than *-one* throughout the time period, which supports the observation on the role of oral genres. On the other hand, the results on sociolinguistic embedding do not support association of *-body* with men. The results clearly show that *-body* is more frequently used by women than by men in this study. Put together, *-body* develops into a more vernacular, informal variant in correspondence roughly a century prior to literate texts, and there exists no evidence of greater association of *-body* with men in CEECE.

As for social and regional variation, evidence of the early vernacular associations of *-body* remains scarce in correspondence data. The results illustrate that it is closely associated with the highest social layer, i.e. the nobility, up to the mid-century. Similarly, it is firmly established in London a few decades before the other areas, as was also the case in Nevalainen & Raumolin-Brunberg (2003) who observe that the informants from both London and the Royal Court prefer *-body* in the seventeenth century. When one correlates the variant form with writers' years of birth, the results in Figures 9.7–9.8 show that *-body* peaks in the material of those born after the turn of the seventeenth and eighteenth centuries. It starts to decline for the following generations, and *-body* and *-one* enter complementary trajectories.

Overall, the CEECE offers a more detailed picture of sociolinguistic embedding of indefinite pronouns than CEEC-1998 where very little social variation was found. For future studies it offers an interesting material not only for multivariate analyses to measure the impact of the social factors discussed but also for investigating the role of recipients in this change (cf. D'Arcy et al. 2013).

In addition, the social coverage of the CEECE enables an investigation of linguistic conservatism, i.e. zooming to individuals who slow down change by retaining the recessive form longer than the others. As pointed out in the introduction, previous studies associate conservatism with older men and with upwardly mobile men. The results presented here not only support this conclusion but also offer new perspectives to it. They confirm gender stratification as men continue using the forms in *-man* as a minor variant throughout the century, and they also use the independent forms in statistically higher frequencies than women (Figure 9.6), though the proportions even out as the change progresses.

Figure 9.15 below confirms the pattern by showing the results of cross-tabulations of the two recessive variants and the years of birth for both men and women. It illustrates the proportions of the individuals in whose personal letters the share of these forms is nearing completion.



**Figure 9.15** Year of birth and gender and the use of recessive variants (for each gender, the value shown is the percentage of recessive variants out of all indefinites used by writers of that gender)

The share of women resorting to the recessive variants decreases consistently throughout the time period, and the proportion of conservative men follows. At the end of the period, the shares are equal, and there is one male and one female individual who use these variants. Both belong to nobility: Ernest Hanover (b. 1771) has very few indefinites overall, and (9.13) shows his use of *-man* variant. It is an example of a disappearing variant, and it thus is a borderline case between a gendered reference and a generic one, but has been interpreted as a generic one since he otherwise resorts to *-one* forms, much like the other Hanovers in the CEECE.

(9.13) Adolphus will also write & thank you for your kind present. **No man** ever desired more to return to his family, to his country than I do; it is almost six years that I have not seen a single one out of England except Frederic.

(Ernest Hanover, 1792; GEORGE 4, II, 236)

(9.14) Ly. Worcester was at Devonshire House the night before last looking in great beauty, though I doubt not you have numbers of correspondents who write all that is worth hearing, yet as you can have **none** who would be happier to contribute in any way to your Ladyship or Ld. Stafford's amusement

(Henrietta Ponsonby, 1797; GOWER, 148)

The illustration in (9.14) is from Henrietta Ponsonby (b. 1761). It is noteworthy that the informants in these two illustrations are from writers who are not writing at their old age, but are in their early twenties (Ernest Hanover) or mid-thirties (Henrietta Ponsonby).

I will lastly explore conservatism by extracting individuals in whose materials the share of the recessive variants exceeds 50%.<sup>3</sup> For the variant form *-man*, whose change was completed by the late seventeenth century, I extracted such conservative informants from the entire time period (1680–1800), and for the independent forms, individuals are accounted for only after the turn of the eighteenth century. Overall, there are 19 such individuals out of the total of 268. The great majority of them, 89%, are men (17/19), compared to the overall share of the informants (72%), and no female informants are conservative on the basis of their use of *-man*, but some resort to the independent forms. A few conservative individuals (c. 11%) rose socially, but otherwise socially ascendant individuals are not overrepresented in this group.<sup>4</sup>

The Reverend F. Blomefield (b. 1705) is an interesting informant overall, as he uses *-man* and the incoming *-one* in his letters in the mid-1730s, as shown in (9.15) and (9.16).

(9.15) I have learnd wisdom enough for the future never to part with any drawings to **any man** living but those I design shall do them

(Francis Blomefield, 1736; BLOMEFIELD, 126)

(9.16) I hope you will serve him with such materials as he wants as cheap as **anyone** and have so requested him to come up and furnish himself

(Francis Blomefield, 1735; BLOMEFIELD, 100)

In his later letters in the late 1740 and early 1750s, he turns to *-body*, which overall was the most frequent variant reaching its peak (Figure 9.1).

(9.17) My service to Mr Masters, but I do not know any thing I have of his by me. I should be very glad to see you here this summer, for **nobody** would be more welcome than your self, to your most obliged friend & servant

(Francis Blomefield, 1749; BLOMEFIELD, 250)

There are clearly older informants, i.e. those over the age of 50, such as Isabella Wentworth (b. 1626), amongst the conservative individuals, but there are also informants writing at a relatively young age. The last illustration comes from William Jones, a philologist and a polyglot, who in (9.18) comments on the performance of his young student.

---

3. In addition, the individuals were required to have at least 5 instances of the indefinites in their CEECE materials.

4. Out of the 315 informants in CEECE, 56 (c. 18%) were upwardly mobile individuals.

- (9.18) I cannot omit saying how glad I am that Lord Althorp is put under the *private* tuition of Dr. Sumner, with whom he will learn more than with me, or **any man** in Europe  
 (William Jones, 1770; JONES, I, 73)

Overall, linguistic conservatism is an understudied phenomenon that needs to be explored in more detail. It is clear that materials such as CEECE offer interesting materials for examining it.

## 9.5 Conclusions

This chapter has investigated the sociolinguistic variability of indefinite pronouns with human singular reference in the long eighteenth century. The study has illustrated how two variants (*-body* and *-one*) emerged from the Early Modern English indefinite pronoun paradigms by replacing the forms in *-man* and the independent forms. It has shown that stratified corpora, such as the CEECE, enable a detailed overview of the sociolinguistic embedding of these processes. The processes that have been established consist of two overlapping changes: In the first one, the forms in *-body* replace the *-man* indefinites, and in the second one the independent forms first increase in frequency towards the 17th century but lose ground after the forms in *-one* start increasing.

These results have fine-tuned the recent findings presented in D'Arcy et al. (2013) and have shown that the stylistic differences between the two incoming variants, *-one* and *-body*, started to develop in correspondence data a century prior to more formal texts. The results also show that the emergences of both *-one* and *-body* were promoted by the highest social layers and women also played an active role in adopting them before men. This chapter has also complemented the previous knowledge on linguistic conservatism and has shown that it is men who prolong linguistic changes, thus confirming the conclusions from previous studies.

## Appendix

**Table 9.2** The total absolute figures for the four variant forms in CEECE

	-BODY	-MAN	-ONE	Indep.	Total
1680–1699	59	16	29	66	170
1700–1719	157	32	30	71	290
1720–1739	133	34	33	22	222
1740–1759	157	19	77	20	273
1760–1779	145	24	50	25	244
1780–1800	285	42	179	22	528
Total	936	167	398	226	1,727

## Ongoing change

### The diffusion of the third-person neuter possessive *its*

Minna Palander-Collin

#### 10.1 Introduction

*Its* has been available as the third-person neuter possessive singular determiner at least from the beginning of the seventeenth century, and by the 1650s, *its* had gained the dominant position (Nevalainen & Raumolin-Brunberg 1994b: 176). This study shows that the use of *its* further increased in the eighteenth century, reaching c. 80% of the paradigm towards the end of the century. *Of it* is the other main variant during the time period as shown in (10.1):

- (10.1) I promis'd you an account of Sherborne, before I had seen it, or knew what I undertook. I imagin'd it to be one of those fine old Seats of which there are Numbers scatter'd over England. But this is so peculiar and **its Situation** of so uncommon a kind, that it merits a more particular description. The House is in the form of an H. **The body of it**, which was built by Sir Walter Rawleigh, consists of four Stories, with four six-angled Towers at the ends.

(Alexander Pope, 1717?; POPE, II, 236)

The diffusion of *its* is an example of a change that has now been going on for four centuries and is not fully completed even today. Figure 13.7 suggests that in Present-day English *its* (10.2) is far more frequent, but *of it* (10.3) is still used.

- (10.2) Deputy Secretary of Defense Paul Wolfowitz stated that the Agreed Framework “does not solve the North Korean nuclear problem” but “simply postpones that problem and may, in the process, make **its solution** ultimately more difficult.”  
(COCA, 2004)
- (10.3) New Delhi has felt that it is largely neglected by major powers, who tend to see India as a problem of nonproliferation and not a solution **of it**. (COCA, 2000)

In this chapter, I shall study the eighteenth-century spread of *its* using both real- and apparent-time data. The analysis of the social embedding of the change focuses on the major social factors explored in this book, including social rank and gender. In

addition, the analysis explores the extent of regional variation during the period and finally focuses on individual writers as “conservative” or “progressive” language users. Although grammatical factors are often emphasized in the connection of possessives, this chapter focuses on the impact of social factors. The importance of grammatical factors is likely to diminish during the eighteenth century as the increasing form, *its*, must have simultaneously acquired a broad grammatical scope. However, in the analysis of one individual writer, the results show that his “conservative” tendencies can be connected to the grammatical distribution of *its* vs. *of it* in his letters.

One of the methodological problems with this study was the scarce occurrence of the third-person neuter possessives, which may have skewed social rank and gender analyses. In comparison to some other features studied in this book, the third-person neuter possessives are relatively infrequent overall (N = 1435), and they seem to be grammatically marginal as they were only found in the letters of roughly half of the writers in the corpus. It seems that thousands of running words are usually needed from a writer in order to capture him or her using *its* or other variant forms. We can, nevertheless, argue that the spread of *its* proceeded both as a generational and a communal change and that during different phases of the change the extent of social variation differed. In this case, the period of rapid change in the latter half of the eighteenth century, which took place after a lengthy stable period, is characterized by significant social variation.

## 10.2 The third-person neuter possessive singular paradigm

The first citation of *its* in the *Oxford English Dictionary* dates to 1598 when the word was used in a definition of John Florio’s Italian-English dictionary. The first instances of *its* in the *Helsinki Corpus* were found in a handbook (1615) and a travel book (1630) (Nevalainen & Raumolin-Brunberg 1994b: 195). The dictionary posits that the form originated in colloquial language in the South of England, around London and Oxford, towards the end of the sixteenth century. At the beginning of the seventeenth century the paradigm still consisted of several variant forms including *his*, *it*, *the*, *thereof*, *of it* and *its* (for a detailed account of paradigm developments, see Nevalainen & Raumolin-Brunberg 1994b: 172–179; see also Grund 2011 for *he* and *it* in inanimate use in Late Middle English). *His* and *it* disappeared first from standard usage during the early decades of the seventeenth century. Other forms continued to coexist, but *its* and *of it* soon became the main variants. In present-day usage *the* can still indicate a neuter possessive relationship when the speaker avoids committing himself either to personal or nonpersonal pronominal use, for example, when referring to animals, but this usage falls outside the scope of

this study. The introduction of *its* can be related to the loss of grammatical gender in English, through which the association of *his* with neuter became more difficult and the paradigm was left with a gap.

We can find evidence to support the colloquial origins of *its* in the *Helsinki Corpus* data (1500–1710) where *of it* was first preferred in the oral genres instead of *thereof* and *of the same*, which were typical of legal language. *His* was only marginally attested in oral genres. After the introduction of *its*, *of it* began to diminish in oral genres and *its* clearly increased (Nevalainen & Raumolin-Brunberg 1994b: 198). In absolute numbers, the third-person neuter possessives are infrequent in the Early Modern English part of the *Helsinki Corpus* (N = 397 in the prefinal version of the corpus) and about 20% of all the instances are found in oral genres (Nevalainen & Raumolin-Brunberg 1994b: 197). This scarcity of neuter possessives in oral genres in the *Helsinki Corpus* is not entirely surprising since these grammatical forms occur somewhat sporadically also in personal letters.

Although linguistic factors are not comprehensively covered in this study, their role cannot be ignored as a factor in the development of the paradigm. For example, the reflexive use where the possessive refers to the subject of the clause seems to have favoured prenominal variants as in Example (10.4), whereas postnominal variants were rarely used reflexively (Nevalainen & Raumolin-Brunberg 1994b: 191–192). However, the association of prenominal variants *his*, *it* and *its* with reflexive use seems to have diminished from 1500 to 1710 (Nevalainen & Raumolin-Brunberg 1994b: 192). This type of development and the generalization of grammatical contexts of *its* would be expected with the relative increase of the word.

- (10.4) A Breakfast at Hillington will take **its** place Tuesday Morning instead, for I shall take my nights Lodging Monday at Missenden in favour of my saddle Horses.  
(Hester Grenville, 1754?; PITT 2, F28R)

The strong association of reflexive use with prenominal possessives seems to suggest that the pre-head alternative was to be expected even after the demise of *his*. Nevalainen & Raumolin-Brunberg (1994b: 192) suggest that the reflexive use may have served as a route for *its* to enter the language although the form was extensively used also in nonreflexive contexts. Examples (10.5) and (10.6) below illustrate *its* in nonreflexive use.

The semantic role of the neuter possessive is another factor that seems to have favoured either post- or pre-head forms to some extent. The semantic role of Possessor in particular favoured prenominal *his*, *it* and *its*, whereas the role of an affected or effected Object typically occurred with postnominal *thereof*, *of it* and *of the same* (Nevalainen & Raumolin-Brunberg 1994b: 193–194). These two semantic roles are illustrated with the form *its* in (10.5) and (10.6) respectively.

- (10.5) I pass'd by Boughton and show'd itt to Mr. Lewis and described **itt's** beauties, and he is much charm'd with itt. (Allen Bathurst, 1719; WENTWORTH 2, 447)
- (10.6) We have some here who begin already to talk of a new war, but surely they must either be very ignorant of the condition of their Country or else maliciously bent upon **its** ruin. (Allen Bathurst, 1714; WENTWORTH 2, 412)

In a later analysis the three semantic roles of the possessive, introduced in Nevalainen & Raumolin-Brunberg (1994b), will be used and they are thus briefly exemplified here. The possible semantic roles of the possessives include Agents, Possessors and Others. "Agents are instigators of the action identified by the verb or verbal noun" (Nevalainen & Raumolin-Brunberg 1994b: 193) as in (10.7), where *its appearance* could be paraphrased as "the volume appears".

- (10.7) The truth is, I have been such a Verse-maker myself, and so busy in preparing a Volume for the press, which I imagine will make **its appearance** in the course of the Winter, ... (William Cowper, 1781; COWPER W, I, 531)

Possessors are the semantic prototype of possessives and they broadly indicate something belonging or pertaining to something or somebody as in (10.8).

- (10.8) I quarrel'd most bitterly with it [= Peace] at first, finding nothing in **the terms of it** but disgrace and destruction to Great Britain. (William Cowper, 1783; COWPER W, II, 115)

'Others' include roles that are not Agents or Possessors. Nevalainen & Raumolin-Brunberg (1994b: 193), whose grammatical analysis employed a much more extensive data set than the one studied here, claim that the most frequent role in this category is an affected or effected Object. This is illustrated in (10.9), where both cases correspond to the object in a nominalization ("packing it", "carrying it").

- (10.9) I have felt an impatience to receive it [= my Desk] that you yourself have taught me, and now think it necessary to let you know that it is not come, lest it should perhaps be detained in London by the negligence of some body to whom you might entrust **the packing of it or its carriage** to the Inn. (William Cowper, 1785; COWPER W, II, 413)

In this study, the variant forms searched for in the eighteenth-century letters include four possible variants of the third-person neuter possessive paradigm: (1) *its* NP, (2) *the* NP *of it*, (3) *the* NP *thereof*, and (4) *the* NP *of the same*. In other words, the noun phrase (NP) heads in variants (2)–(4) had to be marked definite with the definite article *the* in order to be included. As *the* NP *of the same* did not produce any relevant hits, the construction was left out of further analysis. *The* NP *thereof* was still used in the seventeenth-century data, but its use rapidly decreased in the

eighteenth century, when only a handful of hits were found. Thus, the variation in the neuter possessive paradigm is primarily between *its* and *of it*.

Following the criteria set in Nevalainen & Raumolin-Brunberg (1994b), cases of *its* with a human reference (10.10) and followed with non-finite clauses (10.11) were left out. Cases of *of it* as quantitative partitives (10.12), prepositional constructions (10.13) and fixed phrases (10.14) were also left out.

(10.10) Tis with the greatest pleasure I congratulate my dear girl on her safe deliverance and her fine boy, for which I am most thankfull to Almighty God, and hope we three shall make each others days happier, for shall doat of my boy more for **its** name sake than my own (Joseph Banks 2, 1717?; BANKS, 164)

(10.11) Yesterday the Lord L-t of Ireland came into the House and spoke against **itts** being extended to Ireland (Allen Bathurst, 1714; WENTWORTH 2, 390)

(10.12) Now if you will be satisfied to make a division of the moor or common that lyes betwixt Salter's Syke (the first boundary) and the Redwells, and take the half **of it** by way of addition to your first boundary, I hope my Lord and the freeholders may be brought to acquiesce in it.  
(John Eden, 1719?; CLAVERING, 186)

(10.13) I shall take the greatest care **of it**. (Thomas Percy, 1762; WARTON, 134)

(10.14) It leaves us entirely to ye World to make ye most **of it**.  
(Edward Young, 1742?; YOUNG, 136)

These criteria guarantee that *its* and *of it* form a linguistic variable and that the results are comparable to those of Raumolin-Brunberg (1998) in particular as she used correspondence data extracted from the *Corpus of Early English Correspondence*, which is collected on similar sociolinguistic principles as the *Corpus of Early English Correspondence Extension*.

### 10.3 Earlier sociolinguistic research

Earlier sociolinguistic studies on the paradigm developments of the neuter possessive include Raumolin-Brunberg (1998) and Nevalainen & Raumolin-Brunberg (2003), who used correspondence data extracted from the *Corpus of Early English Correspondence* and covered the period from 1620 to 1680. These studies serve as a direct point of comparison for my study and continue the time span from 1680 to 1800. In addition, Nevalainen & Raumolin-Brunberg (1994b) mapped the paradigm developments of the third-person neuter possessives in the *Helsinki Corpus* between 1500 and 1710 as a standardization process and the results of the current analysis are compared to their grammatical findings in particular.

Nevalainen & Raumolin-Brunberg's research on both corpora testify to a rapid spread of *its* in the first half of the seventeenth century. Linguistic factors may explain the pace of change as *its* may have been a particularly suitable form to fill an empty slot in the paradigm after the use of *his* as the neuter possessive declined. On the other hand, social and economic conditions in the seventeenth century may have contributed to the rise of *its* as the instability of the period may have caused an increase of weak ties in social networks. The Civil War in 1642–1646 in particular was a period of unrest when people came into contact with a lot of different speakers. This development contributed to the number of weak ties and may thus have accelerated linguistic changes (Nevalainen & Raumolin-Brunberg 1994b: 201, Raumolin-Brunberg 1998). Raumolin-Brunberg's (1998: 373) analysis of *its*, however, does not fully support the Civil War hypothesis as the rise in the use occurred already during the war although newly-adopted forms usually find their way on paper with some delay.

More specifically, Raumolin-Brunberg (1998) has shown that the frequency of *its* increased during the seventeenth century from 5% in 1620–1639, to 20–25% in 1640–1659 and up to 40% in 1660–1681. Her study was based on correspondence data allowing methodological experiments with apparent time to find out how the change proceeded on the level of language users. Her findings suggest that the spread of *its* was both a communal change, where all the members of the community adopt a new form simultaneously, and a generational change, where subsequent generations increase their use (Labov 1994: 84). In the apparent-time analysis, Raumolin-Brunberg (1998: 372–373) studied letters written between 1620 and 1681 by individuals born over a period spanning from pre-1580 to post-1640. Her results show that the oldest generation born before 1580 – and before the first attestations of *its* – did not use the new possessive determiner at all. The subsequent generations then increased their use of *its*. In the early period between 1620 and 1639 the users of *its* were middle-aged and younger people, but during the following period from 1640 to 1659 the old and middle-aged generations increased their use. This finding can be interpreted as a generational change. However, Raumolin-Brunberg's (1998: 372) results also show indications of a communal change as the age cohorts of those born in 1580–1599 and 1600–1619 increase their use of *its* when they grow older.

Moreover, social stratification was observed in the seventeenth century (Raumolin-Brunberg 1998: 378); during the period 1640–1659 gentry was leading the diffusion so that *its* accounted for 38% of the neuter possessives in their letters. During the period 1660–1681 the tables were turned as non-gentry (64%), professionals (50%), and social climbers (48%) were leading the diffusion of *its*, but gentry was then lagging behind (16%). This type of pattern where non-gentry and professionals lead the change is observed in the current analysis of the eighteenth-century data. Patterns of gender variation in the seventeenth century suggest that women's

use of *its* exceeded that of men in the period 1660–1681 (Raumolin-Brunberg 1998: 375). However, the number of occurrences is relatively low, which may explain the fact that the eighteenth-century data explored in this study shows a lower use of *its* in women's letters, and women's use of *its* exceeds men only towards the end of the eighteenth century.

## 10.4 Results

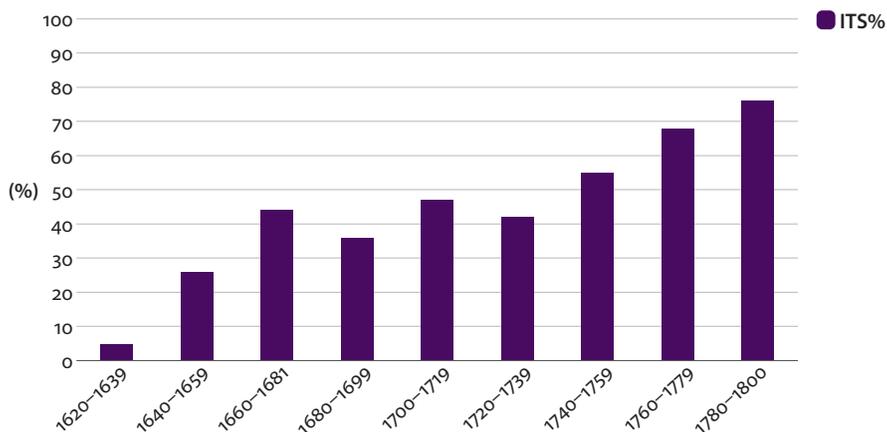
### 10.4.1 Time course of change

The time course of the change from 1680–1800 is shown in Table 10.1, including the three variants *its* NP, *the* NP *of it*, and *the* NP *thereof*. Figure 10.1 additionally combines the percentages of *its* given in Raumolin-Brunberg (1998: 382) for the years 1620–1681 with the time period studied here. As a result of the two analyses we see the S-shaped diffusion of the new form *its* from the incipient stage of the change (1620–1639), through the new and vigorous stage (1640–1659), to the mid-range (1660–1759) and finally to the change nearing completion (1760–1800). These stages of the change are based on Labov (1994: 79–83) and Nevalainen & Raumolin-Brunberg (2003: 55), who use the following classification and proportions for the incoming form: incipient (below 15%), new and vigorous (15–35%), mid-range (36–65%), nearing completion (66–85%) and completed change (over 85%) (see also 5.1.2 for periodizing processes of change).

**Table 10.1** Absolute and relative frequencies (%) of *its*, *of it*, and *thereof*

	1680–1699	1700–1719	1720–1739	1740–1759	1760–1779	1780–1800
<i>its</i> NP	48 (36%)	63 (47%)	32 (42%)	175 (55%)	191 (68%)	373 (76%)
<i>the</i> NP <i>of it</i>	68 (51%)	64 (47%)	42 (55%)	138 (44%)	88 (31%)	120 (24%)
<i>the</i> NP <i>thereof</i>	18 (13%)	8 (6%)	2 (3%)	3 (1%)	1 (0%)	1 (0%)
Total	134 (100%)	135 (100%)	76 (100%)	316 (100%)	280 (100%)	494 (100%)

The beginning stages of the spread of *its* were rapid as claimed, for example, in Nevalainen & Raumolin-Brunberg (2003: 63), but the mid-range lasted for about a century. Another sharp slope upwards in the S-curve occurs during the latter half of the eighteenth century from 1740 to 1800, but the change did not quite reach the 85% proportion of the paradigm, regarded as a threshold for a completed change. A comparison of the S-curve of Figure 10.1 against the longer time span of the Google Ngram in Figure 13.7 suggests that it took some fifty years after the end point of the current study for *its* to reach the peak in the 1850s. The Google Ngram is based on unanalysed data and cannot be taken at face value, but it nevertheless illustrates broad trends.



**Figure 10.1** The spread of the third-person possessive *its* (%) 1620–1800. Percentages for the first three periods come from Raumolin-Brunberg (1998: 382)

#### 10.4.2 Age

To show more specifically how the change proceeded in time, Table 10.2 presents the percentages of *its* in real time and apparent time. The apparent-time development can be followed in the vertical columns and the real time in the horizontal columns. For this analysis only the writers with known dates of birth were included. Since the birth dates for high-ranking individuals are more often known than those of common people or even professionals, the analysis is more skewed towards the upper social ranks than the other analyses. Women were included in the analysis summarized in Table 10.2 whenever their birthdates were known.

Similar tendencies to Raumolin-Brunberg (1998) can be observed in Table 10.2, which confirms the real-time increase presented in Table 10.1 and Figure 10.1 but which additionally allows us to observe a generational change as subsequent generations increase their use. The clearest jumps in Table 10.2 occur between the generations born in 1620–1639 and 1640–1659, and again between the generations born in 1680–1699 and 1700–1719.

These generational changes seem further to relate to developments observed in real time, showing how speakers enter the speech community with a characteristic frequency of a linguistic form but then increment on it during their life spans. Figure 10.1 shows a sharp real-time rise in the use of *its* from 1620–1639 (6%) to 1640–1659 (26%). In the data explored here, the writers born in those two periods and writing later in 1680–1699 seem to reflect the stage of *its* in their infancy respectively, but both generations have also increased their use when they have grown older. The 1620–1639 age cohort have *its* in 25% of the third-person

**Table 10.2** Apparent-time analysis of the use of *its* (%) 1680–1800. Absolute numbers are given in Table A in the Appendix

Time of writing	1680–99	1700–19	1720–39	1740–59	1760–79	1780–99	
Year of birth	Percentage (%) of <i>its</i>					Percentage by generation	
1620–39	25	–				29	
1640–59	50	–				48	
1660–79	42	50	–			45	
1680–99		48	43	40		43	
1700–19			–	66	72	88	69
1720–39				60	80	69	71
1740–59					53	81	72
1760–79						74	74
Total	35	48	42	56	70	75	

neuter possessives, whereas the percentage for the 1640–1659 age cohort is 50% (Table 10.2).

Another similar example between the real and apparent times can be found between the periods 1680–1699 and 1700–1719. Figure 10.1 shows a rise from 36% to 47% of *its* between those two real-time periods. The age cohorts born in those periods again show a difference when they write in 1740–1759 (40% vs. 66%). Both generations have increased their use from what they must have learned in their infancy but not to the same level. This evidence seems to say that generational and communal change operate simultaneously.

The age cohort of 1700–1719 in particular shows participation in communal change as they increase their use over time from 66% to 88% (Table 10.2). Although the same cannot be said for the age cohort of 1680–1699 as their percentages decline from 48% to 40%, the general tendency for age cohorts is to increase their use. The age cohorts born in 1660–1679, 1720–1739 and 1740–1759 show further tendencies speaking for a communal change in that these age cohorts increase their use of *its* when they grow older.

Mean frequencies of groups can be misleading especially if there is a lot of subgroup internal variation. As in 7.3.2, this factor was examined using bootstrapping (for quantitative methods, see 5.2). Figure 10.2 shows how the year of birth contributes to the diffusion of *its* with 10 confidence intervals using the pooling method. The figure suggests that the range of variation, and hence uncertainty, is widest in the oldest generation born before 1639, but generally subsequent generations increase their use except for the last two generations (born in 1720–1759 and after 1760) as the confidence intervals overlap. Hence, the generations born

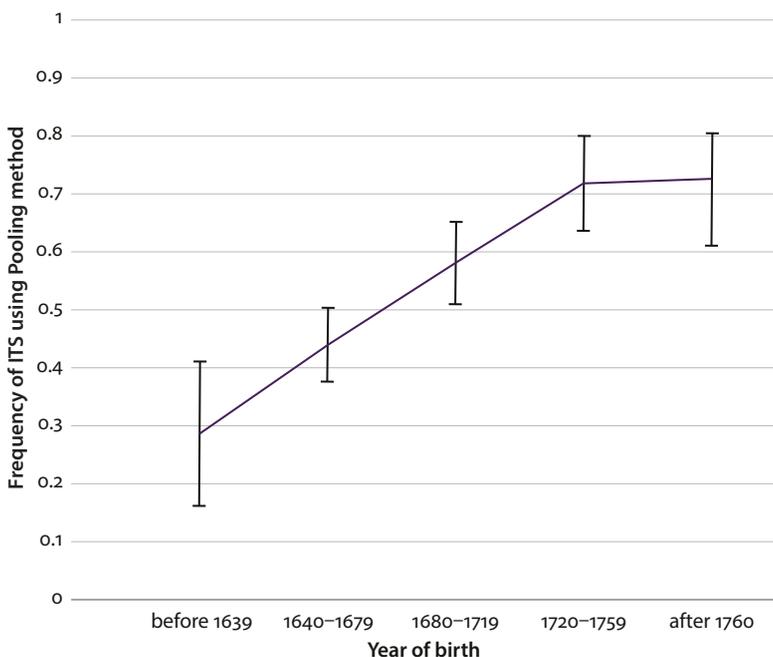


Figure 10.2 Writers' year of birth and the use of *its*, bootstrap estimates using pooling

immediately before and at the beginning of the nearing-completion stage of change do no longer differ from each other but carry on the change equally.

In the next step we could analyse possible life-span changes in the same individual writing over a long period of time (for such a study on three sixteenth- and seventeenth-century gentlemen's language, see Raumlöin-Brunberg 2009). We have included individuals with letters from different phases in their lives in the *Corpus of Early English Correspondence Extension*, but in practice such an analysis proved impossible for any conclusive results as the numbers are low for third-person neuter possessives (see Table B in the Appendix).

#### 10.4.3 Social status variation

The twenty-year periodization employed in Tables 10.1 and 10.2 is replaced in further analysis of social status variation with a more robust periodization. This is due to the nature of the data, which is too scarce for fine-grained divisions according to several factors. The first sixty years in particular are problematic. To find suitable periods, I tried several options. A division according to the percentage proportions of the incoming form *its* at various stages of the change would have produced two time periods as suggested in the previous section: (1) 1680–1759 for the mid-range

of the change and (2) 1760–1800 for the change nearing completion. However, as the period 1740–1759 appears to be characterized by a more prominent use of *its* than previous periods such a bi-polar division did not seem ideal (see Table 10.1 and Figure 10.1). To find diachronically important time periods, I tested the statistical significance of frequency differences between each subsequent twenty-year period. According to chi-square tests the differences between the periods from 1680 to 1739 are not statistically significant, but the differences between 1720–1739 and 1740–1759 and between each subsequent period after this are statistically significant (at least on 5% level). In addition, the simple bi-polar division into the mid-range of the change and the change nearing completion is statistically a highly significant one ( $p < 0.0001$ ) albeit this division seems to miss the first stages of a more rapid diffusion. For the analysis of social variation, I finally decided to lump the first sixty years together (1680–1739) and then present the social embedding of *its* in three twenty-year periods. This was also motivated by the fact that during the long mid-range of the change social variation does not seem to be as salient as during 1740–1759 and afterwards.

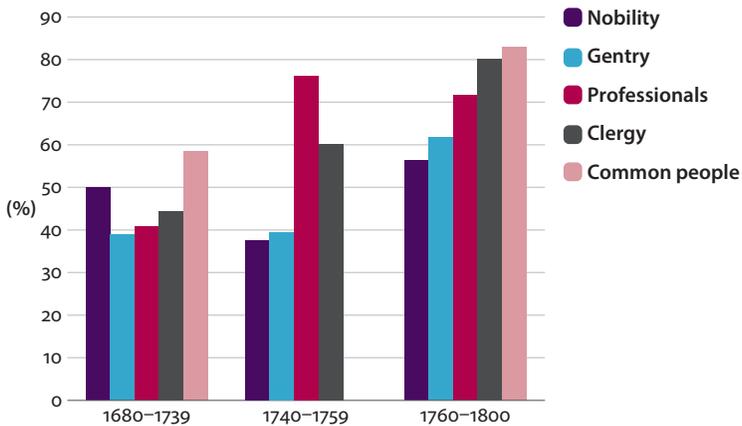
The data is scarce for a detailed analysis of variation based on the writers' social status. Table 10.3 thus represents the data using the same rough social rank division as Nevalainen (Chapter 7 in this volume), including male writers only. In this way the data for the gentry, clergy and professionals are adequate, but in the highest and the lowest ranks of nobility and common people the numbers are still low. Social rank differences are statistically significant according to chi-square tests in the periods 1740–1759, 1760–1779 and 1780–1800.

**Table 10.3** Absolute and relative frequencies (%) of *its* according to the male writers' social status 1680–1800. N is the total of the third-person neuter possessives

Social status	1680–1739	1740–1759	1760–1779	1780–1800
	<i>its</i> /N (%)	<i>its</i> /N (%)	<i>its</i> /N (%)	<i>its</i> /N (%)
Nobility	12/24 (50%)	6/16 (38%)	8/14 (57%)	1/2
Gentry	35/90 (39%)	24/61 (39%)	36/61 (59%)	38/59 (64%)
Professionals	47/115 (41%)	51/67 (76%)	69/83 (83%)	107/163 (66%)
Clergy	23/52 (44%)	53/88 (60%)	24/36 (67%)	45/50 (90%)
Common people	7/12 (58%)		26/29 (90%)	3/6

Figure 10.3 further illustrates social rank differences combining the last two periods as the data for the nobility and common people in 1780–1800 is scarce. The spread of *its* seems to be a change from below rather than from above in the sense that lower ranks lead the change. During the period 1740–1759 professional people are ahead of the change with the clergy in the second place. The nobility and gentry lag behind. Although the number of hits produced by common people is low, *its* is

the prominent form in their letters and they have the highest frequency of all the ranks both in 1680–1739 and 1760–1800. The groups leading the spread of *its* in the eighteenth century are typically people who use language professionally.



**Figure 10.3** Relative frequency (%) of *its* according to the male writers' social status 1680–1800

#### 10.4.4 Gender variation

Men lead the diffusion of *its* for the most of the eighteenth century, whereas women quickly increase their use up to 90% towards the end of the century when *its* begins to be firmly established as the third-person neuter possessive (Table 10.4 and Figure 10.4).

**Table 10.4** Absolute frequency of *its* out of the total of the third-person neuter possessives (N) according to the writers' gender

Gender	1680–1739	1740–1759	1760–1779	1780–1800
	<i>its</i> /N	<i>its</i> /N	<i>its</i> /N	<i>its</i> /N
Men	124/293	134/232	169/231	222/322
Women	18/53	41/84	22/49	151/168

Subgroup internal variation was again tested using bootstrapping. Figure 10.5 suggests that the range of variation, and hence uncertainty, is generally wider in women. As there is the least overlap in confidence intervals between women and men in the periods 1720–1739 and 1760–1779, the gender difference is likely to be the greatest at those times.

Just as social status variation, gender variation is somewhat problematic to establish. In addition to women's lower proportion of *its* up to the period 1760–1779,

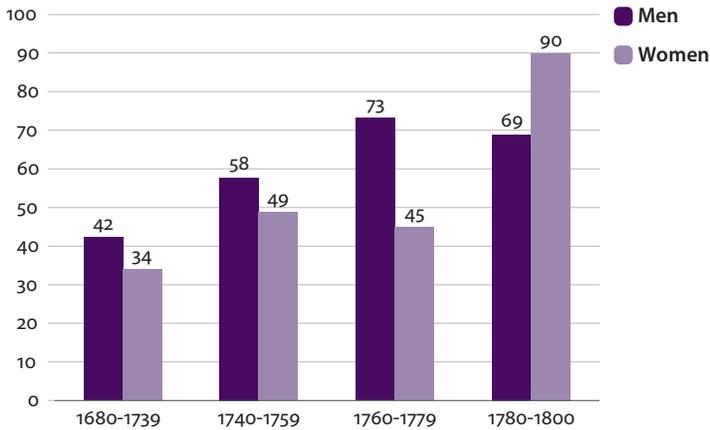


Figure 10.4 Relative frequency (%) of *its* according to the writers' gender

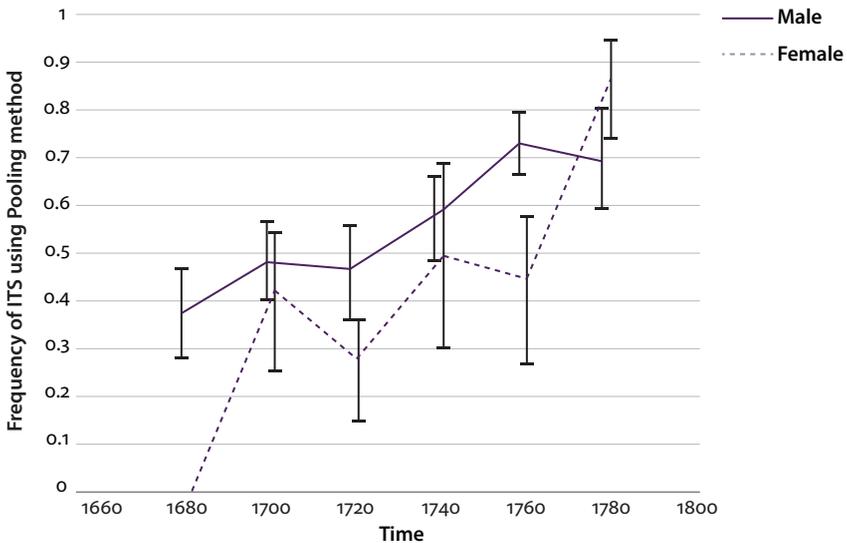


Figure 10.5 Gender variation in the use of *its*, bootstrap estimates using pooling (start of each 20-year period)

men and women differ in the absolute frequency of the third-person neuter possessives. Especially in the late seventeenth- and early eighteenth-century data women use neuter possessives far less frequently than men (Table 10.5). On the one hand, the number of letters and the total word count written by women are far lower than the number of letters or the total word count produced by men. This may affect the results as the neuter possessives overall are a relatively infrequent grammatical form not used in every letter. The corpus data by women can be described as thin in the

sense that most women writers have only a letter or two, which does not seem to be enough for the third-person neuter possessives to occur with any certainty. This situation is remedied only in the last period when there is more data by women writers (see 4.3.2 for gender balance in CEECE). On the other hand, the use of neuter possessives may relate to specific registers, stylistic choices and discourse topics that are perhaps not part of the female sphere at the beginning of the period, but these ideas need to be explored further.

**Table 10.5** Normalized frequency of third-person neuter possessives per 10,000 words according to gender. Absolute frequency in parentheses

Gender	1680–1699	1700–1719	1720–1739	1740–1759	1760–1779	1780–1800
Men	5.4 (145)	5.7 (117)	4.0 (63)	9.6 (274)	8.8 (256)	9.1 (372)
Women	1.5 (4)	3.6 (31)	3.8 (20)	10.2 (92)	9.3 (79)	7.4 (196)

#### 10.4.5 Regional variation

Since it has been claimed that *its* originated in the Southern dialects, regional variation might still account for the use of this form particularly at the beginning of the period as *its* is still only about a hundred years old. The analysis of the writers' regional background focused on individuals whose regional background is known, including women. This is the case for most of the writers, but additionally some are classified regionally as 'court' or 'abroad' and these were left out of this analysis. Similarly, Norfolk did not produce enough data and was left out. The regions included are London, Home Counties immediately around London, North including Cumberland, Durham, Lancashire, Northumberland, Westmorland, and Yorkshire, and 'other' including the rest of the southern regions. Table 10.6 presents the results in three periods combining 1760–1779 and 1780–1800 as the data from the North was so scanty.

**Table 10.6** Absolute and relative frequencies (%) of *its* according to the writers' region. N is the total of the third-person neuter possessives

Region	1680–1739	1740–1759	1760–1800
	<i>its</i> /N (%)	<i>its</i> /N (%)	<i>its</i> /N (%)
North	9/36 (25%)	7/10 (70%)	12/17 (71%)
London	84/187 (45%)	81/164 (49%)	272/358 (76%)
Home Counties	18/43 (42%)	54/82 (66%)	104/163 (64%)
Other non-north	27/60 (45%)	30/53 (57%)	102/128 (80%)

The results show that dialectal differences still exist in 1680–1739 but they are soon levelled out. The dialectal differences are not otherwise statistically significant except for the North (25% of *its*) vs. other regions (42–45%) during the first period. The North is not very well represented in the data as the actual number of hits for *its* is only 9 for the first period and 7 and 12 for the latter periods, but we can take this evidence as support for the claims that *its* originated in the South and was widely spread there by 1680–1739 as *its* covered over 40% of the possessive paradigm. By the end of the eighteenth century *its* is the major variant in the whole country with percentages ranging from 64% to 80%. Already in the mid-century (1740–1759) *its* dominates the paradigm with percentages ranging between 49–70%. Geographical diffusion thus precedes social embedding here.

#### 10.4.6 Conservative/progressive individuals?

The hits of *its* and *of it* are scattered so that most individual writers have no hits at all or only a few hits. The analysis of outliers either as conservative or progressive language users in the case of this change therefore focuses on 40 individual writers with ten or more examples of the variable. Table 10.7 includes these writers and indicates the percentage of *its* in their usage. In this case “progressive” writers use *its* more frequently than the period average, and “conservative” writers less. The writers representing the period average have been included in the table as “average”.

In some cases, the age of the writer may explain the innovative or conservative tendencies. For example, in 1680–1699 and 1700–1719 Humphrey Wanley, born in 1672, is an advanced user of *its*, whereas older writers Samuel Pepys, born in 1633, and Daniel Defoe, in 1659, are lagging behind the period average. But older writers may also show progressive tendencies like John Evelyn, Charles Burney or Samuel Johnson. Sarah Lennox, Erasmus Darwin, and Humphrey Wanley have several hits of *its* and *of it* in two consecutive periods. Humphrey Wanley seems to maintain the level of *its* in time, whereas Sarah Lennox seems to increase her use. Erasmus Darwin, on the other hand, apparently becomes more conservative in time. These individuals could be studied further to see whether we are indeed dealing with lifespan changes or if something else is going on. For instance, numbers are still relatively low, which may simply explain such individual variation.

For individual analysis, I singled out William Cowper with letters written in the last period 1780–1800 in Table 10.7 as his use of *its* (43%) can be characterized as “conservative” in comparison to the period average (76%) although he was born as late as 1731 and this generation overall has 71% of *its* on average. Cowper is an interesting test case for several reasons. Methodologically, we can assume that with the number of hits he produced we get a relatively representative view of his usage.

Table 10.7 Individual writers (with their social status) compared to the average frequency (%) of *its* in the subperiod

1680–1699 (ITS 36%)	1700–1719 (ITS 47%)	1720–1739 (ITS 42%)	1740–1759 (ITS 55%)	1760–1779 (ITS 68%)	1780–1800 (ITS 76%)
<b>Progressive</b>	<b>Progressive</b>	<b>Progressive</b>	<b>Progressive</b>	<b>Progressive</b>	<b>Progressive</b>
John Evelyn (GLP) 60% (9/15) b. 1620	Henry Liddell (GL) 71% (10/14) b. 1673	Alexander Pope (Prof) 54% (7/13) b. 1688	Spencer Cowper (CU) 91% (10/11) b. 1713	Ignatius Sancho (O) 100% (13/13) b. 1729	Samuel Johnson (Prof) 100% (15/15) b. 1709
Humphrey Wanley (Prof) 54% (7/13) b. 1672	Humphrey Wanley (Prof) 50% (10/20) b. 1672		Thomas Gray (Prof) 87% (13/15) b. 1716	Charles Burney (Prof) 94% (17/18) b. 1726	Frances Burney (Prof) 100% (63/63) b. 1752
Humphrey Prideaux (CU) 45% (5/11) b. 1648	Isabella Wentworth (N) 50% (5/10) b. 1653		Edward Young (CL) 81% (21/26) b. 1683	Richard Champion (M) 93% (13/14) b. 1743	Hester Piozzi (GL) 96% (27/28) b. 1741
			Robert Dodsley (Prof) 70% (19/27) b. 1703	Erasmus Darwin (Prof) 88% (15/17) b. 1731	
			Hester Grenville (GU) 69% (11/16) b. 1720		
			Elizabeth Carter (CL) 69% (18/26) b. 1717		
<b>Average</b>	<b>Average</b>	<b>Average</b>	<b>Average</b>	<b>Average</b>	<b>Average</b>
	Joseph Addison (GLP) 46% (6/13) b. 1672		William Pitt (GL) 56% (15/27) b. 1708	Richard Hurd (CU) 70% (7/10) b. 1720	Sarah Lennox (N) 76% (16/21) b. 1745
			Richard Hugh (CL) 53% (10/19) b. 1720	David Garrick (GL) 63% (10/16) b. 1717	Augustus Hanover (R) 75% (9/12) b. 1773
					George IV (R) 73% (8/11) b. 1762
<b>Conservative</b>	<b>Conservative</b>	<b>Conservative</b>	<b>Conservative</b>	<b>Conservative</b>	<b>Conservative</b>
Samuel Pepys (Prof) 14% (3/22) b. 1633	Daniel Defoe (Prof) 35% (7/20) b. 1659		Henry Purefoy (GL) 30% (3/10) b. 1697	Thomas Twining (CL) 58% (7/12) b. 1734	Erasmus Darwin (Prof) 71% (10/14) b. 1731
			Thomas Secker (CU) 9% (1/11) b. 1693	Elizabeth Draper (Prof) 58% (14/24) b. 1744	Mary Wollstonecraft (Prof) 65% (11/17) b. 1759
			Mary Wortley Montagu (N) 27% (10/37) b. 1689	Roger Newdigate (N) 57% (8/14) b. 1719	William Jones (GUP) 58% (7/12) b. 1746
				Jeremy Bentham (GLP) 43% (6/14) b. 1748	Jeremy Bentham (GLP) 54% (15/28) b. 1748
				Sarah Lennox (N) 23% (2/13) b. 1745	William Cowper (Prof) 43% (29/67) b. 1731

In addition to the data included in Table 10.7, he has letters in the corpus in two earlier periods, and his letters in 1760–1779 produced 6 more hits, giving the total of 73 examples, shared between 33 *its* and 40 *of it*. As a well-educated professional from Buckinghamshire, Cowper also represents a social group and region that in general were leading the diffusion of *its*. He is thus exceptional in comparison to his peers.

Since grammatical factors referred to in Section 10.2 have not been taken into account so far, I decided to see whether the two factors that have been found to correlate with the use of pre- and post-head forms can be used to account for the variation in Cowper's usage. In other words, we can hypothesise on the basis of Nevalainen & Raumolin-Brunberg (1994b) that (1) reflexive use favours the pre-nominal *its* and non-reflexive use the postnominal *of it*, and that (2) the semantic role of Other favours the postnominal *of it*.

The impact of a grammatical factor is particularly clear in the case of the post-head possessive *of it*, as 37 (93%) of the 40 cases are non-reflexive, i.e. the possessive refers to something else than the subject of the clause. *Its*, on the other hand, has a broader grammatical scope as it is used both reflexively (12 = 36%) and non-reflexively (21 = 64%). In the post-head (non-reflexive) cases, one type of clausal position stands out. Namely, many examples include the possessive right at the beginning of the clause in the first (subject) noun phrase as in (10.15). In such cases the choice of the pre-head variant would place *its* as the first word in the clause (*Its purport was to inform ...*), but *its* is not found in this position in Cowper's usage.

(10.15) The purport of it was to inform him that there are very near 1200 Lace Makers  
in this Beggarly Town, ... (William Cowper, 1780; COWPER W, I, 363)

Semantic roles of the possessive did not produce such clear differences. Both *its* and *of it* predominantly express possession (58% vs. 68%), and both occur in the Agent (8 = 24% vs. 4 = 10%) and Other (6 = 18% vs. 9 = 23%) roles. In comparison to the findings of Nevalainen & Raumolin-Brunberg (1994b), this result may indicate that the grammatical and semantic scope of *its* has widened together with its diffusion.

## 10.5 Normative grammars

*Its* did not apparently attract much commentary but was adopted in grammar books with the rate of change; in some cases the third-person neuter paradigm was not even included in grammars (Nevalainen & Raumolin-Brunberg 1994b: 180–181). *Its* attracted some normative commentary, but it concerned the status of contracted forms more generally. Namely, some eighteenth-century grammars suggested that *it's* is the possessive form of *it*, not a contraction of *it is*, which should be *'tis* (Sundby, Björge & Haugland 1991: 164).

