

Advances in Historical Sociolinguistics

7

Exploring Future Paths for Historical Sociolinguistics

EDITED BY

Tanja Säily, Arja Nurmi,
Minna Palander-Collin
and Anita Auer

John Benjamins Publishing Company

This volume explores potential paths in historical sociolinguistics, with a particular focus on the inter-related areas of methodological innovations, hitherto un- or under-explored textual resources, and theoretical advancements and challenges. The individual chapters cover Dutch, Finnish and different varieties of English and are based on data spanning from the fifteenth century to the present day. Paying tribute to Terttu Nevalainen's pioneering work, the book highlights the wide range and complexity of the field of historical sociolinguistics and presents achievements and challenges of interdisciplinary collaboration. The book is of interest to a wide readership, ranging from scholars of historical linguistics, sociolinguistics, corpus linguistics and digital humanities to (advanced) graduate and postgraduate students in courses on language variation and change.

"The work of Professor Terttu Nevalainen has inspired countless researchers in the sphere of historical sociolinguistics, the area of research in which she has consistently been the most significant pioneer. Her research and thinking have now inspired this brilliant *Historical Sociolinguistics* which sets out exciting new directions in this vitally important and impressive field."

Peter Trudgill

ISBN 978 90 272 0086 0



9 789027 200860

John Benjamins Publishing Company

AHS
7

SÄILY, MURMI, PALANDEER-COLLIN & AUER

Exploring Future Paths for Historical Sociolinguistics



Exploring Future Paths for Historical Sociolinguistics

Edited by

Tanja Säily

University of Helsinki

Arja Nurmi

University of Tampere

Minna Palander-Collin

University of Helsinki

Anita Auer

Université de Lausanne

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.7

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2017041505 (PRINT) / 2017050572 (E-BOOK)

ISBN 978 90 272 0086 0 (HB)

ISBN 978 90 272 6481 7 (E-BOOK)

© 2017 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

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

Exploring Future Paths for Historical Sociolinguistics

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 Tiekens-Boon van Ostade

Leiden University

Donald N. Tuten

Emory University, Atlanta GA

Wim Vandenbussche

Vrije Universiteit Brussel

Anna Verschik

Tallinn University

Volume 7

Exploring Future Paths for Historical Sociolinguistics

Edited by Tanja Säily, Arja Nurmi, Minna Palander-Collin and Anita Auer

Acknowledgements

We would like to thank the authors, who were enthusiastic from the start, wrote excellent chapters and met their deadlines diligently. The anonymous reviewers, in both the peer review organised by the editors and that by John Benjamins, gave invaluable feedback and were prompt in doing so. Taru Lilja was a great help with copy-editing the chapters. Marijke van der Wal, co-editor of the AHS series, enabled us to turn the book from proposal into publication in a timely fashion, going above and beyond the call of duty with her efforts, as did Anke de Looper at Benjamins. The editing of this volume was supported by the Academy of Finland (*Reassessing Language Change: The Challenge of Real Time*, project number 276349, *Multilingual Practices in the History of Written English*, project number 258434, and *Democratization, Mediatization and Language Practices in Britain, 1700–1950*, project number 295381). Finally, our warmest thanks are due to Terttu Nevalainen, whose ongoing contributions to historical sociolinguistics provided the inspiration for this volume.

Helsinki, Tampere & Lausanne
May 2017

Tanja Säily
Arja Nurmi
Minna Palander-Collin
Anita Auer

Table of contents

Acknowledgements	vii
The future of historical sociolinguistics? <i>Tanja Säily, Arja Nurmi, Minna Palander-Collin & Anita Auer</i>	1
Part I. Methodological innovations	
Exploring part-of-speech frequencies in a sociohistorical corpus of English <i>Tanja Säily, Turo Vartiainen & Harri Siirtola</i>	23
Reading into the past: Materials and methods in historical semantics research <i>Susan Fitzmaurice, Justyna A. Robinson, Marc Alexander, Iona C. Hine, Seth Mehl & Fraser Dallachy</i>	53
Ireland in British parliamentary debates 1803–2005: Plotting changes in discourse in a large volume of time-series corpus data <i>Helen Baker, Vaclav Brezina & Tony McEnery</i>	83
Discord in eighteenth-century genteel correspondence <i>Minna Nevala & Anni Sairio</i>	109
Part II. New data for historical sociolinguistic research	
Competing norms and standards: Methodological triangulation in the study of language planning in nineteenth-century Finland <i>Taru Nordlund & Ritva Pallaskallio</i>	131
Relativisation in Dutch diaries, private letters and newspapers (1770–1840): A genre-specific national language? <i>Andreas Krogull, Gijsbert Rutten & Marijke van der Wal</i>	157
“A graphic system which leads its own linguistic life”? Epistolary spelling in English, 1400–1800 <i>Samuli Kaislaniemi, Mel Evans, Teo Juvonen & Anni Sairio</i>	187
Part III. Theory: Bridging gaps, new challenges	
Historical sociolinguistics and construction grammar: From mutual challenges to mutual benefits <i>Martin Hilpert</i>	217