## 10.6 Conclusion

This chapter explored the spread of *its* in eighteenth-century data, providing a continuation for earlier studies mapping the spread of *its* from the early stages of change to the young and vigorous stage in the seventeenth century. At the beginning of the period studied here, the form was about a hundred years old and at the mid-range of change. During the latter half of the eighteenth century (1760–1800) we can say that the change was nearing completion, but the final completed stage in the use of the form was reached a bit later in the mid-nineteenth century. *Its* seems to have spread regionally to different parts of the country before the more rapid rise of the form from 1740 onwards. The apparent time analysis suggested that both generational and communal change operated simultaneously at times, but the generations born immediately before and at the beginning of the final rapid rise no longer differed from each other. This period of nearing-completion is characterised by social status and gender variation, with professional ranks and men leading the change except for the very end of the eighteenth century when women's use exceeds that of men.

## Appendix

**Table A.** Apparent time analysis. Hits of *its* of the total of the third-person neuter possessive paradigm (N)

Time of writing	1680–99	1700–19	1720–39	1740–59	1760–79	1780–99	
Year of birth	<i>its</i> /N						Generation total
1620–39	17/68	7/16					24/84
1640–59	17/34	2/6					19/40
1660–79	11/26	35/70	3/14				49/110
1680–99		15/31	18/42	40/101	6/8		79/182*
1700–19			9/16	103/156	48/67	21/24	181/263*
1720–39				30/50	92/115	125/182	247/347
1740–59					43/81	157/195	200/276
1760–79						59/80	59/80
Total	45/128	59/123	30/72*	173/307*	189/271*	362/481	

\* indicates a statistically significant difference at a 5% level or less between the adjacent cells.

**Table B.** Individual use of *its*. Hits of *its* of the total of the third-person neuter possessive paradigm (N)

Time of writing	1680–99	1700–19	1720–39	1740–59	1760–79	1780–1800
Writer (date of birth)	<i>its</i> /N (word-count)					
Humphrey Wanley b. 1672	7/13 (11,302)	10/20 (18,923)				
Edward Young b. 1683		0/0 (1,796)	3/6 (6,180)	21/26 (13,458)	4/5 (2,621)	
Alexander Pope b. 1688		3/4 (5,525)	7/13 (15,699)	0/3 (6,182)		
Mary Wortley Montagu		5/9 (15,188)	1/2 (8,319)	10/37 (43,367)	0/1 (3,950)	
b. 1689? Roger Newdigate			1/1 (2,740)	3/5 (9,629)	8/14 (15,030)	0/0 (2,655)
b. 1719 David Garrick b. 1717			1/2 (5,375)	1/4 (15,234)	10/16 (22,880)	
Spencer Cowper b. 1713			0/0 (932)	10/11 (8,217)	0/1 (1,038)	
Thomas Secker b. 1693 Richard			1/1 (1,210)	1/11 (14,552)	0/0 (363)	
Hurd b. 1720			2/2 (1,026)	10/19 (24,450)	7/10 (6,659)	4/4 (1,006)
Robert Dodsley b. 1703 Thomas				19/27 (15,612)	1/1 (1,651)	
Gray b. 1716			1/1 (4,262)	13/15 (13,982)	13/16 (10,965)	
Samuel Johnson b. 1709			1/2 (1,469)	5/7 (7,243)	4/6 (6,642)	15/15 (10,352)
Elizabeth Carter b. 1717 George			2/6 (5,775)	18/26 (14,468)		2/3 (5,256)
III b. 1738				0/0 (2,313)	6/8 (10,003)	4/7 (12,117)
Edward Gibbon b. 1737				1/2 (5,650)	4/6 (11,634)	2/2 (9,256)
Jeremiah Bentham b. 1712				5/5 (1,402)	1/1 (1,769)	0/1 (1,346)
William Cowper b. 1731				0/0 (332)	4/6 (10,572)	29/67 (45,350)
William Mason b. 1725 Joseph				5/6 (9,419)	3/3 (3,923)	5/7 (4,619)
Warton b. 1722? Thomas				0/0 (2,497)	0/0 (1,813)	0/0 (347)
Warton b. 1728 Erasmus				1/1 (1,245)	2/3 (4,478)	2/2 (6,419)
Darwin b. 1731 Frances Burney					15/17 (5,748)	10/14 (12,614)
b. 1752					1/2 (5,765)	63/63 (44,781)
Mary Wollstonecraft b. 1759					1/4 (3,200)	11/17 (28,709)
Sarah Lennox b. 1745					3/13 (31,032)	16/21 (34,645)
Charles Burney b. 1726					17/18 (19,255)	14/14 (20,196)



## Incipient and intimate

### The progressive aspect

Anni Sairio

#### 11.1 Introduction

On 10 September 1793, 18-year-old William Clift wrote to his older sister Elizabeth from London, where he was training as an apprentice:

as I expected Mrs Hunter to go to Cornwall every day I defer'd writing as I had not the Cash to pay the postage for I have been taking in some novels that are publishing in weekly numbers at sixpence each I quite forgot to save any money to pay the postage. I have been saving every farthing I have been able to scrape together since Christmas last [...] <sup>1</sup> (William Clift, 1793; CLIFT, 77)

The youngest son of a Cornish miller, William Clift (1775–1849) was apprenticed to the surgeon and anatomist John Hunter in 1792, and after Hunter's death he became the first Conservator of the Hunterian Museum of anatomical specimens in London in 1799 (Austin ed. 1991: 1). During the final years of the eighteenth century, Clift was a young and ambitious social riser, working his way in London to the professional ranks. He also represents in many ways a typical user of the progressive (or the *be+ing* construction) in the CEEC Extension: he is a letter-writer with lower-rank background who corresponds with a close family member at the end of the century. William Clift and his sister Elizabeth Clift (1757–1818), the recipient of William's letters in the corpus, are exceptionally frequent users of the progressive which is generally characterised as an oral-like, informal feature. On the other hand, William's use of the progressive passival (*novels that are publishing*) illustrates that prolific though he is and therefore certainly an innovator, he is not innovative in terms of syntax; the progressive passive (*novels that are being published*), one of the few grammatical innovations of the Late Modern period, does not appear in William's letters at all. In fact, the progressive passive occurs in the CEEC

---

1. The novels that William Clift had purchased included Henry Fielding's *Tom Jones* and *Joseph Andrews*, Oliver Goldsmith's *Vicar of Wakefield*, Tobias Smollett's *The Adventures of Roderick Random*, Laurence Sterne's *Tristram Shandy*, and Daniel Defoe's *Robinson Crusoe* (Austin ed. 1991: 77).

Extension only once. According to Anderwald (2012: 30), the nineteenth-century development of the passive and its complete takeover from the passival results from the increase of the progressive itself, and Pratt & Denison (2000) document the progressive passive in the late eighteenth-century Southey-Coleridge circle as ‘radical experimentation’. The passive form did not yet have a place in the general world of eighteenth-century letter-writing. Even though the progressive was in many ways a mature construction in this period, it underwent more substantial change in terms of frequency and function during the nineteenth and the twentieth centuries.

This chapter examines large-scale sociolinguistic variation of the progressive in eighteenth-century letters, with focus on gender, social rank, and register (here understood as the relationship between the correspondents). The influence of gender in use of the progressive has been previously discussed in for example Arnaud (1998, 2002), Fitzmaurice (2004), Smitterberg (2002) and Kranich (2010), and that of gender, social rank, and register in my own small study of eighteenth-century letters (Sairio 2009), but social parameters surrounding the use of the progressive in Late Modern English have not been systematically examined in large historical data.

The progressive takes the following tenses in the eighteenth century:

- Present progressive: ‘I am now rambling from Place to Place’  
(John Gay, 1719; GAY, 36)
- Past progressive: ‘A tall lazy villain was bestriding his poor beast’  
(Ignatius Sancho, 1777?; SANCHO, 102)
- Present perfect progressive: ‘I have been a hunting with Mr Varny’  
(Lady Sarah Lennox, 1762; LENNOX, I, 127)
- Past perfect progressive: ‘Just the night before God took her from me, she had been discoursing with one of our sisters’  
(Winifred Thimelby, 1680?; TIXALL, 102)
- Future progressive with a modal: ‘When summer approaches I shall be inventing Schemes for that purpose’  
(Lady Mary Wortley Montagu, 1759?; MONTAGU, III, 196)

In terms of tense and both literal and figurative contexts of use, the Late Modern English progressive was already a mature construction, but still infrequently used. However, quantifying this feature is not unproblematic. It is difficult to evaluate whether the number of all verb phrases remains diachronically stable, whether the proportion of progressives actually increases with regard to all possible variants, and what circumstances enable the progressive to be used: in other words, normalised frequencies need to be interpreted with healthy caution (Smitterberg 2005, Aarts, Close & Wallis 2010: 154–155; see also Section 5.3 on researching linguistic forms that lack a variable). Kranich (2010: 13) points out that the frequency of progressives per verb phrase would provide a more exact measurement than normalising the absolute frequencies of the progressive, and Smitterberg (2005) has

tagged the *Corpus of Nineteenth-Century English* (CONCE) in order to investigate the progressive in relation to the non-progressive VPs. CEECE is not tagged, so this study relies on normalised frequencies and the method used by Säily & Suomela (2009) which examines whether significant sociolinguistic variation actually appears; see also Vartiainen, Säily & Hakala (2013) for an application of the beanplot method in the analysis of pronoun frequencies in the CEEC.

The chapter is outlined as follows. Section 11.2 provides an introduction and an overview of the progressive and its previous historical research. Section 11.3 presents the results of the corpus analysis. In Section 11.4 the influence of gender is examined, Section 11.5 considers the influence of social rank, and Section 11.6 the influence of register. Section 11.7 examines the outliers, and Section 11.8 concludes the findings. I do not provide semantic analysis of the progressive or an account of its morphosyntactic variation, amply discussed in Hundt (2004), Smitterberg (2005) and Kranich (2010).

## 11.2 The progressive in Late Modern English

The origins of the progressive are not entirely clear. It has been argued to be a native development in English (Visser 1963–1973), to have developed under Latin influence (Mossé 1938), or to have resulted from Celtic contacts (e.g. Braaten 1967; in Hickey 2012: 501–502). The equivalent of *be+ing* was used already in Old English. The progressive is found in all Celtic languages, and Celtic constructions in fact precede the English *be+ing* usage; the extensive use of the progressive in some regional varieties may thus be linked to Celtic influence and the frequent use of the progressive in Celtic Englishes (Filppula & Klemola 2012: 1688, 1691–1694; see also Filppula 2003). During the Late Modern period the progressive increased considerably, a development which has continued throughout the nineteenth and twentieth centuries (see Beal 2004, Hundt 2004, Núñez-Pertejo 2004, Strang 1982, Denison 1993, Rissanen 1999, Aarts, Close & Wallis 2010, Kranich 2010, Anderwald 2012, and Mair 2006). Due to this expanding use in the nineteenth and twentieth centuries both in terms of frequency and function, we have categorized the eighteenth-century progressive as belonging to the early stages of change.

Smitterberg (2005: 67) concludes on the basis of nineteenth-century genre divergence that “the progressive is an oral rather than a literal feature, as the construction is decidedly more common in popular than in specialized genres”. Kranich (2010) provides a thorough account of the functions and developments of the progressive in various genres. Her survey of ARCHER-2 indicates that the progressive was particularly favoured in drama, private letters, and fiction, “which are much more concerned with what is or was going on at a specific moment in time” (2010: 106). The nineteenth-century progressive occurs most often in letters,

a genre where it also increases over time (Smitherberg 2002). The increase of the progressive can be regarded as grammaticalisation, but Kranich (2010) points out that this increase does not extend across all linguistic contexts as hypothesized by Strang (1982); instead, the progressive continued to be used mostly in the present or past tense. In ARCHER-2, the seventeenth and eighteenth-century increase in the aspectual function has resulted in an overall increase, and Kranich (2010: 252) suggests that between 1650 and 1800 “the grammatical function crystallizes”. In the late eighteenth century and the nineteenth century, “the more clearly grammatical status of the construction leads to its extension across the verbal paradigm”, including the progressive passive (Kranich 2010: 252). The focus of this study is thus on a period when the progressive was starting to be a rich and mature feature, but it was not a part of the verb repertoire to the extent that it is today.

In Late Modern English, the progressive was used predominantly in an aspectual or objective function (Kranich 2010: 168–169, see also Wright 1994, Killie 2004, and Sairio 2009: 186 for the *Bluestocking Corpus*). Aspectual or objective progressives refer to factual, physical, and dynamically advancing events and are relatively time-dependent. The subjective progressive, on the other hand, represents more figurative expressions of belief or attitude (see Kranich 2008, Fitzmaurice 2004). Example (11.1) includes one objective and two subjective progressives:

- (11.1) But here I am writing nonsense when I should be thanking you seriously for Your £100 & sending you security. Voila Don! here it is.  
(William Mason, 1771; GRAY, 1182)

The objective construction *I am writing* informs of a factual event in motion, and the modal construction *I should be thanking & sending you security* has a more figurative, attitudinal meaning. In Example (11.2), Sir William Jones describes his philological pursuits in Calcutta:

- (11.2) I read and write Sanscrit with ease, and speak it fluently to the Brahmans, who consider me as a Pandit; but I am now only gathering flowers: the fruit of my Indian studies will be a complete Digest of Law, which a number of Pandits employed, at my instance, by the Government, are now compiling  
(William Jones, 1788; JONES, II, 813)

*I am now only gathering flowers* is a figurative present-tense progressive with a first-person subject and an adverbial, and it illustrates emotive state. The second progressive (*which a number of Pandits ... are now compiling*) has an aspectual function and it informs of events currently in motion.

Contemporary metacommentary of the progressive shows that seventeenth and eighteenth-century grammarians understood it quite poorly (Wright 1994: 471),

but their comments were mild and generally positive, which extends to the nineteenth century as well (Anderwald 2012). Anderwald (2012: 36) suggests that the positive evaluations of the progressive in nineteenth-century grammars (“*Propriety, harmony and precision, force, emphasis and nicety*”) reflects the “very slow character of this linguistic change” – it appears that contemporaries did not perceive this development as a change in progress, and this lack of awareness explains the unusually positive evaluations of a changing feature. Lenience and neutrality did not extend to the new passive construction (*the house is being built*), but the most vocal protests regarding the “moral deficiency” of the passive progressive did not take place until the nineteenth century, and they are thus beyond the scope of this study (Anderwald 2012: 36; see also Rissanen 1999: 218; Beal 2004: 78, 81; Visser 1973: 2013–2023).

The short-lived progressive passival developed in the seventeenth century (Denison 1998). It is an active construction used to express passive meaning (Kranich 2010: 116), and examples in the CEEC Extension include the following:

- (11.3) But y<sup>e</sup> sight best pleased me was y<sup>e</sup> cartoons by Raphael, w<sup>ch</sup> are far beyond all y<sup>e</sup> paintings I ever saw. **They** are brought from y<sup>e</sup> Tower and hung up ther, and **are copying for my Lord Sunderland.**

(Charles Hatton, 1697; HATTON 2, 229)

- (11.4) **Grand preparations are making for Prince of Oranges Wedding**, which they say will be the 10<sup>th</sup> of this Month, it should have been sooner, but that he was taken ill at the Dutch chappel, (David Garrick, 1733; GARRICK, I, 8)

Beal (2004: 80) points out that Samuel Johnson disapproved of the progressive passival and offered as an alternative the construction prefixed by *a*:

The grammar is now printing, brass is forging ... in my opinion a vitious expression probably corrupted from a phrase more pure but now somewhat obsolete: a printing, a forging ... (quoted in Beal 2004: 80)

The *a*-prefixed progressive was already old-fashioned and in decline in the course of the eighteenth century. In CONCE, the progressive passive takes over from the passival form in the course of the nineteenth century, although both are extremely rare (Smutterberg 2005: 128). Arnaud’s corpus of late eighteenth-century and nineteenth-century private letters (reanalysed by Smutterberg 2005) does not contain passive progressives before 1800 (Smutterberg 2005: 129), and in the CEEC Extension the only instance of the new progressive passive is from 1780:

- (11.5) If they can but drive them to give him the Marines, **which it seems are being kept *in peto* for Sr Hugh**, it will be doing 2 right things at once.

(Lady Sarah Lennox, 1780; LENNOX, I, 302)

Used by an aristocratic woman in a letter to a close friend towards the end of the century, this single incident echoes Smitterberg's (2005) and Hundt's (2004) suggestion that the passive form emerged initially in informal contexts, such as private letters and diaries.

The next section presents the findings of sociolinguistic variation of the progressive in CEECE.

### 11.3 Diachronic developments in CEECE

In eighteenth-century letters, the progressive is a low-frequency item (1,741 tokens in total), and it increases moderately throughout the period from 4.45 to 10.88 (/10,000). Table 11.1 shows its distribution in 20-year time periods, and Figure 11.1 illustrates its development in terms of normalised frequencies. Figure 11.2 shows how many percent of the words in each person's letters are progressives per twenty-year period.<sup>2</sup> I have examined present- and past-tense progressives as two separate categories due to their high frequency in the data; perfect-tense progressives and modal usage have been grouped together into the so-called 'complex' tense.

Table 11.1 Frequencies of the progressive (N and /10,000)

1680–1699	1700–1719	1720–1739	1740–1759	1760–1779	1780–1800
158 4.45	166 5.94	141 6.61	227 6.09	311 7.98	738 10.88

In Figure 11.2, each thin horizontal line represents the normalised frequency of progressives in one person's letters, while the thick horizontal line indicates the median of the normalised frequencies (see Chapter 5). These median frequencies indicate that there is a drop in the tokens between 1720 and 1760. The decrease between 1700–1719 and 1720–1739 is not significant, but the increase between 1760–1779 and 1780–1800 is. The slight mid-century gap is shown also in Figure 11.1, but the earlier gap which the beanplots reveal in the period 1720–1739 disappears when the figures are normalised as a pooled average per period. The mid-century gap seems to result from a temporary drop in present tense progressives (Figure 11.4). But the progressive is clearly on the increase, given that its higher frequencies in 1780–1800 appear both in normalised figures and the beanplots. Moreover, the Wilcoxon rank-sum test indicates that the difference between the last two periods is statistically significant. The final period also stands out according to the permutation testing method.

2. Beanplot and cucumiform figures generously provided by Tanja Säily.

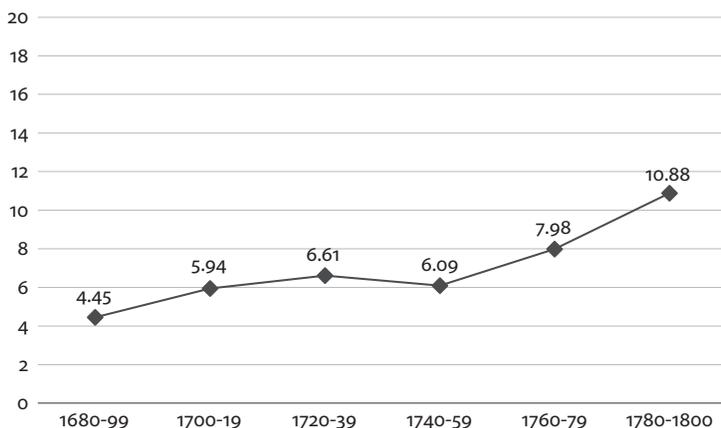


Figure 11.1 Normalised frequencies of the progressive /10,000 words

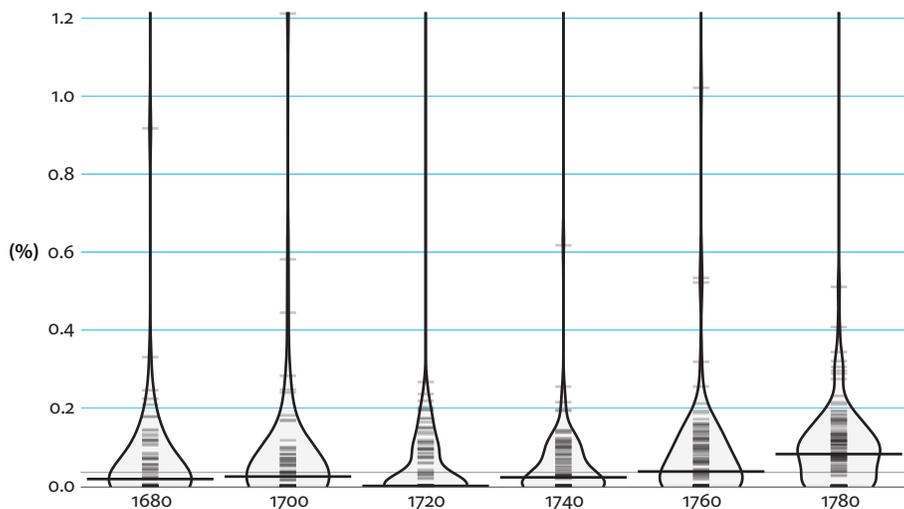
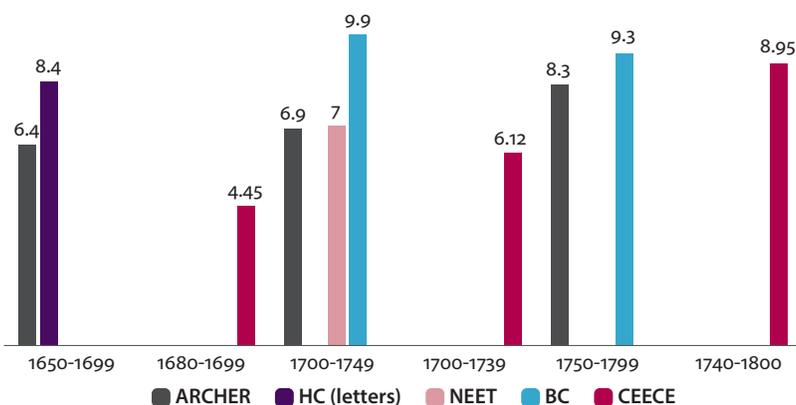


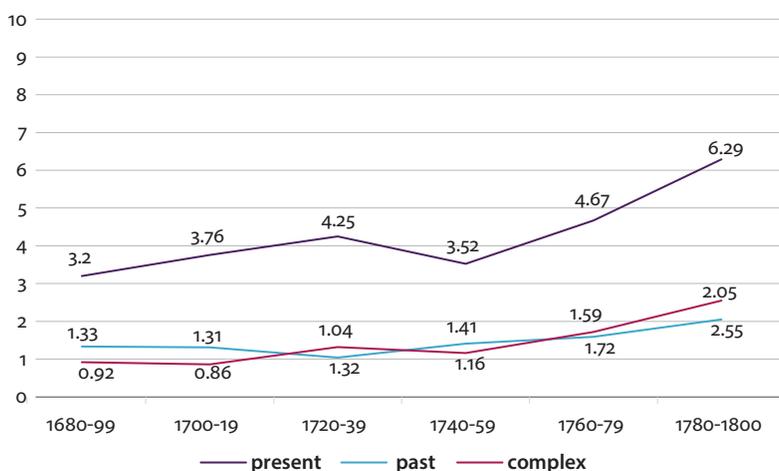
Figure 11.2 Beanplotting the progressive: percentages per individuals and time periods

The 1680–1719 period in CEECE shows lower frequencies compared to ARCHER (Hundt 2004) and the seventeenth-century letters in the *Helsinki Corpus* (Núñez-Pertejo 2004) (Figure 11.3). The informal context of private letter-writing would seem to predict higher figures, but this does not take place. The *Bluestocking Corpus*, a small letter corpus which consists of private correspondence within a social network (Sairio 2009), contains a higher frequency of the progressive than CEECE, but the difference evens out at the end of the century.



**Figure 11.3** Normalised frequencies of the Late Modern progressive: CEECE and previous studies (Hundt 2004, Núñez-Pertejo 2004, Sairio 2009)

In CEECE, the present tense is the most common context for the progressive (Figure 11.4). The other tenses appear in modest, very slowly increasing numbers.



**Figure 11.4** Diachronical developments in tense

The mid-century drop in present-tense progressives seems to explain the overall decrease that takes place during that period (Figures 11.1 and 11.2). Sociolinguistic analysis in Section 11.5 suggests that this decrease results from a temporary drop in material by lower-ranking writers. The following sections present the results for gender, social rank, and register variation.

## 11.4 Gender

Previous research on gender variation in the history of progressive suggests that this construction does not display consistent trends in time. Men seem to be more prone to use the subjective progressive in early eighteenth-century letters (Fitzmaurice 2004), whereas in nineteenth-century letters progressives are more frequent in women's letters (Smutterberg 2005: 79–82; Arnaud 1998) and in the eighteenth-century Bluestocking letters gender is not a relevant variable (Sairio 2009). Kranich (2010: 105–106) suggests that the subject matter of the letters could explain the variation, and that emotional involvement might lead to higher frequencies of *be+ing*. In CEECE, clear gender-related patterns do not emerge. Normalised frequencies suggest that men use the progressive more than women up until the end of the eighteenth century (Figure 11.5, Table 11.2), which is in line with Fitzmaurice's (2004) findings on the subjective progressive. The beanplot in Figure 11.6 shows that men lead women in every time period, but this result is not statistically significant, and women's letters simply do not provide enough material for statistical analysis until the end of century.

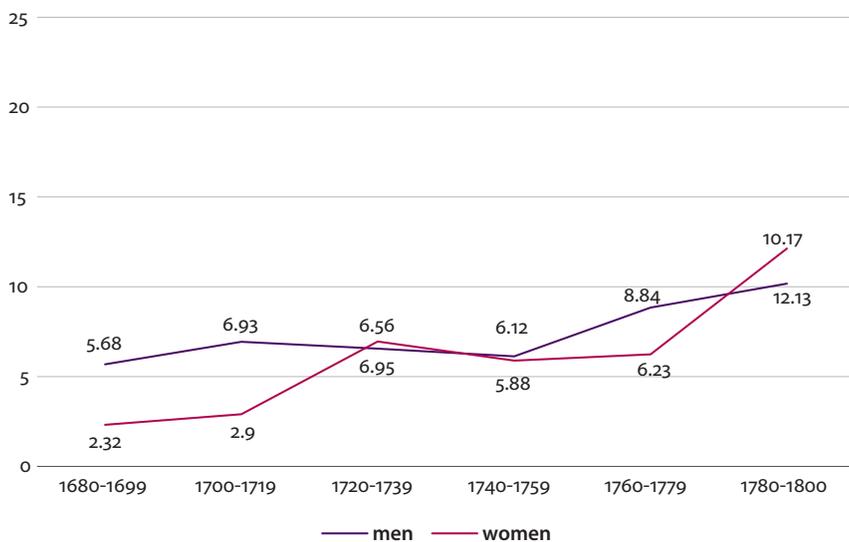
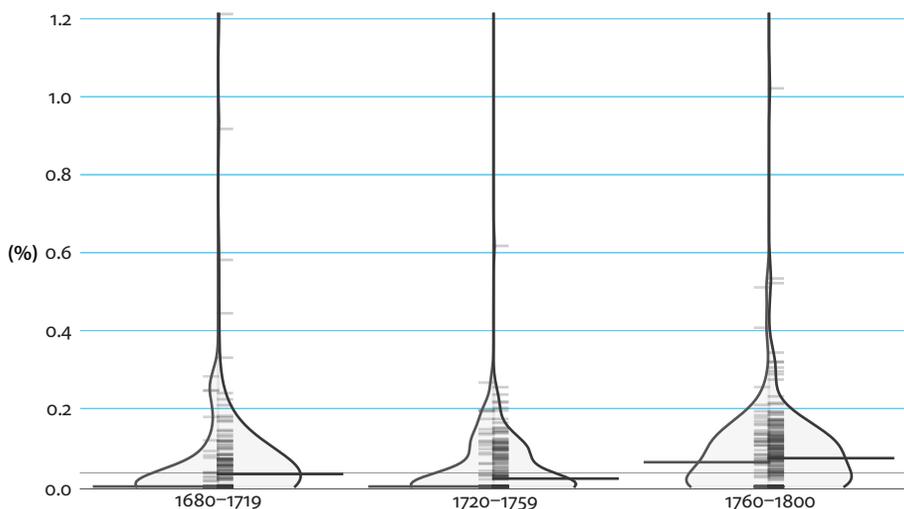


Figure 11.5 Gender variation in 20-year time periods

Table 11.2 The progressive and gender variation (N and /10,000 words)

Gender	1680–1699	1700–1719	1720–1739	1740–1759	1760–1779	1780–1800
Men	152 5.68	141 6.93	104 6.56	211 6.12	152 8.84	415 10.17
Women	6 2.32	25 2.9	37 6.95	53 5.88	34 6.23	323 12.13



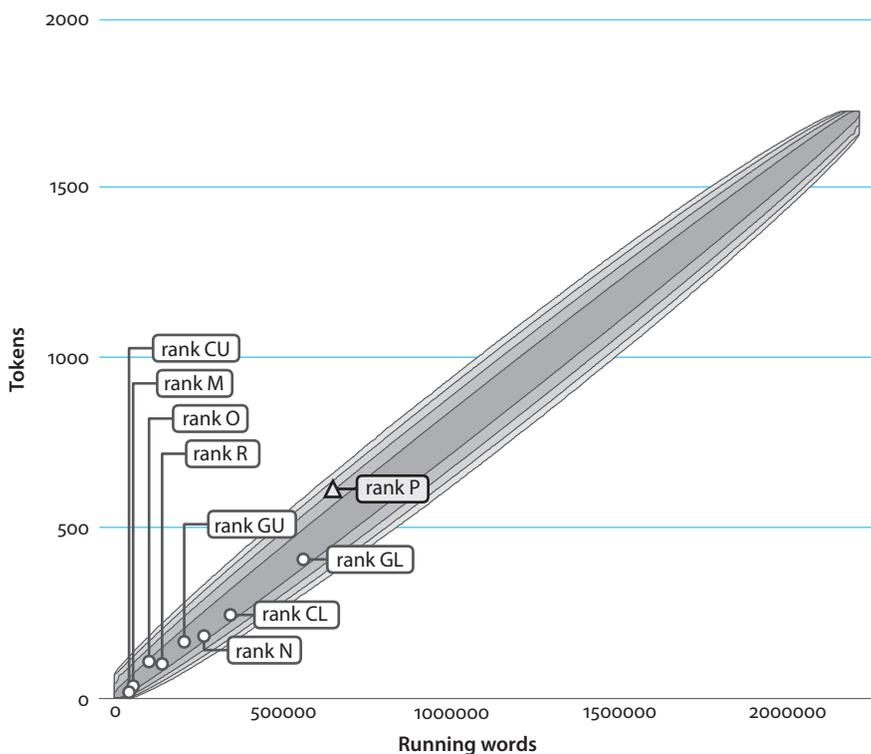
**Figure 11.6** Beanplotted gender variation in forty-year spans (women writers, left / male writers, right)

Women's lead at the end of the century, though statistically insignificant, seems to anticipate the nineteenth-century gender variation; perhaps gender becomes a relevant variable only in the course of the following century, when the change is more strongly under way. Smitterberg (2005: 86) interprets his findings via Labov's (2001: 292–293) hypothesis which links gender, norms, and innovation together, so that the nineteenth-century increase of the progressive can be viewed as innovation and change from below given how it may be considered to advance on the non-progressive constructions. Women's use of the progressive in that period could therefore correspond with Labov's hypothesis of women as linguistic innovators. However, in the eighteenth century the progressive does not yield this type of socially meaningful findings.

As gender variation is not statistically significant, a more fine-grained analysis of gender and tense is unnecessary (see, however, Table 11.3 in the Appendix for the numbers). Suffice it to say that women begin to catch up from 1720 onward in present-tense usage, and they take the lead at the end of the century.

## 11.5 Social rank

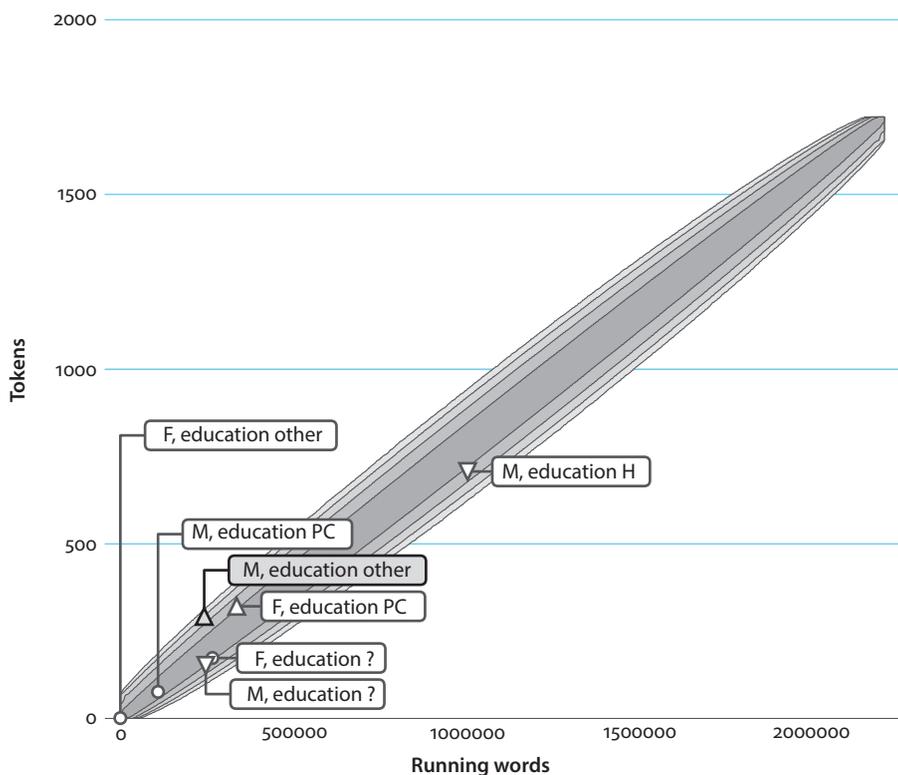
The impact of social rank is more substantial than gender. The progressive is very common in the letters of lower-ranking writers, and professionals stand out from the other ranks as high users (Figure 11.7). Permutation testing shows that the overuse by professionals is statistically significant in the case of the 'complex' progressive (i.e. perfect tense and modal usage combined).



**Figure 11.7** Social rank and the progressive. (CL = clergy lower, CU = clergy upper, GL = gentry lower, GU = gentry upper, M = merchant, N = nobility, O = other, P = professional, R = royalty)

The variable of social status seems to shed light on the early and mid-century gap observed in Table 11.1 and Figure 11.2. Perhaps it also offers some explanation regarding women's infrequent use of the progressive in the beginning of this period. Lower-ranking writers are underrepresented in this early section of CEECE; there are no progressives by lower rank writers in the dip period of 1740–59, which is explained by the temporary decrease of material in those decades. And while the period of 1680–99 provides a wider range of letters by women in terms of rank than 1700–19, neither period contains letters by women of the lowest ranks.

Men who were not highly educated (that is, men who were apprenticed or had only gone through elementary and secondary education) use the progressive in significantly high frequencies (Figure 11.8). This would support the finding that the lower ranks overuse the progressive and that we are looking at change from below. However, when we consider the influence of rank in the last two decades (Table 11.4), singled out from the other time periods because of the overall higher *be+ing* tokens and word counts per rank, we find that at the end of the century the nobility and lower gentry writers use the progressive more or as often as lower rank writers (coded as Other).



**Figure 11.8** The progressive and education of the letter-writers. (F = Female, M = Male, PC = Private Classical education, H = Higher education, including Oxbridge and Inns of Court)

**Table 11.4** The progressive and social rank, 1780–1800

Rank	N	/10,000	Word count
Royalty	79	6.61	119,491
Nobility	84	14.22	59,082
Gentry upper	18	4.54	18,077
Gentry lower	18	12.26	101,939
Clergy upper	0	0	1,729
Clergy lower	81	11.23	72,107
Professionals	283	11.43	247,491
Merchants	0	0	762
Other	68	12.65	53,773

To illustrate the use of the progressive in lower-ranking letters, in Example (11.6) Ignatius Sancho (1729–1780) gives a lively, in-the-moment eyewitness account about the Gordon Riots of June 1780, which he observed from his shop in London.

- (11.6) Lord George Gordon has this moment announced to my Lords the mob – that the act shall be repealed this evening: – Upon this, they gave a hundred cheers – took the horses from his hackney-coach – and rolled him full jollily away: – **They are huzzaing now ready to crack their throats.** *Huzzah.*  
(Ignatius Sancho, 1780?; SANCHO, 232)

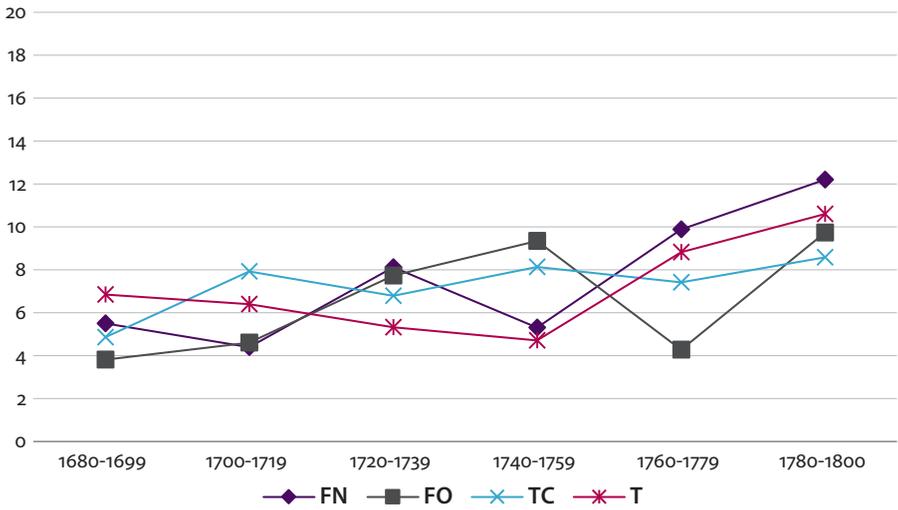
The lower-rank letters are unevenly distributed in the corpus, and there is a lot of inter-decade variation in the tokens. Idiosyncratic preferences are highly visible, and individual writers have considerable influence in the overall results: for example, almost 40% of the progressives in the 1780–1800 Other category appear in Elizabeth Clift’s letters (Table 11.4). When individuals have this much influence in the frequencies of a rare variable, the material inevitably leads towards micro-level sociolinguistic analysis.

Upper clergy writers in CEECE appear to be underusers of the progressive: their letters include a total of 35 tokens in the corpus. We can explore the hypothesis that upper clergymen’s tendency to avoid the progressive may point to genre-internal variation. Kranich’s (2010: 96–103) analysis of the distribution of the progressive across genres in ARCHER-2 and the analysis of Smitterberg (2005), Núñez-Pertejo (2004), Fitzmaurice (2004), Strang (1982) and Arnaud (1973) indicate that the progressive is an oral rather than a literate feature, seemingly preferred in more speech-based and colloquial written genres such as letters, drama, and fiction (Kranich 2010: 102). Eighteenth-century religious texts in ARCHER-2 contain only 17 progressives in all (Kranich 2010: 101). Religious texts are thus an atypical genre for the progressive, so it might follow that upper clergy letters are an unfavourable environment for the progressive to occur. For the lower clergy letters the hypothesis does not apply, as this material include letters by women whose fathers or husbands represent this rank.

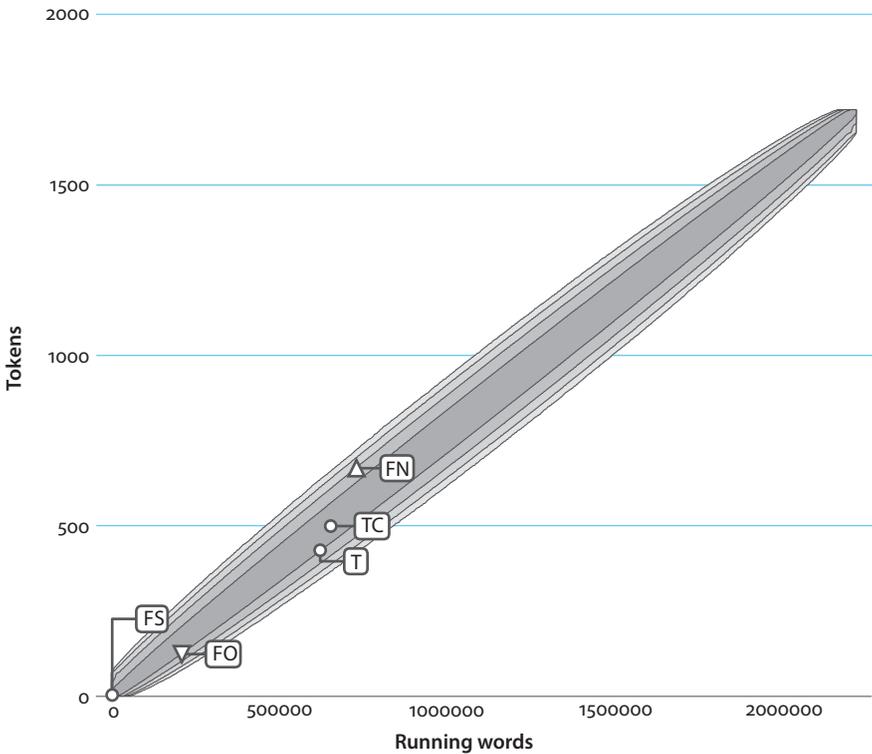
The next section discusses the influence of register, which adds another important dimension to women’s increasing use of the progressive.

## 11.6 Register

Overall, correspondence between nuclear family members (FN in Figures 11.9 and 11.10) is the most common context for the progressive. This would confirm the associations of the progressive with “more spontaneous, unmonitored, colloquial” language use (Kranich 2010: 102). Family communication seems to be the most likely context for this dynamic, oral-like verb construction to occur.



**Figure 11.9** Diachronic register variation (/10,000 words). (FN = Family Nuclear, FO = Family Other, TC = Close Friends, T = Acquaintances)



**Figure 11.10** Register variation and the progressive

Close family correspondence becomes the most common register for the progressive after mid-eighteenth century: prior to the 1760s, register variation is less clear. The most intimate contacts (FN and TC) use progressive significantly more often than more distant writers (Wilcoxon rank-sum test, last 40-year period). The progressive rates for T writers remain consistently low, whereas in other registers this aspect slightly increases over time.

In Example (11.7), Lady Mary Montagu anticipates her daughter's response to the letter she is writing, which builds into an informal, conversational style:

- (11.7) **I fancy you are now saying** – 'Tis a sad thing to grow old. What does my poor mama mean by troubling me with Criticisms on Books that no body but her selfe will ever read over? – You must allow something to my Solitude.

(Lady Mary Wortley Montagu, 1752?; MONTAGU, III, 9)

Towards the end of century, several sociolinguistic developments take place. Women begin to use the progressive more frequently, the construction increases particularly in the present tense, and lower-ranking writers (but also the nobility!) use the progressive in the highest frequencies. In the next section, we arrive at micro-level analysis of the individual letter-writers.

## 11.7 Outliers

William Clift, aged 17 to 24 when he wrote the letters in CEEC Extension, uses the progressive more than any other letter-writer in terms of normalised frequencies (Table 11.5). Also the permutation testing method places him as the highest user of the progressive (Table 11.6), and unlike the other outliers, William Clift uses the progressive in a multitude of tenses, not merely in present tense constructions. It is hardly a coincidence that out of 308 letter-writers in this corpus, the second person on the list of outliers is William's sister Elizabeth (1757–1818). These over-users of the progressive are nuclear family members, originally from Cornwall, with labourer roots (though William was upwardly mobile), and they write during the last decade of the eighteenth century to their siblings. In addition to family register,

Table 11.5 Outliers: highest frequencies of the progressive per tense

	Present	Past	Complex	N / 10,000
William Clift	21	25	21	67 / 23.12
Elizabeth Clift	19	6	1	26 / 18.26
Josiah Wedgwood	23	5	8	36 / 15.74
Mary Wollstonecraft	33	2	11	46 / 14.42
William Jones	31	5	5	41 / 12.36

Table 11.6 Outliers per tense: calculated with permutation testing\*

Tense	Individual	Side	p-value	q-value
ing-complex	Jeremy Bentham	above	0.00298731	1.60119816
ing-past	William Clift	above	0.01589152	4.25892736
ing-complex	William Clift	above	0.01651846	2.95129818666667
ing	William Clift	above	0.01805818	2.41979612
ing-past	Thomas Gray	above	0.0387751	4.15669072
ing-present	M. Wollstonecraft	above	0.04754548	4.24739621333333

\* While the results are not significant after false discovery rate control (see *q*-value), they are indicative of tendencies

their outlier status may also result from regional language patterns in Cornwall and the possible Celtic influence in the West Country region, and probably reflects shared linguistic patterns in the family.

In Table 11.5, we find also Sir William Jones (1746–1794), a notable philologist and member of the Supreme Court in Bengal, Josiah Wedgwood (1730–1795), the founder of the Wedgwood pottery, and Mary Wollstonecraft (1759–1797), writer and women's rights advocate. In addition to William Clift's West Country background, it may be significant that he is the youngest of these people and also twenty years younger than his sister Elizabeth, who is the second youngest outlier. Generational differences may be at work. Wedgwood and Wollstonecraft use the present-tense progressive more than William Clift does, but Clift is more accustomed to using past and perfect progressives than they are and thus has a wider syntactic repertoire.

Table 11.5 includes the individuals whose letters contain the highest number of the progressive in normalised frequencies; Table 11.6 has been compiled using the permutation testing method of the individuals who provide sufficient data for quantifying outlier positions. This list is headed by the philosopher, jurist and social reformer Jeremy Bentham (1748–1832), who did not make the normalised list but shows up in Table 11.6 because of his significantly high use of the progressive in complex tense (i.e. present perfect, past perfect and modal usage denoting the future tense). Another letter-writer who frequently uses past tense progressives is eighteenth-century poet and scholar Thomas Gray (1716–1771), and Mary Wollstonecraft makes the list with her present-tense usage.

Table 11.7 presents a compilation of the outlier findings. The letters of the outliers are written between 1734 and 1800, mainly in the latter half of the century. The oldest (Grey) was born in 1716, and the youngest (William Clift) in 1775. Over half of the progressives in their letters appear in letters to close family members (FN 56%, Table 11.7). Four out of seven outliers represent the professional ranks. These

individuals personify the habitual users of the progressive in eighteenth-century private letters: most of them write during the latter part of the century, particularly the last decade, and represent the low and middling sorts, with William Jones and William Clift as notable social risers. The list includes both men and women, which is illustrative of the insignificant gender variation at the end of the century. Their geographical backgrounds vary, but London is a common denominator to nearly all of them. The Clifts hail from Cornwall, from where William Clift relocated to London; Jeremy Bentham and Thomas Grey were Londoners who spent some years in Oxford and Cambridge respectively; William Jones was a Londoner educated in Oxford, who travelled to India in 1783 and spent the rest of his life there. At the time Mary Wollstonecraft wrote these letters, she was living in London and in France. Josiah Wedgwood lived in Staffordshire in the West Midlands. The vibrant capital city thus makes an appearance in this study.

**Table 11.7** Outliers: timeline, rank, and register

Years of activity	Rank
Grey: 1734–1767	Grey, Wedgwood, Wollstonecraft, WClift: Professional
Bentham: 1761–1800	EClift: Other
Jones: 1765–1793	Jones: Gentry Upper
Wedgwood: 1767–1793	Bentham: Gentry Lower
Wollstonecraft: 1780–1797	
WClift: 1792–1799	<b>Register: % and N</b>
EClift: 1794–1799	FN: 56% (148)
	TC: 24.2% (64)
	T: 19% (50)
	FO: 0.8% (2)

As a final comment, the progressive is present also in non-native written English in the eighteenth century. Joseph Emin (1726–1809), an Armenian soldier devoted to the cause of liberating his native country, was an adult learner of English who found friends and supporters among the British gentry and was sponsored as a cadet officer in the Royal Military Academy (Fisher 2004, Apcar ed. 1918). In 1757, he wrote to his friend Elizabeth Montagu from continental Europe:

(11.8) **I am glad you have been amusing your dearself** seeing different Places I wish it may do you good and add to your Health; but I am sory to find you are so much discouraged for you shall not be my Queen if you don't have as great a Heart as your great Soul [...]

(Joseph Emin to Elizabeth Montagu, 1757, in Apcar ed. 1918: 84)

This informally tinged present-perfect progressive shows that an eighteenth-century L2 writer who had lived and worked in London for years had obtained the *be+ing* construction into his written language. The progressive aspect was thus part of the language repertoire of a non-native speaker.