A lost Canadian dialect: The Ottawa Valley 1975–2013	239
<i>Bridget L. Jankowski & Sali A. Tagliamonte</i>	
“Vernacular universals” in nineteenth-century grammar writing	275
<i>Lieselotte Anderwald</i>	
Revisiting weak ties: Using present-day social media data in variationist studies	303
<i>Mikko Laitinen, Jonas Lundberg, Magnus Levin & Alexander Lakaw</i>	
Index	327

The future of historical sociolinguistics?

Tanja Säily¹, Arja Nurmi², Minna Palander-Collin¹ & Anita Auer³

¹University of Helsinki / ²University of Tampere / ³Université de Lausanne

In this chapter we discuss the current achievements of historical sociolinguistics and highlight new insights provided by the contributions in the volume. Taking the essay by Nevalainen (2015) as a starting point, we consider the themes of crossing boundaries and bridging gaps between different levels of analysis and different paradigms, as well as proposing new paths for historical sociolinguistics as part of the wider field of digital humanities.

Keywords: historical sociolinguistics; language variation and change; layered simultaneity; informational maximalism; multidisciplinarity; digital humanities

1. Introduction

1.1 Purpose of the volume

This collection of articles focuses on three areas that play an important role in advancing research in science – and therefore also in the field of historical sociolinguistics – which are *methodological innovations*, *hitherto un- or under-explored data*, and *theoretical advancements and challenges*. By highlighting these three fundamental areas, the volume traces some of the most recent developments in the field, thereby indicating selected future directions into which historical sociolinguistics are likely to develop, particularly within the wider framework of digital humanities.

All three areas under investigation are inter-related, and each of them may serve as the starting point and/or driving factor of a specific study in the field. Due to the significant developments in digital humanities and its impact on the field of historical sociolinguistics, we start our more detailed overview of these areas with *methodological innovations*. In recent years, new methods in historical sociolinguistics have been closely linked to the developments in digital humanities/computational linguistics. It can be observed that, in some respects, a move has taken place from more philological and qualitative approaches to more expert quantitative approaches and/or combinations between them. Coupled with big data approaches, testing new methodologies is increasingly becoming

the starting point for research. Some entirely new computational methods that can be applied to sociohistorical data and that allow us to shed new light on the interpretation of the data will be presented in selected contributions in this volume (see Section 2.1). Another focus area that has had and continues to have a significant effect on developments in the field of historical sociolinguistics is the investigation of hitherto *un- or under-explored data*. Several contributions in the volume (see Section 2.2) use new data and/or make use of new combinations of data sets to interpret language phenomena from more nuanced perspectives as well as novel combinations of theoretical approaches. Finally, *theoretical advancements*, as well as challenges, can be brought about by using new methods and data, but also by applying previously unrelated theories to historical sociolinguistic data (see Section 2.3).

The volume showcases the wide range as well as the complexity of the field of historical sociolinguistics and re-emphasises the need to reach out to other disciplinary fields, often in the form of actual collaborations between scholars from different disciplines. This will in turn have an impact on the methods applied, the discovery and choice of data and the advancements of theories. English has played an important role early on in the development of the fairly young field of historical sociolinguistics, i.e. since the landmark publication by Romaine (1982). Since then, many researchers working on related topics in different languages have joined forces and have advanced the field through valuable contributions (cf. for instance the establishment of the Historical Sociolinguistics Network and publications, including proceedings, book series, and the journal, that have emerged from this collaboration). In line with this, the studies in this volume are concerned with different languages, including Dutch, Finnish and different varieties of English. What is more, the approaches described and applied in these studies will be valid for and applicable to other languages as well. As regards the temporal coverage of the volume, the contributions work with data spanning from the fifteenth century to the present day. We hope that the insights presented in the volume will significantly facilitate historical sociolinguistic research in the future, and open new avenues and trajectories for research, especially in the context of digital humanities. While the volume separates the approaches according to their main focus on either methods, data or theory, all contributions are concerned with more than one of the main questions of the volume, highlighting the fact that advances in one area are by necessity linked to rethinking and re-evaluating the others.

In its discussion of advances in the field of historical sociolinguistics with regard to method, data and theory, this volume also pays tribute to Terttu Nevalainen's pioneering work in the field (e.g. Nevalainen 2012, 2015; Nevalainen & Raumolin-Brunberg 1996a, 2012, 2017). Several of the contributions pick up on the themes and multidisciplinary ways of working for which Terttu Nevalainen is

well known and respected. For this reason, we want to commence our discussion with Nevalainen's 2015 essay.