## 11.8 Conclusion

In eighteenth-century letters, the progressive appears to be unobtrusive, slowly emerging language change in its early stages. In terms of syntactic forms and pragmatic functions it was already a relatively mature construction, but it simply was not yet used to the extent that it would come to be used. Increase over time was slow and perhaps boosted from the lower strata of society, but not unequivocally. The progressive appears most frequently in family correspondence, and this intimate and familiar context seems to explain why rank and gender alone do not account for the variation. The frequent progressive use by lower-ranking writers, such as the outliers William Clift and Elizabeth Clift, is probably influenced by the oral-like, informal characteristics of this construction, and in the Clifts' case their Cornwall background and connections to possible Celtic influence in the English progressive might be involved. Nevertheless, a close relationship between the writer and the recipient seems to be a key sociolinguistic element.

Future research of the progressive in CEECE should put to the test the hypothesis that women's more frequent use of the progressive may result from their more frequent use of the subjective progressive, while the aspectual progressives would not display a gender difference (Kranich 2010: 233–234). Sociolinguistic analysis of the functional developments of the progressive is yet to be carried out in CEECE.

## Appendix

Table 11.3 Tense and gender: absolute and normalised figures (/10,000)

Present	1680–99	1700–19	1720–39	1740–59	1760–79	1780–1800
Men	90 3.36	86 4.22	68 4.29	98 3.45	152 5.21	211 5.17
Women	2	17 1.97	23 4.32	35 3.88	34 4.00	217 8.15
Past	1680–99	1700–19	1720–39	1740–59	1760–79	1780–1800
Men	36 1.35	31 1.52	15 0.95	43 1.51	51 5.21	86 2.11
Women	3	7 0.81	7 1.31	9 1.00	9 1.06	52 1.95
Complex	1680–99	1700–19	1720–39	1740–59	1760–79	1780–1800
Men	26 0.97	24 1.18	21 1.32	33 1.16	55 1.88	118 2.89
Women	1	1	7 1.31	9 1.00	10 1.18	54 2.03

## Change or variation? Productivity of the suffixes *-ness* and *-ity*

Tanja Säily

### 12.1 Introduction

Adamson (1989: 204) points out that English has a “double lexicon”, in which almost all native words have Romance or Latinate synonyms whose use is sociolinguistically conditioned. Synonymous pairs of this kind can also be found among derivational affixes, such as the nominal suffixes *-ness* and *-ity*. Results from previous diachronic research (Säily & Suomela 2009, Säily 2011, Säily 2016) suggest a stable gendered style in which women use *-ity* significantly less productively than men, while there is no gender difference in the use of *-ness*.

However, in 18th-century letters in the *Corpora of Early English Correspondence* (CEEC) there is no gender difference with either suffix (Säily 2016). Instead, what emerges as significant is an increase in the productivity of *-ity* over time, a trend that is also apparent in the 17th-century section of the CEEC. There also seem to be differences between social classes, so that, e.g., the productivity of *-ity* is significantly high among professionals. To uncover reasons for this variation, the 18th-century section of the CEEC is here subjected to a more fine-grained analysis.

Firstly, the present study analyses the role played by register in terms of the relationship between the sender and the recipient of the letters, comparing it to the effect of social class at the macro level as well as that of individual outliers. Secondly, the study explores the question of whether the variation in the productivity of *-ity* represents a linguistic change or a shift in style, by comparing letter collections from different social backgrounds. Thirdly, the study seeks to explain the disappearance of the overall gender difference in the productivity of *-ity* by analysing semantic change in *-ness* and *-ity* between the 17th and 18th centuries, again examining whether the change is a linguistic or stylistic development. Finally, the study investigates links between normative grammar and the productivity of *-ness* and *-ity*.

The remainder of this chapter is organised as follows. Section 12.2 surveys the theoretical and methodological background to the study, beginning with morphological productivity and proceeding to sociolinguistics. Section 12.3 focuses on previous research on sociolinguistic variation in the productivity of *-ness* and *-ity*.

Based on the preceding sections, Section 12.4 formulates the research questions explored in the study. Section 12.5 presents the results, while Section 12.6 discusses them and concludes the chapter.

## 12.2 Theoretical background

Morphological productivity was defined by Bolinger (1948: 18) as “the statistically determinable readiness with which an element enters into new combinations”. Taking his idea further, several corpus-linguistic measures of productivity were developed by Harald Baayen and his associates in the 1990s and early 2000s. Three of these measures have stood the test of time and are still recommended by Baayen (2009). The measures are based on *type frequency* (the number of different words formed using the element), *token frequency* (the number of all word tokens formed using the element; if a word occurs more than once in the corpus, it is counted each time) and the frequency of hapax legomena, also known as *hapax frequency* (the number of words formed using the element that occur only once in the corpus). The first of Baayen’s measures is *realised productivity*, which is simply defined as type frequency, and assesses the extent of use of the element. The second measure is *potential productivity*, which is hapax frequency divided by token frequency, and estimates the growth rate of the element. The final measure is *expanding productivity*, which is hapax frequency divided by the number of all hapax legomena in the corpus, and estimates the share of the element out of the overall vocabulary growth.

Baayen’s measures are dependent on the size of the corpus, which makes it problematic to compare measurements obtained from differently sized subcorpora representing, e.g., time periods or social categories. Normalising type and hapax frequencies is not recommended because this would presuppose that they grow linearly with the size of the corpus, which is an invalid assumption (see further Säily & Suomela 2009). Säily & Suomela (2009) present a robust method for comparing type and hapax frequencies across subcorpora of varying sizes. Based on accumulation curves and the statistical technique of permutation testing, the method is data-driven, highly visual and provides a built-in measure of statistical significance. The method is described in more detail in Chapter 5 of this volume. Säily & Suomela (2009) and Säily (2011) show that hapax-based measures are unusable in small corpora, so the results presented here are based on type frequency alone, using two measures of corpus size: (1) the number of running words in the corpus and (2) the token frequency of the element (see Säily 2016 for further discussion of the two).

Romaine (1985) is one of the first to consider productivity from a diachronic perspective. Starting from the assumption that change stems from variation, she hypothesises that competing patterns of word-formation establish themselves through

social or stylistic specialisation. Thus, after entering the language in the Middle English period through French loanwords and later through calques on Latin, *-ity* becomes like a more learned and prestigious version of *-ness*, favoured in somewhat different genres and restricted to certain Romance/Latinate bases (Marchand 1969: 312–313, 334–335; Barber 1976; Riddle 1985; Dalton-Puffer 1996: 120, 128; Cowie 1999: 248, 224). In addition, Riddle (1985) argues that the suffixes undergo semantic specialisation, so that *-ity* becomes more and more associated with the meaning ‘abstract or concrete entity’ and *-ness* with that of ‘embodied attribute or trait’. In the Middle English period, there is also regional variation in the productivity of the suffixes (Gardner 2013: Chapter 5). Further sociolinguistic research on the suffixes is surveyed in the next section.

Several findings from present-day sociolinguistics are of possible relevance to this study. Labov (2001: 259) proposes that linguistic change from below follows a *curvilinear pattern*, originating in the interior social classes rather than the highest or lowest-status group. Another study emphasising the role of the middle class is Milroy (1992: 213), whose *anvil-shaped status/solidarity model* suggests that the socially mobile middle classes have the least close-knit social networks and are thus the most open to innovation. Class is not all that matters, however. Wolfson (1990) shows that the language use of middle-class Americans varies depending on register, or more specifically, the social distance between them and their interlocutors, so that minimally and maximally distant relationships cause similar speech behaviour, whereas less stable relationships, such as those between status-equal friends, are markedly different. She calls this the “*bulge*” theory.

The relative importance of social variation (including class-based variation) and situational or register variation has been hotly debated in the literature. Bell (1984: 151) claims that register variation derives from social variation, whereas Finegan & Biber (2001) see register variation, defined in terms of genre, as primary and claim that it predicts social variation. Like Bell (1984), Nevalainen & Raumolin-Brunberg (2003: 191) make a distinction between register variation according to topic and that according to addressee, hypothesising that social variation is more fundamental than register variation according to addressee, which in turn is more fundamental than register variation according to topic. They test the first part of the hypothesis using the *Corpus of Early English Correspondence* and are able to confirm it in the changes they study, which involve the variables *ye/you*; *-th/-s*; the object of the gerund; *mine*, *thine/my*, *thy*; and *the which/which*. However, the social factors they analyse are region and gender, while social class or rank is not included. The question of the relative importance of class- and register-based variation in the history of English thus appears to remain open, certainly in the field of derivational productivity, which presents the additional complication of the absence of an obvious linguistic variable (see Chapter 5).

Another pertinent issue is that of linguistic versus stylistic change. The key question here is whether the variation and change in the frequency of a linguistic feature observed in a corpus is indicative of language change, as in a change in the grammar of a speech community, or whether it could be, e.g., part of a stylistic shift in the genre(s) represented by the corpus (Nevalainen 2008). A major stylistic shift in recent English is the colloquialisation of many genres of written language (Hundt & Mair 1999). However, some written genres, such as scientific writing, display the opposite tendency of becoming more formal (Biber & Finegan 1997). Colloquialisation can be linked to the sociocultural trend of democratisation; other sociocultural trends influencing language use include urbanisation and standardisation (Culpeper & Nevala 2012). As noted in Chapter 3, the standardisation process of the English language reached a key phase in the 18th century with the rise of normative grammar.

As for how to assess whether a change is linguistic or stylistic, Nevalainen (2008: 32, 34) suggests that if the change is linguistic, it should show up in more than one type of material. If the change is stylistic, on the other hand, we should be able to find other linguistic features participating in the stylistic shift in the corpus (see also Leech & Smith 2005: 89). Szmrecsanyi (2016) proposes the use of statistical modelling as a tool for disentangling grammatical and environmental change: if there is statistically significant change in the effect of a language-internal constraint on the choice of a variant, the change is grammatical, i.e., linguistic. As noted above, however, not all changes present an obvious variable to analyse. Another means is to consider the way in which the change spreads (cf. Nevalainen 2008: 34): a linguistic change is more likely to follow Labov's curvilinear pattern and Milroy's anvil model mentioned above.

### 12.3 Previous research

Sociolinguistic variation in the productivity of the nominal suffixes *-ness* and *-ity* has been studied in 17th-century correspondence (Säily & Suomela 2009), 18th-century correspondence and trial proceedings (Säily 2016) as well as various genres of Present-day English (Säily 2011). The main finding has been that of a stable gender difference in the use of *-ity*: its productivity has been significantly low with women in every corpus except for the 18th-century section of the CEEC.

Säily (2011) connects this difference to a wider trend of men tending to use more nouns than women, while women use more pronouns than men. This has been observed for early English correspondence by Säily et al. (2011) and for spoken and written Present-day English by Rayson et al. (1997) and Argamon et al. (2003), respectively. Argamon et al. (2003) and Säily et al. (2011) interpret the tendency

in terms of Biber's (1988) multidimensional analysis of register variation, which considers the co-occurrence patterns of a number of linguistic features in texts. Biber's Dimension 1 is labelled Informational vs. Involved Production. One of the key features of the informational style is a high frequency of nouns, whereas the involved (interactive and affective) style is characterised by, e.g., first- and second-person pronouns. Thus, the gender difference is explained by the notion of gendered styles: men's style is said to be more informational, whereas women's style is more involved.

The idea presented in Säily (2011) is that if women use fewer nouns than men, they will also use nominal suffixes less productively than men. The question then becomes why women only use *-ity* and not *-ness* less productively than men. Säily (2011) suggests that the semantics of *-ness* is in fact well suited to an involved writing style, as *-ness* is often used to describe embodied attributes or traits, whereas *-ity* is generally used for abstract or concrete entities (Riddle 1985). Furthermore, it would seem that *-ness* often co-occurs with possessive pronouns, as in *your kindness* (see also Baeskow 2012). In a study of the multi-genre ARCHER corpus, Cowie (1999: 224, 260–264), too, tentatively connects *-ness* with an involved style and *-ity* with an informative style based on the genres that they prefer (fiction and sermons vs. scientific and medical prose), but argues against the semantic differentiation proposed by Riddle.

As for why there is no gender difference in the use of *-ity* in 18th-century letters, Säily (2016: 145) suggests that they go against the general tendency because “the educated, culturally homogeneous literati overrepresented in the corpus had developed a style of their own, a nominal style shared by both men and women”, referring to Biber & Finegan (1997), who found that the letter-writing style of the 18th century was more involved but also more elaborated than that of the 17th century in ARCHER (see also Cowie 1999: 222). The development of a shared style only became possible with upper-class women's increased access to education, although they still lacked the wealth of opportunities their male counterparts had (Tieken-Boon van Ostade 2010). According to McIntosh (2008: 231), 18th-century British culture could even be called “feminized”, meaning that the feminine virtues of politeness and sensibility were required of anyone wishing to belong to the better sort, and this was also evident in language use. Fashionable literary salons were hosted by women, and there were more female authors than ever.

In 17th-century correspondence, Säily & Suomela (2009) find that the productivity of *-ity* increases significantly over time, while there is no change in the productivity of *-ness* (see Palmer 2009: 282 for evidence of an increase in the productivity of *-ity* as early as the 16th century in the CEEC). Säily (2016) observes a similar trend over time in the 18th century. The increase in the productivity of *-ity* is corroborated by Lindsay & Aronoff's (2013) study of the *Oxford English*

*Dictionary* (OED), which discovers that the number of new *-ity* derivatives begins to increase drastically from the 17th century onwards and has continued to rise ever since. As noted above, Säily (2016) finds no gender differences in the 18th-century section of the CEEC, but does uncover some class-based differences. The present study analyses 18th-century correspondence in more detail, taking into account, e.g., the relationship between the sender and the recipient of the letters. Furthermore, this study takes advantage of the new version of the software used to compute the results (Suomela 2014), which now provides false discovery rate control (Benjamini & Hochberg 1995), enabling a more accurate assessment of statistical significance.

## 12.4 Research questions

This study explores the following research questions.

1. Do the results reported in Säily (2016) on the 18th-century section of the CEEC remain significant after false discovery rate control?
2. What is the relative importance of class-based variation and register variation according to addressee in the change in the productivity of *-ity*?
3. Is the change in the productivity of *-ity* linguistic or stylistic?
4. Is there a change in the semantics of *-ness* and *-ity* between the 17th and 18th centuries, and if so, is it linguistic or stylistic?
5. Does normative grammar have an influence on the productivity of *-ness* and *-ity*?

Question 1 will be dealt with in Sections 12.5.1 and 12.5.2, the latter of which also touches upon question 2. To address questions 2–4, a series of case studies will be conducted in Section 12.5.3, while question 5 will be discussed in Section 12.5.4.

## 12.5 Results

### 12.5.1 Overall trends

In the use of *-ness*, none of the time periods studied differs significantly from the corpus as a whole. In the use of *-ity*, however, there is a clear change over time. With 40-year periods, for example, the productivity of *-ity* is significantly low in the first period, neither low nor high in the middle period, and significantly high in the last period (Figure 12.1), which indicates that the productivity increases over time, confirming the observation by Säily (2016). These results are similar

to those observed in the 17th-century section of the CEEC by Säily & Suomela (2009: 105–106), so the increasing productivity of *-ity* seems to be a long-term trend in early English correspondence.

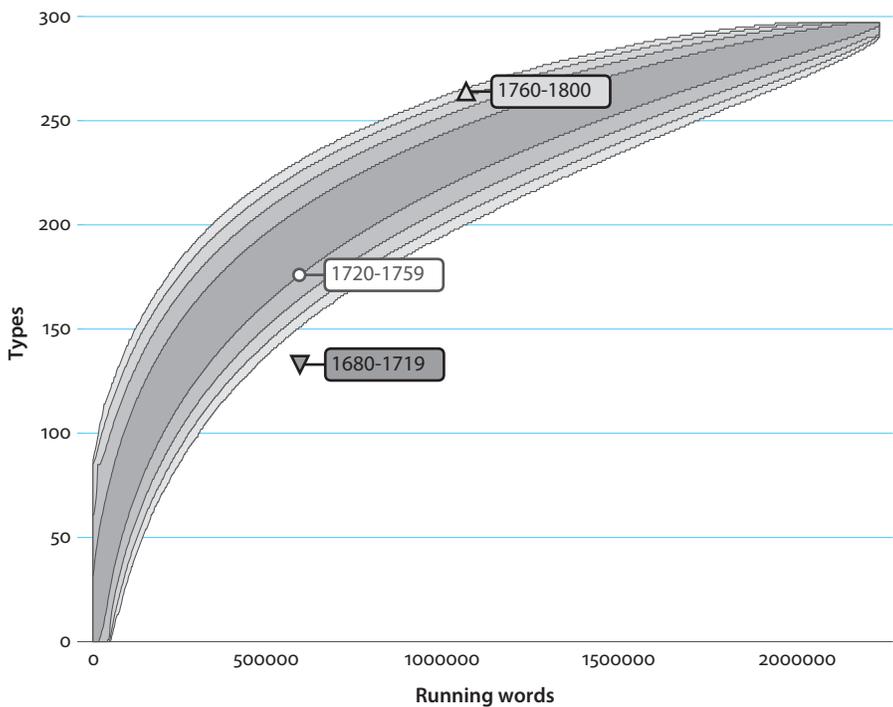


Figure 12.1 Change in the productivity of *-ity* over time, 1680–1800

### 12.5.2 Social categories

As expected, there is no gender difference in the use of *-ness*. Furthermore, this study confirms the observation by Säily (2016) that there is no significant gender difference in the use of *-ity*, either, as shown in Figure 12.2. Given that the productivity of *-ity* seems to have been steadily growing in English correspondence since at least the 17th century and that women used *-ity* significantly less productively than men in the 17th century, the lack of a gender difference in the 18th century must mean that women have caught up with men in the use of *-ity*, rather than men decreasing their usage. Thus, the use of *-ity* undergoes ‘masculinisation’ rather than ‘feminisation’, at least in terms of productivity (but see Section 12.5.3.3 for an analysis of the changing semantics of *-ness* and *-ity* derivatives).

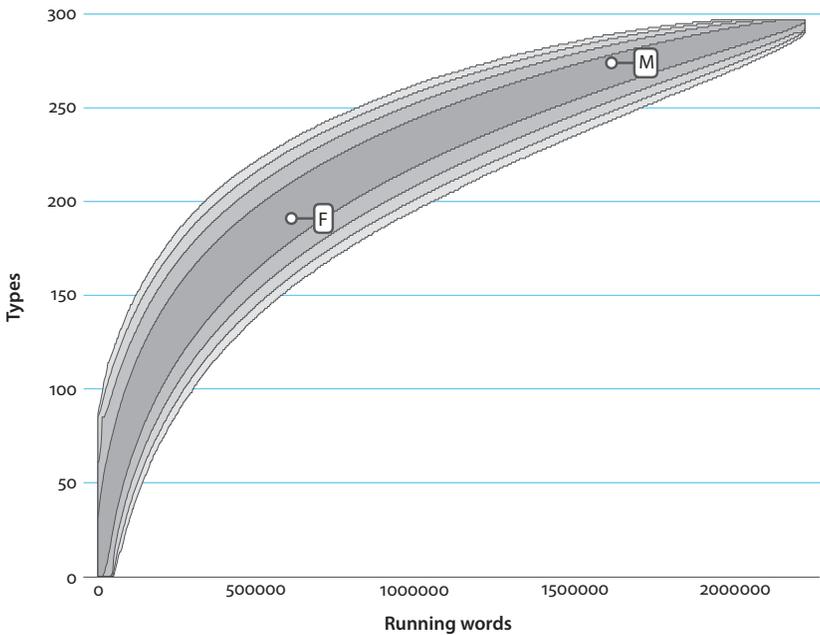


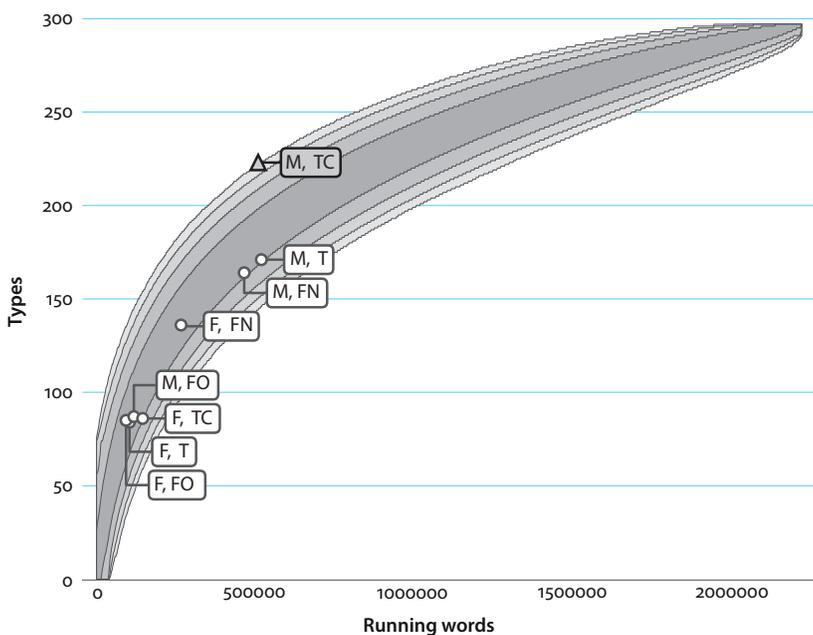
Figure 12.2 Gender and *-ity* types, 1680–1800

The gender difference remains in one register, however. As shown in Figure 12.3, when the data is divided up according to the relationship between the sender and recipient of the letter, the productivity of *-ity* is significantly high in letters written by men to their close friends (again, there are no significant differences in the use of *-ness*). While McIntosh (2008) claims that the masculine virtue of wit was no longer in vogue in the 18th century, these letters indicate otherwise:

- (12.1) I think I am a very tractable sort of a poet. Most of my **fraternity** would as soon shorten the noses of their children because they were said to be too long, as thus dock their compositions in compliance with the opinion of others. I beg that when my life shall be written hereafter my Authorship's **ductility** of temper may not be forgotten. William Cowper to Walter Bagot, 1789; COWPER W, III, 297 (COWPERW\_063)

It was not only male friends who were at the receiving end of such playful language use:

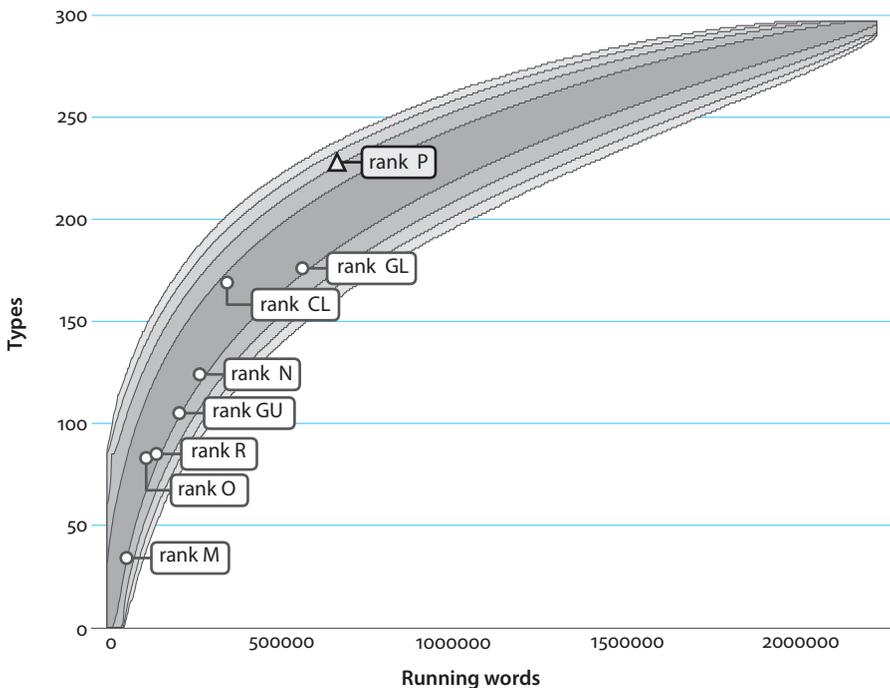
- (12.2) I protest, it is to me the most difficult of things to write to one of your female geniuses – there is a certain degree of **cleverality** (if I may so call it), an easy kind of derangement of periods, a gentleman-like – fashionable – careless – see-saw of dialogue – which I know no more of than you do of cruelty. Ignatius Sancho to Miss Crewe, 1778?; SANCHO, 132 (SANCHO\_019)



**Figure 12.3** Variation in the productivity of *-ity* across subcorpora based on the gender of the letter writer and the relationship between the writer and the recipient, 1680–1800 (FN = nuclear family, FO = other family, TC = close friends, T = other acquaintances). The productivity of *-ity* is significantly high with men writing to close friends ( $p < 0.00006$ )

In some cases, the use of *-ity* is perhaps more creative than productive, as evidenced by the self-conscious parenthetical remark in (12.2) above (for a discussion of creativity and productivity, see Bauer 2001: 62–71). It seems that men wished to amuse their friends with linguistic creativity, and by using the etymologically foreign suffix *-ity*, they could also subtly show off their learnedness.

Who were the people who led the increase in the productivity of *-ity*? Men writing to their close friends are one good candidate; another is the social rank of professionals, such as authors, lawyers, doctors and government officials. Unfortunately, there is too little data for a closer analysis of, say, professional men writing to their close friends in each time period, or even of professionals over time. While Säily (2016) finds that the productivity of *-ity* is significantly high in letters written by professionals in the corpus as a whole, after FDR control the difference is not quite significant, but there does seem to be a tendency towards that direction (Figure 12.4). Another result by Säily (2016) that is not quite significant after FDR control is the highly productive use of *-ness* by clergy. The unproductive use of *-ness* by royalty, however, remains significant (see Säily 2016 for an explanation of these results).



**Figure 12.4** Variation in the productivity of *-ity* across rank-based subcorpora, 1680–1800 (R = royalty, N = nobility, GU = upper gentry, GL = lower gentry, CL = lower clergy, P = professionals, M = merchants, O = other non-gentry). The productivity of *-ity* is almost significantly high with professionals ( $p < 0.0065$ )

### 12.5.3 Case studies

Having reassessed the statistical significance of the results in Säily (2016) and discovered a gender difference in the register of letters written to close friends, we turn to our other research questions. Section 12.5.3.1 focuses on the relative importance of class- and register-based variation in the change in the productivity of *-ity* by analysing outliers. Section 12.5.3.2 explores the question of whether the change is linguistic or stylistic by comparing the language use of a high-status group and a low-status group at the end of the 18th century. Finally, Section 12.5.3.3 moves from productivity to semantics, asking whether there is a change in the semantics of *-ness* and *-ity* between the 17th and 18th centuries which could explain some of the variation and change in their productivity. As these are small pilot studies, further research is needed to confirm and clarify the results.

### 12.5.3.1 *Individual outliers*

From the macro-level study in Section 12.5.2, let us zoom in on sociolinguistic variation at the micro level and analyse individual outliers in the productivity of *-ity*, listed in Table 12.1. For this analysis, the data set was narrowed down to 58 people from whom there was enough data, defined heuristically as a word count amounting to 2% or more of the word count of the (sub)corpus to which they were compared. Out of these 58 people, 22 were outliers in terms of one or more (sub)corpora, with only two of them being overusers of *-ity*, while the rest were underusers, including two members of the Clift family discussed further in the next section. Besides the full corpus, the subcorpora to which the people were compared included individual ranks (professionals, lower clergy and lower gentry, chosen because they provided the largest amount of data and thus a better chance of statistically significant results), individual genders (female, male) and individual relationships with the recipient (nuclear family, close friends, other acquaintances) as applicable, using both definitions of corpus size (running words, suffix tokens).

The two overusers, Eliza Draper (1744–1778) and William Cowper (1731–1800), both belong to the professional class. Cowper was a religious poet, while Draper was the wife of a government official in India and a friend of Laurence Sterne, the author of *Tristram Shandy*. Cowper is an overuser in terms of three corpora when the size of the corpus is measured in running words: the full corpus, the corpus of men's letters and that of letters written to close friends. Draper, on the other hand, is an overuser in terms of other professionals as well as women, again using running words as the measure of corpus size. Thus, both are outliers even within a category that itself uses *-ity* highly productively, i.e., close friends and professionals. Thanks to the method used, it is implausible that these two outliers could be the reason why these categories seem to use *-ity* highly productively – on the contrary, outliers make the confidence intervals wider, so that it becomes more difficult for a category to be significantly different from the corpus as a whole.

The Notes column of Table 12.1 lists possible reasons why these people are outliers. Cowper (see Example (12.1) above) writes relatively late, and as the productivity of *-ity* increases over time, it is natural for his letters to stand out when they are compared to a corpus that includes earlier periods. With Draper, the reason could be a stylistic one, as she seems to prefer an even more Latinate style than most of her peers (12.3). As for the underusers, many of them are from the earlier periods, when the productivity of *-ity* was in general lower. Another explanation for their underuse may be a shared style among friends and family: for instance, Daniel and Henry Fleming (father and son) are both underusers, as are Charles Lennox, Sarah Lennox and Thomas Pelham-Holles (father, daughter and father's close friend), and Elizabeth and William Clift (siblings). Furthermore, *-ity* seems to be underused in business letters and in letters written by lower-class people.

**Table 12.1** *-ity* outliers sorted by 20-year period (**boldface** = overusers, the rest are underusers; \* = also underuses *-ness*). For explanations of the abbreviations, see Figure 12.4 above (CU = upper clergy)<sup>†</sup>

Name		First period	Last period	Social rank	Notes
John	Pinney	1680	1680	CL	early
*Thomas	Dixon	1680	1680	CL	early (but also underuses <i>-ness</i> )
Henry	Fleming	1680	1700	CL	early; cf. Daniel Fleming
Daniel	Fleming	1680	1680	GU	early; cf. Henry Fleming
Samuel	Pepys	1680	1700	P	early
*Humfrey	Wanley	1680	1720	P	fairly early (but also underuses <i>-ness</i> )
*Ann	Clavering	1700	1720	GL	fairly early (but also underuses <i>-ness</i> )
Henry	Liddell	1700	1700	GL	fairly early; lots of business content
*Daniel	Defoe	1700	1720	P	fairly early (but also underuses <i>-ness</i> )
Thomas	Secker	1720	1760	CU	
Elizabeth	Purefoy	1720	1740	GL	lots of business content
Thomas	Pelham-Holles	1720	1740	N	cf. Charles Lennox
Charles	Lennox	1720	1740	N	cf. Thomas Pelham-Holles, Sarah Lennox
Theophilus	Hughes	1740	1760	CL	
<b>Eliza</b>	<b>Draper</b>	1740	1760	P	Latinate style; professional
<b>William</b>	<b>Cowper</b>	1740	1780	P	late; many letters to close friends
Samuel	Crisp	1760	1780	GL	overuses <i>security</i> , <i>opportunity</i> in letters to sister
Sarah	Lennox	1760	1780	N	cf. Charles Lennox
Erasmus	Darwin	1760	1780	P	overuses <i>quantity</i> , <i>society</i>
*Elizabeth	Clift	1780	1780	O	cf. William Clift; low social status
*George	Culley	1780	1780	O	lots of business content; low social status
*William	Clift	1780	1780	P	cf. Elizabeth Clift; lower-class background

<sup>†</sup> The importance of the distribution of the outliers' social ranks should not be overestimated as the table only includes people from whom there was enough data, and these people may represent a skewed rank distribution compared with the corpus as a whole.

(12.3) I say not this, with a view to disqualify and extort refinements or flattery, but from such a consciousness of my own **imbecility** as makes me very serious, when reduced to the **necessity** of self examination – if therefore, you have the **generosity**, which I take you to have, you will rather endeavour to correct my foibles than to add to it by your encomiums –

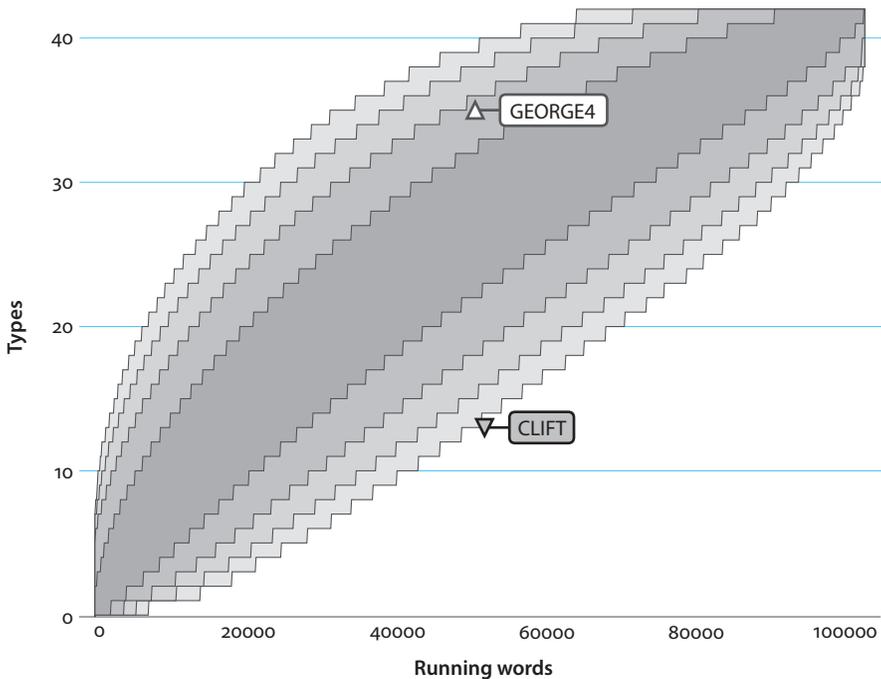
Eliza Draper to John Wilkes, 1775?; DRAPER, 175 (DRAPER\_019)

The fact that the two overusers are both professionals speaks for the importance of class-based variation in the increase in the productivity of *-ity*, even though the general tendency of professionals to overuse *-ity* did not reach significance in our macro-level analysis. On the other hand, Cowper is also an overuser in terms of people writing to their close friends, which reinforces the significant role of register already seen in the macro-level analysis of social categories.

### 12.5.3.2 *Royalty vs. the Clifts*

Let us now shift our focus on two small groups of people at the opposite ends of the social spectrum in the final period of the corpus, 1780–1800 (CEECE collections CLIFT and GEORGE 4). Is there a difference in the productivity of *-ity* in family letters written by the lower-class Clifts on the one hand and by the royal Hanover family on the other? Out of the Cornish Clift siblings, the women, Elizabeth and Joanna, worked as domestic servants, while the men's occupations included steward (the eldest brother John), shoemaker (Thomas), sailor (Robert) and museum curator (the youngest brother William, who received more schooling than his siblings and moved to London). The majority of our royal family letters, too, are between siblings: the crown prince George (later king George IV), his brothers Frederick, William, Edward, Ernest, Augustus and Adolphus, and his sisters Charlotte, Augusta, Elizabeth and Amelia.

Even though neither the royalty nor the lowest, 'other' category differs from the corpus as a whole in Figure 12.4, we do see a significant difference when limiting the comparison to the two smaller groups. As shown in Figure 12.5, the Clifts underuse *-ity* compared with the royalty. In fact, more than two thirds of their instances consist of the most common *-ity* word in the corpus as a whole, *opportunity*, used in a formulaic manner at the beginning of the letters, as in (12.4). Their second most common *-ity* lexeme is *prosperity*, used in closing formulae, and the other two *-ity* words used by Elizabeth Clift, *calamity* and *civility*, are also common in the corpus as a whole. It is only the social climber, William (from whom we have the most data), who uses *-ity* somewhat more diversely, as in (12.5).



**Figure 12.5** Productivity of *-ity* in letters written to nuclear family members in the CLIFT and GEORGE 4 collections, 1780–1800. Owing to lack of data, the samples used here are individual letters rather than people (see Chapter 5)

(12.4) My Dear Brother

I have once more took an **oppurtunity** of writing a few lines to you hoping it will find you and my Brother thos in Good health as this leves me at Present I Bless the Lord for it [...]

Elizabeth Clift to William Clift, 1796; CLIFT, 140 (CLIFT\_043)

(12.5) [...] guided and performed in a wonderfull manner but by what means we shall be for ever ignorant, at least on this side of what we call **eternity**, and by some invisible power – and which power we call God – thus far we may go – but when we come to argue and reason on things which cannot be demonstrated such as the shape, **materiality**, co-partnership and co-this and that, the residence &c of this invisible being [...]

William Clift to Elizabeth Clift, 1797; CLIFT, 157 (CLIFT\_051)

The usage by the royalty is much more diverse: while *opportunity* is their most common *-ity* word, it only comprises 37% of all instances. The suffix is used especially extensively by the princes, who often write to their eldest brother about happenings abroad, as in (12.6). It is also often used to describe human attributes,

such as *anxiety*, (*in*)*civility*, *credulity*, *felicity*, *rationality*, *sensibility*, *sincerity* and *tranquillity*, as in (12.7).

- (12.6) Great many reports are spread about the Peace with the Empire, but as probably you will be much better informed about their **authenticity** than I can, I will say nothing more upon this subject, except that, should a Peace take place soon, I am very much afraid our affairs in Brittany will not succeed.

Prince Adolphus to the Prince of Wales, 1795;

GEORGE 4, III, 83 (GEORGE4\_114)

- (12.7) I hope that my too easy **credulity** has been too forward on this occasion; if so you can but say that I am very *silly*; if not, you may soon guess what I must feel.

Princess Augusta to the Prince of Wales, 1798;

GEORGE 4, III, 442 (GEORGE4\_158)

The results imply that even though the increase in the productivity of *-ity* in English correspondence had been going on for at least two centuries by the end of the 1700s, lower-class informants had not caught up with those above them.<sup>1</sup> Furthermore, the difference between the lowest and highest ranks suggests that Milroy's (1992) anvil model does not fully apply to this change, even if it is led by the middling rank of professionals. Based on this case study, it seems that the change in the productivity of *-ity* is not linguistic in the sense of a change in the grammar of the speech community as a whole; rather, it is connected to changing styles of letter writing among middle- and upper-class users.

### 12.5.3.3 *Semantics*

Our final case study focuses on the semantics of *-ness* and *-ity* in the 17th and 18th centuries. Given that the overall gender difference in the productivity of *-ity* disappears in the 18th century, is it possible that this is because *-ity* is used in the meaning 'embodied attribute or trait' more often in this century, making it a better match to women's involved writing style? Furthermore, does Riddle's (1985) claim that this meaning is more characteristic of *-ness* than of *-ity* hold equally well for both the 17th and the 18th centuries? What about her claim that the meaning 'abstract or concrete entity' is more characteristic of *-ity* than of *-ness*?

As the process of semantic classification is resource intensive, analysing these issues in the entire corpus would be impossible. Therefore, the data set was reduced to two samples, one from 1600–1679 and the other from 1720–1800. Each sample

---

1. A pragmatic reason for the underuse of *-ity* among lower-class users is that the topics they discussed may not have required a frequent and varied use of abstract nouns. In accordance with this hypothesis, the Clifts also underuse *-ness* compared with the royalty.

consisted of 22 men and 8 women who contributed five hapax legomena (words occurring only once in the individual's letters) for both suffixes, amounting to 300 instances for both *-ness* and *-ity*. The proportion of women, c. 25%, was consistent with the full corpus. Hapax legomena were used in order to avoid the most common, lexicalised types, such as *business*. To make the data more homogeneous, the highest and lowest social ranks (royalty; merchants, other non-gentry) were excluded. Within these limits, the people and instances were chosen randomly. However, even this data set turned out to be too large to classify in a reasonable amount of time, so it was further reduced by randomly selecting 10 instances from women and 30 instances from men in both periods, for a total of 80 instances for both *-ness* and *-ity*. While small, this data set still yielded interesting results, lending support to Labov's (1978 [1972]: 204) rule of thumb that a study of 25 people, with as few as five people per category, may be enough to form a general picture of a phenomenon (see also Mannila et al. 2013).

The reason why the classification was so difficult was because the two meanings posited by Riddle (1985) are extremely context-dependent. Riddle (1985: 458) notes that the meanings 'embodied attribute or trait' and 'abstract or concrete entity' closely correspond to the notions 'specific' and 'generic', respectively. In my analysis, it turned out that almost any word could have either meaning, the only way of differentiating between them being whether or not there was something in the context indicating a specific individual or thing whose 'embodied attribute or trait' the word was. Thus, it could be said that the meaning was often vague (as opposed to ambiguous, see Saeed 1997: 60–62) between genericity and specificity. As suggested by Baeskow (2012: 27), the specificity was frequently manifested syntactically by means of a possessive construction, as in *your sauciness* or *the wetness of the weather*, and Cowie (1999: 263) goes so far as to state that this is the only possible criterion for distinguishing between the attribute and entity meanings. However, what should we do with examples like (12.8) and (12.9)?

(12.8) give me leave, Sir to demand of you once more & to demand of you with the last **earnestness** the return of your paternal tenderness, which I have forfeited by the unhappy step I have made.

Edward Gibbon to his father, 1755; GIBBON, I, 5–6 (GIBBON\_004)

(12.9) The Lord Ranelagh having taken upon him to relate what has past between the King and himself I suppose it may bee no **vanity** in mee to tell you the particulars of the reception I had from his Ma<sup>tie</sup> when I attended him at Windsor the 18th of July being the next day after I arrived.

Arthur Capel to his brother, 1676; CEEC: ESSEX, 74 (ESSEX\_015)

It could be argued that uses like the above are metaphorical and extend the meaning into entity territory (cf. Nevalainen & Tissari 2010: 149–151). However, while *earnestness* could be seen as an entity used as an instrument in (12.8), it is also the specific earnestness of Edward Gibbon and could be paraphrased as *I am earnest*. Similarly, even though *vanity* in (12.9) could be regarded as a metaphorical substance, it is clearly the vanity of the writer, Arthur Capel, and the instance could be paraphrased as *I am not vain*. Therefore, I have chosen to include the above in my ‘embodied attribute or trait’ category.

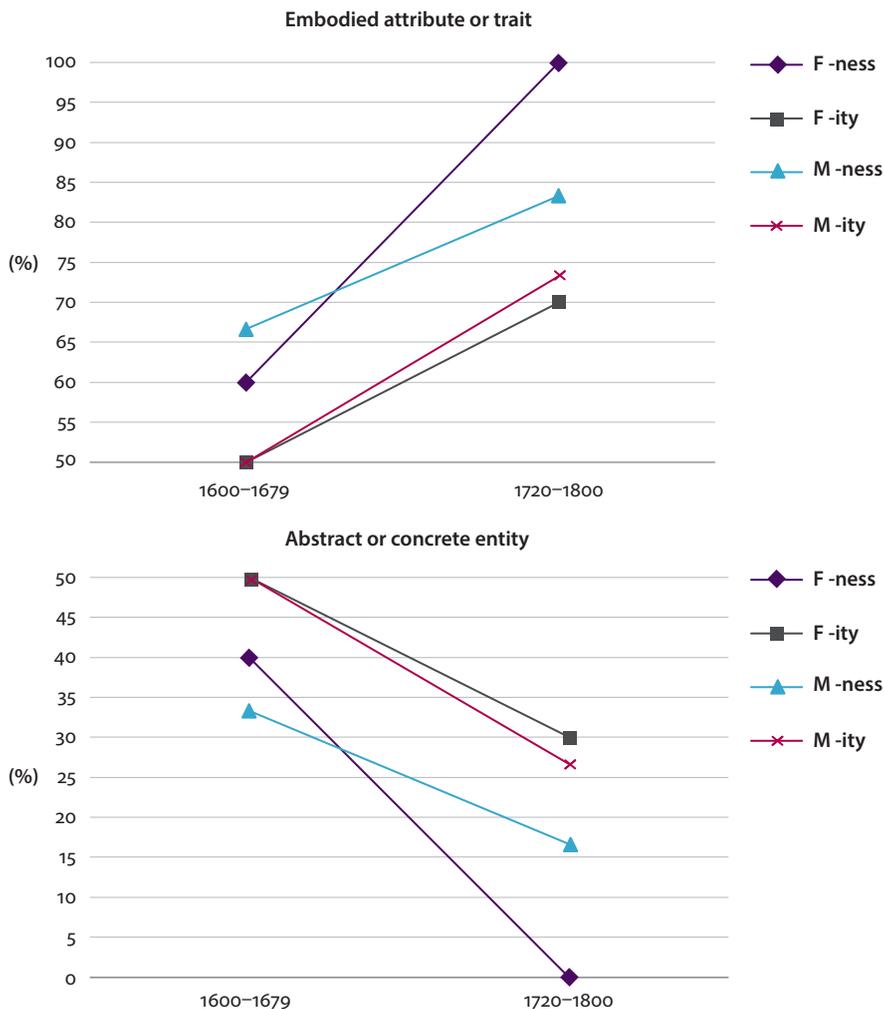
My ‘abstract or concrete entity’ category includes words that can context-independently be regarded as entities, such as *witness* and *university*, as well as polysemous words used in their entity sense, e.g., *rarities* ‘rare objects’ and *absurdities* ‘instances of absurdity’, even if these appear in a possessive construction. Furthermore, the category includes generic or even somewhat specific uses which could not straightforwardly be paraphrased as X IS ADJ, as in (12.10).

(12.10) I am confident you will hate all plundering and vnmercifullness.

Brilliana Harley to her son, who is in the army, 1643;  
CEEC: HARLEY, 206 (HARLEY\_060)

While somewhat idiosyncratic, my categorisation is at least consistent; I have gone through the instances multiple times, and they were sorted in a random order during analysis to avoid any biases. With all this in mind, let us have a look at the results, shown in Figure 12.6. It seems that Riddle (1985) is correct in that the meaning ‘embodied attribute or trait’ is more common with *-ness* than with *-ity*, while the reverse is true for the meaning ‘abstract or concrete entity’, although the difference between the suffixes is not as great as might be expected. Interestingly, however, the proportion of the trait meaning increases over time for both *-ness* and *-ity* and for both genders, while the proportion of the entity meaning decreases. Thus, Riddle’s claim that the entity sense has been gradually increasing for *-ity* and disappearing for *-ness* does not seem to hold for the correspondence genre. The differences between the genders should not be made too much of as the number of instances from women is very low.

To ensure that the results were not due to my classification, I reclassified the instances so that only those which appeared in possessive constructions (and were not context-independently categorisable as entities) were counted towards the meaning ‘embodied attribute or trait’. The main result of an increase in the trait meaning and a decrease in the entity meaning over time was preserved for both suffixes and both genders. For women, the entity meaning was more frequent for *-ness* than for *-ity* in the 17th century, but this could be coincidental as the number of instances was very low.



**Figure 12.6** Semantics of *-ness* and *-ity* in a sample of 17th- and 18th-century letters (F = women, M = men)

18th-century letters do, then, seem to refer to embodied attributes or traits more often than 17th-century letters in the case of both *-ness* and *-ity*, lending support to McIntosh's (2008: 231) claim of the feminisation of 18th-century culture (if we accept the proposition that talking about embodied attributes is part of an involved style, which in the 17th century was especially characteristic of women). It is possible that letter-writing, at least for the upper classes in our corpus, was more about maintaining and building social relationships and identities than about conveying information. Even though the style was more 'nouny' than before, this did not mean that the orientation was more informational; rather, writer and addressee

involvement were expressed in a more elaborated style by both men and women. As mentioned in Section 12.3, this is in accordance with Biber & Finegan's (1997) finding that 18th-century letters are both more involved (Dimension 1) and more elaborated (Dimension 3) than 17th-century letters.

Biber (2001), however, re-computes the dimensions for 18th-century material rather than using those based on present-day material, discovering that the informational vs. involved dimension is basically split in two in the 18th-century material and that the letters in the ARCHER corpus can only be characterised as involved according to the first dimension. While he concludes that 18th-century letters are not "personally involved and interactive" (Biber 2001: 105), this is in my opinion not entirely justified because there are still plenty of involvement features on the first dimension, including first- and second-person pronouns, modal verbs, mental verbs and general emphatics. Moreover, the second dimension seems to be more about orality than about ego/addressee involvement, and only drama scores especially high along the dimension. The range of variation along the second dimension is narrower than that along the first dimension, so the differences between the genres are not as great as Biber's (2001: 105) description might lead us to believe.

#### 12.5.4 Normative grammar

As noted by Yáñez-Bouza in Chapter 3, normative grammar may influence language use. On the other hand, Cowie (2003) argues that proscriptive attempts at standardisation in word-formation could be a reaction to increased productivity. Either way, most 18th-century grammarians do not have much to say about *-ness* and *-ity*. While *-ness* is mentioned by, e.g., Greenwood (1722: 186) as an ending that creates abstract nouns from adjectives, *-ity* and most other suffixes of a foreign origin are not discussed as productive suffixes but as terminations in previously borrowed or derived words.

Kirkby (1746: 61–63, as cited in Görlach 2001: 174) considers *-ness* to be in competition with *-th*, as in *warmness/warmth*, but does not mention *-ity* in this context. Greenwood (1722: 203) discusses *-ity* (conflated with *-ty*) among rules "whereby to know when a Word is derived from the *Latin*, and how it may be made *Latin* again", Metcalfe (1771: 71) in "How do you know when English Words are derived from Words in other Languages?", and Fenning (1771: 98) in "From what Latin words do the English ones that end in *ty* come?". By this, the authors would seem to imply that *-ity* was not or should not be used as a productive suffix, or possibly that their intended audience, who were often children, would not be able to use it in such a way. The first implication is clearly not borne out by the data, as the productivity of *-ity* increases during the period, so it could perhaps be seen as a (very subtle) reaction against the trend. The second implication, too, is doubtful,

as the women in the corpus, who did not as a rule receive a classical education, use *-ity* as productively as men in most registers. Most probably no such implications were intended, and this was simply the ‘received’ way of describing Latinate suffixes in the grammars of the period.

Campbell (1776: 391–393) objects to several long words in *-ness*. Firstly, he discourages the use of terms “composed of words already compounded, where of the several parts are not easily, and therefore not closely united”, such as *bare-faced-ness*, *shame-faced-ness*, *un-success-ful-ness*, *dis-interest-ed-ness*, *wrong-headed-ness* and *tender-hearted-ness* (his hyphenation), a couple of which are 18th-century neologisms according to the OED. “They are so heavy and drawling, and withal so ill compacted, that they have not more vivacity than a periphrasis, to compensate for the defect of harmony.” Secondly, he disapproves of words such as *peremptoriness*, in which the accented syllable is followed by “too many” syllables: “For though these be naturally short, their number, if they exceed two, makes a disagreeable pronunciation.” By contrast, he praises the “easy fluency” of the *-ity* words *levity*, *vanity* and *avidity*, which the OED classifies as loanwords from previous centuries. A quick calculation shows that the proportion of extremely long *-ness* tokens (between 16 and 19 characters) in fact increases over time in the corpus, so it seems that Campbell is unsuccessfully objecting to a widespread phenomenon. The playwright Samuel Crisp, for instance, is happy to use *disinterestedness* four years after the publication of Campbell’s *The philosophy of rhetoric*:

(12.11) Now I am talking of Law matters, it puts me in mind to tell You I have some thoughts of new making my Will; and of joining that honestest of beings, Jack Edison, of Basinghall Street (and whom I have had experience of, between 30 and 40 years) with You as a Joint Executor; being fully convinc’d of his perfect integrity of **disinterestedness**, and likewise of his experience and conduct of business; so that he would prove a very usefull Assistant to You, in Case of need; and exclusive of that as You and I are neither of Us young, I think one would wish the due execution of a Trust, should not hang upon one single Life.

Samuel Crisp to his sister, 1780; CRISP, 55 (CRISP\_013)

It may well be that the productivity of *-ness* on complex bases increases during this period, prompting the response by Campbell, but more research is needed to verify this. The hypothesis is lent some support by Anderson’s (2000) study of the OED (as cited in Palmer 2009: 14), which finds that the productivity of *-ness* on suffixed bases increases in the 17th and 18th centuries and decreases in the 19th century. The decrease probably cannot be attributed to proscription as it occurs particularly on Latinate suffixed bases, whereas Campbell seems to be mostly objecting to (compound) bases containing native suffixes.

## 12.6 Conclusion

The results of this study indicate that the productivity of *-ity* continues to increase significantly in 18th-century correspondence and that women have mostly caught up with men in the use of the suffix. It seems that the change is led by the professional class, following Labov's (2001) curvilinear pattern and Milroy's (1992) anvil model. Deviating from the anvil model, however, the lowest classes are lagging behind compared to the highest-status group. Aside from class, register in terms of the relationship between the sender and the recipient of the letter also plays a role: in keeping with Wolfson's (1990) bulge theory, the less stable relationship between friends is a trigger for the productive or creative use of *-ity*. Especially men seem to wish to impress and amuse their friends by using the suffix. Thus, it comes as no surprise that the individuals who use the suffix the most productively, William Cowper and Eliza Draper, belong to the professional class, and one of them is a man writing to his friends.

As class and register seem to be interconnected, it is difficult to weigh their relative importance in the change. One way to do this might be to compare the statistical significance of the macro-level differences. While the subcorpus of men writing to their close friends differs from the corpus as a whole at  $p < 0.00006$ , the subcorpus of professionals differs from the corpus as a whole at a lower level of  $p < 0.0065$ , which is not significant after false discovery rate correction. By this measure, register variation according to addressee is more important than class-based variation in the productivity of *-ity*. However, the measure may be affected by factors like corpus balance.

On the one hand, the increase in the productivity of *-ity* appears to be stylistic: it seems to be connected to the change towards a more involved and elaborated style in the correspondence genre observed by Biber & Finegan (1997). As noted by McIntosh (1986, 2008), the cultural ideals of politeness and elegance also show up in the language use of the upper classes. That the lowest classes are lagging behind is evident in our case study of the Clift siblings, who use *-ity* significantly less productively than the siblings in the royal family of the time. Lower-class writers may not have had a full command of the intricacies of the changing correspondence genre, nor would this have been necessary if they only wrote to people with whom their relationship was minimally distant (cf. Wolfson 1990). On the other hand, the change in *-ity* could be connected to the overall increase in the productivity of *-ity* observed by Lindsay & Aronoff (2013) in the OED, which would make it at least partly linguistic. This is supported by the fact that the productivity of *-ity* is increasing in 18th-century courtroom discourse as well (Säily 2016: Figures 6–7), although the change there could also be due to the increasing formality of the genre,

i.e., another stylistic trend (Huber 2007). More genres should be studied to reach a firmer conclusion.

As for semantic change, both *-ness* and *-ity* shift towards trait meanings in the usage of both genders between the 17th and 18th centuries with *-ness* in the lead, while the share of entity meanings decreases. The change is probably stylistic and, like the increase in the productivity of *-ity*, connected to the temporary increase in elaboration and involvement in the letter genre discovered by Biber & Finegan (1997). As *-ity* is now used more to express involvement, making it a better match to women's consistently involved style, the overall gender difference in its productivity disappears. Future research should attempt to verify the results of the semantic study using a larger data set. To make this more automatic and thus less resource intensive as well as less dependent on the analyst's intuitions, the change could be studied, e.g., by examining collocates (Renouf 2012) or through topic modelling (Rohrdantz et al. 2012).

Despite the increase in the productivity of *-ity*, 18th-century grammarians do not discuss it as a productive suffix. Campbell (1776) objects to long words in *-ness*, which could possibly be a reaction against an increase in the productivity of *-ness* on complex bases. To determine whether this could be the case, future research should analyse variation in the productivity of *-ness* across different base types in the corpus. The overall productivity of *-ness* does not change in the 18th-century section of the CEEC, nor is there register-based variation in its productivity. Sociolinguistic variation is limited to a near-significant tendency for clergy to overuse *-ness* (cf. its high productivity in sermons discovered by Cowie 1999: 239) and to a significant tendency for royalty to use it less productively, as they tend to repeat the same few *-ness* tokens in fixed expressions (see further Säily 2016).

PART III

## Changes in retrospect



## Zooming out

### Overall frequencies and Google Books

#### 13.1 Normalised frequencies of the phenomena studied

Tanja Säily

In Part II of this volume we examined variation and change in several features from the perspective of historical sociolinguistics. This chapter widens the focus from historical sociolinguistics to frequency change in general: Section 13.1 compares the overall frequencies of the phenomena studied (excluding *thou*, which is too infrequent for quantitative analysis), while Section 13.2 extends the time period studied to the 21st century by looking at the phenomena in the Google Ngram Viewer and critically analysing the usefulness of the big and messy database of Google Books for studies of grammatical variability.

Biber (2010: 242) talks about the “pervasive linguistic characteristics” of registers, meaning their most frequent and typical features. How frequent are the linguistic features we have studied? Figure 13.1 shows that our most frequent feature is the variable *has/hath*, which constitutes between 0.33% and 0.43% of the words in each time period in the CEECE, undergoing fluctuations and an overall slight increase in frequency over time. The second most frequent feature is the suffix *-ity*, which stays within the range of 0.21–0.29%, also increasing over time (note that we are here looking at its token frequency, whereas Chapter 12 analysed its productivity or type frequency). The rest of the features are surprisingly similar in frequency, varying between 0.02–0.11%. We can see that affirmative DO decreases and the progressive aspect increases within this range over time, while the variables of indefinite pronouns and *its/of it/thereof* seem to fluctuate with less of a clear pattern of change.

We can also look at the changes in another way, by breaking down the variables into individual variants and analysing their normalised frequencies alongside those of the changes lacking a variable. Figure 13.2 shows increases in the normalised frequencies of the incoming variants *does*, *-body* and *its* as well as the progressive aspect (we have chosen *does* instead of *has* here because its frequency more closely matches that of the other changes). Interestingly, while *-body* consistently gained ground in terms of its proportion of the variable of indefinite pronouns

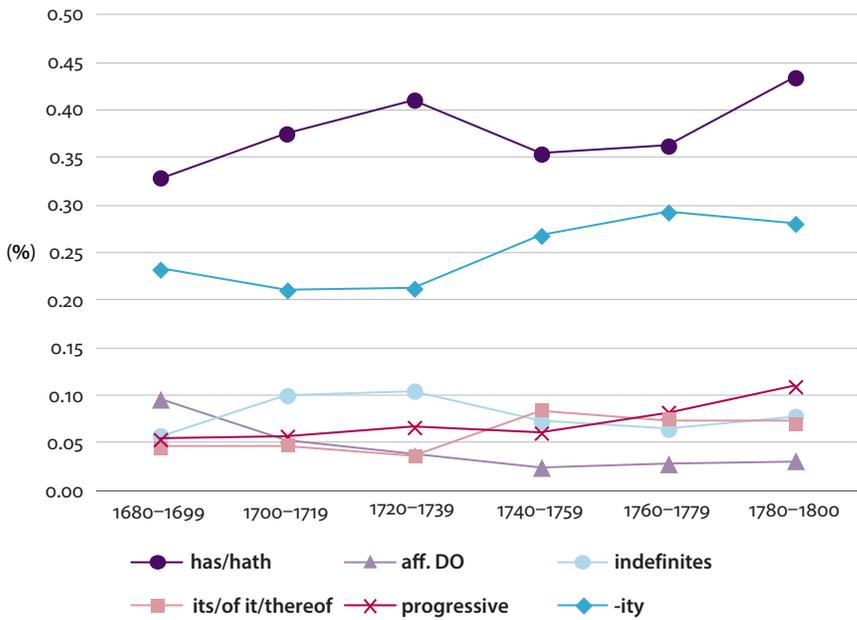


Figure 13.1 Normalised frequencies of the chief variables and other phenomena studied

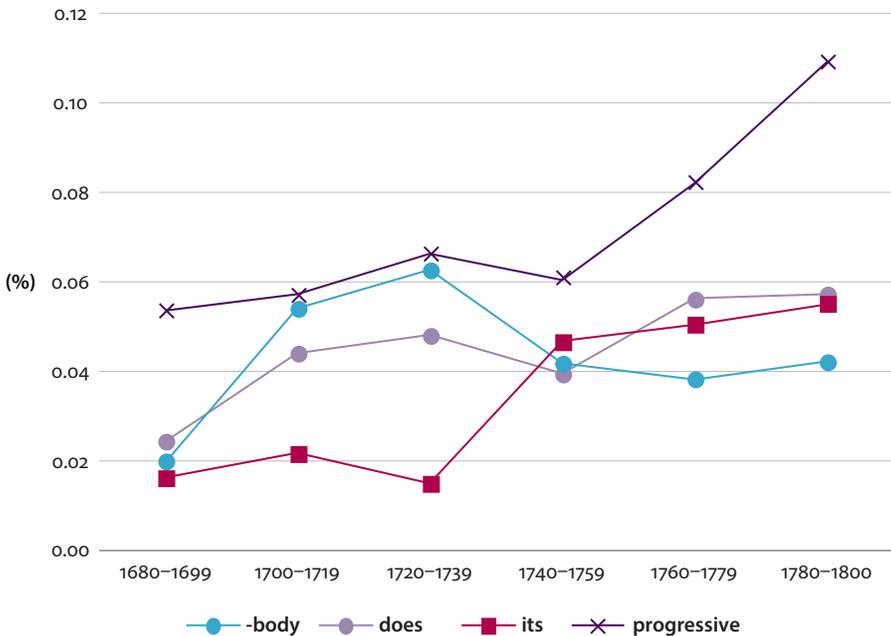


Figure 13.2 Normalised frequencies of three incoming variants and the progressive aspect

with singular human reference (Chapter 9), its normalised frequency does not show a consistent increase as there is a sharp drop between the periods 1720–39 and 1740–59.

We could also compare these frequencies with those found in other sources. The eighteenth century is covered by such materials as the *Old Bailey Corpus* and the *Corpus of Late Modern English Texts*, and it would certainly be useful to complement our data with these larger datasets in the future. As a first step, the next section compares the results we have gained using CEECE with those observable in the massive Google Books database.

### 13.2 Google Books: A shortcut to studying language variability?

Mikko Laitinen and Tanja Säily

This section looks into evidence provided by the big but messy database of Google Books (Google 2013a), which extends diachronically to Early Modern English, covering also the time period examined in this book. The purpose is to explore what types of broad diachronic evidence related to grammatical variability can be obtained from this massive freely available database with respect to the changes studied in this volume, to complement the insights into sociolinguistic variability offered by our specialized and tailor-made corpus in Part II.

The motivation for this type of comparison of big and messy databases with small and tidy corpora comes from two sources. The first is related to Google itself and the effects its search engines are said to have caused on information seeking in general. Google Books is, just like many but not all of the tools provided by Google Inc., an easily available free tool that enables access to a vast amount of information. The success of Google has led to the coining of a popular phrase, the Google generation, which in popular myth refers to the generation born after 1993 and is used to refer to a generation whose first port of call for knowledge is the Google search engine. In a recent report on information seeking patterns, commissioned by the British Library and JISC (Joint Information Systems Committee), the authors explore the myths around the Google generation and point out that “in a real sense, we are all Google generation now” (Rowlands et al. 2008: 301). There now exists convincing evidence from libraries’ deep log statistics that people of all ages use the Internet and its various technologies in surprisingly simple ways. Indeed, the investigation shows that the digital information world is characterized by massive choice, easy access, and simple to use tools, but people “from undergraduates to professors” exhibit “a strong tendency towards shallow, horizontal, ‘flicking’ behavior in digital libraries” (Rowlands et al. 2008: 300).

The second reason stems from a debate on quantities of empirical evidence in English historical linguistics. Some of the articles in a recent handbook (Nevalainen & Traugott, eds. 2012) deal with observing recent grammatical change and focus on the benefits and disadvantages of corpus materials of varying sizes. Davies (2012), a proponent of large corpora, illustrates that the value of large (but highly structured) corpora such as the *Corpus of Historical American English* (COHA) lies in the fact that they enable looking at not only low frequency lexical and semantic changes, but also low/medium frequency lexico-grammatical structures that may indeed reveal the entire life cycle of a change from its incipient stages onwards. He argues that the advantages of a large historical corpus, such as COHA, come from the fact that it is a balanced collection in which the conversion of the materials has been carried out accurately. In addition, it contains an enormous number (100,000+) of texts annotated for year of publication and genre, and all this comes equipped with a web-based search interface that enables a range of searches.

As for the advocates of small and structured corpora, such as the Brown family of corpora or diachronic materials such as ARCHER, Hundt & Leech (2012) show that the particular strengths of smaller materials have to do with careful sampling, exhaustive, qualitative scrutiny of the raw results, whole-text access and accurate metalinguistic information of the materials. They also point out that equidistant observation points of 30 years, roughly equalling to one generation, make these corpora suitable for observing language change. They provide two case studies both of which require careful qualitative scrutiny of the raw results for excluding homonymous forms (the first of relative pronouns in restrictive relative clauses and the second one of *for* as a causal conjunction).

It goes without saying that Google Books and its Ngram Viewer do not represent a structured historical corpus *per se* even though it can be “an incredibly useful tool for looking at the frequency of exact words and phrases” as Davies (2012: 159) points out. It is an unorganized text archive that offers access to massive amounts of language material in a nice, easy-to-use package. This source has been used in a range of studies. They include for instance investigations in variationist sociolinguistics (Tagliamonte 2016), historical pragmatics (Jucker et al. 2012), English historical linguistics (Friginal et al. 2014), and psycholinguistics (Brysbaert, Keuleers & New 2011). In addition, Google Books is recommended as a useful source for trend studies across time in course books in sociolinguistics (cf. Friginal & Hardy 2014: 188).

Since Google Books is one of the closest sources of evidence to hundred-million word mega corpora (such as COHA) and offers easy and open access to hundreds of millions of words of texts of British English (and other English varieties and other languages) that correspond with the time frame of the CEECE, it is clearly justified to ask what kinds of evidence can be found in it by someone interested in language history and language change in social-historical context. And most importantly,

how does this information fare with the evidence from carefully sampled corpus resources, such as the CEECE? Our discussion deals with Google Books, but there are other text mining tools, such as the EEBO N-gram Browser (available at <http://earlyprint.wustl.edu/>), which cover English print culture up to 1700.

The usefulness of Google Books as a source for data is discussed in Nevalainen (2013) who studies the diachronic evolution of three sets of words related to polite society. These “buzz words” (*courteous/courtesy*, *civil/civility*, *polite/politeness*) have distinct histories in the Middle Ages, the Renaissance, and the Enlightenment, and Nevalainen looks into the evidence provided by Google Books’ Ngram Viewer. She points out that one obvious shortcoming in this vast database is the fact that the materials are unevenly distributed and the early periods are considerably smaller than later ones. She concludes that at the time of multiple data sources, i.e. small structured and large corpora, and big/messy databases, linguistic study reaches the best result by triangulating a range of evidence from data sources.

Lijffijt, Säily & Nevalainen (2012) explore culturally loaded words over time, and focus on the frequencies of war-related vocabulary during the seventeenth century. Their results from Google Books show, unexpectedly, that the frequency of the lexeme *war* peaks during the Civil War (1642–1651), but that its frequency continues to grow even after the war toward the end of the century. They conclude that big and messy databases can be useful heuristic tools in diachronic studies, but that the results obtained should be checked against more reliable data using significance testing. They also point out that improving the reliability of the Google Books data requires analysing the actual n-gram frequencies aggregated over longer time periods.

Pechenick et al. (2015) and Kopleinig (2017) draw attention to the scarcity and dubious quality of the metadata associated with Google Books. For instance, Pechenick et al. (2015) show that the 2009 version of the dataset labelled “English Fiction” in fact seems to consist mainly of scientific writing. While the 2012 version represents fiction more accurately, it is clear that the Google Books metadata is still far from perfect. Crucially for diachronic studies, we have observed that the year of publication is often inferred incorrectly, so that a book published in the twentieth century whose topic concerns the sixteenth century may be falsely categorized as a sixteenth-century publication. Furthermore, both Pechenick et al. (2015) and Kopleinig (2017) note that as the composition of the corpus and how it changes over time is largely unknown, it is difficult to discover whether a change in frequency represents actual linguistic or cultural change, or whether it is merely an artefact of the data. Again, triangulation with more reliable sources is called for (Nevalainen 2013).

From these previous studies concentrating primarily on lexical variation, this section shifts the focus towards the focus in the book, *viz.* lexico-grammatical

variability in social context. One way to illustrate our approach is to explore what types of evidence can be extracted from Google Books using a variable which is discussed more extensively in one of the chapters. The variable concerns the various ways of expressing singular human reference using indefinite pronouns. In Early Modern English, the variable consisted of four variant forms, the older *-MAN* and the independent forms which were decreasing in frequency, and the incoming *-ONE* and *-BODY* indefinites, as has been shown by Raumolin-Brunberg & Kahlas-Tarkka (1997). During the 15th–17th centuries, before spelling in general became standardized, the spelling for the determiners (*EVERY*, *ANY*, *SOME*, *NO*) varied considerably. As an illustration from the CEEC, it is easy to find nearly two dozen spelling forms of a determiner form *ANY* which can be used in compound indefinites (e.g. *any*, *ani*, *anie*, *anii*, *anney*, *anny*, *annye*, *anye*, *ene*, *eney*, *enu*, *enni*, *anny*, *eny*, *ony*, *onny*, *onnye*, *onye*). In addition, spelling variation exists in highly literate texts, as shown in (13.1) from the early 17th century:

- (13.1) That was the true light, which lighted **euery man** that cometh into the world  
(*The Authorized Version* of the Bible, John 1:9)

As for the long eighteenth century, the orthographic variation is mainly related to spelling the indefinites as separate units. According to Denison (1998: 101), it is only in the early nineteenth century when the indefinites start appearing as one unit. This means that for a comprehensive overview of the indefinite pronoun use from Early Modern English to the present day, one needs to account for the spelling variants in both joined and separate forms.

The searches in Google Books offer limited tools for accounting for the spelling variation, particularly since many of the spelling variants are extremely low frequency items. The only feasible tool in the search interface is the sum operator “+” that allows combining multiple n-gram series into one, but it cannot be used in case insensitive searches so any query must contain both lower and upper case forms. Despite the limitation, the operator function enables combining the searches for the most common spelling variants (i.e. for the assertive paradigm: *ANY ONE* + *ANYONE*, *ANY MAN* + *ANYMAN*, etc.) and this search needs to be repeated for all four paradigms.

Figure 13.3 illustrates the pooled results depicting the development of the three nominal forms in the indefinite pronouns from the early 17th century onwards. It shows the results of the relative frequency for the compound indefinites with *-ONE*, *-BODY* and *-MAN* endings, but not for the independent forms that are homonymous with pro forms in the negative paradigm (*He asked for money, but I gave him none*).

The diachronic picture that emerges from the search in this large unstructured database is twofold. Firstly, the decrease of the forms in *-MAN*, which had been decreasing in frequency from the 15th century, is also visible in the Google Books

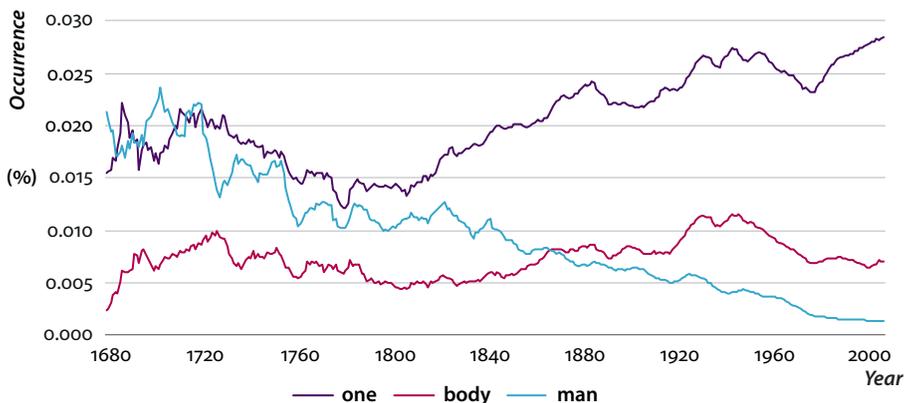


Figure 13.3 -ONE, -BODY and -MAN indefinites in British English Google Ngram Viewer

results. The cross-over from -MAN to the incoming variant forms takes place in the early 18th century. This result is partly similar with the results in Nevalainen & Raumolin-Brunberg (2003: Table 14 in Appendix II) in which the crossover in correspondence data takes place in the mid-17th century. Secondly, what is noteworthy is that the crossover takes place to the -ONE forms rather than to -BODY. Previous studies, which have drawn evidence from smaller corpora which make it possible to circumscribe the variable and enable careful qualitative investigation of each occurrence, have shown that it is the -BODY variant which emerges earlier than -ONE (Raumolin-Brunberg & Kahlas-Tarkka 1997: 68 using the multi-genre *Helsinki Corpus of English Texts*). This observation is most likely brought about by the fact that the Google Books results also include partitive OF-forms in which only -ONE can be used but not -BODY (cf. ANYONE OF THEM, but \*ANYBODY OF THEM). Similarly, the results also include both generic and non-generic cases.

It is obvious that even this diachronic overview of a very simple variable of compound indefinite pronouns from the Google n-grams requires additional corpus evidence, and the results obtained are only partially accurate provided that one knows the limitations in the searches. These limitations relate to excluding (a) partitive structures which skew the picture, and (b) the independent forms that would need to be found through extensive post-editing of the raw results which is not possible in Google Books. So Davies's (2012) conclusion, that Google's resources are inadequate for diachronic linguistics since they only enable looking at the frequencies of exact words and phrases, is fully justified, but it should be added that the results can be used as a confirmatory tool supplementing more fine-grained data.

In the case of the indefinite pronouns with human generic reference, this fine-tuning is related to two issues. The first one comes from Hundt & Leech's (2012) notion of exhaustive scrutiny for qualitative analyses and is clearly shown

in the results above. They include a considerable share of false positives, and their precision is thus limited even though the results are based on a large set of data. The second one is related to the sociolinguistic embedding of the variant forms. Sociolinguistic variation in the indefinite pronoun use is important because the variable consists of two incoming forms (-ONE and -BODY) and two outgoing ones, one of which is semantically loaded for gender-specific references (i.e. -MAN).

In previous studies (Nevalainen & Raumolin-Brunberg 2003), the early modern period variability was found to correlate with writers' gender above all the other social categories. Zooming into sociolinguistic variability and comparing the findings with a range of other structures in Part II of this book has offered answers to whether this development continues in the later stages of the change (cf. Chapter 15 below). In addition, information of sociolinguistic embedding could help to answer the question of how linguistic variables may index social characteristics and reveal how speakers place themselves within the social landscape through their linguistic practices (cf. the third wave of sociolinguistics in Eckert 2012). For instance, information on the frequencies of outgoing variants may enable identifying characteristics such as linguistic conservatism, i.e. speakers who lag behind the others in linguistic change and who polarize the process (see Chapter 14 below). It goes without saying that the evidence from a large unstructured database in this case falls short of such a close analysis.

Another limitation in the big and messy database of Google Books concerns the nature of the data included and the question of its representativeness. While Google Books contains a massive number of books, not all periods are equally well represented in the database. For instance, the British English section in the 2012 version of the database<sup>1</sup> goes back as far as the 16th century, but the amount of data in that century is extremely limited and patchy, so that even the frequency of the most common lexical items in the language, such as *the*, cannot be studied reliably using the Ngram Viewer. This is partly due to the fact that the viewer employs per-year frequencies and lacks the basic feature of pooling the data into longer time periods. As there is data from only 38 years in the 16th century, for the remaining 61 years the frequency of *the*, as well as that of any other item, is zero. If smoothing is applied to the graph (the default smoothing of 3 means that the yearly frequency is averaged across three consecutive years), the smoothed frequency will be erroneously low, because the zeros are not treated as missing data points but as actual frequencies (see Lijffijt et al. 2012: Section 2.2).

The situation with the 17th-century data is better in that every year except for 1601 is represented by at least one book, making it possible to study high-frequency

---

1. Statistics on various versions of the database, as well as the datasets themselves, can be downloaded at <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html>.

items. This is an improvement over the 2009 version of the database, where Lijffijt et al. (2012: Section 2.2) report 15 empty years between 1600 and 1640. From 1680 onwards, which is the starting point of this volume, each year is represented by at least fifteen books. In the 1700s, the amount of data ranges from 64 to 731 books per year for a total of 29,034 books, but rare or poorly dispersed items remain difficult to study using the yearly approach. It is only in the 19th century that the truly big data begins to appear, the 19th–21st centuries being all represented by thousands of books per year for a total of hundreds of thousands of books per century (see Table 13.1).

**Table 13.1** Number of volumes by century in the British English section of the Google Books database (2012 version). Based on yearly totals retrieved from <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html>

Period	Number of books
1500–1599	141
1600–1699	1,645
1700–1799	29,034
1800–1899	286,693
1900–1999	563,980
2000–2009	232,384
<b>Total</b>	<b>1,113,877</b>

The basic Ngram Viewer search enables the user to search for one or more words or phrases (i.e. n-grams), after which the results are displayed as a line graph with time on the *x*-axis and relative frequency on the *y*-axis (Google 2013b). As noted above, words (such as different spellings of the same word) can be combined into a single n-gram series by using the + operator. Phrases may include one instance of the wildcard \*, which stands for any word; the graph will then show results for the top ten words, which is perhaps not what a linguist would want. Furthermore, the \* operator cannot be used word-internally to replace a sequence of characters, which means that the viewer cannot be used to search for affixes. Part-of-speech tags can be used at the end of words (e.g. *cook\_VERB*) or on their own (*\_VERB\_* or *\*\_VERB\_*), but their use is severely restricted as they cannot be utilized in queries that contain a wildcard standing on its own or in queries consisting of more than three words. Similar restrictions apply to the lemma query operator *\_INF* (*cook\_INF* comprises the inflected forms *cook*, *cooks*, *cooking*, *cooked*) and to the dependency relations operator *=>* (*dessert => tasty* will reveal how often *tasty* modifies *dessert*). As pointed out above, case-independent search is only available for simple queries that do not contain wildcards for instance.

To address some of these issues, Mark Davies (2011–, 2014) has developed an advanced interface for the Google Books data. However, the amount of pre-19th century British English data in Davies’s version appears to be considerably smaller than in the official interface, and it is unclear what has been excluded. In Davies’s version, too, there are limitations on what kinds of query operators can be combined. Moreover, the part-of-speech annotation in Davies’s version is based on matching individual words with their most frequent parts of speech in a corpus of present-day American English, which means that it is less reliable than the official annotation, which is based on analysis of the running text of the books (Lin et al. 2012). Because of these limitations, the following analysis will employ the official interface.

Let us now turn to the rest of the changes analysed in this book. How well can they be studied using the Ngram Viewer and its British English component? First, it is clear that the interface is of no use to the study of suffixal productivity in Chapter 12 as searching for affixes is impossible. This also means that the verbal *-s* (Chapter 7) and progressive *-ing* (Chapter 11) can only be studied by searching for some of the most frequent verbs in which they occur, which is obviously inadequate but may reveal something of interest. Our findings are described further in Figures 13.4–13.8.

The Ngram Viewer interface does not enable disambiguating singular and plural *you*, and Figure 13.4 therefore focuses on forms of *thou* alone. The figure shows the overall decline in the use of *thou* towards the present. As expected, the number of instances is very low from the 1680s onwards. The occasional peaks can be explained when we look more closely into the material used by the Ngram Viewer: most of the data come from different prints of the Bible and the works of Shakespeare, both of which traditionally make a distinction between *thou* and *you*.

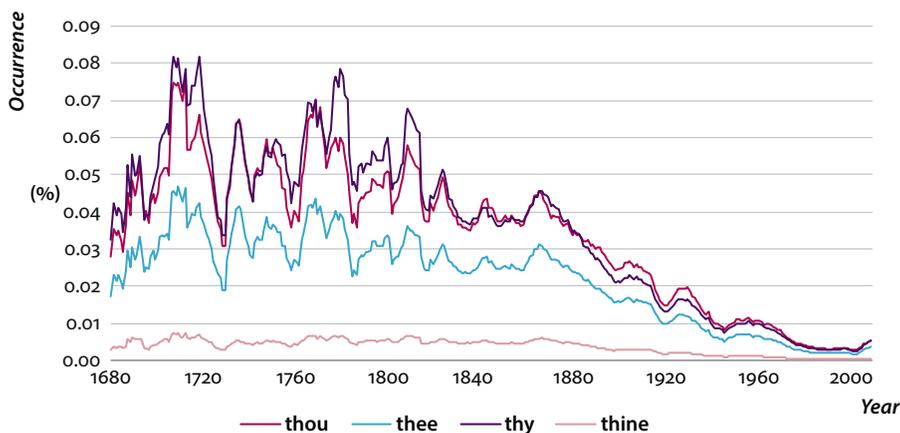


Figure 13.4 Forms of *thou*

Since the interface does not allow searching for affixes, Figure 13.5 focuses on a single frequent item, *has*, vs. the older form, *hath*. The Google Books figures agree with both the EEBO-TCP and CEECE data discussed in Chapter 7 that the crossing-over from *has* to *hath* took place at the very end of the 17th century. However, as expected, the mixed print materials included in the Google Books suggest a much longer tail for the use of the outgoing *hath* variant in the post-1700s than could be seen in personal correspondence.

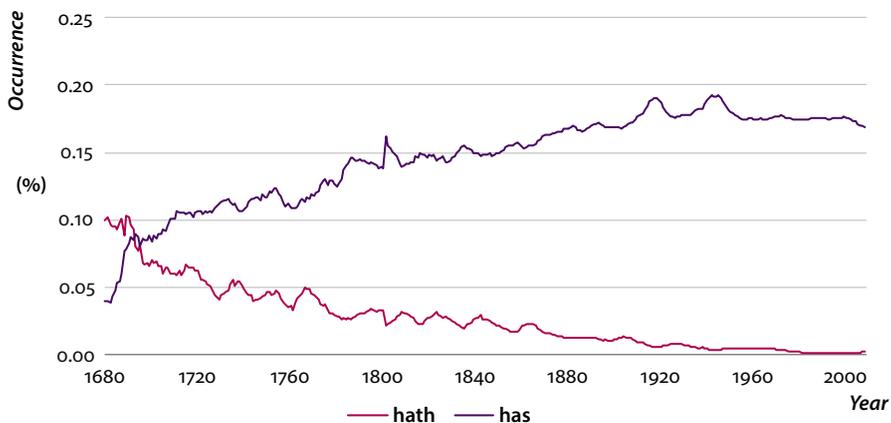


Figure 13.5 Diffusion of verbal -s

As the general case cannot be reliably identified using the Ngram Viewer, Figure 13.6 focuses on *do* followed by five of its most frequent lexical verb companions in

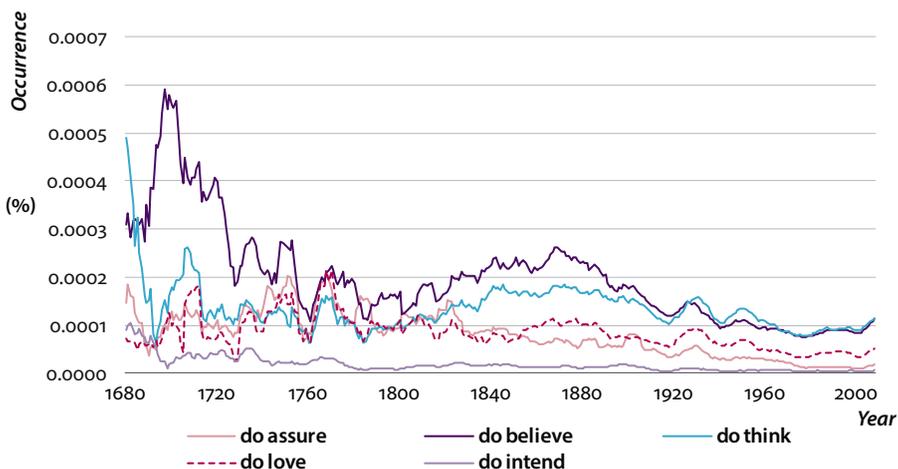


Figure 13.6 Periphrastic DO in affirmative statements

the CEECE. The results from the Ngram Viewer support the general view of the development of periphrastic *DO* in affirmative statements: there is a clearly declining trend from the early eighteenth century onwards. Nevertheless, the construction never goes completely out of style. Biber et al. (1999: 433–434) list several high-frequency verbs that still go with *DO* today (see Section 8.2.3 above). Of the ones tested here, only *think* and *believe* are included in their list, which agrees with the findings in Figure 13.6, as the other three verbs are clearly less frequent by 2008. It can also be argued that *assure* may have been genre-specific even in the eighteenth century; Biber et al. certainly identify register variation in verbs appearing with *DO*. It might be an interesting exercise to trace the development of *DO* with all the verbs listed by Biber et al., but this would undoubtedly leave aside interesting developments with other verbs, genre-specific trends and, of course, cases where there is an intervening adverbial (see Sections 8.2.4 and 8.4.3 above). For cases like periphrastic *DO* in affirmative statements, the Ngram Viewer is more a source of supporting information than even a diagnostic tool.

Since the Ngram Viewer makes it possible to use part-of-speech tags, the search on the third-person neuter possessives in Google Books (Figure 13.7) focuses almost on the same grammatical contexts as Palander-Collin (see Section 10.2 above). The constructions *its N*, *the N of it* and *the N thereof* can be compared, although *the N of it* has to be shortened to *N of it* owing to the limitation of part-of-speech tags to queries of a maximum of three words. The overall trends are highly similar with the findings based on CEECE. In the CEECE, *its N* rises during the eighteenth century and reaches 76% of the paradigm towards the end of the period studied (1780–1800; see Section 10.4.1). Figure 13.7 shows that the frequency of *its N* further increases, as would be expected, and peaks around 1840s.

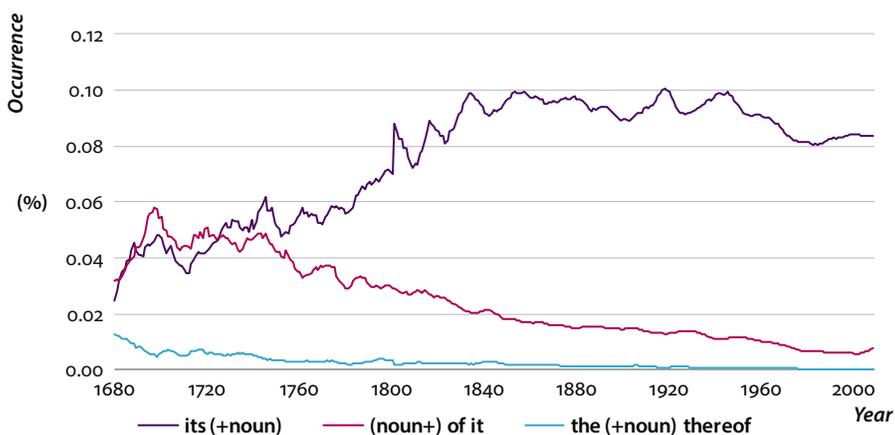
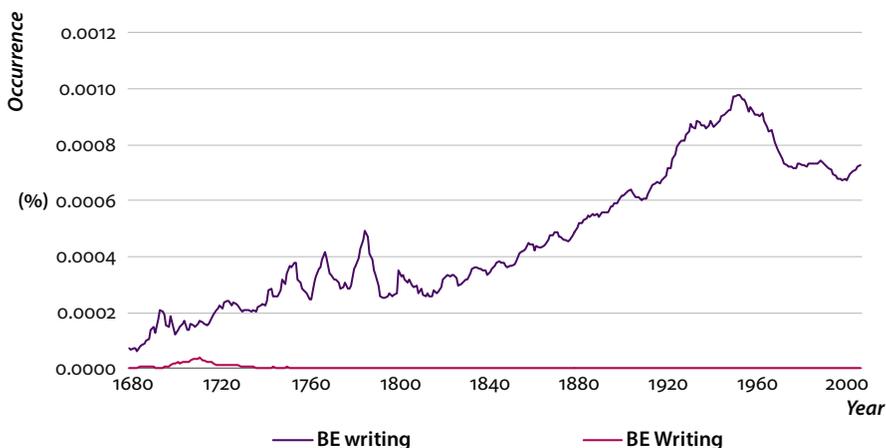


Figure 13.7 Diffusion of *its*

Because the Ngram Viewer interface does not allow affix searches, the changes in the progressive aspect are illustrated in Figure 13.8 through a verb construction that is frequent in letters, namely *BE writing/Writing* (*I am writing*). The numbers are very low, but the increase of this particular verb form follows the observed upward trajectory of the progressive from Early Modern English to the present day (see Kranich 2010: 95 for an overview).



**Figure 13.8** The progressive aspect: *BE writing* (lower- and uppercase spelling)

To conclude, we have seen that Google Books may provide a useful point of comparison or starting point for the study of some grammatical changes. However, the limitations of the interface – and the messy, patchy and uncertain nature of the data and the metadata behind it – mean that Google Books alone is not a sufficient source of data. This is especially true in the case of sociolinguistic research, where we need to know whose language it is that we are analysing, which groups lead the change and which are lagging behind. For this we need carefully-compiled specialist corpora with rich sociolinguistic metadata, such as the CEECE.



## Conservative and progressive individuals

Tanja Säily

This chapter summarises our research on individual outliers, looking for patterns in their behaviour. Are individuals consistently conservative or progressive across changes, or is there variation here as well? Does the nature and stage of the change influence their behaviour? What about the social background of the individuals?

### 14.1 Definition of outlier

First we need to look into how outliers are defined in each individual chapter. Because each change is different in terms of e.g. stage and social meaning, there is variation in what constitutes an outlier.

For **changes nearing completion**, we can often only analyse conservative users as almost everyone else has switched to the incoming variant. In the case of the very infrequent *thou* (Chapter 6), practically anyone who uses it counts as a conservative individual, although the meaning of ‘conservative’ is debatable as the functions of *thou* changed in the 18th century. For verbal *-s* (Chapter 7), conservatives are defined more strictly as those who use 10% or less of the incoming forms *has*, *does* or *says*, and who have a minimum of six instances of the variable. In the case of affirmative *DO* (Chapter 8), which lacks a linguistic variable, individuals from whom there is enough data are compared with the corpus as a whole using permutation testing (see 5.3.2 above). Here, too, the meaning of ‘conservative’ is debatable: while Hester Piozzi uses *DO* a great deal at the end of the 18th century, she does so in a new way, as an early adopter of emphatic *DO*. Finally, for the outgoing indefinite pronouns (Chapter 9), conservatives are defined as those individuals whose use of the recessive variants exceeds 50%.

For **mid-range changes**, both progressive and conservative individuals can often be identified. In the case of *its* (Chapter 10), conservatives use *its* less frequently and progressives more frequently than the period average. Outliers in the use of the increasingly productive *-ity* (Chapter 12) are defined as individuals who differ significantly from the corpus as a whole in terms of permutation testing. As for the

**incipient** progressive aspect (Chapter 11), progressive users are identified by comparing individuals with the corpus as a whole using both normalised frequencies and permutation testing. In the case of the incoming indefinite pronouns *-body* and *-one* (Chapter 9), progressive users are defined as those who have at least ten instances of indefinites and whose use of the variant in question exceeds 50% (for *-body*, the period covered is up to 1759 as its increase had stalled by that point).

## 14.2 Analysis

Individuals who are outliers in terms of more than one change are listed in Table 14.1. Most people are not outliers in terms of multiple changes. Looking at the outliers who are consistently conservative across changes (their names and social metadata are shaded with light grey in the table), we can see that a disproportionate number of them seem to be clergymen. This could be due to their close association with archaic biblical language in their work. Otherwise the distribution of social ranks and genders more or less reflects that of the corpus as a whole. However, the middle rank of professionals is overrepresented among the outliers who are progressive in terms of at least one change, which supports Labov's (2001: 259) hypothesis of a curvilinear pattern of language change.

Based on previous research (e.g. Nevalainen et al. 2011), we might also hypothesise that social aspirers would be more conservative than others, especially with regard to incipient and mid-range changes (cf. Table 14.3), where the incoming form might not yet have established a positive social value and was thus better avoided by someone who wished to be accepted by the upper echelons of society. This, however, is not borne out by our data as there are both upwardly and downwardly mobile conservative users, as well as those whose social status remains static. On the other hand, the four consistently progressive users (indicated in boldface in Table 14.1) are united by their lack of social mobility as none of them cross the line between gentry and non-gentry. Most of them are professionals, and even though Elizabeth Carter was a clergyman's daughter, she too moved in the same literary circles as the others as a poet, writer and translator. It thus seems that the most progressive users were firmly placed among professionals.

Table 14.1 also shows that individuals may change their outlier status during their lifetime, which can be regarded as another aspect of lifespan change in language use. Physician Erasmus Darwin is among the progressives in the use of *its* earlier in his life but later starts lagging behind, becoming consistently conservative in terms of both *its* and *-ity*. Not only does he lag behind in the communal increase, but the proportion of *its* in his letters actually decreases from the first period to the next, as shown in Chapter 10 above. Conservatism may thus increase with age, to the point of "retrograde change" (Wagner & Sankoff 2011: 304–305).

Table 14.1 Individuals who are outliers in terms of two or more of the changes studied, sorted by gender and period.\*

Outlier	Change											
	Period	Name	Gender	Rank	Social mobility	<i>thou</i>	verbal -s	DO	indefinites	<i>its</i>	progressive aspect	- <i>ity</i>
1700–1719	Isabella Wentworth	F	N	U					C	P		
1700–1739	Ann Clavering	F	GL	N					C			C
1700–1779	Mary Wortley Montagu	F	N	N				P	P: <i>-body</i>	C		
1720–1800	Elizabeth Carter	F	CL	N					P: <i>-body</i>	P		
1740–1779	Elizabeth Draper	F	P	U						C		P
1760–1800	Frances Burney Sarah	F	P	N					P: <i>-one</i>	P		
1760–1800	Lennox	F	N	N						C	(P: p.pass.)	C
1760–1800	Hester Piozzi	F	GL	N		C		C		P		
1760–1800	Mary Wollstonecraft	F	P	D						C	P	
1780–1800	Elizabeth Clift	F	O	N					P: <i>-one</i>		P	C
1680–1699	Daniel Fleming	M	GU	U			C		C			C
1680–1699	Philip Henry	M	CL	N			C		C			
1680–1699	John Pinney	M	CL	?			C		C			C
1680–1719	Henry Fleming	M	CL	D					(C but early)			C
1680–1719	John Evelyn Samuel	M	GL	N					C	P		
1680–1719	Pepys Humphrey	M	P	U					P: <i>-body</i>	C		C
1680–1739	Prideaux Humfrey	M	CU	U			C		P: <i>-one</i>	P		
1680–1739	Wanley Henry	M	P	N				C	C	P		C
1700–1719	Liddell	M	GL	N						P		C
1700–1739	Daniel Defoe	M	P	N					C	C		C
1700–1739	John Gay	M	P	D			C		P: <i>-body</i>			

(continued)

Table 14.1 (continued)

Period	Name	Gender	Rank	Social mobility	Change						
					<i>thou</i>	verbal -s	DO	indefinites	<i>its</i>	progressive aspect	<i>-ity</i>
1720–1759	Francis Blomefield	M	CL	D		C		C			
1720–1759	Charles Lennox	M	N	N				P: <i>-body</i>			C
1720–1759	Thomas Pelham-Holles	M	N	N				P: <i>-body</i>			C
1720–1779	David Garrick <b>Thomas</b>	M	GL	U	C		P	P: <i>-body</i>			
1720–1779	<b>Gray</b>	M	P	N				P: <i>-body</i>	P	P: past	
1720–1779	Thomas Secker	M	CU	U		C			C		C
1720–1800	<b>Samuel Johnson</b>	M	P	N				P: <i>-body</i>	P		
1720–1800	Roger Newdigate	M	GU	N				P: <i>-one</i>	C		
1740–1779	Theophilus Hughes	M	CL	N		C					C
1740–1800	William Cowper	M	P	N	C			C	C		P
1760–1779	Richard Champion	M	M	N				(C: <i>-man</i> )	P		
1760–1800	Jeremy Bentham	M	P	N					C	P: complex	
1760–1800	Charles Burney	M	P	N	C			P: <i>-one</i>	P		
1760–1800	Samuel Crisp	M	GL	U	C						C
1760–1779	Erasmus Darwin	M	P	N					P		C
1780–1800	Erasmus Darwin	M	P	N					C		C
1760–1800	William Jones	M	GU	U				C	C	P	
1760–1800	Ignatius Sancho	M	O	N	C				P		
1780–1800	William Clift	M	P	U						P	C

\* **Gender:** F = female, M = male. **Rank:** N = nobility, GU = upper gentry, GL = lower gentry, CU = upper clergy, CL = lower clergy, P = professionals, M = merchants, O = other non-gentry. **Social mobility:** U = up, D = down, N = none. **Change:** C = conservative, P = progressive.

Let us next zoom in on the different stages of change. Table 14.2 lists individuals who are outliers in terms of more than one change nearing completion or completed. Most of the outliers are consistently conservative as it was only possible to identify progressive individuals in the case of affirmative DO. The only woman on the list, Hester Piozzi, is conservative in terms of her use of *thou* and affirmative DO, but she uses them quite differently from the earlier centuries, so she could also be called progressive. This is in line with the oft-repeated observation that women tend to be leaders rather than laggards when it comes to language change. David Garrick, the famous Shakespearean actor, uses *thou* as an intimacy marker, but somewhat surprisingly makes little use of affirmative DO in his letters. Three of the conservatives are clergymen: Philip Henry, John Pinney and Francis Blomefield, all in terms of verbal -s and indefinite pronouns (see 15.3.3 below for a possible explanation for this). Professionals and gentry are also represented among the outliers. As with the full data set, social mobility does not seem to have an effect among the conservatives here, either.

Table 14.2 Individuals who are outliers in terms of more than one change nearing completion, sorted by gender and period

Outlier		Change						
Period	Name	Gender	Rank	Social mobility	<i>thou</i>	verbal -s	DO	outgoing indefinites
1760–1800	Hester Piozzi	F	GL	N	C		C	
1680–1699	Daniel Fleming	M	GU	U		C		C
1680–1699	Philip Henry	M	CL	N		C		C
1680–1699	John Pinney	M	CL	?		C		C
1680–1739	Humfrey Wanley	M	P	N			C	C
1720–1759	Francis Blomefield	M	CL	D		C		C
1720–1779	David Garrick	M	GL	U	C		P	
1740–1800	William Cowper	M	P	N	C			C

Table 14.3 shows individuals who are outliers in terms of more than one incipient to mid-range change. Here most of our informants display a mixed profile of conservatism in one change and progressiveness in another, perhaps in part owing to differences in the social meanings of the forms, or how the forms were evaluated by the language users (cf. Nevalainen et al. 2011: 30, 32). For instance, the progressive aspect seems to have been a colloquial feature, whereas *-ity* belonged to a more elevated register, even though it was also increasingly used as a marker of involvement between close friends (see further Chapter 15 below).