1.2 Nevalainen's 2015 essay

Nevalainen (2015) asks the question, "What are historical sociolinguistics?" She begins her discussion with Bell's (2013) list of the paradigms of sociolinguistics: the sociolinguistics of multilingualism (including sociology of language and, increasingly in recent years, critical constructivism), variationist sociolinguistics and ethnographic-interactional sociolinguistics. According to Nevalainen (2015), these can be applied to historical sociolinguistics in varying degrees. The first paradigm could be called *comparative historical sociolinguistics*, a growing area that can be studied on the basis of meta-textual and secondary sources. *Variationist sociolinguistics*, or the study of language variation and change in relation to external factors, is only possible for documented periods of the language in question and requires access to primary textual materials by identifiable individuals and groups. The same applies to *ethnographic-interactional sociolinguistics*, with the additional complication that participant observation is not available, which is why we need to rely on ego-documents such as private letters and diaries. However, Nevalainen points out that pigeonholing research is generally not useful and that it ignores commonalities: many research questions require the combination of a variety of approaches.

This brings Nevalainen to the key point of her essay: the *layered simultaneity* of various micro- and macro-levels of contextual meaning, first discussed by discourse analysts (e.g. Fairclough 1992) and later developed in ethnographic nexus analysis. Nevalainen argues that a holistic perspective accounting for layered simultaneity is especially necessary in the study of the past, which is less known to us. For example, considering the role of communities in language change, we need to account for the micro-level agency of individuals within the community as well as macro-level diffusion across communities, with social networks perhaps forming an intermediate level. As noted by Auer & Hinskens (2005), these levels also represent varying time scales: individuals exhibit short-term accommodation in interaction but also long-term accommodation over their lifespans, while language change may actuate in the interactions between individuals in their social networks, and diffuse over a longer time period to the community at large. Another example is that of layered socio-cultural processes (Culpeper & Nevala 2012), with the action of individuals at the micro-level (e.g. speech acts), mezzo-level activities of local communities (e.g. genres), and macro-level processes associated with broader communities (e.g. ideologies). Nevalainen notes that while all of these may manifest at the micro-level, they also require a macro-level historical

analysis; in addition, all of them may change over time and vary across cultures. Hence, texts need to be analysed “in the context of contemporary social and discursive practices”, in relation to both social structures and active production in interaction (Nevalainen 2015: 252; see also Fairclough 1992: 72).

The idea of layered simultaneity has a number of implications for historical sociolinguistics, as Nevalainen (2015) points out. Most work on layered simultaneity so far has focused on the present. In studies of the past, reconstructing contexts and activities is more difficult and takes place at the intersection of other disciplines, such as paleography, history, discourse analysis and genre studies. This multidisciplinaryity is part of what Nevalainen, following Janda & Joseph (2003: 37), calls the principle of *informational maximalism*. The other part of informational maximalism involves matching parallel data sources, including for example biographies, ego-documents (both manuscripts and various editions thereof), official documents and history writing, especially social, economic and population history. She further notes that access to real time is crucial in diachronic studies as both external circumstances and linguistic forms change at varying time scales.

Nevalainen (2015) also identifies some requirements specific to the study of real-time language change, which is one of the central areas of interest in historical sociolinguistics. First, as the actuation problem of linguistic change usually remains unsolved, *models* are needed that account for the diffusion of change in social interaction. Moreover, she argues for the importance of *baseline evidence*, or “mapping actual processes of change in their different stages at the aggregate level of the community” (Nevalainen 2015: 265). To come up with this evidence, we need both multi-genre and socially stratified corpora, the metadata of which provide some of the layered simultaneities to be considered. Finally, Nevalainen (2015: 266) expresses her hope for “further rapprochement between the history disciplines” in the spirit of informational maximalism, noting that this could take place within the framework of the digital humanities.

This volume is our contribution to the notion of layered simultaneity in historical sociolinguistics. In line with Nevalainen (2015), we strive for informational maximalism in terms of both multidisciplinaryity and multiple data sources. We hope to show that crossing disciplinary boundaries and bridging gaps between different levels of analysis opens up new paths for historical sociolinguistics.

2. New insights

As previously pointed out, albeit several of the contributions could be discussed in any of the focus areas, for the sake of discussion, we divided them up into areas to which we felt they made a major contribution. What follows below is a more

detailed discussion of these three areas and a first introduction to the new insights provided by the contributions in this volume.