**Table 14.3** Individuals who are outliers in terms of more than one incipient to mid-range change, sorted by gender and period

Outlier		Change						
Period	Name	Gender	Rank	Social mobility	incoming indefinites	<i>its</i>	progressive aspect	<i>-ity</i>
1700–1779	Mary Wortley Montagu	F	N	N	P: <i>-body</i>	C		
1720–1800	<b>Elizabeth Carter</b>	F	CL	N	P: <i>-body</i>	P		
1740–1779	Elizabeth Draper	F	P	U		C		P
1760–1800	<b>Frances Burney</b>	F	P	N	P: <i>-one</i>	P		
1760–1800	Sarah Lennox	F	N	N		C	(P: p.pass.)	C
1760–1800	Mary Wollstonecraft	F	P	D		C	P	
1780–1800	Elizabeth Clift	F	O	N	P: <i>-one</i>		P	C
1680–1719	Samuel Pepys	M	P	U	P: <i>-body</i>	C		C
1680–1739	<b>Humphrey Prideaux</b>	M	CU	U	P: <i>-one</i>	P		
1680–1739	Humfrey Wanley	M	P	N		P		C
1700–1719	Henry Liddell	M	GL	N		P		C
1700–1739	Daniel Defoe	M	P	N		C		C
1720–1759	Charles Lennox	M	N	N	P: <i>-body</i>			C
1720–1759	Thomas Pelham-Holles	M	N	N	P: <i>-body</i>			C
1720–1779	<b>Thomas Gray</b>	M	P	N	P: <i>-body</i>	P	P: past	
1720–1779	Thomas Secker	M	CU	U		C		C
1720–1800	<b>Samuel Johnson</b>	M	P	N	P: <i>-body</i>	P		
1720–1800	Roger Newdigate	M	GU	N	P: <i>-one</i>	C		
1740–1800	William Cowper	M	P	N		C		P
1760–1800	Jeremy Bentham	M	P	N		C	P: complex	
1760–1800	<b>Charles Burney</b>	M	P	N	P: <i>-one</i>	P		
1760–1779	Erasmus Darwin	M	P	N		P		C
1780–1800	Erasmus Darwin	M	P	N		C		C
1760–1800	William Jones	M	GU	U		C	P	
1780–1800	William Clift	M	P	U			P	C

However, six individuals (displayed in boldface in Table 14.3) are consistently progressive: poet, translator and writer Elizabeth Carter, writer Frances Burney, dean of Norwich Humphrey Prideaux, poet and literary scholar Thomas Gray, author and lexicographer Samuel Johnson, and musician and author Charles Burney. Carter, the Burneys (daughter and father) and Johnson were contemporaries and moved in the same literary circles in London, whereas Gray was a somewhat more remote literary figure. A social network approach might thus explain some of the progressiveness of Carter, Johnson and the Burneys (cf. Pratt & Denison 2000). As Carter and Johnson were earlier and more central members of the Bluestocking network and the Literary Club (respectively) than the Burneys, this implies that Carter and Johnson would have been more influential in the diffusion of changes as early adopters (e.g. Sairio 2009; Bax 2005; Conde-Silvestre 2012).

While the other professionals favour the indefinites in *-body*, the Burneys prefer *-one*, perhaps as a shared style within the family. All of them favour *its* in addition to the incoming indefinites, and Gray is also an overuser of the past-tense form of the progressive aspect. As with the full data set, none of the consistently progressive users are socially mobile, except for the only non-literary person, Prideaux, who represents an earlier period and is a social aspirer who eventually rose to the rank of upper clergy. His preference for *-one* is in line with the clergy of the time in general, while *its* was preferred in the South, where he spent much of his life.

If we discount Prideaux, the uniting factor behind the consistently progressive individuals thus seems to be a literary profession combined with a relative lack of social mobility. This is also something that distinguishes them from other members of the Bluestocking network included in our corpus who are not consistently progressive in terms of incipient and mid-range changes. For instance, Hester Piozzi belonged to the gentry and married beneath her, and while Mary Wollstonecraft was a professional, she became one through downward mobility, her father having been a gentleman farmer.

On the other hand, three men are or become consistently conservative: author Daniel Defoe and physician Erasmus Darwin, who are professionals, and bishop Thomas Secker, all in terms of *its* and *-ity* (there being no conservatives in the incoming indefinites or progressive aspect). As there are professionals and clergy among the consistently progressive users as well, social rank alone does not explain these results, nor does social mobility. Especially Defoe would rather seem to fit the profile of the consistently progressive individuals, but he is an earlier writer and led quite a different life as a tallow chandler's son and a religious dissenter who spent a considerable amount of time travelling on government business.

Focusing on gender, two of the seven women are consistently progressive, while none are consistently conservative, which points to women's tendency to

lead changes. Moreover, the proportion of progressive women out of all women listed here is greater than the proportion of progressive men out of all men for each change except for *its*, which was indeed led by men for most of the 18th century.

### 14.3 Conclusion

Despite the heterogeneous nature of the changes and the methods used to discover the outliers, some patterns have emerged in the behaviour of conservative and progressive individuals. We have found that most people are not outliers in terms of multiple changes, and when they are, their behaviour may vary from change to change depending on e.g. the stage of the change or the social evaluation of the variants in question. However, we have also discovered some individuals whose outlier status is consistently conservative or progressive. Most of the consistently conservative individuals are men, and clergy are overrepresented among them. The consistently progressive individuals, on the other hand, are most often professionals whose social status remains relatively static, and four of the six individuals who are consistently progressive in terms of incipient and mid-range changes were found to move in the same London literary circles, suggesting a network effect. Perhaps these individuals felt free to experiment with their language use, having less to lose than their upwardly or downwardly mobile peers.

## Changes in different stages

### 15.1 Introduction

Minna Palander-Collin

The linguistic changes that we have analysed in this volume were at different stages in the 18th century: some were practically completed, others in mid-range and one, the progressive aspect, only beginning to spread. Even those changes that reached completion in the course of the long 18th century were at different stages at the beginning of it. They also patterned differently in sociolinguistic terms, or, as the case may be, failed to show any obvious sociolinguistic profile in the data, and variant-specific patterns could emerge within a linguistic variable.

In this chapter we will compare the processes of change we have analysed in individual chapters. Special attention is paid to the time courses of these changes and the sociolinguistic patterns associated with them. We are interested in the extent to which changes that have reached the same stage also pattern in a similar manner socially across time, but our changes show somewhat different stages during the 18th century and direct comparisons are often difficult. Table 15.1 summarises the linguistic features studied and indicates the phase of change in consecutive periods as well as those social factors that proved relevant in each period (for the phase of change, see 5.1.2). All studies on linguistic features in Chapters 6–12 looked at gender and social class as far as it was possible with the data available. If gender and/or social class were relevant, i.e. statistically significant, factors in a given period, they have been marked down in the table. If they are included in parentheses, there seems to be a tendency where gender and/or social class has some impact on linguistic variation but a clear pattern or statistical significance cannot be discerned. As the two linguistic forms *thou* and *do*, which are ‘completed’ for the whole of the 18th century, are so infrequent in the data, nothing very conclusive can be said about the impact of gender or social class. Most likely, they are not important but what emerges instead in both cases is a register related use. The progressive, *-ness* and *-ity*, on the other hand, are difficult to assign a stage of change as they are not measured within a linguistic variable.

Table 15.1 Summary of linguistic features according to the stage of change and social variation (gender, social status)

	1680–1699	1700–1719	1720–1739	1740–1759	1760–1779	1780–1800
	stage of change	stage of change	stage of change	stage of change	stage of change	stage of change
	social variation	social variation	social variation	social variation	social variation	social variation
<i>thou</i>	<b>nearing completion</b>	<b>completed</b> (shift from status related markedness to register related)	<b>completed</b> (shift from status related markedness to register related)	<b>completed</b> (shift from status related markedness to register related)	<b>completed</b> (shift from status related markedness to register related)	<b>completed</b> (shift from status related markedness to register related)
periphrastic <i>do</i> <i>-man</i>	<b>nearing completion</b> (register) <b>completed</b> (gender; primarily used by men), social status (upper classes leading) <b>nearing completion</b> gender (men leading)	<b>completed</b> (register) <b>completed</b> (gender; primarily used by men), social status (upper classes leading) <b>completed</b> gender (gender; women leading)	<b>completed</b> (register) <b>completed</b> (gender; primarily used by men), social status (upper classes leading) <b>completed</b> gender (gender; women leading)	<b>completed</b> (register) <b>completed</b> (gender; primarily used by men), social status (upper classes leading) <b>completed</b>	<b>completed</b> (register) <b>completed</b> (gender; primarily used by men), social status (upper classes leading) <b>completed</b>	<b>completed</b> (register) <b>completed</b> (gender; primarily used by men)
<i>says</i>	<b>mid-range</b> gender (women leading)	<b>completed</b> gender (gender; women leading)	<b>completed</b> gender (gender; women leading)	<b>completed</b>	<b>completed</b>	<b>completed</b>
<i>does</i>	<b>mid-range</b> gender (women leading), social status (upper classes leading)	<b>nearing completion</b> gender (women leading), social status (upper classes leading)	<b>nearing completion</b> gender (women leading), social status (upper classes leading)	<b>completed</b>	<b>completed</b>	<b>completed</b>
<i>has</i>				<b>completed</b> (social status)	<b>completed</b> (social status)	<b>completed</b> (social status)

	1680–1699	1700–1719	1720–1739	1740–1759	1760–1779	1780–1800
	stage of change	stage of change	stage of change	stage of change	stage of change	stage of change
	social variation	social variation	social variation	social variation	social variation	social variation
<i>its</i>	<b>mid-range</b> social status, gender n/a	<b>mid-range</b> social status, (gender)	<b>mid-range</b> social status, gender	<b>mid-range</b> social status, (gender)	<b>nearing completion</b> social status, gender	<b>nearing completion</b> social status, (gender)
<i>-body</i>	<b>mid-range</b> gender (women leading), social status (upper classes leading)	<b>mid-range</b> gender (women leading), social status (upper classes leading)	<b>mid-range</b> social status (upper classes leading)	<b>mid-range</b> social status (upper classes leading)	<b>gender (women leading)</b>	<b>mid-range</b> gender (women leading)
<i>-one</i>	<b>incipient</b> social status	<b>incipient</b> social status	<b>new and vigorous</b> social status	<b>new and vigorous</b> social status	<b>new and vigorous</b> social status	<b>new and vigorous</b> social status
<i>progressive</i>	<b>incipient?</b> (gender; men leading), social status	<b>incipient?</b> (gender; men leading), social status (middle classes leading)	<b>incipient?</b> social status (middle classes leading)	<b>rising=new and vigorous?</b> social status (middle classes leading)	<b>rising=new and vigorous?</b> (gender; men leading), social status (middle classes leading)	<b>rising=new and vigorous?</b> (gender; women leading), social status (middle classes leading)
<i>-ness</i>	<b>stable</b> no gender difference	<b>stable</b> no gender difference	<b>stable</b> no gender difference	<b>stable</b> no gender difference	<b>stable</b> no gender difference	<b>stable</b> no gender difference
<i>-ity</i>	<b>lower productivity</b> combined gender and register difference	<b>lower productivity</b> combined gender and register difference	<b>mid productivity</b> combined gender and register difference	<b>mid productivity</b> combined gender and register difference	<b>higher productivity</b> combined gender and register difference	<b>higher productivity</b> combined gender and register difference

## 15.2 From incipient to mid-range and beyond

Minna Palander-Collin, Mikko Laitinen, Anni Sairio and Tanja Säily

This section first looks at changes that range from incipient (below 15%) and new and vigorous (15–35%) to mid-range (36–65%), with some nearing completion (66–85%) but not reaching that stage (over 85%) by the end of the 18th century. The changes in these stages include the progressive, the indefinite pronouns ending in *-body* and *-one*, and *its*. The final section focuses on changes in derivational productivity and the difficulty of determining these processes in terms of distinct stages.

### 15.2.1 Time courses of change

The three changes we are comparing here can be regarded as change in progress, but they are still different in many ways as we will illustrate below. The changes in *-body*, *-one* and *its* can be measured as variables and we can say that by the end of the 18th century *-body* had reached mid-range, *-one* was still in a new and vigorous stage, whereas *its* was already nearing completion. The progressive, on the other hand, cannot be treated in terms of a linguistic variable and its frequencies are measured in normalized frequencies. It is therefore difficult to say which stage of change the form had reached by 1800.

Though the progressive was a low-frequency phenomenon, it became significantly more common over the 18th century as it climbed up from 4.45 (/10,000) at the beginning of the century to 10.88 at the end. It was used predominantly in the present tense. The progressive passive was not observed in the CEECE apart from the isolated case (described by Pratt and Denison 2000 as radical experimentation in Late Modern English), so all grammatical forms were not yet attested and the change was still ongoing. Compared to the 18th century, the progressive increased more vigorously during the 19th century; however, according to Anderwald (2012: 36), the positive evaluations of *be+ing* in the nineteenth century suggest that it was changing at a slow pace. Contemporaries did not perceive it as change in progress, and there thus appears to be an element of quiet stability in its increase.

The compound indefinite pronouns *-body* and *-one* first emerged in Middle English and their development has to be viewed in the complex grammatical context of other competing indefinite pronouns. Previous historical sociolinguistic investigations have shown that two of the compound indefinite pronoun variants were on the increase in correspondence data by the late 17th century (Nevalainen & Raumolin-Brunberg 2003). On the one hand, *-body* had started to increase in the second half of the 17th century and was new and vigorous (15–35%), replacing the

independent forms as the most frequent variant by the early 18th century. On the other hand, the share of *-one* remained lower than *-body* throughout the decades. There are two possible language internal factors for the early dominance of *-body*. As shown by Raumolin-Brunberg & Kahlas-Tarkka (1997: 74), its introduction to all indefinite pronoun series was much faster than that of *-one*. Additionally, despite the fact that the meanings of both the variants denote singularity, the semantic weight of *-one* seems to have prolonged the grammaticalization process more than that of *-body*. These factors seem to play a role in its early success.

The diachronic trajectories in the results above show a marked cross-over in which the independent forms lose out to *-body* in the late 17th century as the main variant form in correspondence. The indefinites in *-body* undergo a period of vigorous growth and the change reaches a mid-range stage by the first decades of the 18th century. However, their increase is stalled and the share of *-body* starts to decline by the mid-century. In this process, they become stylistically marked as more informal and casual, associated more with spoken genres than written. The forms in *-one* remain minor variants until the early 18th century once the decline of *-man* to a minor variant is completed. As pointed out above, the incipient stages of this decline take place a century earlier in correspondence than in the literate texts investigated in D'Arcy et al. (2013). Their results show that the share of *-body* started to decrease in edited prose in the mid-19th century in a process in which the more ubiquitous *-one* became the prestige form used more frequently in formal and literate genres. All in all, the main forms remain variable in correspondence data at the end of the 18th century as *-body* is the dominant form, and a mid-range variant (36–65%), and *-one* reaches the new and vigorous stage by 1800.

In comparison to *-body* and *-one*, the progression of *its* is much faster. A possible explanation may lie in the relative simplicity of the linguistic variable in the third-person neuter possessive in comparison to indefinite pronouns. With regard to *its* and the main variant form *of it*, previous corpus studies indicate that *its* had been available as the third-person neuter possessive singular determiner at least from the beginning of the 17th century, and already by the 1650s *its* had gained the dominant position (Nevalainen & Raumolin-Brunberg 1994: 176). The use further increased in the 18th century, reaching c. 80% share of the variable towards the end of the century, so that by 1760–1800 the change was nearing completion. The final completed stage was reached somewhat later in the mid-nineteenth century, but just as *-body* and *-one* continue as variant forms in Present-day English, *its* and *of it* variation still exists.

### 15.2.2 Sociolinguistic patterns

All three variables are relatively similar with regard to their absolute frequencies. The frequencies vary between some 1,400 and 1,700 instances in the CEECE data. These raw frequencies make possible nuanced sociolinguistic analyses, as seen in the preceding chapters, but we soon encounter the problem of vanishing evidence when we increase the number of independent variables. In addition, it is important to point out that language internal grammatical factors also played a role in the development of all of the features with certain grammatical environments favouring a specific variant form. The most salient sociolinguistic variables differed in each case and in different phases of the change.

Rank and register were shown to have the most significant influence in the use of the progressive. Middle class writers, specifically the professionals, were ahead of the other ranks throughout the century, which suggests that this quiet, seemingly inconspicuous change was led from below during the Late Modern period. The final decades of the century show that the usage increased throughout the social strata; at this point the rank difference started to even out. The progressive was also particularly frequent in familiar communication, correspondence between close family members (FN). This seems to confirm the associations of the progressive as a more “spontaneous, unmonitored, colloquial” language feature (Kranich 2010). As for gender, this was not a significant variable. Women’s letters do not provide enough data of this low-frequency item until the end of the century, at which point women take the lead.

When we examine the overusers of the progressive, these outliers represent writers of lower and middle class background, both men and women, some of them social risers, who are active during the latter part of the century and who focused their use of this feature in their letters to close family members and close friends. The outliers thus epitomize the general trends that were observed: increased activity in usage during the latter part of the 18th century, in a familiar register, and largely as change from below.

With regard to the forms in *-body* and *-one*, the sociolinguistic stratification in CEEC is such that the change is led by women in the 17th century, but there are no significant correlations with writers’ social status. In addition, the earlier results indicate that *-body* was more frequently associated with the South (London, the Royal Court, and East Anglia), whereas *-one* was dominant in the North. However, the size of CEEC makes it difficult to study a low-frequency variable in general, and the larger size of the CEECE offers more insights of social stratification and socio-cultural context for understanding this variable. If we use indirect evidence, the examination of the 18th-century grammars suggests that the changes in the

indefinites seem to have taken place below the level of linguistic awareness as they are not commented upon in grammars, but the corpus results question this observation.

The results show that the change towards *-body* was clearly promoted by women, but there are no gender-related patterns found for *-one* in correspondence data. This observation is confirmed both by the correlational results and the non-parametric bootstrapping evidence. This evidence here does not corroborate some previous observations that the forms in *-body* would have at the early stages been associated with men. In fact, evidence of vernacular associations of *-body* remains scarce in correspondence data. The results illustrate that it is closely associated with the highest social layer, i.e. the nobility, up to the mid-century. Similarly, it is firmly established in London a few decades before the other areas, as was also the case in Nevalainen & Raumolin-Brunberg (2003). When one correlates the variant forms with writers' years of birth, the results show that *-body* in the early decades was a generational change as it peaks in the letters of those individuals who were born after the turn of the 17th and 18th centuries. It starts to decline for the following generations, and *-body* and *-one* enter complementary trajectories.

These results suggest that the spread of *-one*, once it enters the stage of vigorous growth in the mid-century, is a change from above in which the driving force is the highest social layer. The *-body* forms in the first period are clearly associated with London, but this distinction disappears during the course of the century.

The incoming *its* seems to have spread regionally to different parts of the country before the more rapid rise of the form from 1740 onwards, when the dialectal differences level out. Just like the spread of the progressive, the development of *its* seems to be a change from below rather than from above in the sense that lower ranks lead the change. The diffusion is led by men for the most of the eighteenth century, but women quickly increase their use up to 90% towards the end of the century when the variant is firmly established as the third-person neuter possessive. Both generational and communal change operate simultaneously as different generations increased their use from what they must have learned in their infancy but not to the same level. The generations born immediately before and at the beginning of the final rapid rise no longer differed from each other.

In this data set, we have some possibilities of observing the significance of social variation in different stages of change as we have two changes in mid-range: *-body* in 1682–1800 and *its* in 1680–1759. In both cases we found gender and social variation during this phase. In the case of *its* it was not constantly significant, but it is difficult to say whether this is an artefact of varying quantities of data from different social ranks and genders in different subperiods.

### 15.2.3 Issues of change in productivity

The discussion so far has ignored the change in the productivity of *-ity*. This is because the nature of the change differs from the others in that the stage of the change is more difficult to evaluate. As with the progressive, there is no clear linguistic variable involved; furthermore, it is even more unclear what would constitute a “completed” change in terms of productivity.

What we may be able to observe is the stage at which an affix becomes productive, which in the case of *-ity* is in the Middle English period (Dalton-Puffer 1996: 106–107; Gardner 2013: 108–111; but cf. Dalton-Puffer 1994). At first the suffix occurs in loanwords from French, after which it starts to be perceived as a word-formational element in English, probably first by bi- or trilingual (English–French–Latin) individuals, who would have most often been highly educated men. However, a second important stage in its development takes place in the 16th century, when *-ity* is increasingly used in calques on Latin and from there develops its automatic productivity on bases in *-able* (Marchand 1969: 312–314; see, however, Dalton-Puffer 1996: 107 for earlier formations on *-able*). Again, we may assume the change to have been led by men with a classical education.

In the correspondence genre, we have evidence of a continuous growth in the productivity of *-ity* in the 17th and 18th centuries, possibly led by the middling rank of professionals (Säily & Suomela 2009; Chapter 12 above). In the 17th century, women are lagging behind, but by the 18th century, they have mostly caught up with men, with the exception of letters written to close friends, in which register men exhibit a more creative and playful use of *-ity*. The lowest classes, however, are still lagging behind in the 18th century. While the change may be linked to the overall increase in the productivity of *-ity* observed by Lindsay & Aronoff (2013) in the OED, part of it may also be due to stylistic change in middle- and upper-class letter-writing practices (cf. Biber & Finegan 1997).

Comparing this change with the three discussed above, we can see that similar social categories are at play: social rank, gender and register may all affect productivity. As a stylistic choice, *-ity* in the 18th century is both elevated (as a Latinate, “learned” suffix) and an involvement feature; as such, it is perhaps more akin to the outgoing second-person singular pronoun *thou*, discussed in the next section.

### 15.3 From nearing completion to completed

Terttu Nevalainen, Mikko Laitinen, Minna Nevala and Arja Nurmi

This section discusses the final stages of four processes of change, showing that, while variation may complicate any attempts at a general sociolinguistic typology of linguistic change, it nonetheless allows us to discern certain patterns that changes display when they are nearing completion and become completed. A recessive variant typically undergoes register change but it can also continue to display gender and social status variation (e.g. *hath*, indefinite *-man* compounds). Moreover, outgoing forms can be grammatically repurposed (*do*) and assigned new socio-pragmatic functions (*thou*).

#### 15.3.1 Time courses of change

Looking at four sets of declining expressions, we found that the use of the second-person pronoun *thou* had largely receded by the end of the 17th century and was confined to specific registers in the next. In fact, estimating the number of *thou* users by the century, no quantitative change was found between the 17th and 18th century in either of the two letter recipient categories considered, family (6%) and close friends (2–3%). There was, however, a noticeable drop in the frequency of the use of the pronoun in letters to family members in the 1700s compared to the previous century (less than one instance per 10,000 words).

A declining trend could also be traced in the frequency of periphrastic *do* in affirmative statements in the 18th century. Although a significant alternative, it had never become the majority choice as a semantically empty tense carrier. The last two decades of the 17th century continued at about the same level of use as the previous decades but the 18th-century data showed a rapid decline in the normalized frequencies of the outgoing form. These frequencies were approaching those reported for Present-day spoken usage in several studies, suggesting that *do* use had already reached its current low-level plateau. However, the levels at which affirmative *do* was used were roughly ten times higher compared to *thou*, which suggests that there was a distinct difference in the relative pervasiveness of these two recessive features – but not necessarily their relative salience, as we will suggest below.

Verbal *-s* spread by means of lexical diffusion, and the most frequent item, *have*, was the last to display the incoming form, the auxiliary *have* being even slower than the main verb. Of the three verbs studied, the incoming forms *has* and *does* were in mid-range in the last two decades of the 17th century, while *says* was already nearing completion. Having passed the 85% mark, the change was completed with

*does* and *says* in the first two decades of the 18th century, whereas *has* was only nearing completion at the time, and went to completion in the course of the next 20 years. The average number of individuals preferring the outgoing form *hath* was in the order of 5% during the 18th century, dropping from about one third in the last two decades of the 17th century to 10% in the first two decades of the 18th century, and declining steadily after that.

The linguistic variable of indefinite pronouns with singular human reference is complex in that, apart from one outgoing (*-man*) and two incoming compound variants (*-body* and *-one*), it involves a set of outgoing independent forms (*some*, *any*, *every*, *each*). The independent forms continued to decline in the course of the long 18th century from about 30% at the beginning to below 10% at the end, while *-man* compounds lingered on at a 10% frequency throughout the century. As the use of the two recessive sets of variants declined, the variable was reduced to competition between the two compound alternatives, *-body* and *-one*. As many as two thirds of the individuals studied preferred the outgoing forms at the turn of the 18th century but their proportion was reduced to one third from the mid-century on.

### 15.3.2 Sociolinguistic patterning of recessive variants

As noted above, the use of the recessive second person singular pronoun *thou* was confined to letters exchanged by nuclear family members and, to a lesser extent, by close friends. Moreover, as a form of address, *thou* could be used to index status and power relations, for example, by parents addressing their children and a husband addressing his wife. A similar trend emerged with periphrastic *do* in affirmative statements, as it tended to be used more in correspondence with intimates than with more distant recipients. However, the trend was not statistically significant, nor were the writer's gender or social status, except for the slight overuse of *do* in affirmative statements by the upper gentry. That, too, may be due to other factors, such as the dominance of that group in the first 20-year period studied.

The relationship between the writer and the recipient correlated to some extent with the outgoing *hath* variant of *have* in the first period, although other social variables emerged as more relevant, as was also the case with the outgoing indefinite pronoun variants. As to gender variation, the verbal *-th* and the independent indefinite pronouns and especially *-man* compounds continued to be used longer by men than women, confirming the often noted tendency of male conservatism in linguistic change. Social status variation was also detected with these outgoing forms. The *hath* variant was used longer by members of the clergy and of the lower ranks than by other social strata. The *-man* compounds were also overused by clergymen and, towards the end of the 18th century, relatively high levels were attested

among the gentry. However, taken together, the outgoing indefinite variants were attested longest among the professionals and members of the clergy.

Sociolinguistic patterning of other kinds could also be detected with the outgoing variants of the singular indefinite pronouns with human reference, which were the recessive forms to take the longest to go out of use in the 18th century. When the writers' years of birth were taken into account, the decline of the recessive variants proved generational rather than communal and the systematic gender difference prevailed. As to regional tendencies, the independent forms were preserved longer in the North than in London and elsewhere in the South.

### 15.3.3 Changing indexicalities

The diffusion of linguistic features can, but need not, be connected with the speakers' awareness and marked by changes in the ways in which the incoming and outgoing forms are evaluated. Evaluation typically involves specialization along dimensions such as colloquial – formal/obsolete, local – supralocal, and vernacular – standard. All the recessive features that we have discussed were canonized in biblical use, the *King James Bible* (1611) and the *Book of Common Prayer* (1662), but the ways in which they were indexed in the course of the 18th century did not follow one single pattern.

With *thou*, uses related to register came largely to replace those related to status during the 18th century. Biblical and literary quotes were used to create and maintain interpersonal relations on a more distant level, while creative use of *thou* could license interpersonal intimacy and familiarity between friends. The relative overuse of the indefinite *-man* compounds and the third-person *hath* by members of the clergy, also points to conscious awareness of register associations and formality distinctions but not of the same kind as were created using *thou*. A second-person pronoun, *thou* could index the writer's involvement with the addressee ("other-involvement"), while, as third-person markers, *-man* compounds and *hath* could not.

*Hath* must have been associated with educated written usage by some writers, but it possibly represented a locally valued regional variant for others. The long history of verbal *-s* provides a good example of a series of changing evaluations with varying degrees of diatopic (regional) and diatypic (register) differentiation over time. Contact between the northern *-(e)s* and the southern *-(e)th* created both regional and register differentiation in the south as well as in the north when *-(e)th* spread there especially in writing in the 15th and 16th century (Moore 2002, Nevalainen & Raumolin-Brunberg 2003: 177–180). However, comments such as those made by Gil (1619) on *has* being a northern form suggest a continued regional

bias in favour of the retention of *hath* in the south in the early 17th century. Some vestiges of regional distinctions can be detected in individual usage in the CEECE in the 18th century although they are neither frequent nor systematic.

Although the use of affirmative periphrastic *do* was frequent compared to *thou*, for example, its social evaluation is more difficult to judge. One can infer from the linguistic environments it inhabits that ego-involvement is one function that supported the outgoing form: it was preferred with first-person subjects and verbs expressing speech-acts, cognitive process and emotion, in that order. However, there was also a large group of other verbs that co-occurred with noun-phrase subjects and could not be analysed in these terms. Another explanation, often given in studies of *do* in affirmative statements, is also offered here, namely, that the weight of an NP subject could trigger the use of *do* to increase the weight of the VP.

What was the role of the 18th-century normative grammar in these various processes? Chapter 3 and our surveys of the individual changes suggest that it was indirect at best. Poplack, Van Herk & Harvie (2002: 94) summarize in general terms the way in which 18th- and 19th-century grammars treated the outgoing forms that we have discussed:

[G]rammars frequently mentioned the co-existence of a high-status innovation and an older variant with a long tradition in educated or formal use. Forms such as *thou*, third-person *-eth* or second-singular *-est*, and unstressed periphrastic *do* in the present and preterite, were all probably near-moribund in common usage for several centuries before grammars were willing to dispense with a reference to them, or to euthanise them as ‘solemn’ or ‘ancient’.

## A wider sociolinguistic perspective

Terttu Nevalainen

This book was partly intended as a follow-up to the work that the authors had carried out on the Early Modern English period from the 15th century onwards, using the same corpus of personal correspondence sampled on sociolinguistic principles. Our approach was therefore motivated by our previous findings and the sociolinguistic generalizations based on them. This final chapter puts our results in a wider sociolinguistic perspective.

### 16.1 Rate and phase of change

The key empirical issues we focus on include the temporal trajectories and phases of linguistic changes. In sociolinguistics in general, the stage of development constitutes an important dimension in interpreting patterns of linguistic variation associated with a change in progress (Tagliamonte 2012: 61–62). In our studies, the changes analysed in terms of linguistic variables are shown to follow the S-curve pattern. Processes that are not readily analysable in such terms can also be traced over time. Table 15.1 (above) surveys the real-time trajectories of the processes that we have studied. Using the model of periodization introduced in 5.1.2, all the major stages from incipient to completed emerge from the data, but not all of them with any one process.

Comparing the progress of the two changes that are **incipient** in the 1680s and 1690s, the progressive and the indefinite pronouns in *-one*, we find that neither of them progresses past the new and vigorous stage during our study period.<sup>1</sup> In the case of *-one* compounds, the new and vigorous stage runs from the 1720s onwards, while with the progressive the incipient stage lasts longer, the frequency of the construction only picking up in the latter half of the century.

Four processes are in **mid-range** at the beginning of our study period: the diffusion of verbal *-s* to *do* and *have*, the generalization of the neuter possessive *its*

---

1. Anni Sairio's estimate of the periodization of the progressive compares her findings from the CEECE with other studies that analyse the process in real time (see 11.3 and 15.2).

and the rise of the indefinite *-body* compounds. The *-body* compounds stand out from the rest in that they remain stable throughout the long 18th century, while *its* is nearing completion in the last two 20-year periods of the century. The diffusion of *has* reaches the nearing-completion stage at the beginning of the 18th century and is completed in the mid-century. The spread of *does* is the most rapid of the four processes as it goes from mid-range at the end of the 17th century to completion in the first decades of the 18th century.

Three of the processes studied are **nearing completion** towards the end of the 17th century: the diffusion of verbal *-s* to *say*, and the recession of *thou* and of affirmative periphrastic *do*. All three are completed in the early decades of the 18th century. Although the latter two are not accounted for here by means of linguistic variables, their frequencies fall low enough to justify the conclusion that their alternative expressions must be considered the quantitative norm in the 18th century (8.2.1, Nevala 2002). The retraction of the indefinite *-man* compounds had already reached **completion** by the beginning of our study period, when the combined frequency of the alternative expressions exceeded the 85% threshold in the data (Chapter 9, Appendix).

Except for the tail end of the lexical diffusion of verbal *-s*, the pace of linguistic change does not strike one as particularly rapid in the long 18th century. What we rather see are long periods of relative stability, processes nearing completion being completed, and those in mid-range moving to near-completion. Even changes that were incipient at the beginning of the period only moved to the new and vigorous stage in the course of the century. The *-man* and *-body* indefinites remained stable throughout the period within the indefinite paradigm, and so did the productivity of the nominal suffix *-ness*.

We could account for the situation in broad sociolinguistic terms by suggesting that there were no major “punctuating events”, such as massive population movements in the wider society at the time (cf. Labov 2001: 314). In the 18th century, there was, for example, no Civil War of the kind witnessed in the 17th century, which gave rise to extensive dialect contact and, as Raumolin-Brunberg (1998) suggests, acceleration of linguistic change. Hence we should perhaps not expect to see any sudden accelerating effects on linguistic changes in progress either. More research is naturally needed to support this provisional conclusion on the “slow” 18th century.<sup>2</sup> Furthermore, as was shown in detail in Chapter 15, we cannot attribute

---

2. Another selection of linguistic changes could naturally have prompted a different general impression. However, there is also evidence on other 18th-century processes that leads to a similar conclusion, for example, the slow decline of subject–verb nonconcord with plural existentials, and of negative concord in lower-class male writers in the CEECE (Nevalainen 2006 and 2009, respectively).

relative stability to most of these processes before the long 18th century. Not even the suffix *-ness* qualifies as a representative of diachronic stability (cf. Säily 2014).

We may of course wonder whether the socio-cultural circumstances in the 18th century could have impacted on linguistic change. As pointed out by Yáñez-Bouza in Chapter 3, the social developments of the time, including improved literacy and the industry of print culture, resulted in heightened language awareness. The rising ideology of standardization was inimical to change, especially to lower-class usages gaining wider currency (3.4.2). This intellectual climate may have had a role to play in slowing down the rate of linguistic change in general. In the next sections, social variation will be brought to bear on the issue.

## 16.2 Social patterns

Our earlier work on Early Modern English, discussed in Chapter 1, supports the variationist sociolinguistic approach in that it shows that social factors correlate with the diffusion of morphological and morphosyntactic changes in real time. Two sociolinguistic variables, gender and social status, were examined in all the studies in this volume. Register variation according to the recipient was also addressed with most linguistic changes, as were their regional origins and diffusion. Age-related patterns were analysed with changes in mid-range. The following sections compare our findings with the sociolinguistic generalizations based on these social variables (for a research summary, see Tagliamonte 2012: 25–35).

### 16.2.1 Gender

Gender has proved a robust social variable in studies of language variation, and its association with prestige has been characterized as a sociolinguistic universal (Hudson 1996: 202). In language change, female advantage has been found to be more common than male advantage (Tagliamonte 2012: 63). Although often combined in variationist studies, gender difference can occur independently from differences in socio-economic status. Labov (2001: 322), for example, notes that in the early stages of change, “gender operates as an independent and powerful factor”.

Statistically significant gender differences are also recorded with the processes of change studied in this volume. They are observed with all the changes studied except those that were practically completed by the 18th century, i.e., the choice between *thou* and *you* and the use of affirmative periphrastic *do*. The suffixes *-ness* and *-ity* also fail to show a general gender difference in the 18th century.

But in most cases the difference emerges. With verbal *-s*, a significant female advantage is detected even in the final phase of the process, which women writers

completed on average twenty years earlier than male writers (7.3.2). This pattern, women being roughly a generation ahead of men, is a common conclusion in present-day variationist studies of sound changes (Labov 2001: 304–306).

In the indefinite pronoun paradigm, the decline of the independent forms and *-man* compounds is consistently promoted by female writers. As to incoming forms, a distinct female advantage is found with the *-body* compounds, which occupy a mid-range position in the data throughout the 18th century. By contrast, the incoming indefinites in *-one* show no gender difference as they move from the incipient stage on to the new and vigorous phase in the course of the study period (9.3.1). Here we find a female retreat from a male-dominated variant with the *-man* compounds but not the reverse, male recession from a female-led change with *-body* compounds, as Labov (2001: 315–319) observed with certain sound changes.

Tracing a change in its early stages presents problems because the incoming form is a rare minority variant and the researcher is bound to encounter the problem of “vanishing reliability” (5.2). Moreover, because of their differential access to education and literacy, male and female writers are not equally represented in our data sources (4.3.2). Combined with a low overall frequency of the linguistic feature, these factors diminish our access to the social roots of linguistic innovations and chances to test empirical generalizations such as gender operating “as an independent and powerful factor” in the early stages of change (Labov 2001: 322).

Data imbalance can be partly responsible for the male advantage observed in the early diffusion of the progressive, but this may not be the whole explanation. Once the process begins to gain ground, its gender alignment changes to female leadership at the end of the 18th century (11.4). Male advantage also appears with the neuter determiner *its* as it advances from mid-range to near completion in the course of the 18th century. As with the progressive, it is only towards the end of the century that female writers become the more frequent users of the incoming form (10.4.4).

Following Tagliamonte’s research summary (2012: 61–62), it would be tempting to account for these gender switches as female correspondents’ sensitivity to the two features becoming part of polite language towards the end of the century. If this were the case, this interpretation would agree with the first part of Labov’s Gender Paradox (2001: 292–293) that “women conform more closely than men to sociolinguistic norms that are overtly prescribed”. However, the progressive and *its* had both supralocalized by that time and neither of them was overtly prescribed by contemporary commentators.<sup>3</sup>

---

3. There were other features, such as the singular *you was*, whose diffusion was curbed by proscription and heightened social awareness. However, the results in Laitinen (2009: 208–211) indicate that male writers showed higher frequencies of the standard form *you were* than female writers as the century wore on. See also Yáñez-Bouza (2016: 170–172).

This is also the case with the other female-led features, verbal *-s* and *-body* indefinites, which were established earlier in the century without attracting prescriptive remarks from contemporary commentators. Furthermore, in these cases the female advantage goes further back in time and can be interpreted as a sign of the gender difference observed with supralocalizing features in general (Nevalainen & Raumolin-Brunberg 2003: 122–125). As these features continued to diffuse across the language community in the 18th century unaffected by overt prescription, these data rather support the latter part of Labov’s generalization (2001: 293) that women conform less than men to sociolinguistic norms when they are not overtly prescribed.

### 16.2.2 Social status

The linguistic processes we have studied mostly concern syntactic and morphological variation. Apart from distinguishing traditional regional dialects (Trudgill 1999), syntactic and morphological variables function as widespread class markers in present-day English (e.g. Chambers 2009: 56–57). Since we study social variation in language change, we do not interpret the correlations we observe automatically in terms of class markers, which are typically found with stable linguistic variables. One of our aims is to determine (a) whether a change diffuses from the lower social classes to the higher, or (b) *vice versa*, or (c) whether its diffusion begins in the middle of the socio-economic hierarchy. When a change is a conscious, superimposed one, it tends to originate in the highest social orders, but when it progresses from below the level of social awareness, variation studies and social network analyses would support the third alternative.<sup>4</sup>

This third option is identified by Labov (2001) as the *Curvilinear Principle*. Based on his findings on ongoing sound changes, he concludes that linguistic change spreading from below the level of social awareness “originates in a central social group, located in the interior of the socio-economic hierarchy” (2001: 188). James Milroy (1992: 213) talks about the “anvil model” of status and solidarity, suggesting that the highest and lowest social groups are characterized by dense and multiplex social networks that also serve as linguistic norm-enforcement mechanisms, whereas the middle groups are less constrained, and more open to linguistic influences from the outside. We will refer to Labov’s findings because he covers the

---

4. However, to be generalized in the language community, a change from below also needs to make its way to the upper social orders at some point; see e.g. Nevalainen & Raumolin-Brunberg (2003: 150) for Early Modern London and Lodge (2004: 149) for 17th-century Paris. For a richly illustrated application of the social network approach to Late Modern English, see Tieken-Boon van Ostade (2009: 100–118).

lifespan of linguistic changes rather than concentrating solely on their origins. The changes we have studied also align differently along the social hierarchy depending on the stage they have reached.

**Completed** changes, such as the second-person pronoun *you* and periphrastic *do*, no longer display any systematic social status differences in our 18th-century data. **Nearing completion**, verbal *-s* in *has* shows a nearly monotonic social class distribution with the lowest social classes being the slowest to complete the change (Table 7.2). The same is true of the decline of the independent and *-man* indefinites, which is led by the nobility with the gentry following suit (Figure 9.9.) These patterns agree in principle with Labov's observation (2001: 186) that the most advanced sound changes resemble stable linguistic variation, although they are less sharply stratified socially.

We would expect to find the curvilinear pattern in **mid-range changes** (Labov 2001: 187). The pattern does emerge with the neuter possessive *its* when male writers are compared across the social hierarchy: in mid-range, the change is led by professionals, with the clergy following closely behind. However, the indefinite pronouns in *-body* are first and foremost promoted by the nobility, followed by the gentry, which would qualify this process as a change led by the highest social groups. But the pattern is not monotonic: the clergy are the last to adopt the *-body* forms and, until the second half of the 18th century, prefer forms in *-one* instead (Figures 9.10 and 9.11). These findings may be related to the stylistic and semantic values associated with the two forms (15.2 and 16.3).

As Labov (2001: 186) points out, it is difficult to gain a clear view of the social locations of **incipient changes**. This is also true of the progressive aspect. Overall, it is the professionals who use the progressive more than the other social categories in the 18th century, especially in its 'complex' tenses (Figure 11.7). But when the progressive begins to gain ground in the final decades of the century, we find that only royalty and the upper gentry use it remarkably less than the other social categories, including the lowest (Table 11.4).

The Curvilinear Principle gains further support from the analysis of outliers, where professionals are found to be overrepresented among those who are in the vanguard of at least one change (14.2). Professionals are also found to be particularly frequent users of the derivational suffix *-ity* (Figure 12.4), whereas representatives of the lowest classes underuse the suffix in comparison to the highest (Figure 12.5). Since in this case we are talking about increased productivity rather than the diffusion of a grammar change, it is possible to interpret such findings as style-shifting among the language-conscious classes (see further 16.3, below).

Overall, "the interior of the socio-economic hierarchy" (Labov 2001: 188) proved a significant influence in linguistic change in our 18th century data. However, despite the view, suggested in 2.1, that the 18th century was probably

closer to the modern era than to the previous centuries in social terms, we are left with the question of how these middle classes and the professionals, in particular, compare with the middling ranks of earlier times and with those who are attributed middle-class status today.

Social mobility can provide an interesting point of comparison. In Nevalainen & Raumolin-Brunberg (2003: 137, 154) we found it justified to analyse successful social aspirers as a separate social category in the Tudor era, but *individual social mobility* did not correlate consistently with either linguistic progressiveness or conservatism in the current 18th-century data (14.2). At the same time, it is harder to evaluate changes in *group mobility*. It is in evidence in many social groups, such as liberal arts professions, which diversified and moved up in the social hierarchy in the course of the long 18th century (2.4). These professional groups are well represented in the CEECE, especially in the second half of the 18th century (Figure 4.4).<sup>5</sup> In sum, although we have evidence in support of the Curvilinear Principle in our data, broad structural similarities should not mask content differences in diachronic generalizations.

### 16.2.3 Region

A particular challenge for the study of real-time linguistic change is tracing it back to its origins. With comparable data from different regional varieties, the dialectal origins of linguistic features and dialect contact can be explored using comparative sociolinguistic methods based on variationist statistical techniques (e.g. Tagliamonte 2012: 167–173).

Since most of the changes we investigated had passed the mid-range and advanced towards completion, they had already supralocalized to the extent that it was not possible to reach back to their regional origins in the 18th century materials. Instead, data from the previous centuries was drawn on to throw more light on their regional provenance. Recessive features persisted longest in traditional regional dialects and thus fell outside the scope of our data, although outliers provided some data on potential regional usages.

For mid-range changes we could present comparative information on the regional distribution of the competing variant forms. The incoming indefinite pronouns in *-body* were associated with the capital region, and the neuter determiner *its* as well as the indefinites in *-one* also showed southern as opposed to northern regional associations in our 18th-century data.

---

5. Connected with internal migration, the rise of the professional classes also tallies with the relative prominence of London in the letter writers' domiciles (Figure 2.1 and Table 4.2).

#### 16.2.4 Real and apparent time

Sociolinguistic research usually makes a distinction between *real-time* and *apparent-time* approaches, the latter serving, as Tagliamonte (2012: 43) puts it, as a synchronic surrogate for chronological (real) time. In contemporary sociolinguistics, studies of real-time variation and change are still rare but are gaining ground as a result of *trend* studies, which add data to earlier analyses from new generations, and *panel* studies, which follow the same individuals through the lifespan (e.g. Labov 1994: 85–112, Wagner & Buchstaller 2017).

The fact that the year of birth is available for most writers included in the CEECE makes it possible to add an apparent-time perspective to studies of real-time processes of change. An extended apparent-time analysis of the indefinite pronouns (9.3.2) traces the new and recessive variants according to speaker age. It shows that the dominant variants for the oldest age group, who were born before 1639, are the independent forms and those ending in *-body*, while the youngest age group, those born after 1760, predominantly use forms in *-body* and *-one*. Hence the increase in the frequency of *-one* is clearly due to the youngest age groups.

A similar apparent-time pattern of *generational* change emerges in the diffusion of the possessive *its* (10.4.2). At the same time, three age cohorts out of four also display a *communal* pattern, increasing their use of the incoming form in the period between 1740 and 1800, when the change advances from mid-range towards completion. This is not unexpected for morphological changes in general. According to Kerswill's research summary (1996: 200), morphologically conditioned changes and lexical productivity are among the second-dialect features that can be acquired throughout the *lifespan*. This was also the outcome of the Early Modern English case studies in Nevalainen & Raumolin-Brunberg (2003, Chapter 5).

All the letter writers in the CEECE are adults but some of them wrote over a longer period of time, thus making it possible to examine whether they changed their usage over their lifespan. The study of the possessive *its* suggests three possibilities: a writer's use of the incoming form can remain stable, or it can either increase or decrease over time. Unlike the aggregate findings on generational and communal processes of change, the individual patterns are based on relatively small numbers and the observed variation may therefore stem from factors other than lifespan change (10.4.6).

Baxter & Croft (2016: 169) account for patterns of lifespan variation in terms of accommodation and argue that speakers who accommodate more readily to their interlocutors will exhibit more continuous lifespan change, whereas those who accommodate less readily will display more sudden changes, or indeed no change at all. The latter is the case with the mostly conservative clergymen in our data (Table 14.1). However, as Baxter & Croft (2016: 170) admit, their model does not

account for socially structured variation, which can relate to variation and changes in social identity and social networks across the speaker's lifespan.<sup>6</sup> Nevalainen, Raumolin-Brunberg & Mannila (2011) also consider the overall rate of change as a factor contributing to individual alignment with an ongoing process. But such variation can be further related to the social evaluation of linguistic features undergoing change. We will turn to this topic next.

### 16.3 Social evaluation and register

Although linguistic variation is not always thought of in social semiotic terms, Eckert (2012: 94), for one, argues that it can potentially express “the full range of a community's social concerns”. Our perspective on social evaluation is diachronic and we would expect to see its manifestations in “repertoires used in utterances by particular sociohistorical populations” (Agha 2007: 149). The linguistic changes we discuss involve processes as a result of which morphological and syntactic variables come to be associated with certain situated meanings as part of the repertoires of individuals and groups of language users. Our particular challenge has been to both observe and analyse how incoming and outgoing expressions are socially re-evaluated over time.<sup>7</sup>

Except for the passival use of the progressive, none of the incoming forms and constructions that we have studied was negatively evaluated by 18th-century normative grammarians. The recessive variants, by contrast, were variously characterized as formal or solemn or occasionally, as in the case of *thou*, as vulgar and rude. However, normative grammar is not our only source – or even the most reliable one – of social evaluation, which is evidenced at various levels of language use ranging from the preferences associated with particular social groups to individual linguistic choices in a given communicative context.

Our point of entry to social evaluation is through **register** differences, where *registers* can be broadly understood in Agha's terms as linguistic repertoires that particular sociohistorical populations recognize and connect to specific indexical

---

6. Social networks and lifespan change in the EModE part of the CEEC are discussed in the second edition of Nevalainen & Raumolin-Brunberg (2003), published in 2017 (Chapters 10 and 11).

7. This challenge is of course confronted in historical studies of any period. For example, Nevalainen & Raumolin-Brunberg (2003, Chapter 1) showed that the 16th-century use of negative concord cannot be judged according to the indexicalities associated with the construction in Present-day English.

values (2007: 81, 149).<sup>8</sup> We have studied potential register differences by analysing the social relations between the writers and recipients of letters. As linguistic differences are detected in the ways in which individuals perform in different social settings, the degree of social distance between the correspondents is expected to be reflected in the degree of formality of their language use (4.4.2).<sup>9</sup> Our findings bear this out but also show how interpersonal relations can be contextually monitored by the choice of given linguistic features and constructions.

Beginning with low-frequency verbal processes, we find that both incoming and outgoing features can be associated with interactions between social intimates. When the progressive aspect gains momentum and yields statistically significant results, it is shown to be used more by nuclear family members and close friends than by more distant correspondents (11.6). The same is true of the use of affirmative periphrastic *do*, which retreats towards the intimate end of the scale of social relations before fading out altogether in its earlier unstressed functions (Figure 8.2). At the same time, the registers associated with such interactions also provide the context for the main present-day uses of *do* as a sign of emphasis, contrast and emotion.

A divided profile emerges with the outgoing third-person *hath*, which is associated with both distance and intimacy. By using the recessive form, socially (upwardly and downwardly) mobile men could create a distancing effect even when writing to their immediate family and close friends. As *hath* persists longest with clergymen in the 18th century, the form could index the writer's social standing by signalling his linguistic conservatism.

There is explicit evidence for enregisterment of the styles associated with the incoming and outgoing variants of *have*. Sir Thomas Browne, a medical doctor and writer, consistently used *hath* in his letters to his eldest son but chose *has* when writing to his daughter; both of them were grown-ups. Browne combined *hath* with linguistic features of detachment, and *has* with those of involvement (7.4.2). By contrast, no such stylistic polarization is found in the correspondence of John Clift,

---

8. Agha (2007: 168) specifies that registers arise from the process of *enregisterment*: "A register exists as a bounded object only to a degree set by sociohistorical processes of enregisterment, processes whereby its forms and values become differentiable from the rest of language (i.e., recognizable as distinct, linked to typifiable social personae or practices) for a given population of speakers." Rather than clearly bounded objects, in our context registers represent repertoire fragments of a register indexing certain social concerns (cf. Agha 2007: 149). See also Eckert (2012: 96–97) and Johnstone (2016).

9. Different approaches make similar distinctions. Biber (2001) distinguishes a fundamental oral vs. literate dimension in a range of 18th-century and present-day registers, based on a multidimensional analysis of a large number of linguistic features in multigenre corpora. Similarly, the much cited model by Koch & Oesterreicher (1985) discusses the language-external criteria that distinguish between the language of immediacy and the language of distance (*Sprache der Nähe* vs. *Sprache der Distanz*).

a working-class West Country man, who systematically used *hath* when writing to his brother at the end of the 18th century. As the language of his letters is otherwise characterized by features of involvement, *hath* could have indexed his local roots. *Hath* was still in regional use in Cornwall at the time.

In the 18th century, the outgoing personal pronoun *thou* is rare but can still be found in correspondence between family members and close friends. Its functions range from signalling intimacy, and so diminishing the power difference between parent and child, for example, to constructing social identity and regulating the distance between adult correspondents. The writer's educated and cultured identity could be indexed by the use of Biblical and literary quotations. On the other hand, the correspondents' close relationship could be foregrounded by addressing the other party with *thou* in creative expressions of affection or banter, as was the case with Ignatius Sancho's letters to Laurence Sterne (6.5).

The letter writer's social identity as a witty person was also indexed by creative use of the suffix *-ity* in neologisms such as *cleverality*. Male writers overused the suffix when writing to their close friends (Figure 12.3). The fact that the suffix increased in productivity as the 18th century wore on was possibly an indicator of changing styles of letter writing among middle- and upper-class writers (12.5.3.2). For example, the overuse of *-ity* by Eliza Draper, another friend of Laurence Sterne, combines with features of a Latinate style.

An ingroup/outgroup contrast is likely to have been involved in the falling out of use of the indefinite pronouns in *-man*. The transparent male associations of the *-man* compounds were probably connected with their demise being led by female writers (Figure 9.6). These outgoing forms remained longest in the repertoires of the clergy and, until the mid-18th century, the varied class of professionals. But the clergy also promoted the incoming indefinites in *-one*, initially rejecting the forms in *-body*. It can be speculated that, for the clerics, these preferences echoed the Biblical register and formed part of their ingroup repertoire (9.3.2).

By contrast, the incoming indefinites in *-body* may long have had urban, "polite society" associations as they diffused from the capital region (Figure 9.13). After the Restoration of the monarchy in 1660, the social prestige of the Royal Court would have added to the attraction of the capital (cf. 2.6). It is therefore not unexpected to find that the form was particularly frequent in the repertoires of the royalty and the nobility, who spent long periods in the capital for both business and pleasure.

Something we have not discussed in detail are wholesale repertoire shifts associated with a writer's exposure to "polite society". A case in point is William Clift, who, having moved from Cornwall to pursue a career in London, abandoned his local and nonstandard linguistic features even when writing to his sister Elizabeth (3.4.3). Despite this, he continued as a frequent user of features of intimacy and involvement, such as the incoming progressive, a practice reciprocated by Elizabeth (11.7).

We hope to have shown that it would hardly be possible to approach individual language users' social evaluation of linguistic choices without access to the practices of the contemporary language community at large. In the light of this, one can interpret some of the outgoing features as widely enregistered stylistic resources indexing authority and religious and literary values (cf. Chapter 2, Polite society and rhetoric). However, as this brief survey has shown, some of the same linguistic features could also serve to index proximity and intimacy.

We found that many of the linguistic resources we studied were used to index the same kinds of common social concerns as present-day variant realizations of phonological variables, namely, social distinctions such as formal/informal, higher status/lower status, older/younger, rural/urban, and ingroup/outgroup (Labov 2001: 437, 463, 511; Tagliamonte 2012: 35). In the diachronic context of language change, the social meanings of linguistic variables are not fixed but are liable to change over time. Semantically underspecified, linguistic features can also be flexibly linked to different synchronous groups of language users, their social personae and practices (cf. Eckert 2012: 87, Agha 2007: 168). As such, these connections may remain local and transient. In general, our findings agree with Labov's observation (1994: 82–83) that the level of social awareness is maximal for changes coming to completion.

Finally, as shown by the study of the possessive determiner *its* in Chapter 10, not all linguistic variables are used to index social evaluation. It also appears from studies of present-day phonological variation and sound-changes that most speakers identify as *in-betweens* with respect to the social polarities discussed above. In other words, language users attend to these polarities to different degrees, as did the Jocks and Burnouts in Eckert's Detroit study (Eckert 2000: 139–141, 2012: 92–93; Agha 2007: 8–9, 150; Nevalainen, Raumolin-Brunberg & Mannila 2011). We have witnessed this range of variation – narrowing down as a change advances towards completion – both between and within social groups and in individual linguistic practices.

#### 16.4 The problem of continuation

In the register analysis we moved closer to the contexts of use of the variant forms and expressions. However, since the processes we have investigated represent changes in progress, this analysis would not give us access to the *actuation* of language change, the transition of a linguistic innovation from the individual to the community.<sup>10</sup> We should therefore rather consider what Labov calls the *problem*

---

10. This aspect of linguistic change normally escapes observation, but lexical innovations especially in the written medium can throw light on the beginnings of a change.

of *continuation* and look for general linguistic principles that are likely to “form the favorable undercurrent, or perhaps prevailing wind” for changes in progress (Labov 2001: 466, 499).

Some of the linguistic influences that may be propelling the changes we have studied are associated with long-term typological trends in English. Based on a systematic analysis of corpus data, Szmrecsanyi (2009, 2012) demonstrates a distinct development in English from increased *analyticity* beginning in Middle English, back to more *syntheticity* from the Early Modern English period onwards. In our selection of 18th-century linguistic changes, syntheticity also emerges as the stronger trend.

Szmrecsanyi et al. (2016) show that this development is marked in *grammar*, for example, in the increased use of the synthetic possessive *s*-form (*the New Year’s message*) at the expense of the *of*-paraphrase (*the message of the New Year*), a trend strongly reinforced by the increase in the NN alternative (*the New Year message*). In our data this trend was witnessed in the generalization of the determiner *its* as opposed to the analytic *of it* construction in the 18th century.

Moreover, including *lexical processes* in diachronic accounts of analyticity and syntheticity, as Danchev (1992: 34–35) suggests linguists should do, would indicate that English has gained substantial synthetic resources over time. After the transition from the earlier stem-based morphology to word-based morphology, a large number of derivational affixes such as *-ity* have been introduced into the language through borrowing. These are alternation-free and more transparent than the stem-based word-formation mechanisms found in Old English, but nevertheless they contribute to the synthetic resources of the language. In our data, the two derivational suffixes, *-ness* and *-ity*, show no signs of losing ground in the 18th century. Compounding is another synthetic word-formation process. It is illustrated by the rise of the indefinite pronouns in *-one* and *-body*, which superseded not only the *man*-compounds but also the more numerous simple forms.

These synthetic processes were counterbalanced in grammar by the introduction of the progressive aspect, which represents an analytic change proper. But an analytic process could also be stalled, as happened in the failure of the *do* periphrasis to spread to affirmative statements. Although studies like those by Szmrecsanyi take major steps forward in charting typological trends over time, clearly more research is needed to quantify the respective gains and losses of the two opposing tendencies (Danchev 1992).

Another typological process relevant in our context is the *Great Gender Shift*, the shift in English from grammatical to notional gender. It can be seen in the demise of the earlier neuter determiner *his* and the rise of *its*. The process must also have partly led to the loss of *-man* compounds through the strengthening of natural gender associations with the form, and thus giving rise to changes in its social evaluation. A similar development took place in the replacement of the personal

subject relative *which* by *who*. How this semantically-driven change was manifested in the language community is a sociolinguistic issue. Adamson (2007: n.p.) concludes her discussion of the topic by singling out the 18th century as a tipping point in the process:

In the long history of the Great Gender Shift in English, the eighteenth century marks a distinctive new phase, the moment when, in Labovian terms, “change from below” gave way to “change from above” and the process of change became affected by various kinds of metalinguistic awareness.

Adamson attributes this process to a change in the contemporary philosophy of language, implemented in pedagogical grammars, which promoted the notion of natural gender “as part of the wider belief that language should (and could) reflect the order of things in the world”.

### 16.5 Historical backprojection?

It goes without saying that more data on both language use and language users is needed to refine the findings discussed in this volume. Labov (2010: 368) notes that “the larger the size of the speech community involved, the more difficult it is to account for uniform patterns of linguistic change”. This problem of accountability is unavoidable in real-time studies of language change, where the size of even one and the same community fluctuates with time and is therefore difficult to estimate.

The method of *backprojection* has been used in historical population statistics, for example, by Wrigley & Schofield (1981) and Oeppen (1993) to arrive at an estimate of the population of England for the pre-census era. Moving in five-year steps from 1871 to 1541, the authors used an algorithm to derive estimates of population size, age structure and net migration based on information on births, deaths and the Census of 1871. Some of these estimates are presented in Section 2.3 (Figure 2.1). Demographic backprojections such as these can provide useful support for historical sociolinguistic studies but similar methods of estimation are not necessarily available across the history disciplines.

As far as language change is concerned, the *Principle of Contingency* states that “[f]actors determining the course of linguistic change are drawn from a pattern of social behavior that is not linked in any predictable way to the linguistic outcome” (Labov 2001: 503). Although historical sociolinguists can rely on the Principle of Uniformitarianism, and use the present to explain the past (Bergs 2012), it may not be possible to develop adequate models of backprojection for social behaviour. Instead, we can draw on the available evidence – linguistic, social and cultural – to identify cultural patterns invoked by social factors. Suggestions for systematic

multi-layered frameworks of socio-cultural contextualization of language use are presented, for example, in Wood (2004), Culpeper & Nevala (2012) and Nevalainen (2014b). They all approach the issue by crossing disciplinary boundaries in search for diachronic generalizations.

One candidate for a generalization is gender bias in language variation and change as a corollary of the tradition of distinct male and female social roles in all known societies (Coates 2004: 187, Romaine 2005: 109–111).<sup>11</sup> Typically, however, changing social circumstances are likely to be responsible for many of the long-term differences we have observed in the social patterning of linguistic changes. They include variation in social and regional mobility and contact between speakers, differences in the life styles of different social groups as well as in their access to the material and non-material culture of their time. Although it is not easy to factor these macro-level issues systematically into real-time studies of language change, they can provide new hypotheses for data-based exploration and explanatory information for interpreting the results obtained.

Another general point to be made is the relevance of cumulative research in the field of sociolinguistics. Here it is easy to agree with Holmes and Meyerhoff (2005: 16), who argue that “[w]e will make greater progress if we seek to accommodate insights from a variety of sources, rather than dismissing in a blinkered and unreflecting manner, results from currently unfashionable paradigms”. The methods and approaches that neighbouring disciplines can contribute to historical sociolinguistic research will alleviate what Labov (1994: 11) identified as the *Historical Paradox*, the situation where we know that the past was different from the present, but do not know how different it was. He suggests that solutions to this paradox should be sought by the application of multiple methods:

Particular problems must be approached from several different directions, by different methods with complementary sources of error. The solution to the problem can then be located somewhere between the answers given by the different methods. In this way, we can know the limits of the errors introduced by the Historical Paradox, even if we cannot eliminate them entirely. (Labov 1994: 25)

In search of motivations for the sociolinguistic patterns detected, the studies in this volume have moved from generalizations pertaining to the language community at large to the level of the individual language user. In doing so, we have turned to social and cultural history, sociopragmatics and discourse analysis for further insight.

---

11. Labov (2010: 368) argues more generally that gender patterns in language change are prototypical of broad cultural patterns. He further suggests that cultural patterns could also be associated with long-standing perceptions of urban and regional varieties, and traces the negative status of the New York City vernacular to the early 19th century.

The methods we have applied vary from simple ways of calculating frequencies to more sophisticated statistical techniques. By indeed “accommodating insights from a variety of sources”, and using several different methods, we hope to have reduced at least some of the errors introduced by the Historical Paradox. It remains for future research to determine, for example, exactly how slow – or rapid – linguistic change was in the 18th century.

## References

- Aarts, Bas, Joanne Close & Sean Wallis. 2010. Recent changes in the use of the progressive construction in English. In Bert Cappelle & Naoaki Wada (eds.), *Distinctions in English grammar: Offered to Renaat Declerck*, 48–167. Tokyo: Kaitakusha.
- Aarts, Bas, Maria José López-Couso & Belén Méndez-Naya. 2012. Late Modern English syntax. In Alexander Bergs & Laurel Brinton (eds.), *Historical linguistics of English*, vol. 2, 869–887. Berlin & New York: Mouton de Gruyter.
- Adamson, Sylvia. 1989. With double tongue: Diglossia, stylistics and the teaching of English. In Michael Short (ed.), *Reading, analysing and teaching literature*, 204–240. London: Longman.
- Adamson, Sylvia. 1999. Literary language. In Roger Lass (ed.), *The Cambridge history of the English language*, vol. III: 1476–1776, 539–653. Cambridge: Cambridge University Press. doi:10.1017/CHOL9780521264778.008
- Adamson, Sylvia. 2007. Prescribed reading: Pronouns and gender in the eighteenth century. *Historical Sociolinguistics and Sociohistorical Linguistics* 7. [http://www.hum2.leidenuniv.nl/hsl\\_shl/Adamson.htm](http://www.hum2.leidenuniv.nl/hsl_shl/Adamson.htm).
- Agha, Asif. 2007. *Language and social relations*. New York: Cambridge University Press.
- Anderson, Karen. 2000. *Productivity in English nominal and adjectival derivation*, 1100–2000. Perth: University of Western Australia PhD dissertation.
- Anderwald, Lieselotte. 2012. Clumsy, awkward or having a peculiar propriety? Prescriptive judgments and language change in the 19th century. *Language Sciences* 34(1). 28–53. doi:10.1016/j.langsci.2011.06.002
- Anderwald, Lieselotte. 2016. *Language between description and prescription: Verb categories in nineteenth-century grammars of English*. Oxford: Oxford University Press.
- Ando, Sadao. 1976. *A descriptive syntax of Christopher Marlowe's language*. Tokyo: University of Tokyo Press.
- Apcar, Amy (ed.). 1918. *Life and adventures of Emin Joseph Emin, 1726–1809*. Calcutta: The Baptist Mission Press.
- Ard, Josh. 1982. Auxiliary *do*: Support or emphasis? *Linguistics* 20. 445–466. doi:10.1515/ling.1982.20.5-6.445
- Argamon, Shlomo, Moshe Koppel, Jonathan Fine & Anat Rachel Shimoni. 2003. Gender, genre, and writing style in formal written texts. *Text* 23(3). 321–346. doi:10.1515/text.2003.014
- Arkell, Tom. 2006. Illuminations and distortions: Gregory King's scheme calculated for the year 1688 and the social structure of later Stuart England. *The Economic History Review* N.S. 59. 32–69. doi:10.1111/j.1468-0289.2005.00330.x
- Arnaud, René. 1973. *La forme progressive en Anglais du XIX<sup>e</sup> siècle*. Lille: Service de Reproduction de Thèses Université Lille III.
- Arnaud, René. 1998. The development of the progressive in 19th century English: A quantitative study. *Language Variation and Change* 10(2). 123–152. doi:10.1017/S0954394500001265

- Arnaud, René. 2002. *Letter-writers of the Romantic Age and the modernization of English: A quantitative historical survey of the progressive*. <http://web.univ-pau.fr/saes/pb/bibliographies/A/arnaud/romanticletterwriters.pdf>. (2007.)
- Auer, Anita. 2008. The letter which that I wrote: Self-corrections in Late Modern English letters. In Marina Dossena & Ingrid Tieken-Boon van Ostade (eds.), *Studies in Late Modern English correspondence: Methodology and data* (Linguistic Insights Studies in Language and Communication 76), 213–234. Bern: Peter Lang.
- Auer, Anita & Victorina González-Díaz. 2005. Eighteenth-century prescriptivism in English: A re-evaluation of its effects on actual language usage. *Multilingua* 24(4). 317–341. doi:10.1515/mult.2005.24.4.317
- Austin, Frances. 1973. Epistolary conventions in the Clift family correspondence. *English Studies* 54. 9–22, 129–140. doi:10.1080/00138387308597547
- Austin, Frances. 1991. *The Clift family correspondence 1792–1846*. Sheffield: CECTAL.
- Austin, Frances. 1994. The effect of exposure to Standard English: The language of William Clift. In Dieter Stein & Ingrid Tieken-Boon van Ostade (eds.), *Towards a Standard English, 1600–1800*, 285–313. Berlin & New York: Mouton de Gruyter.
- Austin, Frances. 2000. Letter writing in a Cornish community in the 1790s. In David Barton & Nigel Hall (eds.), *Letter writing as a social practice*, 43–61. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/swll.9.04aus
- Authorized Version = *The Holy Bible, containing the Old and New Testaments, translated out of the original tongues: And with the former translations diligently compared and revised by His Majesty's special command. Authorized King James version*. Oxford: Oxford University Press. <http://quod.lib.umich.edu/k/kjv/>. (15 June, 2013.)
- Baayen, R. H. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, 899–919. Berlin: Mouton de Gruyter. doi:10.1515/9783110213881.2.899
- Bækken, Bjørg. 1998. *Word order patterns in Early Modern English with special reference to the position of the subject and the finite verb* (Studia Anglistica Norvegica 9). Oslo: Novus Press.
- Bækken, Bjørg. 1999. Periphrastic *do* in Early Modern English. *Folia Linguistica Historica* XX. 107–128.
- Bækken, Bjørg. 2002. *Yet this follie doth many times assault the brauest minds*: Affirmative declarative *do* in 17th-century English. *Nordic Journal of English Studies* 1. 317–337.
- Bækken, Bjørg. 2003. *Word order in 17th century English. A study of the stabilisation of the XSV pattern* (Studia Anglistica Norvegica 12). Oslo: Novus Press.
- Baeskow, Heike. 2012. *-Ness* and *-ity*: Phonological exponents of *n* or meaningful nominalizers of different adjectival domains? *Journal of English Linguistics* 40(1). 6–40. doi:10.1177/0075424211405156
- Bailey, Richard W. 2010. Variation and change in eighteenth-century English. In Raymond Hickey (ed.), *Eighteenth-century English: Ideology and change*, 182–199. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511781643.011
- Baines, Paul. 2004. *The long 18th century*. London: Hodder Arnold.
- Baker, Robert. 1770. *Remarks on the English language, in the nature of Vaugelas's remarks on the French*. London.
- Barber, Charles. 1976. *Early Modern English*. London: André Deutch.
- Bauer, Laurie. 2001. *Morphological productivity* (Cambridge Studies in Linguistics 95). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486210

- Bax, Randy C. 2005. Traces of Johnson in the language of Fanny Burney. *International Journal of English Studies* 5(1). 159–181.
- Bax, Randy C. 2008. *Foolish, foolisher, foolishest*: Eighteenth-century English grammars and the comparison of adjectives. In Ingrid Tieken-Boon van Ostade (ed.), *Grammars, grammarians and grammar-writing in eighteenth-century England*, 279–288. Berlin & New York: Mouton de Gruyter.
- Baxter, Gareth & William Croft. 2016. Modeling language change across the lifespan: Individual trajectories in community change. *Language Variation and Change* 28(2). 129–173.  
doi:10.1017/S0954394516000077
- Bayly, Anselm. 1758. *An introduction to languages, literary and philosophical*. London.
- Bayly, Anselm. 1772. *A plain and complete grammar of the English language*. London.
- Beal, Joan C. 2004. *English in modern times 1700–1945*. London: Arnold.
- Beal, Joan C., Carmela Nocera & Massimo Sturiale (eds.). 2008. *Perspectives on prescriptivism*. Bern: Peter Lang.
- Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2). 145–204.  
doi:10.1017/S004740450001037X
- Benjamini, Yoav & Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1). 289–300.
- Bergs, Alexander. 2004. Address pronouns in Late Middle English. In Alicia Rodríguez-Alvarez & Francisco Alonso-Almeida (eds.), *Voices on the past: Studies in Old and Middle English language and literature*, 127–138. Coruna: Netbiblo.
- Bergs, Alexander. 2005. *Social networks and historical sociolinguistics: Studies in morphosyntactic variation in the Paston Letters (1421–1503)*. Berlin: Mouton de Gruyter.  
doi:10.1515/9783110923223
- Bergs, Alexander. 2012. The Uniformitarian Principle and the risk of anachronisms in language and social history. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*, 80–98. Malden, MA: Blackwell.  
doi:10.1002/9781118257227.ch5
- Bestgen, Yves. 2014. Inadequacy of the chi-squared test to examine vocabulary differences between corpora. *Literary and Linguistic Computing* 29(2). 164–170. doi:10.1093/lc/fqt020
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.  
doi:10.1017/CBO9780511621024
- Biber, Douglas. 1995. *Dimensions of register variation*. Cambridge: Cambridge University Press.  
doi:10.1017/CBO9780511519871
- Biber, Douglas. 2001. Dimensions of variation among 18th-century speech-based and written registers. In Hans-Jürgen Diller & Manfred Görlach (eds.), *Towards a history of English as a history of genres* (Anglistische Forschungen 298), 89–109. Heidelberg: C. Winter.
- Biber, Douglas. 2010. What can a corpus tell us about registers and genres? In Anne O’Keeffe & Michael McCarthy (eds.), *The Routledge handbook of corpus linguistics*, 241–254. New York: Routledge. doi:10.4324/9780203856949.ch18
- Biber, Douglas & Edward Finegan. 1997. Diachronic relations among speech-based and written registers in English. In Terttu Nevalainen & Leena Kahlas-Tarkka (eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen* (Mémoires de la Société Néophilologique de Helsinki 52), 253–275. Helsinki: Société Néophilologique.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.

- Bingham, William. 1867. *A grammar of the English language: For the use of schools and academies*. Philadelphia.
- Black, Jeremy. 2008. *Eighteenth-century Britain, 1688–1783*, 2nd edn. Houndmills, Basingstoke: Palgrave.
- Black, Joseph et al. (eds.). 2006. *The broadview anthology of British literature: The restoration and the eighteenth century*. Peterborough, ON: Broadview Press.
- Blake, Norman F. 1992. The literary language. In Norman F. Blake (ed.), *The Cambridge history of the English language, vol. II: 1066–1476*, 500–541. Cambridge: Cambridge University Press. doi:10.1017/CHOL9780521264754.007
- Bolinger, Dwight L. 1948. On defining the morpheme. *Word* 4. 18–23. doi:10.1080/00437956.1948.11659323
- Bolinger, Dwight. 1976. The *in*-group: *One* and its compounds. In P. A. Reich (ed.), *The second LACUS forum 1975*, 229–237. Columbia, SC: Hornbeam Press.
- Bonnell, Thomas F. 2009. The reprint trade. In Michael F. Suarez & Michael L. Turner (eds.), *The Cambridge history of the book in Britain, vol. 5, 1695–1830*, 699–709. Cambridge: Cambridge University Press. doi:10.1017/CHOL9780521810173.039
- Boulton, Jeremy. 2000. London 1540–1700. In Peter Clark (ed.), *The Cambridge history of urban Britain, vol. 2, 1540–1840*, 315–346. Cambridge: Cambridge University Press. doi:10.1017/CHOL9780521431415.016
- Braaten, Björn. 1967. Notes on continuous tenses in English. *Norsk Tidsskrift for Sprogvidenskap* 21. 167–180.
- Brant, Clare. 2006. *Eighteenth-century letters and British culture*. London & New York: Palgrave Macmillan.
- Brezina, Vaclav & Miriam Meyerhoff. 2014. Significant or random? A critical review of sociolinguistic generalisations based on large corpora. *International Journal of Corpus Linguistics* 19(1). 1–28. doi:10.1075/ijcl.19.1.01bre
- Brightland, John & Charles Gildon. 1711. *A grammar of the English tongue*. London.
- Brittain, Lewis. 1788. *Rudiments of English grammar*. Louvain.
- Brown, Roger & Albert Gilman. 1960. The pronouns of power and solidarity. In Thomas A. Sebeok (ed.), *Style in language*, 253–276. Cambridge, MA: M.I.T. Press.
- Brysbaert, Marc, Emmanuel Keuleers & Boris New. 2011. Assessing the usefulness of Google Books' word frequencies for psycholinguistic research on word processing. *Frontiers in Psychology* 2, 27. doi:10.3389/fpsyg.2011.00027
- Buchanan, James. 1770. *A plan of an English grammar-school education*. Edinburgh.
- Burke, Peter. 2000. A civil tongue: Language and politeness in Early Modern Europe. In Peter Burke, Brian Harrison & Paul Slack (eds.), *Civil histories: Essays presented to Sir Keith Thomas*, 31–48. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198207108.003.0002
- Burles, Edward. 1652. *Grammatica Burlesca, or a new English grammar*. (Facsimile edition, R. C. Alston (ed.) *English linguistics 1500–1800*. Menston: Scholar Press, 1971.)
- Burn, John. 1766. *A practical grammar of the English language*. Glasgow.
- Burnley, David. 2003. The T/V pronouns in later Middle English literature. In Irma Taavitsainen & Andreas H. Jucker (eds.), *Diachronic perspectives on address term systems (Pragmatics & Beyond New Series 107)*, 27–45. Amsterdam: John Benjamins. doi:10.1075/pbns.107.03bur
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511750526

- Byrne, Muriel St Clare. 1964. The foundations of Elizabethan language. In Allardyce Nicoll (ed.), *Shakespeare in his own age* (Shakespeare Survey 17), 223–239. Cambridge: Cambridge University Press.
- Campbell, George. 1776. *The philosophy of rhetoric* [...]. Volume 1. London: printed for W. Strahan; and T. Cadell, in the Strand; and W. Creech at Edinburgh.
- Cannadine, David. 2000. *Class in Britain*. London: Penguin Books.
- Card, William, Raven I. McDavid Jr., & Virginia McDavid. 1984. Dimensions of usage and dictionary labelling. *Journal of English Linguistics* 17. 57–74. doi:10.1177/007542428401700105
- Carretta, Vincent. 2004. Sancho, (Charles) Ignatius (1729?–1780). *Oxford dictionary of national biography*. Oxford: Oxford University Press. doi:10.1093/ref:odnb/24609
- Carter, Philip. 2001. *Men and the emergence of polite society, 1660–1800*. London: Pearson Education.
- CEEC = *Corpora of Early English Correspondence*. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg et al. at the Department of Modern Languages, University of Helsinki. <http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/>. (5 May, 2014.)
- CEECE = *Corpus of Early English Correspondence Extension*. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg, Samuli Kaislaniemi, Mikko Laitinen, Minna Nevala, Arja Nurmi, Minna Palander-Collin, Tanja Säily and Anni Sairio at the Department of Modern Languages, University of Helsinki.
- Chambers, J. K. 2009. *Sociolinguistic theory: Linguistic variation and its social significance*. Revised edition. Chichester & Malden, MA: Blackwell.
- Cheshire, Jenny. 2013. Grammaticalisation in social context: The emergence of a new English pronoun. *Journal of Sociolinguistics* 17(5). 608–633. doi:10.1111/josl.12053
- Claridge, Claudia & Merja Kytö. 2014. *I had lost sight of them then for a bit, but I went on pretty fast*: Two degree modifiers in the Old Bailey Corpus. In Irma Taavitsainen, Andreas H. Jucker & Jukka Tuominen (eds.), *Diachronic corpus pragmatics*, 29–52. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/pbns.243.05cla
- Coates, Jennifer. 2004. *Women, men, and language: A sociolinguistic account of gender differences in language*, 3rd edn. London: Pearson Education.
- COCA = Davies, Mark. (2008–) The Corpus of Contemporary American English (COCA): 400+ million words, 1990–present. Available online at <http://www.americancorpus.org/>.
- Cohen, Michèle. 2006. 'A little learning': The curriculum and the construction of gender difference in the long eighteenth century. *British Journal for Eighteenth-Century Studies* 29(3). 321–335. doi:10.1111/j.1754-0208.2006.tb00651.x
- Conde-Silvestre, Juan Camilo. 2012. The role of social networks and mobility in diachronic sociolinguistics. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*, 332–352. Chichester: Wiley-Blackwell. doi:10.1002/9781118257227.ch18
- CoRD = *Corpus Resource Database*. Compiled by the Research Unit for Variation, Contacts and Change in English (VARIENG), University of Helsinki. <http://www.helsinki.fi/varieng/CoRD/>.
- Coote, Charles. 1788. *Elements of the grammar of the English language*. London.
- Corfield, Penelope J. 1991. Class by name and number in eighteenth-century Britain. In Penelope J. Corfield (ed.), *Language, history and class*, 101–130. Oxford: Blackwell.
- Cowie, Claire. 1999. *Diachronic word-formation: A corpus-based study of derived nominalizations in the history of English*. Cambridge: University of Cambridge PhD dissertation.

- Cowie, Claire. 2003. "Uncommon terminations": Proscription and morphological productivity. *Italian Journal of Linguistics* 15(1). 17–30.
- Cressy, David. 1980. Literacy and society in England and beyond. In David Cressy (ed.), *Literacy and the social order. Reading and writing in Tudor and Stuart England*, 175–189. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511560484.009
- Culpeper, Jonathan & Merja Kytö. 2010. *Early Modern English dialogues: Spoken interaction as writing*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan & Minna Nevala. 2012. Sociocultural processes and the history of English. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English* (Oxford Handbooks in Linguistics), 365–391. Oxford: Oxford University Press.
- D'Arcy, Alexandra, Bill Haddican, Hazel Richards, Sali A. Tagliamonte & Ann Taylor. 2013. Asymmetrical trajectories: The past and present of *-body/-one*. *Language Variation and Change* 25(3). 287–310. doi:10.1017/S0954394513000148
- Dahl, Torsten. 1956. Linguistic studies in some Elizabethan writings. II The auxiliary *do*. *Acta Jutlandica* XXVIII. 3–104.
- Dalton-Puffer, Christiane. 1994. Productive or not productive? The Romance element in Middle English derivation. In Francisco Fernández, Miguel Fuster & Juan José Calvo (eds.), *English Historical Linguistics 1992: Papers from the 7th International Conference on English Historical Linguistics, Valencia, 22–26 September 1992* (Current Issues in Linguistic Theory 113), 247–260. Amsterdam: John Benjamins.
- Dalton-Puffer, Christiane. 1996. *The French influence on Middle English morphology: A corpus-based study of derivation* (Topics in English Linguistics 20). Berlin: Mouton de Gruyter. doi:10.1515/9783110822113
- Danchev, Andrej. 1992. The evidence for analytic and synthetic developments in English. In Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen (eds.), *History of Englishes: New methods and interpretations in historical linguistics*, 25–41. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110877007.25
- Daunton, M. S. 1995. *Progress and poverty: An economic and social history of Britain 1700–1850*. Oxford: Oxford University Press.
- Davies, Mark. 2011–. *Google Books Corpus*. Based on Google Books n-grams (Michel et al. 2011). <http://googlebooks.byu.edu>.
- Davies, Mark. 2012. Some methodological issues related to corpus-based investigations of recent syntactic changes in English. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 157–174. Oxford: Oxford University Press.
- Davies, Mark. 2014. Making Google Books n-grams useful for a wide range of research on language change. *International Journal of Corpus Linguistics* 19(3). 401–416. doi:10.1075/ijcl.19.3.04dav
- Daybell, James (ed.). 2001. *Early modern women's letter writing, 1450–1700*. Basingstoke: Palgrave.
- Daybell, James. 2006. *Women letter-writers in Tudor England*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199259915.001.0001
- De Smet, Hendrik. 2013. *Spreading patterns: Diffusional change in the English system of complementation*. Oxford & New York: Oxford University Press.
- Denison, David. 1985. The origins of periphrastic *do*: Ellegård and Visser reconsidered. In Roger Eaton, Olga Fischer, Willem Koopman & Frederike van der Leek (eds.), *Papers from the 4th International Conference on English Historical Linguistics*, 45–60. Amsterdam: John Benjamins.
- Denison, David. 1993. *English historical syntax: Verbal constructions* (Longman Linguistics Library). London: Longman.

- Denison, David. 1998. Syntax. In Suzanne Romaine (ed.), *The Cambridge history of the English language, IV: 1776–1997*, 92–329. Cambridge: Cambridge University Press.
- Denison, David. 2003. Log(ist)ic and simplistic S-curves. In Raymond Hickey (ed.), *Motives for language change*, 54–70. Cambridge: Cambridge University Press.  
doi:10.1017/CBO9780511486937.005
- Deumert, Ana & Wim Vandenbussche (eds.). 2003. *Germanic standardizations: Past to present* (Impact: Studies in Language and Society 18). Amsterdam & Philadelphia: John Benjamins.  
doi:10.1075/impact.18
- Dickens, Charles. 1859. *A tale of two cities*. London: Chapman & Hall.
- Diller, Hans-Jürgen, Hendrik De Smet & Jukka Tyrkkö. 2011. A European database of descriptors of English electronic texts. *The European English Messenger* 19. 21–35.
- Dilworth, Thomas. 1751. *A new guide to the English tongue*. 13th edn. London.
- Dossena, Marina (ed.). 2014. *Transatlantic perspectives on Late Modern English*. Amsterdam & Philadelphia: John Benjamins.
- Dossena, Marina, John G. Newman & Sylwester Łodej (eds.). 2014. *Token: A Journal of English Linguistics* (Volume 3): *Special issue on Late Modern English*.
- Dryden, John. 1672a. *The conquest of Granada by the Spaniards in two parts*. London: T. N. for Henry Herringman.
- Dryden, John. 1672b. *Defence of the Epilogue to 'The conquest of Granada'*. London.
- Duncan, D. 1731. *A new English grammar*. London.
- ECEG = Eighteenth-Century English Grammars database. 2010. Compiled by Nuria Yáñez-Bouza (Manchester) and María E. Rodríguez-Gil (Las Palmas de Gran Canaria). <http://www.manchester.ac.uk/ECEG>. (September 2012.)
- Eckert, Penelope. 2000. *Linguistic variation as social practice*. Malden, MA & Oxford: Blackwell.
- Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41. 87–100.  
doi:10.1146/annurev-anthro-092611-145828
- Edwards, Paul & Polly Rewt (eds.). 1994. *The letters of Ignatius Sancho*. Edinburgh: Edinburgh University Press.
- Ellegård, Alvar. 1953. *The auxiliary do, the establishment and regulation of its use in English* (Gothenburg Studies in English 2). Stockholm: Almqvist & Wiksell.
- Elphinson, James. 1765. *The principles of the English language digested*. London.
- Elsaß, Stephan. 2012. Between linguistic creativity and formulaic restriction: Cross-linguistic perspectives on nineteenth-century lower class writers' private letters. In Marina Dossena & Gabriella Del Lungo Camiciotti (eds.), *Letter writing in Late Modern Europe*, 45–64. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/pbns.218.04els
- Elsaß, Stephan, Nils Langer, Joachim Scharloth & Wim Vandenbussche (eds.). 2007. *Germanic language histories 'from below' (1700–2000)* (Studia Linguistica Germanica 86). Berlin & New York: De Gruyter. doi:10.1515/9783110925463
- Emsley, Clive, Tim Hitchcock & Robert Shoemaker. 2013. London history: A population history of London. *Old Bailey proceedings online*. <http://www.oldbaileyonline.org>, version 7.0. (13 May, 2013.)
- Fairman, Tony. 2012. Letters in mechanically-schooled language. Theories and ideologies. In Marina Dossena & Gabriella Del Lungo Camiciotti (eds.), *Letter writing in Late Modern Europe*, 205–227. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/pbns.218.12fai

- Fairman, Tony. 2015. Language in print and handwriting. In Anita Auer, Daniel Schreier & Richard J. Watts (eds.), *Letter writing and language change*, 53–71. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139088275.005
- Farnworth, Richard. 1655. *The pure language of the spirit of truth set forth for the confounding false languages; acted out of pride, ambition and deceit, or, thee and thou in its place is the proper language to any single person whatsoever: Proved by severall examples, etc.* London.
- Farro, Daniel. 1754. *The royal universal British grammar and vocabulary.* London.
- Feather, John. 1985. *The provincial book trade in eighteenth-century England.* Cambridge: Cambridge University Press. doi:10.1017/CBO9780511552892
- Fell, John. 1784. *An essay towards an English grammar.* London.
- Fenning, Daniel. 1771. *A new grammar of the English language; or an easy introduction to the art of speaking and writing English with propriety and correctness: [...].* London: printed for S. Crowder.
- Ferguson, Charles. 1996. Variation and drift: Loss of agreement in Germanic. In T. Huebner (ed.), *Sociolinguistic perspectives: Papers on language in society, 1959–1994*, 241–260. Oxford & New York: Oxford University Press.
- Filppula, Markku. 2003. More on the English progressive and the Celtic connection. In Hildegard Tristram (ed.), *The Celtic Englishes III*, 150–168. Heidelberg: Winter.
- Filppula, Markku & Juhani Klemola. 2012. English in contact: Celtic and Celtic Englishes. In Alexander Bergs & Laurel Brinton (eds.), *English historical linguistics, vol. 2, 1687–1703.* Berlin & Boston: Mouton de Gruyter.
- Finegan, Edward. 1998. English grammar and usage. In Suzanne Romaine (ed.), *The Cambridge history of the English language, 1776–1997*, vol. IV, 536–588. Cambridge: Cambridge University Press.
- Finegan, Edward & Douglas Biber. 2001. Register variation and social dialect variation: The Register Axiom. In Penelope Eckert & John R. Rickford (eds.), *Style and sociolinguistic variation*, 235–267. Cambridge: Cambridge University Press.
- Finkenstaedt, Thomas. 1963. *You und thou: Studien zur Anrede im Englischen.* Berlin: Mouton de Gruyter. doi:10.1515/9783110822588
- Fisher, Michael. 2004. *Counterflows to colonialism: Indian travellers and settlers in Britain, 1600–1857.* Ohio: Oberlin College.
- Fitzmaurice, Susan. 1998. The commerce of language in the pursuit of politeness in eighteenth-century England. *English Studies* 79. 309–328. doi:10.1080/00138389808599136
- Fitzmaurice, Susan. 2000. The *Spectator*, the politics of social networks, and language standardisation in eighteenth-century England. In Laura Wright (ed.), *The development of Standard English, 1300–1800: Theories, descriptions, conflicts*, 195–218. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511551758.012
- Fitzmaurice, Susan. 2004. The meanings and uses of the progressive construction in an early eighteenth-century English network. In Anne Curzan & Kimberly Emmons (eds.), *Studies in the History of the English Language II*, 131–173. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110897661.131
- Fitzmaurice, Susan. 2012. Social factors and language change in eighteenth-century England: The case of negative concord. *Neuphilologische Mitteilungen* 113. 293–321.
- Fogg, Peter W. 1792–96. *Elementa Anglicana; or, the principles of English grammar displayed and exemplified.* 2 vols. London.
- Fogg, Peter Walkden. 1796. *Elementa Anglicana.* Stockport.
- Forgeng, Jeffrey L. 2007. *Daily life in Stuart England.* Westwood, CT: Greenwood Press.

- Fox, George, John Stubs & Benjamin Furley. 1660. *A battle-door for teachers & professors to learn singular & plural; you to many, and thou to one*. London.
- Frank, Thomas. 1985. The rise of do-support in Early Modern English: A reappraisal. *Revista Canaria de Estudios Ingleses* 10. 1–29.
- Friginal, Eric & Jack A. Hardy. 2014. *Corpus-based sociolinguistics: A guide for students*. New York: Routledge.
- Friginal, Eric, Marsha Walker & Janet Beth Randall. 2014. Exploring mega-corpora: Google Ngram Viewer and the Corpus of Historical American English. *EuroAmerican Journal of Applied Linguistics and Languages* 1(1). 48–68. doi:10.21283/2376905X.1.4
- Gachelin, Jean-Marc. 1997. The progressive and habitual aspects in non-standard Englishes. In Edgar Werner Schneider (ed.), *Englishes Around the World*, vol. 1, 33–46. Amsterdam: John Benjamins. doi:10.1075/veaw.g18.07gac
- Gardner, Anne. 2013. *Derivation in Middle English: Regional and text type variation*. Zurich: University of Zurich PhD dissertation.
- Gerner, Jürgen. 1996. *Untersuchungen zur Funktion des emphatischen do im Englischen*. Frankfurt am Main: Peter Lang.
- Gil, Alexander. 1619. *Logonomia Anglica*. London: J. Beale.
- González-Díaz, Victorina. 2008. On normative grammarians and the double marking of degree. In Ingrid Tieken-Boon van Ostade (ed.), *Grammars, grammarians and grammar-writing in eighteenth-century England*, 289–310. Berlin & New York: Mouton de Gruyter.
- Google. 2013a. Google Books Ngram Viewer. <https://books.google.com/ngrams/>.
- Google. 2013b. About Ngram Viewer. <https://books.google.com/ngrams/info>.
- Görlach, Manfred. 1988. The study of Early Modern English variation – the Cinderella of English historical linguistics? In Jacek Fisiak (ed.), *Historical dialectology, regional and social*, 211–228. Berlin: Mouton de Gruyter.
- Görlach, Manfred. 1991. *Introduction to Early Modern English*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166010
- Görlach, Manfred. 2001. *Eighteenth-century English* (Sprachwissenschaftliche Studienbücher). Heidelberg: Winter.
- Greaves, Paul. 1594. *Grammatica Anglicana*. London. (Facsimile edition, R. C. Alston (ed.) *English linguistics 1500–1800*. Menston: Scolar Press, 1969.)
- Greenwood, James. 1722 [1711]. *An essay towards a practical English grammar*, [...], 2nd edn. London: printed for John Clark.
- Gries, Stefan Th. & Martin Hilpert. 2010. Modeling diachronic change in the third person singular: A multifactorial, verb- and author-specific exploratory approach. *English Language and Linguistics* 14(3). 293–320. doi:10.1017/S1360674310000092
- Grund, Peter. 2011. The science of pronominal usage: *He* and *it* in conference to inanimate entities in Late Middle English texts on alchemy. *Journal of English Linguistics* 39(4). 335–358.
- Hammond, Samuel. 1750. *A new introduction to learning; or, a sure guide to the English pronunciation and orthography*. Nottingham: Creswell.
- Hay, Douglas & Nicholas Rogers. 1997. *Eighteenth-century English society: Shuttles and swords*. Oxford: Oxford University Press.
- Helsinki Corpus = *The Helsinki Corpus of English Texts* (1991). Department of Modern Languages, University of Helsinki. Compiled by Matti Rissanen (Project leader), Merja Kytö (Project secretary); Leena Kahlas-Tarkka, Matti Kilpiö (Old English); Saara Nevanlinna, Irma Taavitsainen (Middle English); Terttu Nevalainen, Helena Raumolin-Brunberg (Early Modern English).

- Hickey, Raymond. 2010a. Timeline for the eighteenth century. In Raymond Hickey (ed.), *Eighteenth-century English: Ideology and change*, 139–159. Cambridge: Cambridge University Press.
- Hickey, Raymond (ed.). 2010b. *Eighteenth-century English: Ideology and change*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511781643
- Hickey, Raymond. 2012. Early English and the Celtic hypothesis. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*. Oxford: Oxford University Press.
- Hinneburg, Alexander, Heikki Mannila, Samuli Kaislaniemi, Terttu Nevalainen & Helena Raumolin-Brunberg. 2007. How to handle small samples: Bootstrap and Bayesian methods in the analysis of linguistic change. *Literary and Linguistic Computing* 22. 137–150. doi:10.1093/lc/fqm006
- Hodson, Jane. 2008. Joseph Priestley's two *Rudiments of English grammar*: 1761 and 1768. In Ingrid Tiekens-Boon van Ostade (ed.), *Grammars, grammarians and grammar-writing in eighteenth-century England*, 177–190. Berlin & New York: Mouton de Gruyter.
- Holmes, George & Daniel Szechi. 1993. *The age of oligarchy. Pre-industrial Britain 1722–1783*. London & New York: Longman.
- Holmes, Janet & Miriam Meyerhoff. 2005. Different voices, different views: An introduction to current research in language and gender. In Janet Holmes & Miriam Meyerhoff (eds.), *The handbook of language and gender*, 1–17. Malden, MA & Oxford: Blackwell.
- Holmqvist, Erik. 1922. *On the history of the English present inflections, particularly -th and -s*. Heidelberg: Carl Winters Universitätsbuchhandlung.
- Honegger, Thomas. 2003. 'And if ye wol nat so, my lady sweete, thanne preye I thee, [...]': Forms of address in Chaucer's *Knight's Tale*. In Irma Taavitsainen & Andreas H. Jucker (eds.), *Diachronic perspectives on address term systems (Pragmatics & Beyond New Series 107)*, 61–84. Amsterdam: John Benjamins. doi:10.1075/pbns.107.05hon
- Hope, Jonathan. 1993. Second person singular pronouns in records of early modern 'spoken' English. *Neuphilologische Mitteilungen* 94(1). 83–100.
- Horne Tooke, John. 1798. *The diversions of Purley*, vol. 1. 2nd edn. London.
- Hornsey, John. 1793. *A short English grammar in two parts: Simplified to the capacities of children*. York.
- Huber, Magnus. 2007. The Old Bailey Proceedings, 1674–1834: Evaluating and annotating a corpus of 18th- and 19th-century spoken English. In Anneli Meurman-Solin & Arja Nurmi (eds.), *Annotating variation and change (Studies in Variation, Contacts and Change in English 1)*. Helsinki: VARIENG. <http://www.helsinki.fi/varieng/series/volumes/01/huber/>.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hudson, Richard A. 1996. *Sociolinguistics*, 2nd edn. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166843
- Hundt, Marianne. 2004. Animacy, agentivity and the spread of the progressive in Modern English. *English Language and Linguistics* 8(1). 47–69. doi:10.1017/S1360674304001248
- Hundt, Marianne & Geoffrey Leech. 2012. "Small is beautiful": On the value of standard reference corpora for observing recent grammatical change. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 175–188. Oxford: Oxford University Press.