2.1 Methodological innovations

Previous research in historical sociolinguistics has tended to focus on language variation and change using variationist methods borrowed from present-day sociolinguistics (e.g. Nevalainen & Raumolin-Brunberg 2003). In the absence of a clear linguistic variable, researchers have resorted to normalised frequencies and simple hypothesis testing as in diachronic corpus linguistics (*ibid.*). Comparative historical sociolinguistics and ethnographic-interactional approaches have had their own, often more qualitative methods. All approaches have touched upon Labov's famous bad-data problem (1994: 11): there are typically only written materials, which are scarce and not representative enough, and our knowledge of the contexts in which they were produced is incomplete (see also Section 1.2 above).

To alleviate the bad-data problem and other issues in historical sociolinguistics, more advanced quantitative methods have gradually been developed, many of them in multidisciplinary projects headed by Terttu Nevalainen. Hinneburg et al. (2007) and Mannila et al. (2013) have investigated better methods of handling small sample sizes, such as bootstrapping. Another trend has been to improve ways of accounting for variability across individuals (e.g. Nevalainen et al. 2011). In hypothesis testing, this has meant the adoption of so-called dispersion-aware tests (Säily 2014: 46), such as the t-test, the Wilcoxon rank-sum test and tests based on the statistical technique of resampling, including the bootstrap test (Lijffijt et al. 2012, 2016; Säily & Suomela 2009). The trend has also spread to statistical methods in present-day sociolinguistics (e.g. Brezina & Meyerhoff 2014; Tagliamonte & Denis 2014). Visualisation techniques, too, are improving, as simple line graphs are increasingly complemented by graphs that reveal the variability within time periods and social groups, such as beanplots (Säily et al. 2011; Vartiainen et al. 2013; Nevalainen et al. forthcoming).

These methods thus facilitate the holistic perspective of layered simultaneity by providing simultaneous access to the individual and to the community. A further step in this direction are interactive visualisation tools (e.g. Siirtola et al. 2014, 2016; Mäkelä et al. 2016), which connect texts, metadata, statistical analyses and visualisations in an exploratory interface that enables effortless movement between various levels of analysis. Similar efforts are being made in related areas of historical linguistics, such as diachronic corpus linguistics (Hilpert 2011), historical semantics (Rohrdantz et al. 2011, 2012), historical discourse analysis (Lyding et al. 2012) and even literary studies (Hope & Witmore 2010), often using larger data sets with poorer social metadata. Nevertheless, both big and rich data

hold potential for historical sociolinguistics, as also shown by the contributions to this volume.

The methodological part of this volume strives to further improve and facilitate research in historical sociolinguistics. In response to Nevalainen's (2015) call for multidisciplinary, the volume seeks to answer the following questions: Which state-of-the-art statistical and visual methods could be relevant to historical sociolinguistics, and what kinds of methods may be drawn from related disciplines? Moreover, thinking of metadata as a bearer of layered simultaneities, how may we better handle the combination of data and (socio)linguistic metadata?

In a collaboration between linguists and a visualisation expert, Säily, Vartiainen & Siirtola (this volume) address the issue of combining textual data, linguistic annotation and social metadata in a large-scale exploration of variation and change in part-of-speech (POS) frequencies in the *Parsed Corpus of Early English Correspondence* (c.1410–1681). As more and more richly annotated corpora are becoming available, Säily et al. conduct a timely methodological investigation into the extent to which POS annotation can be used as a tool for historical sociolinguistics, tracing not only genre evolution but also sociolinguistic variation and change at a higher level than that of individual linguistic variables. Their exploration is data-driven but also tests the hypothesis of colloquialisation in the letter genre, providing baseline evidence for further research, as called for by Nevalainen (2015). While their choice of visualisation is the line graph (or a regression line based on a scatter plot), which is arguably the simplest alternative for visualising a large number of different categories, the data behind the graphs is not based on aggregate mean frequencies but accounts for variability across individuals. Moreover, Säily et al. complement their quantitative analysis by close reading and a discussion of the relevant social contexts at various levels of granularity. They conclude that POS ratios, explored through simple visualisations and combined with qualitative analysis, can be a useful tool for achieving an overview of sociolinguistic variation and change in a corpus.

The insights presented by Fitzmaurice et al. (this volume) come from the field of historical semantics. Fitzmaurice et al. study big data in historical linguistics while taking historical and social contexts into account. Furthermore, they combine the massive textual source of EEBO-TCP with human-curated data from the *Historical Thesaurus*, bringing together corpus linguistics and lexicography. Their key methodological insight is a data-driven, bottom-up investigation of conceptual change: their “discursive concepts” are not word-based, enabling researchers to uncover historical dependencies and sociolinguistic relations unconstrained by their own modern worldview. The potential of the methods for historical sociolinguistics is made even greater by the fact that they can be easily applied to other data sets.