- Hundt, Marianne & Christian Mair. 1999. 'Agile' and 'uptight' genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2). 221–242. doi:10.1075/ijcl.4.2.02hun
- Ihalainen, Ossi. 1976. Periphrastic *do* in affirmative sentences in the dialect of East Somerset. *Neuphilologische Mitteilungen* LXXVII. 608–622.
- Ihalainen, Ossi. 1982. On the notion "possible grammatical change": A look at a perfectly good change that did not quite make it. *Studia Anglica Posnaniensia* 15. 3–11.
- Ihalainen, Ossi. 1994. The dialects of England since 1776. In Robert Burchfield (ed.), *The Cambridge history of the English language*, vol. V: *English in Britain and overseas: Origins and developments*, 197–274. Cambridge: Cambridge University Press.
- Immel, Andrea. 2009. Children's books and school books. In Michael F. Suarez & Michael L. Turner (eds.), *The Cambridge history of the book in Britain*, vol. 5, 1695–1830, 736–749. Cambridge: Cambridge University Press. doi:10.1017/CHOL9780521810173.042
- Joby, Chris. 2014. Third-person singular zero in the Norfolk dialect: A re-assessment. *Folia Linguistica Historica* 35. 135–172.
- Johnson, Samuel. 1755. *A Dictionary of the English Language*, vol. 2. London.
- Johnstone, Barbara. 2016. Enregisterment: How linguistic items become linked with ways of speaking. *Language and Linguistics Compass* 10. 632–643. doi:10.1111/lnc3.12210
- Jones, Hugh. 1724. *An accidence to the English tongue*. London.
- Jucker, Andreas H. 2000a. *History of English and English historical linguistics*. Stuttgart: Ernst Klett.
- Jucker, Andreas H. 2000b. *Thou* in the history of English: A case for historical semantics or pragmatics? In Christiane Dalton-Puffer & Nikolaus Ritt (eds.), *Words: Structure, meaning, function: A festschrift for Dieter Kastovsky*, 153–163. Berlin: Mouton de Gruyter. doi:10.1515/9783110809169.153
- Jucker, Andreas H., Irma Taavitsainen & Gerold Schneider. 2012. Semantic corpus trawling: Expressions of "courtesy" and "politeness" in the Helsinki Corpus. In Carla Suhr & Irma Taavitsainen (eds.), *Developing corpus methodology for historical pragmatics* (Studies in Variation, Contacts and Change in English 11). Helsinki: VARIENG. [http://www.helsinki.fi/varieng/series/volumes/11/jucker\\_taavitsainen\\_schneider/](http://www.helsinki.fi/varieng/series/volumes/11/jucker_taavitsainen_schneider/).
- Kaislaniemi, Samuli. 2006. *The Corpus of Early English Correspondence* Extension and Supplement. Poster presented at ICAME 27 (the 27th conference of the International Computer Archive for Medieval and Modern English), 24–28 May, Espoo, Finland. Available on CoRD: [http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/Kaislaniemi\\_posteri\\_96dpi.pdf](http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/Kaislaniemi_posteri_96dpi.pdf).
- Kampstra, Peter. 2008. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software* 28. Code Snippet 1, <http://www.jstatsoft.org/v28/c01/>. doi:10.18637/jss.v028.c01
- Kekäläinen, Kirsti. 1971. *Aspects of style and language in the English and Scottish popular ballads*. Helsinki: University of Helsinki licentiate thesis.
- Kerswill, Paul. 1996. Children, adolescents, and language change. *Language Variation and Change* 8. 177–202. doi:10.1017/S0954394500001137
- Kilgariff, Adam. 2001. Comparing corpora. *International Journal of Corpus Linguistics* 6(1). 97–133. doi:10.1075/ijcl.6.1.05kil
- Killie, Kristin. 2004. Subjectivity and the English progressive. *English Language and Linguistics* 8(1). 25–46. doi:10.1017/S1360674304001236
- Kirkby, John. 1746. *A new English grammar, or, guide to the English tongue, with notes*. London: for R. Manby & H. S. Coy (Facs. ed. Menston: Scolar, 1971, EL 297).

- Klein, Lawrence E. 1994. "Politeness" as linguistic ideology in late seventeenth- and eighteenth-century England. In Dieter Stein & Ingrid Tiekens-Boon van Ostade (eds.), *Towards a Standard English, 1600–1800*, 31–50. Berlin & New York: Mouton de Gruyter.
- Klein, Lawrence E. 2002. Politeness and the interpretation of the British eighteenth century. *The Historical Journal* 45. 869–898. doi:10.1017/S0018246X02002716
- Kleiner, Brian & Dennis R. Preston. 1997. Discourse disputes: How come you do do like you do. *Folia Linguistica* 31. 105–131. doi:10.1515/flin.1997.31.1-2.105
- Knox, Vicesimus. 1789. *Liberal education*, 10th edn., 2 vols. London.
- Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36. 15–46.
- Koplenig, Alexander. 2017. The impact of lacking metadata for the measurement of cultural and linguistic change using the Google Ngram data sets: Reconstructing the composition of the German corpus in times of WWII. *Digital Scholarship in the Humanities* 32(1). 169–188.
- Kranich, Svenja. 2008. Subjective progressives in seventeenth- and eighteenth-century English. Secondary grammaticalization as a process of objectification. In Marina Dossena, Richard Dury & Maurizio Gotti (eds.), *English Historical Linguistics 2006. Selected Papers from the Fourteenth International Conference on English Historical Linguistics (ICEHL 14)*, vol. 1, 241–256. Amsterdam: John Benjamins.
- Kranich, Svenja. 2010. *The progressive in Modern English: A corpus-based study of grammaticalization and related changes* (Language and Computers: Studies in Practical Linguistics 72). Amsterdam: Rodopi. doi:10.1163/9789042031449
- Kurki, Tommi. 2005. *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen: Kielenmuutosten seuraamisesta ja niiden tarkastelussa käytettävistä menetelmistä* [Real-time language change of group and individual: On following linguistic changes in apparent time and real time, and on the methods used in the study of language change]. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kytö, Merja. 1993. Third-person present singular verb inflection in early British and American English. *Language Variation and Change* 5. 113–139. doi:10.1017/S0954394500001447
- Kytö, Merja (comp.). 1996. *Manual to the diachronic part of the Helsinki Corpus of English Texts. Coding conventions and lists of source texts*, 3rd edn. Helsinki: Department of English, University of Helsinki.
- Kytö, Merja & Matti Rissanen. 1997. Introduction: Language analysis and diachronic corpora. In Raymond Hickey, Merja Kytö, Ian Lancashire & Matti Rissanen (eds.), *Tracing the trail of time: Proceedings from the second Diachronic Corpora workshop*, 9–22. Amsterdam: Rodopi.
- Labov, William. 1978 [1972]. *Sociolinguistic patterns*. Oxford: Basil Blackwell.
- Labov, William. 1994. *Principles of linguistic change*, vol. 1: *Internal factors* (Language in Society 20). Oxford, UK & Cambridge, USA: Blackwell.
- Labov, William. 2001. *Principles of linguistic change*, vol. 2: *Social factors* (Language in Society 29). Oxford: Blackwell.
- Labov, William. 2010. *Principles of linguistic change*, vol. 3: *Cognitive and cultural factors* (Language in Society 39). Malden, MA & Oxford: Wiley-Blackwell. doi:10.1002/9781444327496
- Laitinen, Lea & Taru Nordlund. 2012. Performing identities and interaction through epistolary formulae. In Marina Dossena & Gabriella Del Lungo Camiciotti (eds.), *Letter writing in Late Modern Europe*, 65–88. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/pbns.218.05lai

- Laitinen, Mikko. 2002. Extending the *Corpus of Early English Correspondence* to the 18th century. *Helsinki English Studies* 2. <http://blogs.helsinki.fi/hes-eng/volumes/volume-2-corpora-in-todays-english-studies>.
- Laitinen, Mikko. 2007a. Explaining present-day variation: Agreement patterns of *he* and *they* with singular antecedents. *Neuphilologische Mitteilungen* 108(4). 537–562.
- Laitinen, Mikko. 2007b. *Agreement patterns in English. Diachronic corpus studies on common-number pronouns*. Helsinki: Société Néophilologique.
- Laitinen, Mikko. 2009. Singular YOU WAS/YOU WERE variation and English normative grammars in the eighteenth century. In Arja Nurmi, Minna Nevala & Minna Palander-Collin (eds.), *The language of daily life in England (1400–1800)*, 199–217. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/pbns.183.12lai
- Langford, Paul. 1998 [1989]. *A polite and commercial people: England, 1727–1783*. Oxford: Clarendon Press.
- Langford, Paul. 2002. The uses of eighteenth-century politeness. *Transactions of the Royal Historical Society* 12. 311–331. doi:10.1017/S0080440102000129
- Lass, Roger (ed.). 1999a. *The Cambridge history of the English language, vol. III: 1476–1776*. Cambridge: Cambridge University Press.
- Lass, Roger. 1999b. Phonology and morphology. In Roger Lass (ed.), *The Cambridge history of the English language, vol. III: 1476–1776*, 56–186. Cambridge: Cambridge University Press.
- Leech, Geoffrey. 1993. 100 million words of English. *English Today* 9(1). 9–15. doi:10.1017/S0266078400006854
- Leech, Geoffrey & Nicholas Smith. 2005. Extending the possibilities of corpus-based research on English in the twentieth century: A prequel to LOB and FLOB. *ICAME Journal* 29. 83–98.
- Lijffijt, Jefrey. 2013. *Computational methods for comparison and exploration of event sequences*. Helsinki: Aalto University PhD dissertation. <http://urn.fi/URN:ISBN:978-952-60-5475-9>.
- Lijffijt, Jefrey, Terttu Nevalainen, Tanja Säily, Panagiotis Papapetrou, Kai Puolamäki & Heikki Mannila. 2016. Significance testing of word frequencies in corpora. *Digital Scholarship in the Humanities* 31(2). 374–397.
- Lijffijt, Jefrey, Tanja Säily & Terttu Nevalainen. 2012. CEECing the baseline: Lexical stability and significant change in a historical corpus. In Jukka Tyrkkö, Matti Kilpiö, Terttu Nevalainen & Matti Rissanen (eds.), *Outposts of historical corpus linguistics: From the Helsinki Corpus to a proliferation of resources* (Studies in Variation, Contacts and Change in English 10). Helsinki: VARIENG. [http://www.helsinki.fi/varieng/series/volumes/10/lijffijt\\_saily\\_nevalainen/](http://www.helsinki.fi/varieng/series/volumes/10/lijffijt_saily_nevalainen/).
- Lin, Yuri, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman & Slav Petrov. 2012. Syntactic annotations for the Google Books Ngram Corpus. In *Proceedings of the ACL 2012 system demonstrations*, 169–174. Stroudsburg, PA: Association for Computational Linguistics.
- Lindert, Peter H. & Jeffrey G. Williamson. 1982. Revising England's social tables 1688–1812. *Explorations in Economic History* 19. 385–408. doi:10.1016/0014-4983(82)90009-2
- Lindsay, Mark & Mark Aronoff. 2013. Natural selection in self-organizing morphological systems. In Nabil Hathout, Fabio Montermini & Jesse Tseng (eds.), *Morphology in Toulouse: Selected proceedings of Décembrettes 7* (LINCOS Studies in Theoretical Linguistics 51). Munich: Lincom.
- Lodge, R. Anthony. 2004. *A sociolinguistic history of Parisian French*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486685

- Los, Bettelou. 2002. The loss of the indefinite pronoun *man*. Syntactic change and information structure. In Teresa Fanego, María José López-Couso & Javier Perez-Guerra (eds.), *English historical syntax and morphology. Selected Papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000*, 181–202. Amsterdam & Philadelphia: John Benjamins.
- Loughton, William. 1734. *A practical grammar of the English tongue*. London.
- Loughton, William. 1744. *A practical grammar of the English tongue*, 4th edn. London.
- Lowe, Solomon. 1737. *English grammar reformd into a small compass and easy method for the readier learning and better understanding of the English tongue*. London.
- Lowth, Robert. 1762. *A short introduction to English grammar: With critical notes*. London.
- Lowth, Robert. 1763. *A short introduction to English grammar: With critical notes*, 2nd edn. Dublin: Saunders.
- Lutz, Angelika. 1998. The interplay of external and internal factors in morphological restructuring: The case of *you*. In Jacek Fisiak & Marcin Krygier (eds.), *Advances in English historical linguistics (1996)*, 189–210. Berlin: Mouton de Gruyter. doi:10.1515/9783110804072.189
- Mair, Christian. 2006. *Twentieth-century English: History, variation and standardization*. Cambridge: Cambridge University Press.
- Mann, H. B. & D. R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* 18(1). 50–60. doi:10.1214/aoms/1177730491
- Mannila, Heikki, Terttu Nevalainen & Helena Raumolin-Brunberg. 2013. Quantifying variation and estimating the effects of sample size on the frequencies of linguistic variables. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 337–360. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511792519.023
- Marchand, Hans. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach*, 2nd edn. Munich: C.H. Beck'sche Verlagsbuchhandlung.
- Martin, Benjamin. 1754. *An introduction to the English language and learning in three parts*. London.
- Martinet, André. 1964. *Elements of general linguistics*. London: Faber.
- Martínez-Insua, Ana E. & Javier Pérez-Guerra. 2006. 'There's Bjørg': On *there*-sentences in the recent history of English. In Leiv Egil Breivik, Sandra Halverson & Kari E. Haugland (eds.), *'These things I write vnto thee ...' Essays in honour of Bjørg Bækken*, 189–211. Oslo: Novus Press.
- McIntosh, Angus, M. L. Samuels & Michael Benskin. 1986. *A linguistic atlas of Late Mediaeval English*, vol. I. Aberdeen: Aberdeen University Press.
- McIntosh, Carey. 1986. *Common and courtly language: The stylistics of social class in 18th-century British literature*. Philadelphia: University of Pennsylvania Press.
- McIntosh, Carey. 1998. *The evolution of English prose 1700–1800: Style, politeness, and print culture*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511582790
- McIntosh, Carey. 2008. British English in the long eighteenth century (1660–1830). In Haruko Momma & Michael Matto (eds.), *Blackwell companion to the history of the English language*, 228–234. Oxford: Wiley-Blackwell.
- Metcalfe, Lister. 1771. *The rudiments of the English tongue; or, the principles of English grammar*, [...] 2nd edn. Newcastle: printed by T. Saint, for J. Wilkie, London.
- Michael, Ian. 1970. *English grammatical categories and the tradition to 1800*. Cambridge: Cambridge University Press.
- Michael, Ian. 1987. *The teaching of English from the sixteenth century to 1870*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139165853
- Michael, Ian. 1997. The hyperactive production of English grammars in the nineteenth century: A speculative bibliography. *Publishing History* 41. 23–61.

- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014). 176–182. doi:10.1126/science.1199644
- Milroy, James. 1992. *Linguistic variation and change: On the historical sociolinguistics of English* (Language in Society 19). Oxford: Blackwell.
- Milroy, James & Lesley Milroy. 2012 [1985]. *Authority in language: Investigating language prescription and standardisation*, 4th edn. London & New York: Routledge & Kegan Paul.
- Minugh, David. 1999. What aileth thee, to print so curiously? Archaic forms and contemporary newspaper language. In Irma Taavitsainen, Gunnel Melchers & Päivi Pahta (eds.), *Writing in non-standard English*, 47–54. Amsterdam: John Benjamins.
- Moore, Colette. 2002. Writing good Southerne: Local and supralocal norms in the Plumpton letter collection. *Language Variation and Change* 14. 1–17. doi:10.1017/S0954394502141019
- Mossé, Fernand 1938. *Histoire de la forme périphrastique être + participe présent en germanique*. Paris: Klincksieck.
- Mugglestone, Lynda. 2003 [1995]. *‘Talking proper’: The rise of accent as social symbol*, 2nd edn. Oxford: Clarendon Press.
- Munck, Thomas. 2000. *The enlightenment: A comparative social history 1721–1794*. London: Arnold.
- Murray, Lindley. 1795. *English grammar: Adapted to the different classes of learners*. London.
- Mustanoja, Tauno F. 1960. *A Middle English syntax. 1: Parts of speech*. Helsinki: Société Néophilologique.
- Nevala, Minna. 2002. *Your moder sent a letter to the: Pronouns of address in private correspondence from Late Middle to Late Modern English*. In Helena Raumolin-Brunberg, Minna Nevala, Arja Nurmi & Matti Rissanen (eds.), *Variation past and present: VARIENG studies on English for Terttu Nevalainen* (Mémoires de la Société Néophilologique de Helsinki 61), 135–159. Helsinki: Société Néophilologique.
- Nevala, Minna. 2004a. Accessing politeness axes: On forms of address and terms of reference in early English correspondence. *Journal of Pragmatics* 36(12). 2125–2160. doi:10.1016/j.pragma.2004.02.001
- Nevala, Minna. 2004b. *Address in early English correspondence: Its forms and socio-pragmatic functions* (Mémoires de la Société Néophilologique de Helsinki 64). Helsinki: Société Néophilologique.
- Nevala, Minna. 2009. *Friends will be friends? The sociopragmatics of referential terms in early English letters*. In Andreas H. Jucker, Daniel Schreier & Marianne Hundt (eds.), *Corpora: Pragmatics and discourse. Papers from the 29th international conference on English language research on computerized corpora*, 81–101. Amsterdam: Rodopi.
- Nevala, Minna & Arja Nurmi. 2013. *The Corpora of Early English Correspondence* (CEEC400). In Anneli Meurman-Solin & Jukka Tyrkkö (eds.), *Principles and practices for the digital editing and annotation of diachronic data* (Studies in Variation, Contacts and Change in English 14). Helsinki: VARIENG. [http://www.helsinki.fi/varieng/series/volumes/14/nevala\\_nurmi/](http://www.helsinki.fi/varieng/series/volumes/14/nevala_nurmi/)
- Nevalainen, Terttu. 1996. Gender difference. In Terttu Nevalainen & Helena Raumolin-Brunberg (eds.), *Sociolinguistics and language history: Studies based on the Corpus of Early English Correspondence*, 77–91. Amsterdam & Atlanta: Rodopi.
- Nevalainen, Terttu. 2000. Processes of supralocalisation and the rise of Standard English in the Early Modern period. In Ricardo Bermúdez-Otero, David Denison, Richard M. Hogg & C. B. McCully (eds.), *Generative theory and corpus studies: A dialogue from 10 ICEHL*, 329–372. Berlin: Mouton de Gruyter. doi:10.1515/9783110814699.329

- Nevalainen, Terttu. 2006. Negative concord as an English “vernacular universal”: Social history and linguistic typology. *Journal of English Linguistics* 34(3), 257–278. doi:10.1177/0075424206293144
- Nevalainen, Terttu. 2008. Variation in written English: Grammar change or a shift in style? In Susan Kermas & Maurizio Gotti (eds.), *Socially-conditioned language change: Diachronic and synchronic insights*, 31–51. Lecce: Edizioni del Grifo.
- Nevalainen, Terttu. 2009. Number agreement in existential constructions: A sociolinguistic study of 18th-century English. In Markku Filppula, Juhani Klemola & Heli Paulasto (eds.), *Vernacular universals and language contact*, 80–102. New York & London: Routledge.
- Nevalainen, Terttu. 2013. English historical corpora in transition: From new tools to legacy corpora? In Paul Bennett, Martin Durrell, Silke Scheible & Richard J. Whitt (eds.), *New methods in historical corpora* (Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 3), 37–53. Tübingen: Narr.
- Nevalainen, Terttu. 2014a. Sociohistorical analysis. In Janet Holmes & Kirk Hazen (eds.), *Research methods in sociolinguistics: A practical guide*, 93–106. Malden, MA: Wiley-Blackwell.
- Nevalainen, Terttu. 2014b. What are historical sociolinguistics? *Journal of Historical Sociolinguistics* 1(2), 243–269. doi:10.1515/jhsl-2015-0014
- Nevalainen, Terttu. 2015. Descriptive adequacy of the S-curve model in diachronic studies of language change. In Christina Sanchez-Stockhammer (ed.), *Can we predict linguistic change?* (Studies in Variation, Contacts and Change in English 16). Helsinki: VARIENG. <http://www.helsinki.fi/varieng/series/volumes/16/nevalainen/>.
- Nevalainen, Terttu & Elizabeth Closs Traugott (eds.). 2012. *The Oxford handbook of the history of English*. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199922765.001.0001
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 1994a. Sociolinguistics and language history: The Helsinki Corpus of Early English Correspondence. *Hermes: Journal of Linguistics* 13, 135–143.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 1994b. *Its strength and the beauty of it*: The standardization of the third person neuter possessive in Early Modern English. In Dieter Stein & Ingrid Tieken-Boon van Ostade (eds.), *Towards a Standard English, 1600–1800*, 171–183. Berlin & New York: Mouton de Gruyter.
- Nevalainen, Terttu & Helena Raumolin-Brunberg (eds.). 1996a. *Sociolinguistics and language history: Studies based on the Corpus of Early English Correspondence*. Amsterdam & Atlanta: Rodopi.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 1996b. The Corpus of Early English Correspondence. In Terttu Nevalainen & Helena Raumolin-Brunberg (eds.), *Sociolinguistics and language history: Studies based on the Corpus of Early English Correspondence*, 39–54. Amsterdam & Atlanta: Rodopi.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2000a. The changing role of London on the linguistic map of Tudor and Stuart England. In Dieter Kastovsky & Arthur Mettinger (eds.), *The history of English in a social context: A contribution to historical sociolinguistics*, 279–337. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110810301.279
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2000b. The third-person singular -(e)s and -(e)th revisited: The morphophonemic hypothesis. In Christiane Dalton-Puffer & Nikolaus Ritt (eds.), *Words: Structure, meaning, function; A festschrift for Dieter Kastovsky* (Trends in Linguistics 130), 235–248. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110809169.235
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical sociolinguistics: Language change in Tudor and Stuart England* (Longman Linguistics Library). London: Pearson education.

- Nevalainen, Terttu, Helena Raumolin-Brunberg & Heikki Mannila. 2011. The diffusion of language change in real time: Progressive and conservative individuals and the time-depth of change. *Language Variation and Change* 23(1). 1–43. doi:10.1017/S0954394510000207
- Nevalainen, Terttu, Helena Raumolin-Brunberg & Peter Trudgill. 2001. Chapters in the social history of East Anglian English: The case of the third-person singular. In Jacek Fisiak & Peter Trudgill (eds.), *East Anglian English*, 187–204. Cambridge: Boydell & Brewer.
- Nevalainen, Terttu & Matti Rissanen. 1986. Do you support the do-support? Emphatic and non-emphatic do in affirmative statements in present-day spoken English. In Sven Jacobson (ed.), *Papers from the Third Scandinavian Symposium on Syntactic Variation. Stockholm, May 11–12, 1985*, 35–50. Stockholm: Almqvist & Wiksell International.
- Nevalainen, Terttu & Matti Rissanen. 2002. *Fairly pretty* or *pretty fair*? On the development and grammaticalization of English downtoners. *Language Sciences* 24. 359–380. doi:10.1016/S0388-0001(01)00038-9
- Nevalainen, Terttu & Gijbert Rutten (eds.). 2012. *Neuphilologische Mitteilungen* CXIII(3). Special issue on comparative historical sociolinguistics.
- Nevalainen, Terttu & Ingrid Tieken-Boon van Ostade. 2006. Standardisation. In Richard M. Hogg & David Denison (eds.), *A history of the English language*, 271–311. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511791154.006
- Nevalainen, Terttu & Heli Tissari. 2010. Contextualising eighteenth-century politeness: Social distinction and metaphorical levelling. In Raymond Hickey (ed.), *Eighteenth-century English: Ideology and change* (Studies in English Language), 133–158. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511781643.009
- Nicholson, James. 1793. *The rudiments or first principles of English grammar*. Newcastle: M. Angus.
- Núñez-Peretejo, Paloma. 2004. *The progressive in the history of English* (LINCOM Studies in English Linguistics 06). Munchen: Lincom.
- Nurmi, Arja. 1996. Periphrastic do and be + ing: Interconnected developments? In Terttu Nevalainen & Helena Raumolin-Brunberg (eds.), *Sociolinguistics and language history: Studies based on the Corpus of Early English Correspondence* (Language and Computers: Studies in Practical Linguistics 15), 151–165. Amsterdam: Rodopi.
- Nurmi, Arja. 1998. *Manual for the Corpus of Early English Correspondence Sampler CEECS*. Department of English. University of Helsinki. <http://clu.uni.no/icame/manuals/>
- Nurmi, Arja. 1999. *A social history of periphrastic do* (Mémoires de la Société Néophilologique de Helsinki 56). Helsinki: Société Néophilologique.
- Nurmi, Arja, Minna Nevala & Minna Palander-Collin (eds.). 2009. *The language of daily life in England (1400–1800)*. Amsterdam: John Benjamins. doi:10.1075/pbns.183
- Nurmi, Arja & Päivi Pahta. 2010. Preacher, scholar, brother, friend: Social roles and code-switching in the writings of Thomas Twining. In Päivi Pahta, Minna Nevala, Arja Nurmi & Minna Palander-Collin (eds.), *Social roles and language practices in Late Modern English* (Pragmatics & Beyond NS 195), 135–162. Amsterdam & Philadelphia: John Benjamins.
- Nurmi, Arja & Minna Palander-Collin. 2008. Letters as a text type: Interaction in writing. In Marina Dossena & Ingrid Tieken-boon van Ostade (eds.), *Studies in Late Modern English correspondence: Methodology and data* (Linguistic Insights Studies in Language and Communication 76), 21–49. Bern: Peter Lang.
- OED = *Oxford English Dictionary*, 2nd edn. 1989. OED Online. Oxford University Press. <http://www.oed.com/>. (5 May, 2014.)
- Oeppen, Jim. 1993. Back projection and inverse projection: Members of a wider class of constrained projection models. *Population Studies* 47(2). 245–267. doi:10.1080/0032472031000146996

- Ogura, Mieko & William S.-Y. Wang. 1996. Snowball effect in lexical diffusion: The development of -s in the third-person singular present indicative in English. In Derek Britton (ed.), *English Historical Linguistics 1994: Papers from the 8th International Conference on English Historical Linguistics*, 119–141. Amsterdam & Philadelphia: John Benjamins.
- Oldireva-Gustafsson, Larisa. 2002. Variation in usage and grammars: The past participle forms of *write* in English 1680–1790. *Historical Sociolinguistics and Sociohistorical Linguistics* 2. [http://www.hum2.leidenuniv.nl/hsl\\_shl/Oldireva%20G.htm](http://www.hum2.leidenuniv.nl/hsl_shl/Oldireva%20G.htm). (July 2012.)
- Olsen, Kirstin. 1999. *Daily life in eighteenth-century England*. Westwood, CT: Greenwood Press. OTA = University of Oxford Text Archive. University of Oxford. <http://ota.ahds.ac.uk>.
- Pahta, Päivi, Minna Nevala, Arja Nurmi & Minna Palander-Collin (eds.). 2010. *Social roles and language practices in Late Modern English* (Pragmatics & Beyond NS 195). Amsterdam & Philadelphia: John Benjamins. doi:10.1075/pbns.195
- Palander-Collin, Minna & Mikko Hakala. 2011. Standardizing the Corpus of Early English Correspondence (CEEC). Poster presented at ICAME 32 (32nd conference of the International Computer Archive of Modern and Medieval English), 1–5 June 2011, Oslo, Norway. Available on CoRD: <http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/standardized.html>.
- Palander-Collin, Minna & Minna Nevala. 2010. Reporting and social role construction in eighteenth-century personal correspondence. In Päivi Pahta, Minna Nevala, Arja Nurmi & Minna Palander-Collin (eds.), *Social roles and language practices in Late Modern English* (Pragmatics & Beyond NS 195), 111–133. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/pbns.195.06pal
- Palander-Collin, Minna, Minna Nevala & Anni Sairio. 2013. Language and identity in letters. In Jukka Tyrkkö, Olga Timofeeva & Maria Salenius (eds.), *Ex philologia lux. Essays in honour of Leena Kahlas-Tarkka* (Mémoires de la Société Néophilologique de Helsinki 90), 291–311. Helsinki: Société Néophilologique.
- Palmer, Chris C. 2009. *Borrowings, derivational morphology, and perceived productivity in English, 1300–1600*. Ann Arbor, MI: University of Michigan PhD dissertation.
- Parker, Richard G. & Charles Fox. 1835 [1834]. *Progressive exercises in English grammar*, 3rd edn. Boston/New York.
- Pechenick, Eitan Adam, Christopher M. Danforth & Peter Sheridan Dodds. 2015. Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLOS ONE* 10(10). e0137041. doi:10.1371/journal.pone.0137041
- Percy, Carol. 2008. Mid-century grammars and their reception in the *Monthly Review* and the *Critical Review*. In Ingrid Tieken-Boon van Ostade (ed.), *Grammars, grammarians and grammar-writing in eighteenth-century England*, 125–142. Berlin & New York: Mouton de Gruyter.
- Percy, Carol. 2009a. Periodical reviews and the rise of prescriptivism: The *Monthly* (1749–1844) and *Critical Review* (1756–1817) in the eighteenth century. In Ingrid Tieken-Boon van Ostade & Wim van der Wurff (eds.), *Current issues in Late Modern English*, 117–150. Bern: Peter Lang.
- Percy, Carol. 2009b. Learning and virtue: English grammar and the eighteenth-century girls' school. In Mary Hilton & Jill Shefrin (eds.), *Educating the child in Enlightenment Britain: Beliefs, cultures, practices*, 77–98. Surrey: Ashgate.
- Poplack, Shana, Gerard, Van Herk & Dawn Harvie. 2002. 'Deformed in the dialects': An alternative history of non-standard English. In Richard Watts & Peter Trudgill (eds.), *Alternative histories of English*, 87–110. London & New York: Routledge.

- Porter, Roy. 2000. *Enlightenment: Britain and the creation of the modern world*. London: Palgrave Macmillan.
- Pratt, Linda & David Denison. 2000. The language of the Southey-Coleridge circle. *Language Sciences* 22. 401–422. doi:10.1016/S0388-0001(00)00013-9
- Prest, Wilfrid (ed.). 1987. *The professions in early modern England*. London: Croom Helm.
- Quackenbos, George P. 1868 [1862]. *An English grammar*. New York.
- Quirk, Randolph. 1971. Shakespeare and the English language. In Kenneth Muir & Samuel Schoenbaum (eds.), *A new companion to Shakespeare studies*, 67–83. Cambridge: Cambridge University Press.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Raumolin-Brunberg, Helena. 1994. The development of the compound pronouns in *-body* and *-one* in Early Modern English. In Dieter Kastovsky (ed.), *Studies in Early Modern English* (Topics in English Linguistics 13), 301–324. Berlin: Mouton de Gruyter. doi:10.1515/9783110879599.301
- Raumolin-Brunberg, Helena. 1998. Social factors and pronominal change in the seventeenth century: The Civil War effect. In Jacek Fisiak & Marcin Krygier (eds.), *Advances in English historical linguistics*, 361–388. Berlin & New York: Mouton de Gruyter.
- Raumolin-Brunberg, Helena. 2009. Lifespan changes in the language of three early modern gentlemen. In Arja Nurmi, Minna Nevala & Minna Palander-Collin (eds.), *The language of daily life in England (1400–1800)*, 165–196. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/pbns.183.10rau
- Raumolin-Brunberg, Helena & Leena Kahlas-Tarkka. 1997. Indefinite pronouns with singular human reference. In Matti Rissanen, Merja Kytö & Kirsi Heikkonen (eds.), *Grammaticalization at work: Studies of long-term developments in English* (Topics in English Linguistics 24), 17–86. Berlin: Mouton de Gruyter. doi:10.1515/9783110810745.17
- Raumolin-Brunberg, Helena & Terttu Nevalainen. 2007. Historical sociolinguistics: The *Corpus of Early English Correspondence*. In Joan C. Beal, Karen P. Corrigan & Hermann L. Moisl (eds.), *Creating and digitizing language corpora*, vol. 2, *Diachronic databases*, 148–171. Houndmills: Palgrave-Macmillan. A pre-print version is available at <http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/generalintro.html>.
- Raven, James. 2009. The book as a commodity. In Michael F. Suarez & Michael L. Turner (eds.), *The Cambridge history of the book in Britain, vol. 5, 1695–1830*, 85–117. Cambridge: Cambridge University Press. doi:10.1017/CHOL9780521810173.005
- Rayson, Paul, Geoffrey Leech & Mary Hodges. 1997. Social differentiation in the use of English vocabulary: Some analyses of the conversational component of the British National Corpus. *International Journal of Corpus Linguistics* 2(1). 133–152. doi:10.1075/ijcl.2.1.07ray
- Renouf, Antoinette. 2012. A finer definition of neology in English: The life-cycle of a word. In Hilde Hasselgård, Jarle Ebeling & Signe Oksefjell Ebeling (eds.), *Corpus perspectives on patterns of lexis* (Studies in Corpus Linguistics 57), 177–208. Amsterdam: John Benjamins.
- Riddle, Elizabeth M. 1985. A historical perspective on the productivity of the suffixes *-ness* and *-ity*. In Jacek Fisiak (ed.), *Historical semantics; historical word-formation* (Trends in Linguistics: Studies and Monographs 29), 435–461. Berlin: Mouton de Gruyter. doi:10.1515/9783110850178.435
- Rissanen, Matti. 1985. Periphrastic *do* in affirmative statements in Early American English. *Journal of English Linguistics* 18(2). 163–183. doi:10.1177/007542428501800205

- Rissanen, Matti. 1989. Three problems connected with the use of diachronic corpora. *ICAME Journal* 13. 16–19.
- Rissanen, Matti. 1997. Whatever happened to the Middle English indefinite pronouns? In Jacek Fisiak (ed.), *Studies in Middle English linguistics*, 513–529. Berlin: Mouton de Gruyter. doi:10.1515/9783110814194.513
- Rissanen, Matti. 1999. Syntax. In Roger Lass (ed.), *The Cambridge history of the English language, vol. III: 1476–1776*, 187–331. Cambridge: Cambridge University Press.
- Robins, Robert H. 1986. The evolution of English grammar books since the Renaissance. In Gerhard Leitner (ed.), *English reference grammar. Languages and linguistics writers and readers*, 292–306. Tübingen: Niemeyer.
- Rohrdantz, Christian, Andreas Niekler, Annette Hautli, Miriam Butt & Daniel A. Keim. 2012. Lexical semantics and distribution of suffixes: A visual analysis. In Miriam Butt, Sheelagh Carpendale, Gerald Penn, Jelena Prokić & Michael Cysouw (eds.), *Proceedings of the EACL 2012 joint workshop of LINGVIS & UNCLH*, 7–15. Stroudsburg, PA: Association for Computational Linguistics.
- Romaine, Suzanne. 1985. Variability in word formation patterns and productivity in the history of English. In Jacek Fisiak (ed.), *Papers from the 6th International Conference on Historical Linguistics*, Poznań, 22–26 August 1983 (Current Issues in Linguistic Theory 34), 451–465. Amsterdam: John Benjamins.
- Romaine, Suzanne (ed.). 1998. *The Cambridge history of the English language, vol. IV, 1776–1997*. Cambridge: Cambridge University Press.
- Romaine, Suzanne. 2005. Variation in language and gender. In Janet Holmes & Miriam Meyerhoff (eds.), *The handbook of language and gender*, 98–118. Malden, MA & Oxford: Blackwell.
- Rowlands, Ian, David Nicholas, Peter Williams, Paul Huntington, Maggie Fieldhouse, Barrie Gunter, Richard Withey, Hamid R. Jamali, Tom Dobrowolski & Carol Tenopir. 2008. The Google generation: The information behaviour of the researcher of the future. *Aslib Proceedings* 60(4). 290–310. doi:10.1108/00012530810887953
- Rutten, Gijsbert. 2009. Grammar to the people. The Dutch language and the public sphere in the 18th century. With special reference to Kornelis van der Palm. *Beiträge zur Geschichte der Sprachwissenschaft* 19. 55–86.
- Rutten, Gijsbert & Marijke J. van der Wal. 2014. *Letters as loot: A Sociolinguistic approach to seventeenth- and eighteenth-century Dutch*. Amsterdam & Philadelphia: John Benjamins.
- Rutten, Gijsbert, Rik Vosters & Wim Vandenbussche (eds.). 2014. *Norms and usage in language history, 1600–1900: A sociolinguistic and comparative perspective*. Amsterdam & Philadelphia: John Benjamins.
- Rydén, Mats & Sverker Brorström. 1987. *The be/have variation with intransitives in English: With special reference to the Late Modern period*. Stockholm: Almqvist & Wiksell International.
- Saeed, John I. 1997. *Semantics* (Introducing Linguistics 2). Oxford: Blackwell.
- Säily, Tanja. 2011. Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations. *Corpus Linguistics and Linguistic Theory* 7(1). 119–141. doi:10.1515/cllt.2011.006
- Säily, Tanja. 2014. *Sociolinguistic variation in English derivational productivity: Studies and methods in diachronic corpus linguistics* (Mémoires de la Société Néophilologique de Helsinki XCIV). Helsinki: Société Néophilologique. <http://urn.fi/URN:ISBN:978-951-9040-50-9>.
- Säily, Tanja. 2016. Sociolinguistic variation in morphological productivity in eighteenth-century English. *Corpus Linguistics and Linguistic Theory* 12(1). 129–151.

- Säily, Tanja, Terttu Nevalainen & Harri Siirtola. 2011. Variation in noun and pronoun frequencies in a sociohistorical corpus of English. *Literary and Linguistic Computing* 26(2). 167–188. doi:10.1093/lc/fqr004
- Säily, Tanja & Jukka Suomela. 2009. Comparing type counts: The case of women, men and *-ity* in early English letters. In Antoinette Renouf & Andrew Kehoe (eds.), *Corpus linguistics: Refinements and reassessments* (Language and Computers: Studies in Practical Linguistics 69), 87–109. Amsterdam: Rodopi. doi:10.1163/9789042025981\_007
- Sairio, Anni. 2009. *Language and letters of the Bluestocking network: Sociolinguistic issues in eighteenth-century epistolary English* (Mémoires de la Société Néophilologique de Helsinki 75). Helsinki: Société Néophilologique.
- Sairio, Anni & Minna Nevala. 2013. Social dimensions of layout in eighteenth-century letters and letter-writing manuals. In Anneli Meurman-Solin & Jukka Tyrkkö (eds.), *Principles and practices for the digital editing and annotation of diachronic data* (Studies in Variation, Contacts and Change in English 14). Helsinki: VARIENG. [http://www.helsinki.fi/varieng/series/volumes/14/sairio\\_nevala/](http://www.helsinki.fi/varieng/series/volumes/14/sairio_nevala/). (26 June, 2015.)
- Sandhu, Sukhdev S. 1998. Ignatius Sancho and Laurence Sterne. *Research in African Literature* 29(4). 88–105.
- Schütze, Carson T. 2004. Synchronic and diachronic microvariation in English *do*. *Lingua* 114. 495–516. doi:10.1016/S0024-3841(03)00070-6
- Schwartz, Leonard. 2000. London 1700–1840. In Peter Clark (ed.), *The Cambridge history of urban Britain, vol. 2, 1540–1840*, 641–672. Cambridge: Cambridge University Press. doi:10.1017/CHOL9780521431415.025
- Shoemaker, Robert. 2004. *The London mob: Violence and disorder in eighteenth-century England*. London: Hambledon and London.
- Skedd, Susan. 1997. Women teachers and the expansion of girl's schooling in England, c.1760–1820. In Hannah Barker & Elaine Chalus (eds.), *Gender in eighteenth-century England*, 101–125. London: Longman.
- Smith, Adam (1762) [1983]. *Lectures on rhetoric and belles lettres. The Glasgow edition of the works and correspondence of Adam Smith*. Vol. 4. Ed. by J. C. Bryce. Oxford: Oxford University Press.
- Smitterberg, Erik. 2002. *The progressive in 19th-century English: A process of integration*. Uppsala, Sweden: Uppsala University PhD thesis.
- Smitterberg, Erik. 2005. *The progressive in 19th-century English: A process of integration* (Language and Computers: Studies in Practical Linguistics 54). Amsterdam & New York: Rodopi.
- Stein, Dieter. 1990. *The semantics of syntactic change: Aspects of the evolution of do in English*. Berlin: Mouton de Gruyter. doi:10.1515/9783110846829
- Steinbach, Susie. 2005. *Women in England 1750–1914: A social history*. London: Phoenix.
- Stone, Lawrence. 1969. Literacy and education in England, 1640–1900. *Past and Present* 42. 69–139. doi:10.1093/past/42.1.69
- Straaijer, Robin. 2010. Prescription or practice? *Be/have* variation with past participles of mutative intransitive verbs in the letters of Joseph Priestley. In Ursula Lenker, Judith Huber & Robert Mailhammer (eds.), *English historical linguistics 2008. Selected papers from the fifteenth International Conference on English Historical Linguistics (ICEHL 15), Munich, 24–30 August 2008*, 63–78. Amsterdam & Philadelphia: John Benjamins.
- Strang, Barbara M. H. 1970. *A history of English*. London: Methuen.

- Strang, Barbara. 1982. Some aspects of the history of the *be+ing* construction. In John Anderson (ed.), *Language form and linguistic variation*, 427–474. Amsterdam: John Benjamins.  
doi:10.1075/cilt.15.22str
- Suarez, Michael F. 2009. Toward a bibliometric analysis of the surviving record, 1701–1800. In Michael F. Suarez & Michael L. Turner (eds.), *The Cambridge history of the book in Britain, vol. 5, 1695–1830*, 39–65. Cambridge: Cambridge University Press.  
doi:10.1017/CHOL9780521810173.003
- Sundby, Bertil, Anne Kari Bjørge & Kari E. Haugland. 1991. *A dictionary of English normative grammar 1700–1800*. Amsterdam & Philadelphia: John Benjamins.
- Suomela, Jukka. 2014. *types2*: Type and hapax accumulation curves. Computer program. Zenodo. doi:10.5281/zenodo.9868, <http://users.ics.aalto.fi/suomela/types2/>.
- Svartvik, Jan & Hans Lindquist. 1997. One and body language. In Udo Fries, Viviane Müller & Peter Schneider (eds.), *From Ælfric to the New York Times: Studies in English corpus linguistics*, 11–20. Amsterdam & Atlanta: Rodopi.
- Sweet, Henry. 1903. *A new English grammar* II. Oxford: Oxford University Press.
- Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change* 21. 319–353. doi:10.1017/S0954394509990123
- Szmrecsanyi, Benedikt. 2012. Analyticity and syntheticity in the history of English. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 654–665. Oxford & New York: Oxford University Press.
- Szmrecsanyi, Benedikt. 2016. About text frequencies in historical linguistics: Disentangling environmental and grammatical change. *Corpus Linguistics and Linguistic Theory* 12(1). 153–171.
- Szmrecsanyi, Benedikt, Douglas Biber, Jesse Egbert & Karlien Franco. 2016. Toward more accountability: Modeling ternary genitive variation in Late Modern English. *Language Variation and Change* 28(1). 1–29. doi:10.1017/S0954394515000198
- Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA & Oxford: Wiley-Blackwell.
- Tagliamonte, Sali A. 2016. *Teen talk: The language of adolescents*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139583800
- The Art of Letter-Writing, divided into two parts. The first, containing rules and directions for writing letters on all sorts of subjects: with a variety of examples, Equally elegant and instructive. The second, a collection of letters on the most interesting occasions in life, etc.* (1762). London: T. Osborne.
- Tieken-Boon van Ostade, Ingrid. 1985. *Do*-support in the writings of Lady Mary Wortley Montagu: A change in progress. *Folia Linguistica Historica* VI. 127–151.
- Tieken-Boon van Ostade, Ingrid. 1987. *The auxiliary do in eighteenth-century English. A Sociohistorical-linguistic Approach*. Dordrecht: Foris.
- Tieken-Boon van Ostade, Ingrid. 2000. Normative studies in England. In Sylvain Auroux, E. F. Konrad Koerner, Hans-Josef Niederehe & Kees Versteegh (eds.), *History of the language sciences*, vol. I, 876–887. Berlin & New York: Walter de Gruyter.
- Tieken-Boon van Ostade, Ingrid. 2005. Of social networks and linguistic influence: The language of Robert Lowth and his correspondents. In Juan Camilo Conde-Silvestre (ed.), *Sociolinguistics and the history of English*. Special issue of *International Journal of English Studies* 5. 135–157.

- Tieken-Boon van Ostade, Ingrid. 2006. Eighteenth-century prescriptivism and the norm of correctness. In Ans van Kemenade & Bettelou Los (eds.), *The handbook of the history of English*, 539–557. London: Blackwell. doi:10.1002/9780470757048.ch21
- Tieken-Boon van Ostade, Ingrid (ed.). 2008. *Grammars, grammarians and grammar-writing in eighteenth-century England*. Berlin & New York: Mouton de Gruyter.
- Tieken-Boon van Ostade, Ingrid. 2008a. Grammars, grammarians and grammar-writing: An introduction. In Ingrid Tieken-Boon van Ostade (ed.), *Grammars, grammarians and grammar-writing in eighteenth-century England*, 1–14. Berlin & New York: Mouton de Gruyter.
- Tieken-Boon van Ostade, Ingrid. 2008b. The 1760s: Grammars, grammarians and the booksellers. In Ingrid Tieken-Boon van Ostade (ed.), *Grammars, grammarians and grammar-writing in eighteenth-century England*, 101–124. Berlin & New York: Mouton de Gruyter.
- Tieken-Boon van Ostade, Ingrid. 2008c. Reception and the market for grammars: Introduction. In Ingrid Tieken-Boon van Ostade (ed.), *Grammars, grammarians and grammar-writing in eighteenth-century England*, 79–80. Berlin & New York: Mouton de Gruyter.
- Tieken-Boon van Ostade, Ingrid. 2009. *An introduction to Late Modern English*. Edinburgh: Edinburgh University Press.
- Tieken-Boon van Ostade, Ingrid. 2010. Eighteenth-century women and their norms of correctness. In Raymond Hickey (ed.), *Eighteenth-century English: Ideology and change* (Studies in English Language), 59–72. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511781643.005
- Tieken-Boon van Ostade, Ingrid. 2014. *In search of Jane Austen. The language of the letters*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199945115.001.0001
- Trudgill, Peter. 1999. *The dialects of England*, 2nd edn. Oxford: Blackwell.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Tyrkkö, Jukka. 2013. Notes on eighteenth-century dictionary grammars. *Transactions of the Philological Society* 111(2). 179–201. doi:10.1111/1467-968X.12031
- Vartiainen, Turo, Tanja Säily & Mikko Hakala. 2013. Variation in pronoun frequencies in early English letters: Gender-based or relationship-based? In Jukka Tyrkkö, Olga Timofeeva & Maria Salenius (eds.), *Ex philologia lux: Essays in honour of Leena Kahlas-Tarkka* (Mémoires de la Société Néophilologique de Helsinki XC), 233–255. Helsinki: Société Néophilologique.
- Visser, F. 1969. *An historical syntax of the English language*. III Part I. Leiden: E.J. Brill.
- Visser, Federicus Th. 1963–73. *An historical syntax of the English Language*. 4 vols. Leiden: E.J. Brill.
- Vorlat, Emma. 1996. Lindley Murray's prescriptive canon. In Ingrid Tieken-Boon van Ostade (ed.), *Two hundred years of Lindley Murray*, 163–182. Münster: Nodus Publikationen.
- Wagner, Suzanne Evans & Isabelle Buchstaller. 2017. *Panel studies of variation and change*. London: Routledge.
- Wagner, Suzanne Evans & Gillian Sankoff. 2011. Age grading in the Montréal French inflected future. *Language Variation and Change* 23(3). 275–313. doi:10.1017/S0954394511000111
- Wales, Kathleen M. 1983. *Thou and you in Early Modern English: Brown and Gilman re-appraised*. *Studia Linguistica* 37(2). 107–125. doi:10.1111/j.1467-9582.1983.tb00316.x
- Walker, Terry. 2007. *Thou and you in Early Modern English dialogues*. Amsterdam: John Benjamins. doi:10.1075/pbns.158
- Wallis, John. 1653. *Grammatica linguae Anglicanae*. Oxford.

- Watts, Richard J. 1999. Language and politeness in early eighteenth century Britain. *Pragmatics* 9(1). 9–20. doi:10.1075/prag.9.1.02wat
- Watts, Richard J. 2002. From polite language to educated language: The re-emergence of an ideology. In Richard Watts & Peter Trudgill (eds.), *Alternative histories of English*, 155–172. London & New York: Routledge. doi:10.4324/9780203468005
- Watts, Richard J. 2008. Grammar writers in eighteenth-century Britain: A community of practice or a discourse community? In Tieken-Boon van Ostade (ed.), 37–56.
- Webster, Noah. 1789. *Dissertations on the English language*. Boston.
- White, James. 1761. *The English verb*. London.
- Whyman, Susan. 2009. *The Pen and the people: English letter writers 1660–1800*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199532445.001.0001
- Wilcoxon, Frank. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6). 80–83. doi:10.2307/3001968
- Wolfson, Nessa. 1990. The bulge: A theory of speech behavior and social distance. *Penn Working Papers in Educational Linguistics* 2(1). 55–83.
- Wood, Johanna L. 2004. Text in context: A critical discourse analysis approach to Margaret Paston. In Terttu Nevalainen & Sanna-Kaisa Tanskanen (eds.), *Letter writing. Special issue of Journal of Historical Pragmatics* 5(2). 229–254.
- Wright, Susan. 1994. The mystery of the modal progressive. In Dieter Kastovsky (ed.), *Studies in Early Modern English*, 467–485. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110879599.467
- Wrightson, Keith. 2002. *Earthly necessities: Economic lives in Early Modern Britain 1470–1750*. London: Penguin Books.
- Wrigley, Edward A. 1987. *People, cities and wealth: The transformation of traditional society*. Oxford: Blackwell.
- Wrigley, Edward A. 2004. *Poverty, progress, and population*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511616365
- Wrigley, Edward A. 2010. *Energy and the English industrial revolution*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511779619
- Wrigley, Edward A. & Roger C. Schofield. 1981. *The population history of England 1541–1871: A reconstruction*. London: Edward Arnold.
- Yáñez-Bouza, Nuria. 2016. Early and Late Modern English grammars as evidence in English historical linguistics. In Merja Kytö & Päivi Pahta (eds.), *The Cambridge handbook of English historical linguistics*, 164–180. Cambridge: Cambridge University Press.
- Yáñez-Bouza, Nuria & David Denison. 2015. Which comes first in the double object construction? *English Language and Linguistics* 19(2). 247–268. doi:10.1017/S136067431500012X
- Yáñez-Bouza, Nuria & María E. Rodríguez-Gil. 2013. Special volume on *English Grammar Writing in the Eighteenth Century*. *Transactions of the Philological Society* 111(2). 141–258. doi:10.1111/1467-968X.12027
- Ziegeler, Debra. 2004. Reanalysis in the history of *do*: A view from construction grammar. *Cognitive Linguistics* 15. 529–574. doi:10.1515/cogl.2004.15.4.529

# Appendix: Editions in the *Corpora of Early English Correspondence*

Arja Nurmi and Samuli Kaislaniemi

This is a list of all editions used for compiling those parts of the *Corpora of Early English Correspondence* used in the studies in this volume (see Chapter 4). Each entry indicates which subcorpus it belongs to: *Corpus of Early English Correspondence* (CEEC) or *Corpus of Early English Correspondence Extension* (CEECE). In addition, the collections from CEEC that have been included in the two published versions, the *Corpus of Early English Correspondence Sampler* (CEECS, released 1998) and the *Parsed Corpus of Early English Correspondence* (PCEECE, released 2006) are so indicated.

ADDISON	1699?–1718	14,201 words
Graham, Walter (ed.). 1941. <i>The Letters of Joseph Addison</i> . Oxford: Clarendon Press. [CEECE]		
ARUNDEL	1589–1680	19,202 words
Hervey, Mary F.S. (ed.). 1921. <i>The Life, Correspondence &amp; Collections of Thomas Howard, Earl of Arundel, “Father of Vertu in England”</i> . Cambridge: Cambridge University Press. [CEECE, PCEECE]		
AUSTEN	1796?–1800?	27,955 words
Le Faye, Deirdre (ed.). 1995. <i>Jane Austen’s Letters</i> . 3rd edition. Oxford/New York: Oxford University Press. [CEECE]		
BANKS	1704–1756	39,162 words
Hill, J.W.F. (ed.). 1952. <i>The Letters and Papers of the Banks Family of Revesby Abbey 1704–1760</i> [Lincoln Record Society 45]. Lincoln: Lincoln Record Society. [CEECE]		
BENTHAM J	1745–1800	51,613 words
Sprigge, Timothy L.S. (ed.). 1968. <i>The Correspondence of Jeremy Bentham</i> , vol. 1, 1752–1776; and vol. 2, 1777–1780. London: Athlone Press. [CEECE]		
AND		
Christie, Ian R. (ed.). 1971. <i>The Correspondence of Jeremy Bentham</i> , vol. 3, <i>January 1781 to October 1788</i> . London: Athlone Press. [CEECE]		
AND		
Milne, Alexander Taylor (ed.). 1981. <i>The Correspondence of Jeremy Bentham</i> , vol. 4, <i>October 1788 to December 1793</i> ; and vol. 5, <i>January 1794 to December 1797</i> . London: Athlone Press. [CEECE]		
AND		
Dinwiddy, J.R. (ed.). 1984. <i>The Correspondence of Jeremy Bentham</i> , vol. 6, <i>January 1798 to December 1801</i> . Oxford: Clarendon Press. [CEECE]		

- BLOMEFIELD 1730–1751 13,318 words  
 Stoker, David A. (ed.). 1992. *The Correspondence of the Reverend Francis Blomefield (1705–52)* [Norfolk Record Society 55]. London: Bibliographical Society. [CEECE]
- BOLTON 1695–1700 8,300 words  
 Simon, André L. (ed.). 1928. *The Bolton Letters. The Letters of an English Merchant in Madeira 1695–1714*, vol. 1, 1695–1700. London: T. Werner Laurie. [CEECE]
- BOWREY 1687–1708 19,229 words  
 Temple, Sir Richard Carnac (ed.). 1905. *A Geographical Account of Countries Round the Bay of Bengal, 1669 to 1679, by Thomas Bowrey* [Hakluyt Society Second Series 12]. Cambridge: Hakluyt Society. [CEECE]  
 AND  
 Temple, Sir Richard Carnac (ed.). 1927. *The Papers of Thomas Bowrey 1669–1713* [Hakluyt Society Second Series 58]. London: Hakluyt Society. [CEECE]
- BROWNE 1653–1681 20,778 words  
 Keynes, Geoffrey (ed.). 1964. *The Works of Sir Thomas Browne*, vol. 4, *Letters*. Chicago: The University of Chicago Press. [CEEC, PCEEC]
- BURNEY 1762–1784 39,451 words  
 Ribeiro, Alvaro, S.J. (ed.). 1991. *The Letters of Dr Charles Burney*, vol. 1, 1751–1784. Oxford: Clarendon Press. [CEECE]
- BURNEY F 1770–1800 55,729 words  
 Troide, Lars E. (ed.). 1990. *The Early Journals and Letters of Fanny Burney*, vol. 2, 1772–1777. Oxford: Clarendon Press. [CEECE]  
 AND  
 Troide, Lars E. and Cooke, Stewart J. (eds). 1994. *The Early Journals and Letters of Fanny Burney*, vol. 3, 1778–1779. Oxford: Clarendon Press. [CEECE]  
 AND  
 Hemlow, Joyce, Cecil, Curtis D. and Douglas, Althea (eds). 1972. *The Journals and Letters of Fanny Burney (Madame d'Arblay)*, vol. 1, 1791–1792. Oxford: Clarendon Press. [CEECE]  
 AND  
 Hemlow, Joyce and Douglas, Althea (eds). 1972. *The Journals and Letters of Fanny Burney (Madame d'Arblay)*, vol. 2, *Courtship and Marriage 1793*. Oxford: Clarendon Press. [CEECE]  
 AND  
 Hemlow, Joyce, Boutillier, Patricia and Douglas, Althea (eds). 1972. *The Journals and Letters of Fanny Burney (Madame d'Arblay)*, vol. 3, *Great Bookham 1793–1797*. Oxford: Clarendon Press. [CEECE]  
 AND  
 Hemlow, Joyce (ed.). 1973. *The Journals and Letters of Fanny Burney (Madame d'Arblay)*, vol. 4, *West Humble 1797–1801*. Oxford: Clarendon Press. [CEECE]
- BUTE 1756–1764? 5,372 words  
 Sedgwick, Romney (ed.). 1939. *Letters from George III to Lord Bute 1756–1766*. London: Macmillan. [CEECE]

- CARTER 1738–1800 25,499 words  
Hampshire, Gwen (ed.). 2005. *Elizabeth Carter, 1717–1806: An Edition of Some Unpublished Letters*. Newark: University of Delaware Press. [CEECE]
- CHAMPION 1774–1776 10,790 words  
Guttridge, G.H. (ed.). 1934. *The American Correspondence of a Bristol Merchant 1766–1776. Letters of Richard Champion* [University of California Publications in History 22 (1)]. Berkeley: University of California Press. [CEECE]
- CLAVERING 1705?–1741 72,846 words  
Dickinson, H.T. (ed.). 1967. *The Correspondence of Sir James Clavering* [Surtees Society 178]. Gateshead: Surtees Society. [CEECE]
- CLIFT 1792–1799 52,038 words  
Austin, Frances (ed.). 1991. *The Clift Family Correspondence 1792–1846*. Sheffield: Centre for English Cultural Tradition and Language, University of Sheffield. [CEECE]
- CONWAY 1640–1680 57,946 words  
Nicholson, Marjorie Hope (ed.) and Hutton, Sarah (rev. ed.). 1992. *The Conway Letters. The Correspondence of Anne, Viscountess Conway, Henry More, and Their Friends, 1642–1684*. Oxford: Clarendon Press. [CEEC, PCEEC]
- COWPER S 1732–1764 10,187 words  
Hughes, Edward (ed.). 1956. *Letters of Spencer Cowper, Dean of Durham 1746–74* [Surtees Society 165]. Durham: Andrews & Co. for the Surtees Society; London: Bernard Quaritch for the Surtees Society. [CEECE]
- COWPER W 1759–1799 56,254 words  
King, James and Ryskamp, Charles (eds). 1979–1984. *The Letters and Prose Writings of William Cowper*, vols 1–4. Oxford: Clarendon Press. [CEECE]
- CRISP 1779–1782 18,389 words  
Hutton, William Holden (ed.). 1905. *Burford Papers. Being Letters of Samuel Crisp to his Sister at Burford; and other Studies of a Century (1745–1845)*. London: Archibald Constable & Co. [CEECE]
- CULLEY 1784–1785 25,100 words  
Orde, Anne (ed.). 2002. *Matthew and George Culley. Travel Journals and Letters, 1765–1798* [Records of Social and Economic History New Series 35]. Oxford: Oxford University Press for the British Academy. [CEECE]
- DARWIN 1763–1797 18,362 words  
King-Hele, Desmond (ed.). 1981. *The Letters of Erasmus Darwin*. Cambridge: Cambridge University Press. [CEECE]
- DEFOE 1703–1729? 32,688 words  
Healey, George Harris (ed.). 1955. *The Letters of Daniel Defoe*. Oxford: Clarendon Press. [CEECE]

- DODSLEY 1743–1764 48,691 words  
Tierney, James E. (ed.). 1988. *The Correspondence of Robert Dodsley, 1733–1764*. Cambridge: Cambridge University Press. [CEECE]
- DRAPER 1757?–1775? 22,511 words  
Wright, Arnold and Sclater, William Lutley (eds). 1922. *Sterne's Eliza. Some Account of Her Life in India: With Her Letters Written between 1757 and 1774*. London: William Heinemann. [CEECE]
- DUKES 1732–1750 50,858 words  
McCann, Timothy J. (ed.). 1984. *The Correspondence of the Dukes of Richmond and Newcastle 1724–1750* [Sussex Record Society 73]. Lewes: Sussex Record Society. [CEECE]
- EVELYN 1665–1703 38,929 words  
de la Bédoyère, Guy (ed.). 1997. *Particular Friends. The Correspondence of Samuel Pepys and John Evelyn*. Woodbridge: Boydell Press. [CEECE]
- EVELYN 2 1665–1700 14,378 words  
de la Bédoyère, Guy (ed.). 1997. *Particular Friends. The Correspondence of Samuel Pepys and John Evelyn*. Woodbridge: Boydell Press. [CEECE]
- FLEMING 1650–1680 39,833 words  
Magrath, John Richard (ed.). 1904. *The Flemings in Oxford, Being Documents Selected from the Rydal Papers in Illustration of the Lives and Ways of Oxford Men 1650–1700*, vol. 1, 1650–1680 [Oxford Historical Society 44]. Oxford: Clarendon Press. [CEEC, PCEEC]
- FLEMING 2 1653–1701 76,297 words  
Magrath, John Richard (ed.). 1913–1924. *The Flemings in Oxford, Being Documents Selected from the Rydal Papers in Illustration of the Lives and Ways of Oxford Men 1650–1700*, vol. 2, 1680–1690; and vol. 3, 1691–1700 [Oxford Historical Society 62 and 79]. Oxford: Clarendon Press. [CEECE]
- FLEMING X 1684–1698 14,020 words  
Magrath, John Richard (ed.). 1913–1924. *The Flemings in Oxford, Being Documents Selected from the Rydal Papers in Illustration of the Lives and Ways of Oxford Men 1650–1700*, vol. 2, 1680–1690; and vol. 3, 1691–1700 [Oxford Historical Society 62 and 79]. Oxford: Clarendon Press. [CEECE]
- FOUNDLING 1758–1767 50,221 words  
Clark, Gillian (ed.). 1994. *Correspondence of the Foundling Hospital Inspectors in Berkshire 1757–68* [Berkshire Record Society 1]. Reading: Berkshire Record Society. [CEECE]
- GARRICK 1733–1777 42,832 words  
Little, David M. and Kahrl, George M. (eds). 1963. *The Letters of David Garrick*, vols 1–3. Cambridge, MA: The Belknap Press of Harvard University Press. [CEECE]
- GAY 1705–1731 7,227 words  
Burgess, C.F. (ed.). 1966. *The Letters of John Gay*. Oxford: Clarendon Press. [CEECE]
- GEORGE 3 1765–1783 7,765 words  
Fortescue, John (ed.). 1927–1928. *The Correspondence of King George the Third, from 1760 to December 1783, printed from the original papers in the Royal Archives at Windsor Castle*, vols 1–6. London: Macmillan. [CEECE]





Gardiner, Dorothy (ed.). 1937. *The Oxinden and Peyton Letters 1642–1670. Being the Correspondence of Henry Oxinden of Barham, Sir Thomas Peyton of Knowlton and Their Circle*. London: The Sheldon Press. [CEECE]

PAUPER 1731–1795? 2,220 words

Sokoll, Thomas (ed.). 2001. *Essex Pauper Letters 1731–1837* [Records of Social and Economic History New Series 30]. Oxford/New York: Oxford University Press for the British Academy. [CEECE]

PEPYS 1663–1680 42,476 words

Heath, Helen Truesdell (ed.). 1955. *The Letters of Samuel Pepys and His Family Circle*. Oxford: Clarendon Press. [CEECE, PCEECE]

PEPYS 2 1681–1692 9,435 words

Heath, Helen Truesdell (ed.). 1955. *The Letters of Samuel Pepys and His Family Circle*. Oxford: Clarendon Press. [CEECE]

PEPYS 3 1665–1700 27,126 words

de la Bédoyère, Guy (ed.). 1997. *Particular Friends. The Correspondence of Samuel Pepys and John Evelyn*. Woodbridge: Boydell Press. [CEECE]

AND

Heath, Helen Truesdell (ed.). 1955. *The Letters of Samuel Pepys and His Family Circle*. Oxford: Clarendon Press. [CEECE]

PERROT 1799–1800 8,924 words

Stevens Cox, J. (ed.). 1985. *Letters from Ilchester Gaol, 1799–1800, by Jane Leigh Perrot, Jane Austen's Aunt* [Ilchester and District Occasional Papers 64]. Guernsey: Toucan Press. [CEECE]

PETTY 1676–1681 22,408 words

Marquis of Lansdowne (Petty Fitzmaurice, Henry W.E.) (ed.). 1928/1976. *The Petty-Southwell Correspondence 1676–1687* [Reprints of Economic Classics]. New York: Augustus M. Kelley. [CEECE, PCEECE]

PETTY 2 1682–1687 14,378 words

Marquis of Lansdowne (Petty Fitzmaurice, Henry W.E.) (ed.). 1928/1976. *The Petty-Southwell Correspondence 1676–1687* [Reprints of Economic Classics]. New York: Augustus M. Kelley. [CEECE]

PIERCE 1751–1775 20,354 words

Macdonald, Violet M. (ed.). 1927. *The Letters of Eliza Pierce 1751–1775, with Letters from Her Son Pierce Joseph Taylor a Schoolboy at Eton*. London: Frederick Etchells and Hugh Macdonald. [CEECE]

PINNEY 1679–1706 25,098 words

Nuttall, Geoffrey F. (ed.). 1939. *Letters of John Pinney 1679–1699*. London/New York/Toronto: Oxford University Press. [CEECE]

- PIOZZI 1784–1798 39,572 words  
 Bloom, Edward A. and Bloom, Lillian D. (eds). 1989–1991. *The Piozzi Letters. Correspondence of Hester Lynch Piozzi, 1784–1821*, vols 1–2. Newark: University of Delaware Press; London/Toronto: Associated University Press. [CEECE]
- PITT 1751–1757 9,071 words  
 Lord Grenville (Grenville, William, Baron Grenville) (ed.). 1805. *Letters Written by the Late Earl of Chatham to His Nephew Thomas Pitt, Esq. (Afterwards Lord Camelford) Then at Cambridge*. London/Gottingen: Reprinted for H. Dieterich. [CEECE]
- PITT 2 1754 15,618 words  
 Kaislaniemi, Samuli (ed.). 2004. Letters of William Pitt and Lady Hester née Grenville. The National Archives of the United Kingdom, PRO 30/8 1st Series: Correspondence of William Pitt, 1st Earl of Chatham and Lady Hester (Grenville) his wife (PRO 30/8/5/1: Letters of William Pitt, Earl of Chatham: To Lady Hester before their marriage. PRO 30/8/7: Letters of Lady Hester to her husband). Unpublished. [CEECE]
- POPE 1708–1744? 32,869 words  
 Sherburn, George (ed.). 1956. *The Correspondence of Alexander Pope*, vols 1–4. Oxford: Clarendon Press. [CEECE]
- PORTER 1789–1800? 13,815 words  
 Martin, Joanna (ed.). 1998. *A Governess in the Age of Jane Austen. The Journals and Letters of Agnes Porter*. London/Rio Grande: Hambledon Press. [CEECE]
- PRIDEAUX 1674–1680 7,571 words  
 Thompson, Edward Maunde (ed.). 1875. *Letters of Humphrey Prideaux, Sometime Dean of Norwich, to John Ellis, Sometime Under-Secretary of State, 1674–1722* [Camden Second Series 15]. London: Camden Society. [CEEC, PCEEC]
- PRIDEAUX 2 1681–1722 15,934 words  
 Thompson, Edward Maunde (ed.). 1875. *Letters of Humphrey Prideaux, Sometime Dean of Norwich, to John Ellis, Sometime Under-Secretary of State, 1674–1722* [Camden Second Series 15]. London: Camden Society. [CEECE]
- PUREFOY 1736–1754 28,549 words  
 Eland, G. (ed.). 1931. *Purefoy Letters 1735–1753*, vols 1–2. London: Sidgwick and Jackson. [CEECE]
- ROYAL 3 1676–1680 13,363 words  
 Bathurst, Benjamin (ed.). 1925. *Letters of Two Queens*. London: Robert Holden & Co. [CEEC]
- ROYAL 4 1681?–1684? 4,408 words  
 Bathurst, Benjamin (ed.). 1925. *Letters of Two Queens*. London: Robert Holden & Co. [CEECE]

- SANCHO 1770?–1780? 20,216 words  
 Edwards, Paul and Rewt, Polly (eds). 1994. *The Letters of Ignatius Sancho*. Edinburgh: Edinburgh University Press. [CEECE]
- SECKER 1738–1761 31,661 words  
 Jenkins, A.P. (ed.). 1991. *The Correspondence of Thomas Secker, Bishop of Oxford 1737–58* [Oxfordshire Record Society 57]. Oxford: Oxfordshire Record Society. [CEECE]
- STUBS 1791–1800 4,274 words  
 Ashton T.S. (ed.). 1939/1961. *An Eighteenth-Century Industrialist. Peter Stubs of Warrington 1756–1806*. Manchester: Manchester University Press. [CEECE]
- SWIFT 1712–1740 56,541 words  
 Woolley, David (ed.). 1999–2007. *The Correspondence of Jonathan Swift*, vols 1–4. Frankfurt am Main: Peter Lang. [CEECE]
- TIXALL 1656–1680 11,545 words  
 Clifford, Arthur (ed.). 1815. *Tixall Letters; or the Correspondence of the Aston Family, and Their Friends, during the Seventeenth Century*, vol. 2. London: Longman, Hurst, Rees, Orme and Brown; Edinburgh: Archibald Constable & Co. and John Ballantyne & Co. [CEEC, CEECS, PCEEC]
- TIXALL 2 1684–1686 392 words  
 Clifford, Arthur (ed.). 1815. *Tixall Letters; or the Correspondence of the Aston Family, and Their Friends, during the Seventeenth Century*, vol. 2. London: Longman, Hurst, Rees, Orme and Brown; Edinburgh: Archibald Constable & Co. and John Ballantyne & Co. [CEECE]
- TWINING 1762–1800 57,793 words  
 Walker, Ralph S. (ed.). 1991. *A Selection of Thomas Twining's Letters 1734–1804. The Record of a Tranquil Life*, vols 1–2 [Studies in British History 25]. Lewiston/Queenston/Lampeter: Edwin Mellen Press. [CEECE]
- WANLEY 1694–1726 32,725 words  
 Heyworth, P.L. (ed.). 1989. *Letters of Humfrey Wanley, Palaeographer, Anglo-Saxonist, Librarian, 1672–1726*. Oxford: Clarendon Press. [CEECE]
- WARTON 1745?–1790 31,344 words  
 Fairer, David (ed.). 1995. *The Correspondence of Thomas Warton*. Athens, GA/London: University of Georgia Press. [CEECE]
- WEDGWOOD 1763–1793 35,232 words  
 Farrer, Lady Katherine Euphemia (ed.). 1903–1906/1973. *Letters of Josiah Wedgwood*, vol. 1, 1762 to 1770; vol. 2, 1771 to 1780; and vol. 3, *Correspondence of Josiah Wedgwood 1781–1794, with an Appendix Containing Some Letters on Canals and Bentley's Pamphlet on Inland Navigation*. Manchester: E.J. Morten for the Trustees of the Wedgwood Museum. [CEECE]

WENTWORTH 2

1705–1739

62,223 words

Cartwright, James J. (ed.). 1883. *The Wentworth Papers 1705–1739. Selected from the Private and Family Correspondence of Thomas Wentworth, Lord Raby, Created in 1711 Earl of Strafford, of Stainborough, Co. York.* London: Wyman & Sons. [CEECE]

WOLLSTONECRAFT

1773?–1797

31,909 words

Wardle, Ralph M. (ed.). 1979. *Collected Letters of Mary Wollstonecraft.* Ithaca, NY/London: Cornell University Press. [CEECE]

YOUNG

1707?–1765

25,842 words

Pettit, Henry (ed.). 1971. *The Correspondence of Edward Young 1683–1765.* Oxford: Clarendon Press. [CEECE]

# Index

## A

- Act of Union 13  
actuation 266  
    *see also* language change  
address pronoun, *see* pronoun  
addressee 54–56, 79–80, 199,  
    214–217, 253  
age, *see* socio-demographic/  
    social category  
American War of Independence  
    13, 16  
anvil-shaped status/solidarity  
    model 199, 211, 217, 259  
apparent time 6, 63, 164–167,  
    176, 262  
*ARCHER Corpus* 46, 181–182,  
    185–186, 191, 201, 215  
Austen, Jane 24, 119  
authenticity, *see Corpus of Early  
    English Correspondence  
    Extension*  
*Authorized Version*, *see King  
    James Bible*  
awareness 25, 37, 40–41, 91, 114,  
    183, 253  
    conscious awareness 253,  
        259–260  
    language awareness 10, 27,  
        29, 34, 139, 147, 249, 257,  
        268  
    social awareness 64, 258–  
        259, 266

## B

- backprojection 268  
biblical language 84, 88, 92–93,  
    253  
*Bluestocking Corpus* 185  
-BODY 137–158, 221–222, 226–  
    228, 236–238, 240, 245–249,  
    252, 256, 258–262, 265, 267  
Browne, Thomas 109–113, 264  
“bulge” theory 199, 217

- Burney, Charles 86, 173–174,  
    177, 238, 240–241  
Burney, Frances (Fanny) 24,  
    174, 177, 237, 240–241

## C

- capital region 111, 195, 265  
    *see also* region, London;  
    region, Court  
CEEC, *see Corpus of Early  
    English Correspondence*  
CEECE, *see Corpus of Early  
    English Correspondence  
    Extension*  
Celtic influence 181, 194  
census 17–18, 268  
change from above, *see* language  
    change  
change from below, *see* language  
    change  
Civil War 164, 225  
Civil War hypothesis 164  
class 18–20, 25–26, 35, 47,  
    199, 207–211, 217, 243–245,  
    248–250, 259–261, 265  
    *see also* rank, social class,  
    social order, social rank,  
    social stratification  
    lower class 25–26, 35, 48, 217  
    middle class 15, 27, 30–32,  
        90, 199, 260–261  
    upper class 214  
    working class 20, 39  
Clift, Elizabeth 24, 114, 150,  
    179, 191, 193–196, 207–210,  
    237, 240, 265  
Clift, John 110–114, 209,  
    264–265  
Clift, William 24, 39–40,  
    113–114, 179, 193–196, 207–210,  
    238, 240, 265  
codification, codified language  
    10, 27, 37  
    *see also* standard language,  
    standardization  
coding  
    parameter coding 54  
    text-level coding 54  
colloquial language/variety  
    42, 99, 160–161, 191, 200, 253  
communal change, *see* language  
    change  
consciousness, *see* awareness  
conservatism (linguistic)  
    67, 109–111, 140, 152, 155–158,  
    173–174, 235–242, 252  
Conserving Effect 7  
copyright 45, 57–58  
corpus compilation 45–47  
*Corpus of Early English  
    Correspondence* (CEEC)  
    6, 45–48, 50, 56–57, 62  
*Corpus of Early English  
    Correspondence Extension*  
    (CEECE) 11, 19–20, 24–25,  
    45–58  
    authenticity 54–57  
    balance 47–48  
    collections 47, 295–304  
    coverage 47–54  
    editions 45–46, 50–51,  
        54–56, 295–304  
    external databases 47, 56–57  
    informants 24–25, 45, 47–54  
*Corpus of Late 18c Prose* (CLEP)  
    46  
*Corpus of Late Modern English  
    Texts* (CLMET) 4, 223  
*Corpus of Nineteenth-Century  
    English* (CONCE) 181  
correspondence 5–6, 20–21,  
    25–26, 45–52, 54–57, 61,  
    87–89, 126, 215, 218  
    *see also* letter-writing  
    family correspondence  
        49–50, 56, 83–84, 86, 92,

- 112–114, 191–196, 209–211, 248
- personal correspondence  
5–6, 45–46, 61, 77–78, 94–95, 102, 115–116, 161, 231
- private correspondence 40, 83–84, 181, 183–184, 185, 195
- Cowper, William 83, 85–86, 173–175, 177, 207–209, 217, 238–240
- Curvilinear Principle 199, 259–261
- D**
- Darwin, Erasmus 25, 173–174, 177, 208, 236, 238, 240–241
- data retrieval 58–59, 68
- Defoe, Daniel 24, 173–174, 208, 237, 240–241
- demography 14–18, 268
- detachment 113  
*see also* involvement
- dialect  
capital region 111, 195, 265  
contact 99, 164, 253, 256, 261  
dialectal differences 100–101, 173  
eastern dialect(s) 111  
*see also* region, East Anglia  
mainstream dialect(s) 6, 63, 97–98, 114, 116  
northern dialect(s) 253–254  
*see also* region, North  
regional dialect(s) 111, 172–173, 261  
rural dialect(s) 111  
southern dialect(s) 99, 172–173, 253–254  
*see also* capital region; region, Home Counties  
urban dialect(s) 138
- diffusion of change 3, 7–12, 62–63, 97–108, 114–115, 159, 164–165, 167–170, 173, 175, 231–232, 249, 255–260  
*see also* language change, lexical diffusion
- discourse 23, 84, 94, 119, 122–123
- DOES, DOT, *see* verbal -s
- DO-periphrasis 8, 71–73, 117–135, 237–239, 243–244
- affirmative statement  
8, 71–73, 117–118, 120, 122, 124–126, 128–130, 133–134, 231–232, 251–252, 254, 264
- DO-support 8, 117–118, 121, 133–134
- emphatic 119–123, 127–129, 133–134
- negative statement 8, 118, 121–122, 124
- place of adverbial 124, 132
- subject type 129–130, 133–134
- type of main verb 123, 130–134
- Draper, Eliza 174, 207–208, 217, 237, 240, 265
- E**
- early adopter 128, 235
- Early English Books Online* (EEBO) 101–102
- education 20–21, 27, 30–35, 109, 189–190
- Eighteenth-Century English Grammars* database (ECEG) 28–29, 33, 34n5
- embedding  
linguistic 117  
social 6, 9–10, 63, 126–129, 143–155, 158, 159–160, 169, 173, 228
- English  
Old English 78, 99, 181  
Middle English 7–9, 63, 78–79, 199, 250  
Early Modern English 5–9, 36, 45, 47–48, 78–79, 81, 119, 137, 158, 226  
Late Modern English 4, 40, 47, 180–184, 246  
Present-Day English 117–118, 121–124, 129–130, 132, 134, 138, 152, 159, 200, 247
- Evelyn, John 173–174, 237
- F**
- factor  
*see also* variable
- external factor 39
- grammatical factor 81, 160, 175, 248
- internal factor 247–248
- linguistic factor 3, 161, 164
- phonological factor 99
- social factor 3, 79, 159–160, 199, 243, 257, 268
- family  
nuclear family 56, 83–84, 86, 92, 128, 191–193, 205, 210, 248, 252  
other family 56, 83–84, 192, 205
- formality 79, 90, 97, 200, 217, 253, 263–264
- formulaic language 25–26, 85–86, 209  
closing formula, subscription formula 85, 93
- G**
- gender 6, 33–34, 48–51, 144–147, 155, 187–188, 228, 241–245, 269
- Gender Paradox 258
- grammatical gender 161
- Great Gender Shift 267–268  
variation, *see* variation
- generational change, *see* language change
- genre 4–5, 42, 181–182, 191, 201, 215, 232, 247
- formal genre(s) 138, 200, 217–218
- literate genre(s) 81–82, 91, 138
- oral genre(s) 137, 144, 154, 161
- spoken genre(s) 144, 247
- written genre(s) 61, 130, 200, 217–218, 250
- George III 15, 24, 49, 77, 177
- George IV 49, 77, 174, 209
- gerund 7–8
- Glorious Revolution 13, 15
- Google Books 102, 223–233

- Google Ngram Viewer  
224–225, 227–233
- grammar 25, 27–43, 81–82, 88,  
91, 120–122, 139, 248–249,  
267–268  
*see also* prescriptivism
- normative grammar  
10, 14, 35–43, 114–116, 175,  
215–216, 254, 263
- variable grammar 108–109,  
116
- grammarians 30–42, 81–82, 84,  
91, 114–115, 119–121, 182–183,  
218, 263
- grammaticalization 8, 141,  
182, 247
- H**
- hapax legomena 198, 212
- HAS, HATH, *see* verbal -s
- Helsinki Corpus of English Texts*  
(HC) 45, 161
- Historical Paradox 269–270
- historical sociolinguistics, *see*  
sociolinguistics
- I**
- illiteracy, *see* literacy
- immigration 14, 16
- incipient change, *see* stage/phase  
of change
- incoming variant/form, *see*  
variant
- indefinite pronouns with singular  
human reference, *see* pronoun
- individuals 55, 64–65, 67, 102,  
105–111, 142–143, 154–157,  
168, 173–175, 191, 193–195,  
207–209, 235–242, 266  
*see also* conservatism,  
outliers, progressivism
- industrialization 15–18, 27
- informal language 42, 154–155,  
184–185, 193, 196
- ING, *see* progressive aspect
- innovation, innovative form, *see*  
linguistic innovation
- interior social group 199,  
259–260  
*see also* Curvilinear Principle
- interpersonal identity 91–94
- intimacy 91–94, 116, 127, 196,  
253, 264–266
- inversion 118, 121, 124–125  
*see also* NICE properties,  
word order
- involvement 113, 214–215, 218,  
253–254, 264–265
- ITS 7–8, 10–11, 59, 159–177,  
221–222, 232, 235–238, 240–  
242, 245–247, 249, 255–256,  
258, 260–262, 266–267  
*see also* OF IT, THEREOF
- ITY 68, 70, 197–218, 237–240,  
245, 250, 260, 265, 267
- J**
- Johnson, Samuel 19, 38, 82, 121,  
173–174, 177, 183, 238, 240–241
- K**
- King James Bible* 97, 151, 253
- kinship, kin 56, 83, 92
- L**
- LALME, *see* *Linguistic Atlas of*  
*Late Medieval English*
- language  
community 3, 259  
*see also* speech community
- ideology 21–22, 37, 79, 257
- language change  
*see also* actuation, diffusion,  
stage/phase of change,  
transition
- communal 164, 167, 176,  
249, 262
- from above 151, 169, 249,  
268
- from below 46, 107, 169,  
188–189, 199, 248–249,  
259, 268
- generational 164, 166–167,  
176, 249, 262
- grammatical 4, 179–180,  
200, 224  
*see also*  
grammaticalization
- lacking a variable 68–74,  
221
- lifespans 173, 236–237, 262–263  
*see also* individuals
- phonological, sound  
change 7, 64, 99–100,  
258–260
- retrograde 236
- semantic 141, 199, 201,  
211–215, 218, 268
- time course of change 7–9,  
103–104, 165, 246–247,  
251–252, 260
- language of immediacy, language  
of distance 264n9
- Latin 31–33, 36, 199, 215–216,  
250
- legal language 119–121, 123
- Lennox, Sarah (Sarah Duchess  
Lennox) 24, 117, 173–174,  
177, 208, 237, 240
- letter, *see* correspondence
- letter-writing 20–21, 25–26, 34,  
77, 185, 201, 214–215, 250, 265
- lexical diffusion 97, 99–100,  
103–104, 114–115, 251
- Linguistic Atlas of Late Medieval*  
*English* (LALME) 101
- linguistic  
change, *see* language change
- embedding, *see* embedding
- factors, *see* factor
- formula, *see* formulaic  
language
- innovation 6, 89, 173, 179,  
188, 254, 266
- variability, *see* variation
- variable, *see* variable
- variation, *see* variation
- literacy 20–21, 23, 27, 34–35,  
46, 51
- full literacy 20, 35
- reading skills 20
- writing skills 20, 35
- M**
- macro level 203–205, 209,  
217, 269
- MAN 137–158, 226–228, 244,  
252–253, 256, 258, 265, 267
- map of regions 53
- metadata 5, 56–59, 225, 233

## methods

- accumulation curve 69–72, 198
- averaging of averages 64–67, 102, 105, 108
- Bayesian 65–66
- beanplot 72–73, 127–128, 184–185, 187–188
- bootstrapping 65–67, 106–107, 145–146, 148, 167–168, 170–171
- boxplot 72
- chi-square test 69, 104, 169
- cucumiform plot 70–71
- dispersion 65
- false discovery rate 73–74, 202, 217
- line graph 69, 72
- log-likelihood ratio test 69
- Mann–Whitney U test, *see* Wilcoxon rank-sum test
- mean/normalized frequency 64–65, 68–69, 72–73, 102, 106, 125–126, 167, 172, 180–181, 184–187, 193–194, 221–223, 246, 251
- median frequency 65, 72, 184
- multiple hypothesis testing 73–74
- permutation testing 69–71, 188, 193–194, 198, 235–236
- pooling 64–67, 102, 105–108, 167–168, 171
- scatterplot 72
- statistical significance 69–74, 104, 126–127, 144–151, 169, 187–189, 198, 200–207, 217–218, 243, 248, 257
- token frequency 69, 71, 198
- type frequency 69–70, 198
- VARBRUL, GoldVarb, Rbrul 67n2
- variability-based neighbour clustering (VNC) 62–63
- Wilcoxon rank-sum test 72–74, 127
- micro level 191, 207–209
- migration 14, 16, 54, 261, 268  
*see also* socio-demographic/social category

- mobility, *see* social mobility, socio-demographic/social category
- morphological productivity 69, 198
- multiple negation, *see* negation

## N

- negation 121, 124  
*see also* NICE properties
- double negation 22n5, 38
- negative concord 6, 8, 256n2, 263n7
- NESS 11, 68, 197–218, 245
- network
  - social network 90, 164, 185, 199, 241, 259
  - ties 92, 111, 164
- neuter possessive, third-person 7–8, 159–177, 232, 247, 249  
*see also* ITS, OF IT, THEREOF
- New Rhetoric 25–26
- NICE properties 118, 121
- nominal suffixes 201–202  
*see also* -NESS, -ITY
- non-standard 21, 38, 80, 91, 265
- normative grammar, *see* grammar
- norms 37–39, 41, 90–91, 129, 258–259
- noun 38–39, 68, 130, 133, 162, 200–201, 215, 232

## O

- OF IT 59, 159–163, 165, 173, 175, 232, 247, 267  
*see also* ITS, THEREOF
- Old Bailey Corpus* (OBC) 4, 223
- ONE 137–158, 226–228, 237–238, 240–241, 245–249, 252, 255, 262
- orality, *see* genre
- orthography, *see* spelling
- outgoing variant/form, *see* variant
- outliers 67, 72, 109–114, 127–128, 173, 193–196, 207–209, 235–242, 248, 260  
*see also* individuals
- Oxford English Dictionary* (OED) 80, 160, 201–202, 216

## P

- Parsed Corpus of Early English Correspondence* (PCEEC) 45–46, 57
- part-of-speech tagging 229–230, 232
- Pepys, Samuel 173–174, 208, 237, 240
- periodization 11–12, 61–64, 168, 255
- periphrastic DO, *see* DO-periphrasis
- phonology, phonological change, *see* language change factor, *see* factor variable, *see* variable variation, *see* variation
- Piozzi, Hester (Mrs Thrale) 85, 127–129, 174, 235, 237, 239, 241
- polite society 22, 25–26, 91, 225, 265
- politeness 21–23, 25, 27–29, 30, 79–82, 90, 201, 225, 258
- population growth 15–16
- possessive determiner/pronoun, *see* pronoun
- power 15, 19, 50, 55, 78, 80, 86, 91–94
- prescriptivism 4, 10, 14, 27, 29, 37–41, 259  
grammar, *see* grammar
- prestige 21, 154, 247, 265
- Principle of Contingency 268
- private correspondence, *see* correspondence
- problem of continuation 266–268
- progressive aspect 11, 59, 68, 179–196, 221–222, 233, 237–240, 245–246, 248, 255, 258, 260
- a*-prefixed progressive 183
- progressive passival 179, 183
- progressive passive 42, 179–180, 182–183, 246
- progressivism (linguistic) 67, 173–175, 235–242
- pronoun 7–8, 40, 59, 113, 137, 139, 147, 200–201  
address pronouns 78, 80, 91, 94

- indefinite pronouns with  
 singular human reference  
 7–8, 10, 137–158,  
 221–222, 226–228, 235–241,  
 246–247, 252–253, 255, 258,  
 260–262, 265, 267  
*see also* -BODY, -MAN, -ONE
- possessive pronouns/  
 determiners 7–8, 10–11,  
 59, 113, 164, 201  
*see also* ITS
- second-person pronouns  
 7–8, 10, 77–95, 251–253, 265  
*see also* THOU, YOU
- pronunciation 7, 37, 80, 82, 99,  
 118, 216
- proscription 215–216
- R**
- rank 18–21, 25, 35, 50–52,  
 108–111, 127, 149–152, 169–170,  
 188–191, 195, 205–208, 211,  
 241, 248–250, 260  
*see also* class, social class,  
 social order, social rank,  
 social stratification
- clergy 22, 47, 89, 111, 116,  
 150–152, 191, 205, 218, 236,  
 241, 252–253, 260, 265
- gentry 15, 19, 126, 149–150,  
 164, 236
- lower gentry 189–190
- lower ranks 18, 46, 48,  
 51–52, 78, 189, 191, 196
- merchants 19–21, 24, 50, 112
- middle/middling ranks 15,  
 30, 50, 90, 211, 236, 261
- nobility 52, 108, 149–151,  
 169, 260
- non-gentry 19, 164, 236
- professionals 19, 22, 50,  
 149–152, 164, 188, 194–195,  
 205–207, 209, 217, 236, 241,  
 260–261
- royalty 15, 24, 52, 205,  
 209–211, 218, 265
- upper gentry 126, 134, 252
- upper ranks 51–52, 78, 80,  
 108, 111, 166
- rate of change 7, 42, 63, 255–  
 257, 263
- stability 37, 246, 256–257
- real time 3–4, 6, 41–42, 61, 63–64,  
 166–167, 255, 261–263, 268–269
- region 52–54  
 Court, Royal 21, 265  
 East Anglia 47, 54, 101, 111  
 Home Counties 111, 172  
 London 13, 15–16, 52–54,  
 152–154, 195, 249, 261n5  
 North 100–101, 152–154,  
 172–173, 253
- regional  
 dialect, *see* dialect  
 variation, *see* variation  
 variety 261
- register 25–26, 55–56, 61, 88,  
 111–112, 116, 123, 209, 239, 243–  
 245, 248, 250–251, 253, 263–266
- enregisterment 264, 266  
 variation, *see* variation
- relationship between sender and  
 recipient of the letter 55–56,  
 77, 86, 127, 199, 204–205, 217,  
 252, 264
- repertoire 196, 263–265
- representativeness 47, 102, 228
- Restoration 13, 21, 265
- Royal Court, *see* region
- S**
- s, *see* verbal -s
- S-curve 42, 61, 63, 100, 104,  
 134–135, 165, 255
- salience 42, 134, 169, 248, 251
- sampling 47, 62–67, 69, 102
- Sancho, Ignatius 25, 83, 87–94,  
 174, 191, 238, 265
- SAYS, SAITH, *see* verbal -s
- semantics 141, 161–162, 175,  
 201, 211–215, 218
- sex 33–34  
*see also* gender
- sibilant 99–100, 115
- social  
 aspirer 111, 236, 241, 261  
 awareness, *see* awareness  
 background 22, 55, 90–91,  
 116, 197, 235  
 change 10, 23  
 class 23, 35, 38, 81, 91, 197,  
 199, 243, 259–260  
*see also* class, social  
 stratification  
 conditions 9–10, 269  
 distance 55–56, 80, 92–93,  
 199, 264–265  
 embedding, *see* embedding  
 evaluation 254, 263–266  
 group 35, 62, 90, 149–152,  
 259–261  
*see also* social  
 stratification, interior  
 social group  
 hierarchy 90, 108, 260–261  
 identity 92, 263, 265  
 meaning 235, 239, 266  
 mobility 18–19, 22, 110–111,  
 236–241, 261  
 network, *see* network  
 order 21, 62, 102, 108, 259  
*see also* social stratification  
 pattern 257–263, 269  
 rank 48, 50–52, 108, 169, 188–  
 191, 205–206, 208, 236, 241  
*see also* rank, social  
 stratification  
 relation 55–56, 90, 214,  
 253, 264  
 riser 179, 195, 248  
 role 94, 269  
 status 34, 50–51, 62, 80,  
 91–92, 94, 107–108, 111,  
 126–127, 147–152, 168–170,  
 189, 236, 242, 244–245,  
 248, 252, 259–261  
*see also* social stratification  
 stratification 18–19, 149, 151,  
 164, 248  
 structure 14, 18  
 variable, *see* variable  
 variation, *see* variation  
 society of intimates 116  
 society of strangers 116  
 socio-demographic/social  
 category 17–18, 62, 107–108,  
 149, 203–206, 250, 260–261  
*see also* variable  
 age 109, 127, 140, 147–148,  
 156–157, 164, 166–168, 173,  
 262, 268  
 domicile 53–54  
*see also* region

education, *see* education  
 gender, *see* gender  
 regional mobility 53  
   *see also* migration  
 social mobility, *see* social mobility  
 social status, *see* social status  
 sociohistorical linguistics, *see* sociolinguistics  
 sociolinguistic  
   method 45–46, 261  
   *see also* methods  
   patterns 63, 243, 248–249, 252–253, 257–263, 269  
   universal 257  
 sociolinguistics 4–6, 9, 46, 163, 228, 233, 255–270  
   historical 5, 46, 140, 221, 246, 268–269  
   present-day/modern 6, 9, 45, 199, 262  
   variationist 4, 6, 61, 224, 257–258, 261  
 speech community 166, 211, 268  
   *see also* language community  
 spelling 55, 57–59, 102, 113, 116, 226, 233  
 spoken language 118, 122, 126  
 stage/phase of change 8, 10, 62–64, 134, 165, 176, 181, 239, 243–258  
 incipient 11, 28, 42, 141, 144, 151, 236, 239–242, 246–247, 255–256, 260  
 new and vigorous 63–64, 246–247, 255–256  
 mid-range 104, 108, 165, 168–169, 235–236, 239–242, 246–247, 249, 255–258, 260–261  
 nearing completion 42, 63–64, 104, 108, 149, 155, 168–169, 176, 235, 239, 246–247, 251–252, 256, 260  
 completed 7–8, 63, 104, 106, 108, 141–142, 146–147, 149, 165, 247, 251–252, 256–258, 260  
 late stage 129

standard language 22n5, 37  
   grammar, *see* grammar  
 standardization 22, 27, 37–38, 163, 200, 215, 257  
   *see also* codification, prescriptivism  
 stigmatization 40–41  
 style of writing 25–26, 81–82, 84–86, 88–91, 94, 113–115, 121, 200–201, 207, 214–215, 217–218  
   elaborated 201, 215, 217  
   gendered 197, 200–201  
   informational 201, 214–215  
   involved 201, 211, 214–215, 217–218  
   style-shifting 79, 260  
 stylistic change 38, 200, 217–218, 250  
 supralocalization, supralocal use 253, 258–259, 261

## T

tense 118, 122–123, 134, 180, 182, 184, 186, 188, 193–194, 196, 246, 251  
 -TH, *see* verbal -s  
 THEREOF 11, 59, 160–162, 165, 221–222, 232  
   *see also* ITS, OF IT  
 third-person neuter possessive, *see* neuter possessive  
 third-person singular (suffix) 7–8, 11, 40, 65, 97, 112–115, 130  
   *see also* verbal -s  
 THOU 59, 77–95, 130, 230, 235, 237–239, 243–244, 251–254, 265  
   *see also* YOU  
   discursive use 81  
   status-related contexts 81, 91  
 transition 19, 266–267  
   *see also* language change

## U

Uniformitarianism 9, 268  
 urbanization, urban 15–18, 200, 265–266  
   *see also* dialect

## V

vanishing evidence/reliability 64, 67, 248, 258  
 variable  
   *see also* factor  
   grammar, *see* grammar  
   independent variable 248  
   linguistic variable 5–6, 11–12, 61–62, 67–73, 105, 108–109, 146, 152, 163, 199, 228, 243, 246–248, 252, 255–257, 259, 266  
   phonological variable 63, 266  
   social variable 98, 105, 126, 129, 134, 147, 252, 257  
   speaker variable 61  
     *see also* socio-demographic/social category  
 variant 7, 10–11, 38–42, 59, 64, 99, 101–102, 109n4, 113, 115, 137–162, 165, 173, 175, 180, 200, 221, 226–228, 236, 246–249, 253–254, 258, 261–262, 266  
   incoming 3, 6, 63, 65, 67, 97, 100, 106, 109, 141–144, 153, 158, 160, 164–165, 168, 221–222, 227–228, 235–236, 251–253, 258, 262–263  
   outgoing/recessive 10, 42, 64–65, 67, 97, 104, 109, 111–112, 114–116, 134, 142, 146–147, 149, 152–153, 155–157, 228, 231, 235, 251–254, 262–265  
 variation  
   gender variation 9, 105–107, 164–165, 170–172, 187–188, 196–197, 200–207, 211, 213, 218, 252–253, 257–259  
   linguistic variation 38, 61, 72, 134, 180–181, 184, 200, 207, 218, 221, 223, 228, 243, 255, 260, 263  
   phonological variation 266  
   regional variation 100–101, 111, 116, 129, 138, 152–154, 172–173, 253–254, 261, 265  
   register variation 55–56, 79–80, 94–95, 97, 100, 191–193,

- 199, 201–202, 217–218, 232,  
253–254, 263–264
- social variation 64, 102,  
109, 126–129, 136, 149–152,  
155, 160, 169, 199, 228,  
244–245, 249, 257–259
- variationist approach, *see*  
sociolinguistics
- verbal -s 7–8, 62–63, 65–66, 97–  
116, 221–222, 230–231, 235, 237–  
239, 251–257, 259–260, 264–265
- syncope 99
- zero variant 113
- vernacular 31, 33–34, 112, 137,  
150, 154–155, 249, 253
- visualization, *see* methods
- W**
- Wanley, Humphrey 127, 173–  
174, 177, 208, 237, 239–240
- weak tie, *see* network
- Wollstonecraft, Mary 22–23,  
174, 177, 193–195, 237, 240–241
- WordCruncher 58–59
- word-formation 36, 198, 215,  
250, 267
- word order 36–37, 119, 132  
*see also* inversion, NICE  
properties
- WordSmith 58–59
- written language 116, 122, 126,  
196, 200
- Y**
- YE, YOU 7–8, 78–82, 90–94, 113,  
230, 260  
*see also* THOU

Eighteenth-century English is often associated with normative grammar. But to what extent did prescriptivism impact ongoing processes of linguistic change? The authors of this volume examine a variety of linguistic changes in a corpus of personal correspondence, including the auxiliary *do*, verbal *-s* and the progressive aspect, and they conclude that direct normative influence on them must have been minimal.

The studies are contextualized by discussions of the normative tradition and the correspondence corpus, and of eighteenth-century English society and culture. Basing their work on a variationist sociolinguistic approach, the authors introduce the models and methods they have used to trace the progress of linguistic changes in the “long” eighteenth century, 1680–1800. Aggregate findings are balanced by analysing individuals and their varying participation in these processes. The final chapter places these results in a wider context and considers them in relation to past sociolinguistic work.

One of the major findings of the studies is that in most cases the overall pace of change was slow. Factors retarding change include speaker evaluation and repurposing outgoing features, in particular, for certain styles and registers.

“This book offers important insights into linguistic variation and change in 18th century England. Across morpho-syntactic variables, multiple methods and embedding their analyses in broader context, the authors demonstrate well-known sociolinguistic principles, e.g. women lead change. Yet important nuances emerge, e.g. synthetic resources increase over time (*-ity*, *-ness*). Together, these findings provide a critical real time backdrop for contemporary studies. It’s the next best thing to time travel!”

**Sali A. Tagliamonte**, *University of Toronto*

ISBN 978 90 272 0103 4



9 789027 201034

**John Benjamins Publishing Company**